SYM-AM-19-084

PROCEEDINGS
OF THE
SIXTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

THURSDAY SESSIONS
VOLUME II

Acquisition Research:
Creating Synergy for Informed Change

May 8–9, 2019

Published: April 30, 2019

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

Eliciting Expert Opinion in Acquisition Cost and Schedule Estimating

Major Gregory E. Brown, USAF—serves as an Instructor of Financial Management and Data Science for the Air Force Institute of Technology (AFIT) School of Systems and Logistics. As an instructor, he develops and teaches continuing education courses on acquisition cost analysis, statistics, and data analytics. Prior to becoming an AFIT instructor, Maj Brown worked in multiple cost-analysis positions across the Air Force Lifecycle Management Center, most recently serving as the Cost Analysis Chief for the Fighters and Bombers Directorate. He holds a BA in economics and BS in business-finance from Colorado State University, a graduate certificate in applied statistics from Penn State University, and an MS in cost analysis from the Air Force Institute of Technology. [gregory.brown@AFIT.edu]

Abstract

Despite the emphasis on data and analytics in acquisition cost and schedule estimating, many estimating situations still require eliciting expert opinion from a subject matter expert. This is problematic, as a 2007 RAND report concludes that there is no standard model for seeking expert input for acquisition estimates. Per the report, the DoD's "elicitation methodologies are largely ad hoc, in that they are seldom based on or derived from references to the elicitation literature" (Galway, 2007). In this paper, a popular and commonly cited elicitation model—the Stanford Research Institute (SRI) elicitation model—is presented and adapted to the cost and schedule estimating process. It is posited that the consistent application of a formal model would reduce expert biases and improve the acquisition community's risk and uncertainty analyses. This paper also provides the results of an original meta-analysis of published experiments that examine expert elicitation for business and engineering problems. The data reveals that experts are overconfident and struggle to identify the true range of outcomes for both business and engineering problems. However, using a structured elicitation model, training the expert prior to the elicitation, and providing the expert with feedback are shown to decrease expert overconfidence.

Introduction

Even with ongoing efforts to improve acquisition databases, sometimes the historical data we need for cost and schedule estimates is simply unavailable. In other instances, historical data is available but requires adjustment to account for radical changes in a technology or manufacturing process (Kitchenham et al., 2002). In these instances, analysts may turn to the opinion of experts, using an interview process known as elicitation. Through elicitation, it is possible to tap into the knowledge and experience of engineers, logisticians, and programmers. Utilizing expert elicitation carries risks, however. Without proper guidance, experts may fall victim to cognitive biases, resulting in predictions that are both inaccurate and overconfident.

Academic research has long since recognized the problem of expert biases and began designing elicitation protocols to guard against them in the 1960s. Experiments have shown that by following a structured elicitation framework and providing feedback, the quality of elicitation may be improved. Regrettably, no standard elicitation model currently exists across the DoD cost and schedule estimating community. Instead, RAND notes that the DoD's "elicitation methodologies are largely ad hoc, in that they are seldom based on or derived from references to the elicitation literature." Simply put—analysts are learning to conduct elicitation by trial and error, rather than being guided by a structured model. To compound this problem, RAND notes that elicitations are poorly documented within DoD cost estimates, resulting in elicited estimates that cannot be reviewed and reexamined after the initial estimate is completed (Galway, 2007).

Based on these revelations, it is evident that we—the cost analysis community—need a change in attitude towards elicitation and elicitation training. We wouldn't expect an analyst to construct a parametric model without first receiving education on linear regression methods, so why is the expectation for elicitation any different? We must stop viewing elicitation as an ad hoc art, and instead adopt a more structured, scientific process. Rather than novice analysts learning elicitation through improvisation, we should educate new analysts using those methods that are validated by decades of research from the fields of psychology, behavioral economics, decision analysis, and Bayesian statistics.

To initiate this change, this author proposes a five-step model first introduced by decision analysis researchers at the Stanford Research Institute (SRI). The model is provided within this research paper in a subsequent section. As a caution, it would be unwise to jump directly to the elicitation model without first understanding the fundamentals that shape the model. Thus, this research paper is divided into five sequential sections, with each section building upon knowledge from the prior. To begin, this paper provides a definition for expert elicitation and background on the advantages and disadvantages of elicitation. The next section describes common expert biases so that the cost analyst may better learn to recognize them. Then this author examines whether the Joint Agency Cost Schedule Risk and Uncertainty Handbook (JA CSRUH) heuristic of treating expert intervals as encompassing only 70% of uncertainty is accurate and defendable. Moreover, strategies are provided for controlling expert overconfidence. Next, the paper outlines the SRI elicitation methodology that will serve to further reduce expert bias while also promoting improved documentation of elicitations. The next section introduces methods for adapting the SRI model to elicitations with multiple experts. Finally, the last section provides a summary and recommendations for future elicitation research and change efforts.

When taken in aggregate, it is the author's hope that the research cited within this paper will help promote a change in attitudes toward expert elicitation in the community, so that expert predictions are treated in a similar manner to traditional data and statistical models. Rather than blindly accepting or rejecting expert predictions, analysts should instead adopt a more structured approach that will allow the expert's opinion to be afforded the same level of review and validation that we would demand for any other cost or schedule model.

