RISK.1027

Estimate Accuracy: Dealing with Reality

John K. Hollmann, PE CCE CEP

ABSTRACT—This paper reviews over 50 years of empirical cost estimate accuracy research and compares this reality to common but unrealistic management expectations. The empirically-based accuracy research of John Hackney, Edward Merrow, Bent Flyvbjerg and others on large projects in the process industries is summarized. The paper then highlights risk analysis methods documented in recent AACE Recommended Practices that yield outputs based upon and comparable to empirical reality. Tragically, many cost engineers are facilitating management's collective and sometimes willful biases regarding accuracy by using flawed, unreliable risk analysis methods; those who use empirically valid practices face the fate of Cassandra. The paper is intended as a fundamental reference on the topic of accuracy as well as a call for our profession to use reliable practices and speak the truth to management. Attendees will gain an understanding of estimate accuracy reality, the risks that drive it, management's biases about it, and methods that analyze risks and address the biases in a way that results in more realistic accuracy forecasts, better contingency estimates and more profitable investments.

Table of Contents

List of Tables

List of Figures

Introduction

Accuracy is a measure of how a cost estimate will differ from the final actual outcome. It is also a measure of cost uncertainty or risk (these terms are essentially synonymous in TCM). Empirical estimate accuracy data has been researched for over 50 years [30]. In addition, reasonably reliable practices for quantifying project cost uncertainty have been recommended by AACE and others. However, the level of industry understanding of the reality of accuracy and how well risk analysis methods forecast it is generally poor. Investment decision makers seem particularly unaware of our research and recommended practices. Sometimes they are aware but seem to ignore them. Worse, many cost engineers facilitate management ignorance by standardizing their wishful thinking (i.e., tunneling or neglecting sources of uncertainty) as exemplified by bias towards 10% contingency and +10/-10% range [42]. Poor investment decisions may result from using risk analysis methods that are known to be a "disaster" when systemic risks are present [27].

One researcher said this behavior verges on "criminal" [19]. Cost disasters and criminality are economic in nature, but deadly serious to owners, investors and tax-payers; one must ultimately take responsibility for our role in their economic well-being. To help improve on the situation, this paper surveys the research facts (reality), exposes flawed practices and highlights better practices.

The paper summarizes data from well referenced studies by others; however, the data confirms the author's experience. The author's data and observations are added as well as observations by others. While fact and opinion are mixed, it is hoped that readers will draw the same conclusions as the paper and work to improve the situation.

Studies of Overall Estimate Accuracy

How accurate have cost estimates been for owners? To answer this, references providing empirical data on estimate accuracy and cost uncertainty were sought. This paper focuses on engineering and construction projects in the process (e.g., oil, gas, chemicals, mining, metals, utilities, etc.) and infrastructure (often associated with process plant projects) industries. These are generally characterized by complexity, unique work scopes, design change and sometimes new technology. The chosen references represent academic, research, consulting and industry practitioner sources. Empirical research on defense, aerospace and IT projects was found but excluded; their experience is analogous but more extreme [13,16, and 20].

Estimate accuracy and cost uncertainty data from 12 empirical studies are summarized in Table 1. These include over 1,000 projects with samples ranging from about 20 to 250 projects each. The projects were typically large enough to affect enterprise success (i.e., typically one million US dollars up to megaprojects). The costs studied are the costs to the owners. Study purposes varied; however, the typical the questions were: "*what is the accuracy of our estimates and why*?" in reaction to a perceived preponderance of cost overruns.

Study Attributes				Estimate Accuracy of Sample		
Study Reference	**Projects**	**Reference Point**	**Adjusted?**	**P10 or similar**	**P50 or mean(µ)**	**P90 or similar**
[14] Figure 1	63 Mining and Metals	From Bankable Feasibility	No	~<3>%	+16%	~+70%
[15] Page 8	100 Mining	From Authorized Feasibility	Scope & time	<15>%	0%	+43%
[19] Figure 1	258 Transport	From estimate at "Decision"	Time	~<15>%	~+15%	~+100%
[22] Figure 18.1	22 Process Plants	<500 Rating (assumed funding)	Scope & time	+2%	+10%	+39%
[31] Table 2	167 Road/Rail	Varied reference	No	~<32>%	µ= +15%	~+62%
[36] Table 4.1	47 Mega Process Plant	From start of "Detailed Engr"	Time	<14>%	µ= +88%	+190%
[35] Table 4.3	30 Process New Technology	RAND Class 2	Scope & time	+7%	µ= +28%	+59%
[34] Page I.3.4	56 Hydropower	From "Appraisal"	Time	<15>%	µ= +24%	+65%
[39] Database	188 US Pipeline 2000-2008	From FERC filing	No	<21>%	0%	+34%
[40] Figure 3	Water Projects-5 Aus. States	From Budget	No	µ for best state = +8%		µ for worst state = +80%
[43] Figure 2	36 Refinery Turnarounds	From Budget	Uncertain	+8%	µ= +23%	+38%
[46] Table 1	21 Mining and Metals	From Feasibility	Time	3 of 21 underran	µ= +17%	2 worst µ= +55%

Table 1—Empirical Estimate Accuracy Studies (Typically From the Funding Estimate)

In summary, the approximate range of ranges for accuracy or uncertainty around the reference amounts are as follows:

- P10: -32% to +8% (average about -9%)
- P50 or mean: 0% to +88% (average about 21%)
- P90: +34% to 190% (average about 70%)

The "accuracy" shown is the percentage variation of the final actual cost from the reference estimate. The reference estimate was usually the basis for an actual or *defacto* investment decision by the owner. Estimate names are industry specific; for example, "feasibility" is the funding estimate for mining projects, but not for projects in other industries. The reference estimates usually include contingency; therefore, the accuracies are understated in respect to base estimates without contingency. From the author's experience, the contingency applied at sanction is usually between 5 and 15 percent.

The statistics provided ranged from mean and standard deviations alone to distribution charts or tables or values at various confidence levels. For comparison, the accuracies in table 1 are summarized at *approximate* p10/p50/p90 confidence levels where the "p" value indicates the percentage that underran. If a mean was provided (μ), it is shown as such. If p-values were not provided, they were approximated from the mean and standard deviation assuming a normal distribution; i.e., p90 equals the mean plus 1.28 times standard deviation (the ~ symbol indicates an approximation.) This approximation underestimates the high range when actual/estimate accuracy data is skewed to the high side (i.e., actual data is not normally distributed).

The project samples were not scientifically random, but were not selected specifically because their estimates were inaccurate; the authors generally considered the projects in their samples to be reasonably representative. Studies done in reaction to overruns may be biased toward that experience; however, the number of studies and the variety of industries, regions and project types covered indicate that cost overruns are prevalent for large process industry projects.

The quality of the datasets varied, but in general the authors lament the poor state of historical project records. For many studies, the only reliable data was the cost at the time of project funding approval (i.e., sanction or investment decision) and the cost at completion. However, some studies were corrected for major scope changes and escalation which many practitioners would not expect an estimate to cover.

The key observation is that <u>in no case was the nominal p90 value ever less than +34% of the funding estimate</u> (i.e., about +40 to +50% of the base estimate). Also, the average mean or median overrun is about 21%. This is the best picture we have of reality for large process industry projects with all their imperfections and risks (unfortunately, causal data is lacking).

Arguably, the most notable studies are by John Hackney and Edward Merrow because these are the foundation for process industry phase-gate project systems [22, 35]. However, the studies by Dr. Bent Flyvbjerg are perhaps best known in the popular press [19]. Dr. Flyvbjerg has made the following statements regarding industry estimating practices: "*We conclude that the cost estimates used in public debates, media coverage, and decision making for transportation infrastructure* *development are highly, systematically, and significantly deceptive." "(those) who value honest numbers should not trust the cost estimates presented by infrastructure promoters and forecasters."* He adds, "*institutional checks and balances—including financial, professional, or even criminal penalties for consistent or foreseeable estimation errors—should be developed to ensure the production of less deceptive cost estimates* [19]."

Merrow disagrees with Flyvbjerg in the following: *"There is widely held belief that large public sector projects tend to overrun because the estimates are deliberately low-balled. Our (IPA's) analysis of large private sector projects suggests that no Machiavellian explanation is required. Large projects have a dismal track record because we have not adjusted our practices to fit the difficulty that the projects present* [33]."

Regardless of motives and causes, large process and infrastructure projects (and defense, aerospace and IT) are frequently overrunning our funding estimates and by very large margins. The search found no research that showed otherwise. Further, as "forecasters" (as one is referred to by Flyvbjerg) we are failing to reliably predict the proper point of funding including contingency, but the *range* of project cost uncertainty.

Studies of Estimate Accuracy Progression Versus Level of Scope Definition
It is generally agreed that the less well defined the project scope is, the wider the estimate accuracy range will be. This is a premise of phase-gate project systems. Table 2 summarizes accuracy studies from among they paper's sample that also addressed accuracy and uncertainty at various levels of scope definition *approximated* to AACE classifications (Class 5 to 1).

	Conf. Level	Class 5	Class 4	Class 3	Class 2	Class 1
		Actual % Difference from Estimate (or as stated)				
Hackney [22] -Conventional process plant -Range around the **Base Estimate** -Adjusted for scope and time	P10	+36	+34	+9	+2	-1
	P50	+60	+41	+15	+6	+4
	P90	+80	+45	+21	+12	+11
RAND [35] -Process plant; newer technology -Range around the **Funded Amount** -Adjusted for scope and time	P10	+39	+18	+9	+5	-1
	P50	+100	+60	+28	+20	+7
	P90	+260	+150	+58	+41	+18
Harbuck [23] -Avg. of 3 Transport Study Averages -Range around the **Construction Estimate** (not the final actual costs) -No adjustment for scope & time	Mean	+49	+37	+18	study base	n/a

Table 2 - Empirical Estimate Accuracy Studies (Progression by Level of Scope Definition)

Complicating the comparison, each study has different data attributes and uses different scope definition rating schemes (i.e., AACE classification ratings were not used). However, the author's experience is that process industry funding decisions are being made based on scope definition somewhat better than AACE Class 4, but worse than Class 3. Research indicates that the design development necessary to thoroughly mitigate definitional risk includes *issued-for-design,* signed-off process and instrumentation diagrams (P&IDs) for all process and utility units; the author rarely sees this level of definition at the time of funds authorization [4].

The key observation from Table 2 is that even projects funded on better scope definition (AACE Class 3) tend to be overrun; there is a huge potential for overruns if the scope is more poorly defined than Class 3. Note that the Hackney and RAND models based on this data are available in working Excel tools found at the AACE website (www.aacei.org [11]).

What Our Estimates Say and What Owners Want Are the Same (i.e., Wishful Thinking)

The next question is "are the overrunning projects within the cost range of our risk analyses?" Unfortunately, they usually are not. The author has reviewed many industry risk analyses by owner companies and their EPC contractors and their p90 forecast is rarely great than 30% over the base estimate excluding contingency. Table 3 provides an indicative sample of risk analysis outcomes.