Background

Defining "Expert" and "Elicitation"

So, what is an expert? An *expert* is defined as an individual who has mastered the specialized skills or bodies of knowledge relevant to a particular subject. While the expert doesn't know everything about a subject, it is expected that his or her prediction on a problem is more likely to be correct than that of the public at large. However, being an expert in one field does not make an individual better qualified in unrelated fields. Research finds that experts—even at the PhD level—are no better at predicting outcomes in fields unrelated to their expertise than the general population (McKenzie et al., 2008; Nichols, 2017). Thus, we would not expect a chemical or nuclear engineer to be particularly skilled at estimating lines of code if he or she had never worked in software engineering. Finding the right expert for a given estimate is paramount.

Conversely, *elicit* means "to call forth or draw out (as information or a response)" ("Elicit," 2017). Thus, in an expert elicitation, the cost estimator is asking the expert call forth information from his or her area of expertise. The term *elicit* and *elicitation* are preferred to synonyms such as *interview*, as *elicit* is the most commonly favored term in academic research, beginning with usage in early Bayesian statistics research (e.g., Winkler, 1967) as well as the earliest RAND Delphi Method study (i.e., Brown, 1968).

Relevant Cost Estimating Methodologies

Many cost estimating methodologies are cited across DoD literature: extrapolation from actuals, parametric, analogy, bottom-up engineering, and expert opinion. In this author's experience, at least three of these methodologies will typically require elicitation of an expert. In instances where the cost estimator has no historical data to leverage, the estimator may directly elicit an *expert opinion* from the expert. Alternatively, when only a few historical data points are available—insufficient for a parametric model—the estimator may seek the expert's help in identifying the best *analogy*, to which the expert may subsequently apply a scaling or complexity factor (AFCAH, 2008). *Parametric* models require elicitation as well, as the inputs to the parametric model are seldom known with certainty at the beginning of a project. For example, when employing a parametric software cost estimating model, variable inputs such as source lines of code (SLOC) and code re-use are typically estimated by a technical expert (Jorgenson, 2007). Because these inputs are uncertain during the early phases of a program, applying relevant elicitation protocols can improve the accuracy of the expert's elicited inputs to the model.

Why Elicit an Expert's Opinion?

The DoD recently introduced the Cost Assessment Data Enterprise (CADE), an online database intended to significantly increase the cost analyst's access to cost, schedule, and technical acquisition data. As the CADE platform matures and access to data improves, less time will be spent gathering data for cost and schedule estimates, allowing for the adoption of more innovative and accurate modeling techniques (Watern, 2016). Given the availability of CADE, is the elicitation of experts still relevant to cost estimating?

Elicitation will likely remain relevant for several reasons. Firstly, CADE is focused on collecting data for Acquisition Category (ACAT) Level 1 programs, currently defined by 10 U.S.C. 2430 as having a development budget greater than $480 million or procurement budget greater than $2.79 billion. As a result, smaller programs are not well represented in CADE, and when they are, they will typically have less collected data to leverage for future estimates. Thus, constructing a parametric model for a minor systems modification may not always be feasible. Secondly, changes in technology mean that available historical data may not always be relevant to the current estimating task and may require adjustment by the expert (Kitchenham et al., 2002). For example, a parametric schedule model based on software using a waterfall strategy may require recalibration by an expert before it is used to estimate a project with an agile strategy. Thirdly, even when sufficient analogous data is available to establish a parametric model, meta-analysis suggests that in certain scenarios, experts are just as accurate as parametric models in estimating outcomes. These scenarios are explored next.

Accuracy of Experts Compared to Models

Do models always outperform expert predictions within cost estimates? Jorgenson (2007) reviewed 16 software cost estimating studies that directly compare the accuracy of formal parametric models with that of experts. After aggregating the studies, Jorgenson found that the average accuracy of the expert-derived estimates was higher than for the model-provided estimates for 10 of the 16 studies. Jorgenson's finding contradicts the belief—held by some—that parametric models will always outperform the expert in the context of cost estimating.

When and why might one method outperform the other? Sanders and Ritzman (1991) theorized that models are superior for prediction when using data which is “stable.” As an example from the medical field, a meta-analysis of 136 individual medical studies find that statistical models are more likely to correctly diagnose a medical condition than medical experts (Grove et al., 2000). One particularly notable study is Nashef et al. (1999), who proposed the European System for Cardiac Operative Risk Evaluation (EuroSCORE) model, in which age, gender, pulmonary disease status, and a multitude of cardiac lab values are able to more accurately predict the likelihood of post-operative death or complication than an experienced heart surgeon. In this setting, the data is stable, in that the human body is not significantly changing or evolving. The same predictive relationships built on the initial sample of patient demographics and lab values are expected to remain valid over time. Almost two decades later, the EuroSCORE model remains in use in the United States, Europe, and Japan, and the model continues to be validated by using the populations of different countries (e.g., Shen et al., 2018).

Conversely, Sanders and Ritzman (1991) theorized that experts are superior at prediction in unstable, changing conditions, as one might face when estimating the cost for a new technology with changing cost drivers. In describing the results of his software meta-analysis, Jorgenson (2007) concluded that in research and development, “the technology, the types of software produced, and the production methods, change frequently” (p. 460). This lack of stability, combined with small data sets, makes it difficult to build an accurate statistical model that is not overfitted to the historical data. Unlike the model, the expert is not limited to considering only a few variables, but instead may utilize decades of cumulative experience as well as all available context about the program being estimated. Thus, in some cost estimating scenarios, the expert may have the advantage “in that they typically possess more information and are more flexible in how the information (or lack of information) is processed.”