Project Type	Estimate Class	Preparer	P10	P50	P90	Method	Notes
Mega, Expansion, Refining	Class 4	Owner	<15>%	+13%	+45%	Ranging w/M-C (validated)	Exceeded P90 at next estimate
Large, New, Mining	Class 4	EPC	+7%	+13%	+19%	Ranging w/M-C	Remote, developing country
Small, Revamp, Refining	Class 3	Owner	<3>%	+5%	+13%	Ranging w/M-C	Plant-based project
Mega, New, Metals	Class 4	EPC	+0%	+5%	+13%	Ranging w/M-C	Low wage country
Mega, Expansion, Refining	Class 4	EPC	+2%	+10%	+18%	Ranging w/M-C	Largest in region, low wage country
Mega, Upgrader	Class 4	EPC	<3>%	+12%	+28%	Ranging w/M-C	Remote, severe winter

Table 3 - Reported Accuracy Ranges from Owner and EPC Contractor Risk Analyses

The first project risk analysis shown in Table 3 had a p90 value of +45%; however, the risks on this project were extreme and while the team captured some of them, the range was overrun by the next phase estimate. P50 values are often as little as 5% even for highly risky projects. The author's experience is that despite extensive risk registers and brainstorming sessions, most risk *quantification* is dominated by an estimator's bias in which the team consciously or unconsciously perceives uncertainty in terms of estimate and takeoff assumptions and math (i.e., "estimator's risks"). A high (p90) range of about +30% reflects the perceived worst case uncertainty around quantities, rates, pricing and productivity while unrealistically assuming the scope is fixed, the execution strategy and plan is never changed, no risk events occur and if they do, risk responses are always effective. The result is a range that seems to be what the owner wants to hear.

So the next question is "what does the owner want to hear?" Table 4 provides an indicative sample from different industry segments of owner accuracy range expectations as stated in their phase-gate project scope development processes.

	CLASS 5	CLASS 4	CLASS 3
	AACE 18R-97 RANGE of RANGES		
	-20/50% to +30/100%	-15/30% to +20/50%	-10/20% to +10/30%
COMPANY	**OWNER "TARGETS"** • *Most misquote AACE (AACE has not quoted target ranges for 15 years)* • *NONE state what the confidence interval statistically represents*		
Oil Sands	-30 to +50%	-20 to +30%	+/- 10%
Power	-30 to +50%	-15 to +30%	-5 to +15%
NOC Oil	-/+50%	-/+30%	-/+15%
Mining	-/+50%	-/+25%	-/+10%
Integrated Oil	-15 to +50%	-10 to +30%	-10 to +25%

Table 4 - Owner Phase-Gate Target Accuracy Ranges Vs. AACE Classes and Empirical Studies

The table compares the owner targets to the range-of-ranges in the **AACE Recommended Practice 18R-97** [4]. Is it coincidence that the owner p90 targets in table 4 are about the same as the p90 values estimated in table 3?

By quoting specific accuracy range targets in their processes, owners display a dangerous misunderstanding of risk and estimating. Once a project plan reaches the target level of scope definition (e.g., Class 3), the residual risk and its potential impact is a project scope attribute and no estimator can appreciably improve the accuracy range by doing a "better estimate." For a project with substantial risks (most large projects), the company accuracy ranges in table 4 have no relevance. Unfortunately, targets tend to pre-determine risk analysis outcomes; i.e., they drive the risk analysis outcomes seen in table 3 (owners get what they ask for). Targets are *prima facie* evidence of risk ignorance (tunneling) driven by the inflated expectations that phase-gate processes alone will manage risks.

Further, the communication of targets by many owners is statistically meaningless. First, many misquote AACE Recommended Practices by stating that the targets are "per AACE" when no AACE document includes specific targets [4]. Also, few state the confidence interval represented or the reference value that the range is around (the base or the funded amount?).

To wrap up the target/as-estimated versus actual accuracy discussion, figure 1 shows the averages of table 1 (Reality) and table 3 (As Estimated) as log-normal curves with p10/50/90 values comparable to the table 1 and 3 averages for those confidence levels.

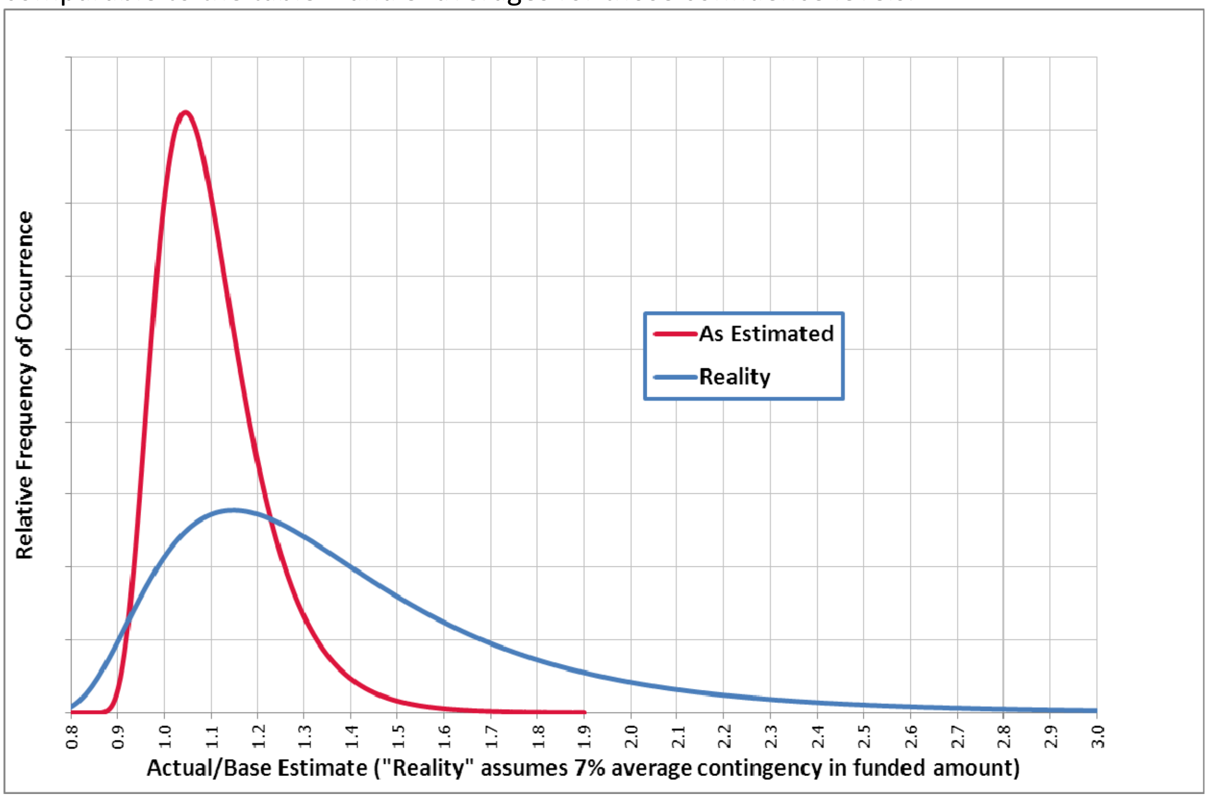

Figure 1 - As Estimated (and Target) Accuracy Vs. Empirical Accuracy at Funding

Several of the studies and the author's experience suggest that the lognormal distribution of actual/estimate data is representative [14,35]. Teams are assuming a p10/p90 accuracy range around the base estimate of about -10/+30% with a p50 value of about 10%, while the reality is closer to -20/+120% with a p50 of about 20%. Arguably, a 20% or even 30% overrun will not render most projects unprofitable; however, a 120% overrun at p90 would.

Challenging the Data : Nowhere to Hide

The following are likely challenges to this paper's findings along with the author's responses:

<u>1) The actual data includes the impact of major scope changes and escalation.</u>

Major scope change and escalation are by definition excluded from contingency [3]. In table 1, 7 of the 12 studies corrected for price changes over time and 3 corrected for major scope changes (i.e., changes to basic project premises such as plant location, product specifications or capacity). Studies show wide accuracy ranges with or without correction. In the author's benchmarking experience, major owner scope change (as opposed to design changes which owner costs must cover) is uncommon (this may be less true for public projects).

2) We cannot forecast volatility and/or black swan events (i.e., unknown unknowns).

While any *one* black swan event is improbable, the probability of *any* black swan or an equivalent confluence or compounding of lesser risks occurring during the extended duration of large projects is likely. The accuracy findings appear to hold for all time periods and regions, in both hot and cold economies. For example, Ernst & Young found that mining projects estimated during hot markets (when market risks were known) were still overrunning during the post-2008 recession; "*Of the companies that reported project overruns publically (between Oct 2010 and March 2011), the average overrun was about 71% of the original project cost estimate*"[37]. Another mining article referenced a series of studies which indicate that overruns have been the norm in every time period since 1965 [38]:

- "A study of 18 mining projects covering the period 1965 to 1981 showed an average cost overrun of 33 per cent compared to feasibility study estimates.
- A study of 60 mining projects covering the period from 1980 to 2001 showed average cost overruns of 22 per cent with almost half of the projects reporting overruns of more than 20 percent.
- A review of 16 mining projects carried out in the 1990s showed an average cost overrun of 25 percent".

Historical experience alone is enough to quantify the probability and impact "unknown-unknowns" as a class. We may not know the risk's name, but we know about what it will cost (i.e., Table 1.)

3) Some systemic risks are difficult to measure and/or politically sensitive.

The tools for rating scope development as well as competency and project system discipline (e.g., weak change management) are well established [4,21,22,35,and 45]. While including "incompetent management" in a risk register is problematic, it is necessary to identify and quantify such risks. The risk analyst must have sufficient independence to do so.

4) Estimating "all" project cost risk is not part of the job (not in my work scope).

If one declared in the *Basis of Estimate* reports that "most significant risks were excluded" and/or "past experience with similar projects was ignored,' this challenge might have some validity. However, in the author's experience, such statements (or confessions) are rarely made. Unfortunately, breaking risk down (e.g., operational, project, strategic, enterprise, contextual, global, background, etc. [42]) and disseminating responsibility for its analysis and quantification is a potential recipe for forecasting failure. Risks interact and often compound and cannot readily be parsed for quantification like elements in a work breakdown.

In summary, it is the author's experience that these "challenges" are usually just reasons for our failings; they do not excuse them. One knows better and the data is clear; with empirical insight added to other methods, risk is always quantifiable albeit imperfectly.

Flawed Practices and Lost Credibility

Flawed practices such as a bias toward estimator's risk, misguided targets, tunneling and parsing risk quantification have been mentioned. The 1990s also brought reengineering and downsizing to the industry with the loss of empirical data and analytical skills. Concurrently, Monte Carlo simulation (MCS) for spreadsheets was introduced which made risk analysis seem simple and doable regardless of skill level. Unfortunately, MCS was applied in "line-item ranging" (as opposed to range estimating) in which the team assigns cost ranges to line-items in their estimate (i.e., contributing to estimator's bias) based on brainstorming, and then runs the MCS, usually without considering line-item dependency [24]. The risks listed in the register (which tend to exclude systemic risks) are not explicitly included in these models. This is the method that research has shown to be a "disaster" for projects with systemic risks [27]. "Line-item ranging" (or activity duration ranging for schedule) fails in part because of faulty application (i.e., no dependencies) but also because brainstorming is unable to elicit the impacts of systemic risks on individual estimate line items or activities, and finally, the impact of risk register events are difficult to ascribe to individual estimate line items in aggregate. This method is not an AACE Recommended Practice.

It is easy to conclude from the research and observations that <u>our risk analyses and contingency estimates are not credible</u> for large process industry projects. Decision analysis expert John Schulyer defines a *credible analysis* as "one that gets used [44]." The following statements by industry executives indicate that our analyses are not useful (self-criticism by owner executives is understandably more difficult to find):

- Schlumberger CEO Andrew Gould stated: *"...while not wishing to embarrass any of my customers, I would add that many greenfield (upstream oil) projects suffer significant cost overruns. Indeed, as a general rule 30 percent of such projects experience budget overruns of 50 percent* [41]."
- Financier Jasper Bertisen of Resource Capital Funds (RCF) had this observation: "*the vast majority of mining projects have been coming in way over budget for the past couple of decades. As a result, RCF now automatically factors in an average cost overrun of 25% when it considers the cost of mining projects* [28]."