Given that neither parametric models nor experts are always the best, some researchers suggest employing an “ensemble” approach, whereby output from the parametric model and output from the expert are combined (i.e., averaged) to reduce estimating error. Over time, theory states that an ensembled estimate will have greater accuracy than either the parametric model or expert alone, assuming that both estimates are unbiased and capture different information. As evidence that ensemble models can be successfully employed in cost estimating, Li et al. (2008) tested the application of Optimal Linear Combining (OLC) to software cost estimating, with the estimates from a parametric software model and expert each weighted based on their expected accuracy. On average, the OLC ensemble increases the accuracy of software cost estimates when compared to the parametric model or expert alone.

Problems with Utilizing Elicitation in Cost Estimating

Despite evidence from Jorgenson (2007) that experts can be as accurate—or more accurate—than models in cost estimating, many decision-makers remain hesitant to make decisions using elicited estimates without traditional data. Why is this?

- Recognizing that experts are prone to both motivational and cognitive biases, the decision-maker may view all elicitations as biased or inaccurate.
- Due to overconfidence, experts have historically been overly precise when estimating prediction intervals, leading the decision-maker to accept more uncertainty and risk than he or she was briefed. However, it is currently not

known how overconfident experts are (i.e., what percentage of uncertainty is actually captured by the expert assisting with DoD cost estimates?) Based on a 1976 study, the JA CSRUH recommends treating the expert's input as only capturing 70% of outcomes.

- No standard elicitation methodology exists within the DoD. As a result, RAND observes that elicitations are often poorly documented within cost and schedule estimates, and it is difficult for more senior reviewers or cost agencies to validate the inputs provided by the expert (Galway, 2007). For decision-makers, the credibility of an elicitation is only as good as the documentation and justification surrounding the expert's estimate.

However, each of these potential problems may be overcome by the research presented in this paper. By following a consistent protocol—such as the SRI elicitation model—and documenting the rationale behind the elicited estimate, it is possible to regain the trust of the decision-maker.

Expert Biases

In the previous section, the definition of expert elicitation was given, and evidence was provided that expert opinion can be as accurate as data-driven models. However, employing expert opinion can be problematic, as experts can be affected by biases—both intentional (i.e., motivational bias) and unintentional (e.g., optimism). These biases may drive the expert to be less accurate within a given estimate. Moreover, biases may cause the expert to consistently underestimate or overestimate a requirement across multiple estimates, resulting in entire product portfolios that are underfunded or overfunded. Although not an exhaustive list, six cognitive biases commonly encountered when eliciting an expert's opinion are summarized.

Motivational Bias

Motivational bias is driven by the expert's desire to influence the decision to his or her own benefit. As notional examples of motivational bias, a program manager may benefit from understating the cost of a new effort in order to secure initial funding or milestone approval. Conversely, an engineer may benefit from overstating the costs for a proposed technical solution that he or she does not support.

Optimism

Individuals assess that they are better than others and less likely than others to experience negative events or outcomes. These individuals will focus on what can "go right" in a project, while believing that nothing could "go wrong." Often, this is driven by a false sense of control over events. As a result, experts who succumb to optimism bias will consistently underestimate task completion times and costs, even when presented with the information that the vast majority of similar tasks have run over both schedule and budget (Flyvbjerg, 2011).

Availability

Availability says that individuals are more likely to recall information that is either recent or made the most significant impression on that individual, while ignoring less impressionable information. As a consequence, experts may base their elicitation on the information that is easiest to recall, rather than taking into account the full range of observations and experience.

Anchoring

Anchoring states that individuals will often use readily available information (e.g., an analogous project) as the initial basis for an estimate, before making further adjustments to account for differences (Spetzler & Stael von Holstein, 1975). However, research experiments have shown that on average, individuals tend to make insufficient adjustments to the initial basis, resulting in the response being "anchored" to the basis (Kahneman & Tversky, 1974). As a result, when using an analogy as basis for an estimate, the expert may fail to fully adjust for the change in complexity between the historical analogy and the new effort.

Unstated Assumptions

The unstated assumptions mode of judgment says that individuals will naturally condition their estimate on unstated assumptions. As a consequence, the elicited distribution will often ignore events which the expert believes he or she is not responsible for considering.

> For example, a cost estimate might be made with the implicit assumptions that the base design will not change. However, the same person, when questioned about the likelihood of the base design's changing, might think such a possibility very likely. (Boyd & Regulinski, 1979)

While assumptions are necessary for a cost estimate, it is important that these assumptions are clearly verbalized by the expert, documented by the cost analyst, and later briefed to the decision-maker.

Overconfidence

Overconfidence states that individuals will believe their point estimate to be a better and more reliable estimate than it really is. As a consequence, the expert will generally understate the uncertainty about a quantity, resulting in a prediction interval that is smaller than it should be.

Eliciting Uncertainty From Experts

Background

When using parametric-based cost estimating relationships, uncertainty about a prediction is calculated in the form of a prediction interval. Assuming the assumptions necessary for linear regression are met (e.g., equal variance of errors and normal distribution of errors at each value of the predictor), there is generally no need to adjust the prediction interval, as it is unbiased. For example, given that a future observation comes from the same population as the sample used to build the parametric model, a 95% prediction interval is expected to contain the future observation 95% of the time.