The prevailing use of flawed analyses has damaged our collective credibility. This will be difficult to remedy because poor practices have become institutionalized. For example, in the mining industry, the author commonly finds companies funding projects at a p80 level of confidence. This has evolved because (as indicated by prior quotation) managers intuitively understand that the p50 values we provide in our estimates are too low (i.e., often <10% contingency on even the riskiest projects) and they feel that the p80 level of about 15 to 20% contingency is more realistic. However, it is "more realistic" because in fact <u>this forecast p80 is the p50 of the "reality" that we fail to predict</u>! Cost engineers who do use realistic risk quantification practices are treated like Cassandra; management will not believe the truth after being fed unreality for decades. The *real* p80 or p90 is likely to be unprofitable; as shown in studies, the *least* p90 capital cost growth is >40 to 50%. If management faced this reality, no project would ever be authorized without stellar scope definition and optimization, top-notch planning, team building, risk management and all of the other best practices we know of. Isn't that the point? Why would anyone facilitate anything less? Why would one let them assume that poor practices are a safe bet when they are courting *disaster*!

The lesson from the empirical history (table 1) and the practice history (table 2) is that one needs to address the entire scope of risks (project-specific, systemic, and escalation) and the empirical "reality" of uncertainty on large process industry projects. Research by others points in the same direction [16,17,18,19,22,27,32,33,42]. AACE is currently developing a Decision and Risk Management Professional (DRMP) Certification that will focus on risk identification and quantification competencies, including AACE Recommended Practices that document reliable methods.

The Project Size Dichotomy

There is less empirical research of small project estimate accuracy because these projects are individually less of a threat to overall profitability and shareholder's perceptions. However, we know that the realities of small and large projects differ; small projects are biased to overestimating and underruns. As stated by one researcher, "*when a project team sets a soft (cost) target, about half of the unneeded funds are usually spent…about 70% of small projects underrun*" [29]. This research also indicated that in small project systems, overruns tend to be punished. To avoid punishment (in less disciplined cultures) teams avoid overruns by including "fat" (i.e., above-the-line contingency) in the base estimate because high visibility contingency is often poorly received by management for any project size. This can bias a company's perception of risk and partly explain their misguided targets.

Few researchers study small projects because not only is record keeping lax, but underruns are rewarded and are not seen as a problem despite being associated with wasteful capital spending. Figure 2 shows a representative distribution of actual/estimate values observed by the author for small project portfolios; often, no projects overrun by more than 10%. In this "cresting wave" pattern, most projects spend all their funds, while some return all or some of the excess; for this outcome, management and/or teams are rewarded. The more that funds are wasted, the sharper the peak between 0.9 and 1.1. Perversely, the more "accurate" the outcome, the less desirable (though best rewarded) it is; underruns and tight accuracy often indicate overfunding and wasted capital rather than excellent project control discipline

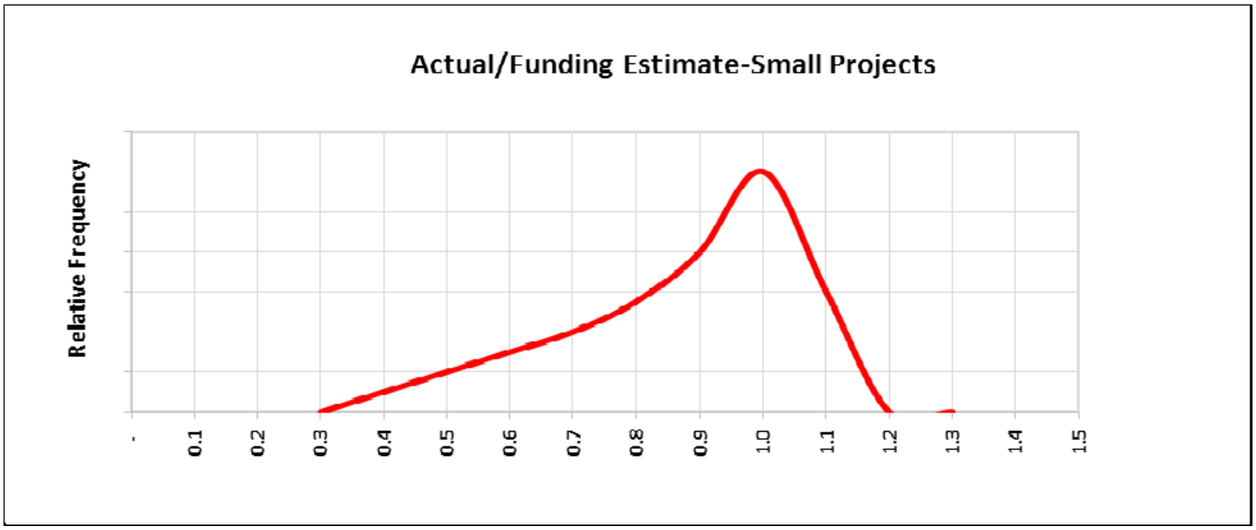

Figure 2 - Actual Cost/Appropriation Estimate For Small Projects

The small vs. large project overrun dichotomy can induce a kind of corporate schizophrenia. Many owner companies have “major project” organizations that are separate from small or plant-based organizations. A newly formed major project group will often inherit the small project system trait of risk-ignorance (expectation of underruns.) They do not appreciate that EPC contractors for major projects prepare base estimates with less fat because reviews expose fat and there is sometimes a bias to keep estimates low to see the project get funded. The combination of small project target-reinforced tunneling and risk-free base estimates is a recipe for overruns on large projects.

Looking at an entire company project portfolio, the combined distribution of accuracy data for small and large projects can look serenely “normal.” Benchmarking data observed by the author indicates that in a population of all project sizes, the P10/P90 range is about +/-20% around the funding estimate as illustrated in figure 3. Given just one distribution, management will be unaware that there are two conflicting realities.

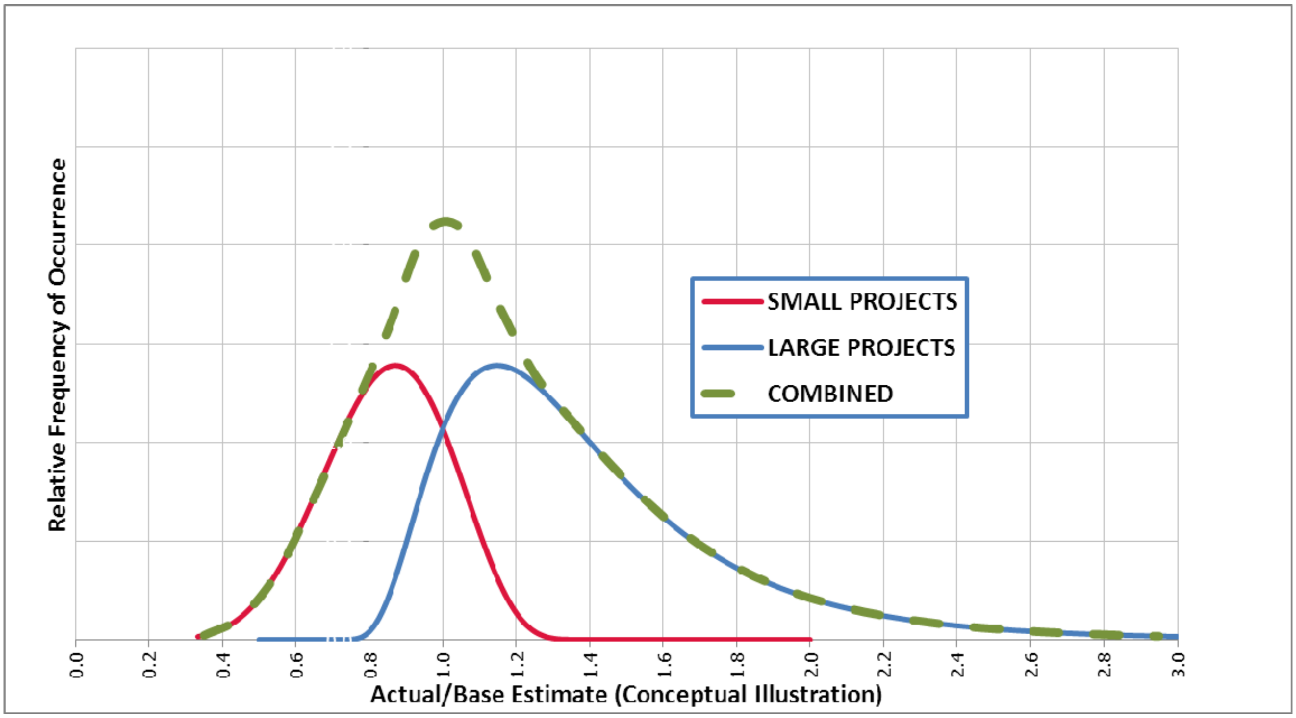

Figure 3 - Balancing Of Over and Underruns When All Project Sizes Are Studied Together

The Measurement Dilemma

Unfortunately, accuracy (i.e., actual/estimated cost) is often misused as a measure of *estimate "quality"* (as in "*a high quality estimate is an accurate estimate*") or estimating performance. This is inappropriate because, as discussed, the only way for an estimator to deliver a targeted accuracy for a given scope is to over-estimate the cost; risk and project performance are not in the estimator's control. Faced with overruns, estimators and the team tend to hide behind the excuses discussed previously. Accuracy should be used to measure the performance of the *risk management process* (not the estimating process) in conjunction with project historical data including causal information so we can improve our risk identification, analysis and quantification, and treatment. Tight accuracy may indicate wasted capital funds; accuracy measures must always be accompanied by measures of project control process discipline and project cost competitiveness (lower absolute costs) or cost bias.

AACE Recommended Practices (RPs)

There is an AACE RP that guides the selection and development of risk quantification and contingency estimating methods [2]. This RP provides "principles" that any method should align with including;

- start with identifying risk drivers;
- link risk drivers and cost/schedule outcomes; and
- employ empiricism.

Note that the previously discussed "line-item ranging" method is not explicitly in accordance with any of above principles.

Risks differ in how they impact project costs and therefore methods vary in how the risks are quantified. To cover the whole scope of risks, AACE has defined a risk breakdown [24] in respect to quantification methods that includes:

- **Project-Specific Risk**: risk affecting the specific project and plan;
- **Systemic Risk**: artifacts or inherent attributes of the system, enterprise or strategy; and,
- **Escalation Risk**: driven by economics (which regionally may involve politics).

Analogies for these risks suggested by others include: operational (project), strategic (enterprise), and contextual (global) risks respectively [42]. Methods that address these risk types can be integrated to generate a "universal" cost risk profile to support decision making. AACE also recommends that cost and schedule risk analysis be integrated.

For each risk type, there are AACE RPs for risk analysis methods that apply as follows ("how-to" descriptions for these methods are covered in the references):

- **Project-Specific Risk**:
  - 41R-08: Risk Analysis and Contingency Determination Using Range Estimating [12].
  - 44R-08: Risk Analysis and Contingency Determination Using Expected Value [9].
  - 57R-09: Integrated Cost and Schedule Risk Analysis Using Monte Carlo Simulation of a CPM Model [8].
  - 65R-11: Integrated Cost and Schedule Risk Analysis and Contingency Determination Using Expected Value [7].
- **Systemic Risk**:
  - 42R-08: Risk Analysis and Contingency Determination Using Parametric Estimating [10].
  - 43R-08: Risk Analysis and Contingency Determination Using Parametric Estimating – Example Models as Applied for the Process Industries [11].
- **Escalation Risk**:
  - 58R-10: Escalation Principles and Methods Using Indices [6].
  - 68R-11: Escalation Estimating Using Indices and Monte Carlo Simulation [5].