However, when relying on expert opinion as the basis for an estimate, the analyst faces the added challenge of generating a prediction interval with the help of the expert. Due to overconfidence, the expert's elicited interval will generally be smaller than the interval representing the true state of the expert's knowledge. Overconfidence can be lessened using techniques that drive the expert to consider the full range of outcomes, but experiments show that these techniques will not completely resolve overconfidence. Moreover, due to "unknown unknowns," it is often not feasible for the expert to imagine all possible outcomes. It is therefore necessary to account for additional uncertainty when modeling inputs elicited from an expert.

The JA CSRUH recognizes this problem and recommends treating the expert's interval as encompassing only 70% of the range of uncertainty. The handbook's 70% heuristic is derived from Capen (1974), who concluded that experts rarely account for more than 60%, and never account for more than 70% of the possible range of outcomes. Capen arrived at his conclusion by surveying 1,000 petroleum engineers who were asked to estimate prediction intervals for 10 generic, encyclopedia-type questions, such as "What is the area of Canada in square miles?"

Based on this author's experience, however, some program managers and decision-makers may question the validity of a heuristic which requires the application of additional uncertainty, increasing the cost or schedule of a program. In turn, the heuristic may be difficult for the analyst to defend, due to the research's age (over 40 years old) and the reality that Capen was asking the engineers to estimate intervals for encyclopedia problems, and not problems directly related to their area of engineering expertise. If the petroleum engineers had instead been asked to generate prediction intervals related to petroleum engineering, would they show less overconfidence and provide more realistic intervals? To help resolve this question and provide the analyst with relevant research to cite when defending their estimate, a meta-analysis is conducted.

Meta-Analysis of Expert-Elicited Intervals

To re-validate Capen's earlier findings, this author searches for additional research that utilizes surveys to assess the accuracy of expert prediction intervals. To best align with the problem types typically encountered in acquisition cost estimating, the search query is designed to capture studies in which either business or engineering experts provide intervals directly related to their field or industry. Studies involving undergraduate students are excluded, but studies involving graduate students (e.g., Goldenson & Stoddard, 2013) are included if it is documented that the graduate students have prior industry experience in their field. To increase the meta-analysis's relevance to cost and schedule estimating, only studies involving the prediction of continuous ranges are included; studies in which experts are asked to estimate probabilities of discrete events (e.g., True or False) are excluded.

After applying the inclusion and exclusion criteria, a total of five studies encompassing 17 total surveys and 21,000 individual predictions are identified. The following are descriptions of the studies:

- Russo and Schoemaker (1992) asked corporate business managers to provide prediction intervals for technical questions related to the managers' own firm and industry (11 aggregated surveys; 7,660 total predictions).
- McKenzie et al. (2008) asked information technology (IT) professionals to provide prediction intervals for IT industry questions (one aggregated survey; 1,720 total predictions).
- Ben-David et al. (2010) asked Chief Financial Officers of major companies to provide prediction intervals for S&P 500 market returns for the following year; the survey is repeated annually over a nine-year period (one aggregated survey; 11,600 total predictions).
- Goldenson and Stoddard (2013) asked graduate students with industry experience to provide prediction intervals for source lines of code (SLOC) and effort in person-years for previously completed software projects based on a description of the software requirements, team size and programming language (three aggregated surveys; 290 total predictions).

- Bar-Yosef and Venezia (2014) asked experienced brokerage analysts to use accounting data for a company to provide prediction intervals for net income, earnings per-share, and share price (one aggregated survey; 30 total predictions).

The results are plotted in Figure 1, and weighted averages for each available increment of confidence are summarized in Table 1. The full results are provided in Appendix A. In viewing Figure 1, the requested prediction interval for the survey is plotted on the x-axis, while the percentage of experts whose response contained the true answer is plotted on the y-axis. The diagonal dashed line represents the calibration line, where a well-calibrated group of experts should fall. For example, if a 50% prediction interval is requested, then approximately 50% of the experts should provide an interval that contained the true response. Additionally, a simple linear regression model is fit to the data and represented by a black line, while a weighted regression model—with survey sample size assigned as the weight—is fit and plotted as a solid blue line. The models are designed to measure whether the confidence level requested from the expert impacts the percentage of correct responses, and whether the confidence level requested from the expert impacts the degree of observed overconfidence.