Regardless of the risk analysis methods used, the findings of this paper suggest that, at a minimum, you always test your p90 outcomes (the "high" scenario given to the business organization to test the robustness of their decision) against the empirical reality. If no other historical data is available, this paper provides actual examples to consider. If your p90 is 25% or less over the base estimate, ask why NO study ever showed less than about 40% for p90; what risks are you missing? what impacts have you underestimated? Finally, and most important, ask "how can one improve project practices and scope in consideration of the risk reality?"

Conclusion

As a student of cost engineering and the editor/lead author of AACE's Total Cost Management Framework process [26], one is dismayed by the extreme disconnect between our practices and the long-known reality as shown in Figure 1. There is an ongoing failure to effectively address the reality of project cost uncertainty and there is a lack of good historical data with causal information. This has led to a credibility crisis. It also raises an ethical question (if not a criminal one per Flyvbjerg); what does it mean if we understand reality but continue to use failed methods known to be contrary to experience to the potential detriment of our employers and clients? The AACE Canon of Ethics (item 2g) states: "*When, as a result of their studies, members believe a project(s) might not be successful…they should so advise their employer or client*" [1]. In respect to large process industry projects, readers of this paper can consider themselves so advised.

There are of course practitioners who do address the entire scope of cost risks (in AACE terms; systemic, project-specific and escalation), capture data and consider the empirical record [16,17,18,21,42, and 43]. However, in the author's experience, the application of robust practices is uncommon. At a minimum, teams should at least test their worst case analysis outcomes against the empirical reality. They should study this paper's references and their own enterprise's historical experience (watching out for the small versus large project behavioral dichotomy). And, they should then seek to improve their practices to improve on past outcomes. The author does not agree with Dr. Flyvbjerg's approach to using empiricism (i.e., "reference class forecasting" [19]) which implies that biases are so intractable that we are doomed to repeat the past.

The paper also points out that companies should not use accuracy as a cost estimating quality measure; it is a risk management and project control process quality measure. Tight accuracy is often an indicator of wasted capital; measures of project control process discipline and project cost competitiveness must accompany accuracy measures.

In summary, this paper references and summarizes over 50 years of empirical cost estimate accuracy research on large projects in the process industries. It shows how this reality compares (or does not compare) to what we say and do. Recommended risk analysis methods have been highlighted. Failed methods are exposed. It is hoped that the facts, observations and opinions brought together here will serve as a valuable reference on the topic of cost accuracy and uncertainty so that we can better speak the truth among ourselves and with management. The path to more realistic uncertainty forecasts, better contingency estimates and more profitable investments is clear and documented by AACE International.

REFERENCES

1. AACE International, *Canon of Ethics*. AACE International, Morgantown, WV, 2012.
2. AACE International, "*Contingency Estimating – General Principles*", **Recommended Practice 40R-08**, AACE International, Morgantown, WV, (latest revision).
3. AACE International, "*Cost Engineering Terminology*", **Recommended Practice 10S-90**, AACE International, Morgantown, WV, (latest revision).
4. AACE International, "*Cost Estimate Classification System – As Applied in Engineering, Procurement, and Construction for the Process Industries,*" **Recommended Practice 18R-97**, AACE International, Morgantown, WV, (latest revision).
5. AACE International, "*Escalation Estimating Using Indices and Monte Carlo Simulation,*" **Recommended Practice (draft) 68R-11**, AACE International, Morgantown, WV, (latest revision).
6. AACE International, "*Escalation Principles and Methods Using Indices,*" **Recommended Practice 58R-10**, AACE International, Morgantown, WV, (latest revision).
7. AACE International, "*Integrated Cost and Schedule Risk Analysis and Contingency Determination Using Expected Value,*" **Recommended Practice (draft) 65R-11**, AACE International, Morgantown, WV, (latest revision).
8. AACE International, "*Integrated Cost and Schedule Risk Analysis Using Monte Carlo Simulation of a CPM Model,*" **Recommended Practice 57R-09**, AACE International, Morgantown, WV, (latest revision).
9. AACE International, "*Risk Analysis and Contingency Determination Using Expected Value,*" **Recommended Practice 44R-08**, (latest revision).
10. AACE International, "*Risk Analysis and Contingency Determination Using Parametric Estimating,*" **Recommended Practice 42R-08**, AACE International, Morgantown, WV, (latest revision).
11. AACE International, "*Risk Analysis and Contingency Determination Using Parametric Estimating – Example Models as Applied for the Process Industries,*" **Recommended Practice 43R-08**, AACE International, Morgantown, WV, (latest revision).
12. AACE International, *"Risk Analysis and Contingency Determination Using Range Estimating,"* **Recommended Practice 41R-08**, AACE International, Morgantown, WV, (latest revision).
13. Arena, Mark, Robert S. Leonard, Sheila E. Murray, and Obaid Younossi, "*Historical Cost Growth of Completed Weapon System Programs*" TR-343, The RAND Corporation, 2006.
14. Bertisen, Jasper and Davis, Graham A., *'Bias and Error in Mine Project Capital Cost Estimation,'* **The Engineering Economist**, 53: 2, 118 — 139, 2008.
15. Biery, Frederick, *"Improving Mineral Project Performance,"* Canadian Institute of Mining, Metallurgy and Petroleum Conference, May 2009.
16. Butts, Glenn and Kent Linton, "*The Joint Confidence Level Paradox; A History of Denial,*" 2009 NASA Cost Estimating Symposium, NASA, April 2009.
17. Curran, Kevin M., "*Value-Based Risk Management (VBRM),*" **Cost Engineering**, AACE International, AACE International, Morgantown, WV, February 2006.
18. Curran, Kevin M and Michael Curran, "*Handling the Truth in Risk Management,*" **2010 AACE International Transactions**, AACE International, Morgantown, WV, 2010.
19. Flyvbjerg, Bent, Mette Skamris Holm and Søren Buhl, "*Underestimating Costs in Public Works Projects; Error or Lie*?" **APA Journal**, 68:3, 279-295, 2002.