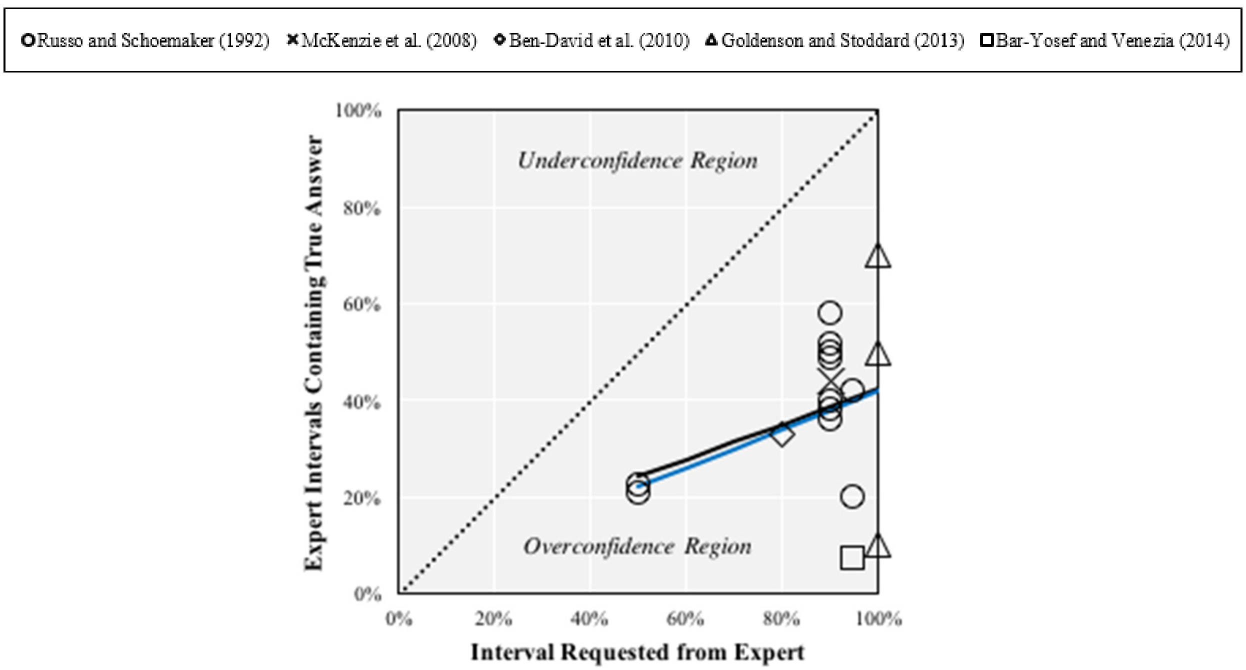

Figure 1. Meta-Analysis Results

After assessing Figure 1, it is observed that for each of the 17 aggregated surveys, the percentage of experts with correct intervals falls below the calibration line. This indicates that experts are overconfident, on average. Moreover, it is observed that at no time do more than 70% of experts in a given survey predict the true response, even when asked to provide a 100% interval, as in Goldenson and Stoddard (2013). Thus, the meta-analysis validates Capen's finding that experts never identify more than 70% of the possible range of outcomes.

Digging deeper, both the simple linear regression and weighted regression models are examined, with the goal of determining if a linear relationship exists between the interval requested and the percentage of experts with the correct answer. Although a positive slope is calculated for the simple linear regression model, it is not statistically significant (p-value = 0.22). The weighted regression model is statistically significant (p-value = 0.04); however, upon closer inspection, this is the result of the model overfitting (i.e., passing directly through) a few influential data points with relatively larger sample sizes. Due to the limited number of surveys, transformations and non-linear methods are not considered. Thus, when considering results from across studies, no regression-based conclusions are drawn about the relationship between the requested expert interval and actual expert interval.

Table 1. Weighted Average of Correct Expert Estimates at Confidence-Level Intervals

Interval Confidence-level Requested from Expert	50%	80%	90%	95%	100%
Expert Intervals Containing Truth (Weighted Avg.)	21%	33%	45%	31%	36%
Aggregated Sample Size	1,600	11,600	5,200	2,610	290
Total Number of Surveys	2	1	8	3	3

However, in viewing the data more subjectively, the experts do appear to drift further away from the calibration line as the confidence-level of the requested interval increases. As indicated in Table 1, at higher confidence levels, a 30% adjustment would be insufficient to capture the true range of uncertainty in all but one case. This raises the question: should the analyst instead be adding greater than 30% uncertainty to the expert's elicited range? This author's experience indicates that an even greater adjustment would face resistance in the acquisition community. Such an extreme adjustment to the expert's cost or schedule estimate may not be palatable to decision-makers and risks offending the expert who provided the elicited input. Thus, strategies for naturally reducing the expert's overconfidence are explored next.

Strategies for Decreasing Overconfidence—Feedback and Formatting

The meta-analysis result raises the concern that assuming that the expert captures 70% of the true responses is itself optimistic. This phenomenon—when observed in other studies—has led researchers to conclude that most individuals have a poor understanding of statistics and prediction intervals (Kahneman, 2011). However, research shows that it is possible to improve the calibration of expert's prediction intervals by focusing on two areas: feedback and elicitation formatting.

As an example of feedback, this author examines Goldenson and Stoddard (2013), a software estimating study previously cited in the meta-analysis. The study consisted of three rounds. During the first round, only 10% of experts provided a prediction interval containing the true requirement, despite being asked to provide a 100% prediction interval. Following round one, feedback was provided to experts that they were overconfident. In turn, 50% of experts identified the true response in round two, and after post-round two feedback, 70% identified the true response in round 3. Thus, providing ongoing feedback to experts does appear to significantly improve the calibration of expert's prediction intervals. Simply making the expert aware that they suffer from overconfidence results in more accurate prediction intervals.

Other elicitation studies show that focusing on the formatting of the elicitation questions can decrease overconfidence. In experiments involving non-experts, the following strategies have been shown to decrease overconfidence:

- Ask for the high and low outcomes prior to asking for the most likely. This format has been shown to decrease overconfidence related to the anchoring bias (e.g., Soll & Klayman, 2004; Speirs-Bridge et al., 2010).
- Allow the expert to select the confidence level of the interval they would like to provide (e.g., 70%). This format decreases overconfidence compared to confidence levels that are pre-specified by the analyst (Teigen & Jorgenson, 2005).
- When appropriate, allow experts to provide intervals based on lower confidence levels. For example, individual experts providing answers corresponding to lower-confidence prediction intervals (e.g., 50% confidence) show less overconfidence than those providing higher-confidence prediction intervals (e.g., 90% confident) (Teigen & Jorgenson, 2005). Using standard formulas from the JA CSRUH's Table 2-8, a lower-confidence interval may be adjusted outward by the analyst to capture 100% confidence.
- For experts who struggle to conceptualize the prediction interval concept, manually walk the expert through the creation of the prediction interval. For example, simplify the prediction interval concept by asking the expert "could the requirement exceed 1,000 hours?" or "what is the probability that the requirement exceeds 1,000 hours?" (Teigen & Jorgenson, 2005).
- Finally, after recording the initial prediction interval, verify the expert's answer by asking the expert to consider why they may be wrong. For example, ask the expert to consider that the true requirement is greater than the upper bound of the prediction interval. Ask: what are a few reasons this could be? What assumptions or considerations may be wrong? Given these erroneous assumptions, was the initial estimate too low? Lastly, ask the expert if they wish to revise the upper bound. Herzog and Hertwig (2009) followed a similar line of questioning in their experiments and discovered that simply questioning the individual's initial conclusion prompts the individual to consider knowledge that was previously overlooked or assumed to be true when constructing the prediction interval.

Elicitation Model for a Single Expert

Many unique interview models have been proposed for gathering expert opinion. However, this author elects to utilize the *Stanford Research Institute (SRI) Elicitation Process* model. The model originated with Spetzler and Stael von Holstein (1975), decision analysis researchers from Stanford University. The SRI model is cited in numerous subsequent research efforts and is regarded by Morgan et al. (1990) as the most popular and influential elicitation model. As presented in Figure 2, the SRI model consists of five sequential phases: motivating, structuring, conditioning, encoding, and verifying.

In viewing Figure 2, it should be stressed that Spetzler and Stael von Holstein did not consider documentation as a separate step. Instead, documentation should be a continual process that takes place throughout each phase of the elicitation model. When writing documentation, the analyst should always strive to communicate the assumptions, rationale, and analogies used to estimate an outcome. This will serve two purposes. Firstly, this will help the analyst better explain the estimate to decision-makers in the event that the expert is not present. Secondly, it is rare that the same expert will be available for consultation throughout a program's life-cycle. Thus, recording the reasoning for the expert's estimate will be useful if a different expert is assigned to the program in the future and the estimate is revisited.

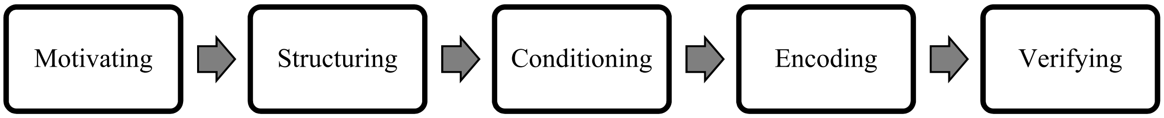

Figure 2. SRI Elicitation Model
(Spetzler & Stael von Holstein, 1975)

Phase 1: Motivating

Unlike data or models, which are at the control of the cost analyst, elicitation requires human interaction. Thus, the motivating stage is intended to introduce the expert to the purpose of the elicitation and establish rapport (Spetzler & Stael von Holstein, 1975, p. 352). Although this phase may seem superfluous, it should not be disregarded. Galway (2007) noted that in the DoD, elicitations are often rushed due to the time constraints and shortages of available experts. Galway's assessment matches with this author's own experience. For most experts, assisting with a cost or schedule estimate is a secondary duty, which takes them away from their primary duty. To motivate the expert and generate "buy in," it is imperative that the expert is made aware of the purpose of their inputs early in the process. Whether the end goal is a major milestone brief or the budgeting of future funding, how will the expert's input help the integrated product team succeed?

This author asserts a secondary focus of the motivating stage should be the education of the analyst, with a focus on achieving a basic technical understanding of the requirement to be estimated, thereby limiting the risk of *hypocognition*. Wu and Dunning (2017) wrote that hypocognition exists when one operates outside his or her conceptual landscape. Hypocognition is problematic as it can limit the ability of two individuals to exchange information. It is therefore imperative that the analyst makes an effort to develop a basic understanding of the requirement or technology to be estimated, as well as its associated terminology, as it will later direct the course of the conversation between the analyst and expert. Without knowing said terminology, the individual will have difficulty receiving and communicating the ideas advanced by the expert during the elicitation process. Moreover, individuals cannot use concepts they do not have or understand to explain phenomena (Levy, 1973). During the briefing stage of the estimate, the analyst—if not conceptually familiar with the requirement or technology being estimated—will risk misrepresenting or distorting the basis of the expert's elicited estimate. Thus, the analyst should ensure they have a working knowledge of the requirement or technology prior to entering the later phases of the elicitation.

Phase 2: Structuring

The purpose of the structuring phase is to define the uncertain quantity (or quantities) that requires expert input. If necessary, the "structure should be expanded as necessary so that the subject does not have to model the problem further before making each judgement" (Spetzler & Stael von Holstein, 1975, p. 353). The typical human can only hold about seven separate pieces of distinct information in their working memory at a time (Miller, 1956). Thus, by simplifying the problem into components or subcomponents, the cost analyst reduces the number of factors that the expert must mentally model when providing an estimate.

The extent of structuring—or breaking down the effort into distinct pieces—should be driven by the basis of the expert's knowledge and any supporting data. Just as estimating at too high a work breakdown structure (WBS) level may reduce precision, attempting to estimate at too low of a WBS level may also insert unnecessary bias or error. Moreover, the cost analyst should generally avoid structuring an estimate so that the expert must provide his or her answer in dollars. Instead, the cost analyst should ask the expert what unit of measure he or she prefers, so that the expert does not have to go through the mental exercise of converting units. The cost analyst may discover that the expert prefers to estimate in hours, full-time equivalents, or SLOC, rather than in dollars.