20. Flyvbjerg, Bent and Alexander Budzier, "*Why Your IT Project May Be Riskier Than You Think,*" **Harvard Business Review**, September 2011.
21. Gough, Martin and Peter Maidment, "*The Stability Model Method of Risk Management and Early Prediction of Project Performance,*" **The Revay Report**; Revay and Associated Limited, February 2006.
22. Hackney, John W. (Kenneth K. Humphreys, Editor), **Control & Management of Capital Projects**, Chapter 18, AACE International, Morgantown, WV, 2002.
23. Harbuck, Robert H, "*Are Accurate Estimates Achievable During the Planning of Transportation Projects?,*" **2007 AACE International Transactions**, AACE International, Morgantown, WV, 2007.
24. Hollmann, John K., "*The Monte-Carlo Challenge: A Better Approach,*" **2007 AACE International Transactions**, AACE International, Morgantown, WV, (latest revision).
25. Hollmann, John K., "*Recommended Practices for Risk Analysis and Cost Contingency Estimating,*" **2009 AACE International Transactions**, AACE International, Morgantown, WV, 2009.
26. Hollmann, John K (editor), **Total Cost Management Framework**, AACE International, Morgantown, WV, 2006.
27. Juntima, Gob and Scott E. Burroughs, "*Exploring Techniques for Contingency Setting,*" **2004 AACE International Transactions**, AACE International, Morgantown, WV, 2004.
28. Kosich, Dorothy, "*Vast majority of mining projects experience steeper cost over-runs,*" MineWeb, 30 Nov 2011.
29. Kulkarni, Phyllis, "*Stop Punishing the Overruns,*" **InSites**, IPA, Inc., Sept 7, 2011.
30. Linder, Stefan, "*Fifty years of Research on Accuracy of Capital Expenditure Project Estimates: A Review of the Findings and their Validity,*" Center for Research in Controllership and Management, WHU - Otto Beisheim Graduate School of Management, Vallendar, Germany. April 2005.
31. Lundberg, Mattias, Anchalee Jenpanitsub and Roger Pyddoke, "*Cost Overruns in Swedish Transport Projects,*" CTS Working Paper 2011:11, Centre for Transport Studies, Stockholm Sweden, 2011.
32. Merrow, Edward W., **Industrial Megaprojects**, John Wiley & Sons, Inc., New York, NY, 2011.
33. Merrow, Edward, W., "*Why Large Projects Fail More Often; Megaproject Failures: Understanding the Effects of Size,*" Presentation to the Joint Meeting of the AACEI National Capital Section and American Society of Mechanical Engineers Section, April 20, 2011.
34. Merrow, Edward W. and Brett R Schroeder, "*Understanding the Costs and Schedule of Hydroelectric Projects,*" **1991 AACE International Transactions**, AACE International, Morgantown, WV, 1991.
35. Merrow, Edward W., Kenneth E. Phillips and Christopher W. Myers, "*Understanding Cost Growth and Performance Shortfalls in Pioneer Process Plants,*" R-2569-DOE, p48, The RAND Corporation, 1981.
36. Merrow, Edward W., Lorraine McDonnell and R. Yilmaz Arguden, "*Understanding the Outcomes of Megaprojects,*" R-3560-PSSP, The RAND Corporation, 1988.
37. Mitchell, Paul, "*Achieving Major Capital Project Effectiveness and Corporate Performance,*" Effective Capital Project Execution: Mining and Metals, EYGH Limited, 2011.

38. Noort, D.J., and C. Adams, “*Effective Mining Project Management Systems*” International Mine Management Conference, Australasian Institute of Mining and Metallurgy, October 2006.
39. Oil & Gas Journal Online Research Center, *US Pipeline Study 2009; Actual vs. Estimate* (database for fee), Penwell Corporation, Houston TX, 2009.
40. Rath, Michael, Robert Spittle, Paul Osman and Nicholas Tawa. “*Major Water Capital Projects: Critical Factors For Improving Outcomes,*” Deloitte Touche Tohmatsu, March 2010.
41. Reddall, Braden (reporting), “*Cost Overruns Common on Oil Projects –Schlumberger,*” Thomson Reuters, **Energy and Oil News**, 2 June, 2011.
42. Rolstadås, Asbjørn, Per Willy Hetland, George Farage Jergeas and Richard E. Westney, *Risk Navigation Strategies for Major Capital Projects: Beyond the Myth of Predictability*, Springer- Verlag London Limited, 2011.
43. Schroeder, Brett and Jan A. Jackson, “*Why Traditional Risk Management Fails in the Oil and Gas Sector: Empirical Front-Line Evidence and Effective Solutions,*” **2007 AACE International Transactions**, AACE International, Morgantown, WV, 2007.
44. Schulyer, John R., “*Investment Decision Making*”, Chapter 14 of **The Engineer’s Cost Handbook**, Editor Richard E. Westney, Marcel Dekker, Inc., New York NY, 1997.
45. Stephenson, H. Lance, “Cost Engineering Maturity Model (CEMM),” **2011 AACE International Transactions**, AACE International, Morgantown, WV, 2011.
46. Thomas, Steve, “*Project Development Costs-Estimates Vs. Reality,*” **Pincock Perspectives**, November 2000.

John K. Hollmann
Validation Estimating, LLC
jhollmann@validest.com