Phase 3: Conditioning

The conditioning phase strives to head off biases and condition the expert to "think fundamentally about his judgement" (Spetzler & Stael von Holstein, 1975 p. 353). In their original paper, Spetzler and Stael von Holstein provided limited detail on the conditioning phase, aside from suggesting that the analyst ask the expert to describe how they go about assigning probabilities. However, later researchers have supplemented the SRI model by recommending that the expert is provided training on cognitive biases and probability distributions. Other authors have suggested putting the expert through a series of warm-up exercises to allow for calibration (Morgan et al., 1990). Based on this author's experience, however, setting aside sufficient time for warm-up exercises or demonstrations may not be feasible for DoD cost and schedule estimates, particularly for routine estimates or smaller programs.

However, this author has found success with utilizing the conditioning phase to introduce the concept of the probability distribution and provide a preview of what to expect in the encoding phase. This author's conditioning protocol consists of three steps. First, the analyst should begin every session with a brief overview of the triangular or beta-PERT distribution, making mention that the expert will be asked to separately provide a low, high, and most likely estimate. Second, explain to the expert that he or she will later be asked to quantify the confidence interval percentage captured by the given low and high estimate. Finally, emphasize that as the expert considers the low, high, and most likely estimate, he or she should verbalize the assumptions and conditions that would lead to the provided outcome.

Phase 4: Encoding

As introduced previously, research suggests we should first ask the expert for the low and high values to avoid the anchoring effect. It is therefore recommended that the analyst first ask the expert for the "low" value, followed by the "high" value. After each value is provided, ensure the expert is verbalizing both the assumptions and events that could lead to that value. If the expert is not being clear—or their response is not understood—the cost analyst should continue to ask "why?" until the analyst is confident that the estimate is justified and can be explained. Only after obtaining the extremes—and their justification—should the cost analyst ask for the "most likely" value. Once again, the most likely value should be accompanied by a rationale that would lead to the most likely outcome.

After recording the range and most likely, ask the expert how confident they are in their low and high values. What percentage of outcomes will fall within the provided range or what percentage will fall outside the range? Alternatively, the analyst can ask the expert what percentage of outcomes will be greater than the high and what percentage will be less than the low. Then use the provided low and high probabilities to calculate the absolute minimum and maximum via the equation in the JA CSRUH's Table 2-8. Going forward, take care to distinguish the expert-provided low and high from the calculated minimum and maximum, which have been expanded to capture 100% confidence.

Phase 5: Verifying

Finally, having recorded the high, low, and most likely values and the rational for those values, the analyst should verify that the expert's judgement has remained consistent. For example, it is possible that during the course of discussion, the expert recalled additional information that may lead him or her to adjust the high and low bounds provided earlier in the elicitation session.

Begin by showing the expert the minimum and maximum values that were calculated when the provided low and high were adjusted to encompass 100% of confidence. Ask the expert whether there are conceivable scenarios that could lead to a value outside of the calculated minimum and maximum bounds. If the expert concedes that a scenario exists, ask if he or she would like to adjust the absolute minimum and maximum. If the expert would prefer not to adjust the bounds, then ask for the probability of an outcome outside of the minimum and maximum bounds and use the provided probability to further adjust the bounds outward. If necessary, repeat this step until the expert is satisfied with the calculated minimum and maximum.

When the expert is satisfied with the absolute minimum and maximum interval, the initial elicitation is completed. At this point, consider applying an additional 30% uncertainty to account for bias and "unknown unknowns." Even when following the SRI protocol, biases will exist in the expert's answer, as it is not possible for the expert to consider all possible outcomes and scenarios, especially those that fall outside of their area of expertise. When determining whether additional uncertainty is warranted, consider that parametric cost models tend to have coefficient of variation (CV) values between 0.15 and 0.35 (Naval Center for Cost Analysis, 2015, p. 32). Also consider which acquisition milestone the expert's estimate is supporting. Elicitations occurring early in the development production life-cycle will have greater uncertainty than those occurring later. Carney (2013) found that at the program level, estimates have CVs of 0.41 to 0.74 at Milestone A, 0.45 to 0.61 at Milestone B, and 0.23 to 0.32 at Milestone C. Thus, an expert-derived prediction interval with a coefficient of variation lower than 0.25 is likely overconfident and could benefit from the inclusion of additional uncertainty.

Adapting the Model for Multiple Experts

Although the SRI model is initially presented as a model for eliciting opinion from a single expert, the encoding phase of the SRI model may be easily adapted to allow for multiple experts. Prior to beginning the elicitation, the analyst must decide how much interaction to allow between experts. Although interaction is beneficial in allowing for the exchange of ideas and assumptions, it also contributes to *groupthink*, a cognitive bias not yet introduced. In groupthink, the position of a few experts leads the entire group to a consensus that does not represent the individual experts' private opinion.

What does literature recommend for controlling groupthink? At one end of the interaction spectrum, an analyst may allow the group to openly discuss the low value, high value, most likely value, and corresponding confidence level without any structure until a consensus is reached for each. However, groupthink is most likely to occur in this scenario. Nearer the other end of the spectrum is the Delphi method, in which experts exchange anonymous written inputs and justification until a consensus is reached. By allowing for anonymous inputs, the risk of groupthink is significantly reduced. Even more extreme, some authors propose not allowing any communication between experts, and instead taking a simple average or weighted average of the experts' individual inputs. In this case, each elicitation is conducted separately with no interaction between the experts, thereby preventing any groupthink.

A more moderate method is the nominal group technique, in which each expert is forced to establish an estimate prior to interacting with other group members. After the initial estimate, each expert presents his or her position, the rationale behind the position, and all relevant assumptions. After the initial positions are revealed, differences between individual estimates are openly discussed in an attempt to reach a consensus. If consensus is not reached, then the divergent position of each expert is averaged (Gustafson et al., 1973). A similar technique, known as "Planning Poker," is commonly utilized for estimating requirements in the agile software community. In planning poker, software engineers assign difficulty to a user story (i.e., software requirement) by simultaneously revealing a poker card with the difficulty number. Each expert then defends the rationale behind his or her initial poker card estimate. After discussions among the experts, subsequent poker rounds—in which each expert may update his or her estimate—are conducted until the estimates converge to the same assigned difficulty value (Cohn, 2012). For both of these methods, the most important step is that each expert is forced to commit to an initial estimate prior to discussions beginning. Committing to an initial estimate prior to group discussion prevents the group from anchoring off the first expert's response and promotes the open exchange of assumptions and ideas across the group.

Conclusion & Recommendations

Research shows that expert opinion can be as accurate as parametric-based methods (Jorgensen, 2007). However, when not properly guided, experts are prone to biases, and liable to be overconfident when estimating the uncertainty surrounding an estimate. To achieve more consistent and accurate results with elicitation, this author advocates for the adoption of a structured elicitation model, such as the SRI model. The model integrates methods—such as first asking the expert for the low and high estimate—that are shown in experiments to naturally reduce human overconfidence. Moreover, adopting a common model will promote more rigorous documentation, so that the expert's opinion may be subjected to the same extent of senior analyst review and verification as traditional data. To further improve the quality of elicitation, two additional recommendations are provided.

As the first recommendation, this author advocates increasing formal training and education on elicitation for new cost and schedule analysts. Most analysts today have learned elicitation via a trial-and-error or ad hoc approach, and not a formal education program. We would not expect an analyst to construct a parametric model without first learning the fundamentals of learning regression, so why are our expectations any different for elicitation? Every new analyst should be given at least a rudimentary introduction to elicitation and provided with a common framework. To assist in guiding new analysts, a checklist that this author has personally used is included in Appendix C. Readers are encouraged to further adapt and improve the checklist for their own uses.

As a second recommendation, further research is needed to determine the accuracy and CV of elicitation-based cost and schedule estimates. The Air Force Life-Cycle Management Center's Cost Research Branch is currently undergoing a project that will examine historical cost growth within program office estimates, with cost estimating methodology being a recorded factor (S. Valentine, personal communication, March 13, 2019). Such a study will prove valuable, as it will establish a statistically-based CV range for expert-elicited estimates, so that we will no longer have to strictly rely on rules-of-thumb, such as adding an additional 30% to expert-estimated uncertainty intervals.

In closing, this is an exciting time to be a cost or schedule analyst. CADE and other acquisition databases are increasing our access to data, allowing for more innovative estimates and analyses. However, even with more data, situations will continue to arise in which we must seek the opinion of an expert. By inserting more structure and discipline into the elicitation process, we can avoid the most common pitfalls of expert opinion, thereby leading to more accurate and reliable cost and schedule estimates.

References

Air Force Cost Analysis Agency. (2007). *Air Force cost risk and uncertainty analysis handbook (CRUH)*. Washington, DC: Author.

Bar-Yosef, S., & Venezia, I. (2014). An experimental study of overconfidence in accounting numbers predictions. *International Journal of Economic Sciences*, *3*(1), 78–89.

Ben-David, I., Graham, J., & Harvey, C. (2010). *Managerial miscalibration*. Retrieved from https://faculty.fuqua.duke.edu/~charvey/Research/Working_Papers/W105_Managerial_miscalibration.pdf

Boyd, D. W., & Regulinski, S. G. (1979). *Characterizing uncertainty in technology cost and performance* (Report 1114). Los Altos, CA: Decision Focus.

Brown, B. B. (1968). *Delphi process: A methodology used for the elicitation of opinions of experts*. Santa Monica, CA: RAND Corporation.

Burgman, M. A. (2016). *Trusting judgements: How to get the best out of experts*. Cambridge, England: Cambridge University Press.

Capen, E. (1976). The difficulty of assessing uncertainty (includes associated papers 6422, 6423, 6424, and 6425). *Journal of Petroleum Technology*, *28*(8), 843–850. doi:10.2118/5579-pa

Carney, S. T. (2013). *Investigation into risk and uncertainty: Identifying coefficient of variation benchmarks for Air Force ACAT I programs* (Master's thesis). Air Force Institute of Technology.

Cohn, M. (2012). *Agile estimating and planning*. Upper Saddle River, NJ: Prentice Hall PTR.

Elicit. (n.d.). In *Merriam-Webster*. Retrieved October 19, 2017, from http://www.merriam-webster.com/dictionary/elicit.

Flyvbjerg, B. (2011). Over budget, over time, over and over again. *Oxford Handbooks Online*.

Galway, L. A. (2007). *Subjective probability distribution elicitation in cost risk analysis: A review*. Santa Monica, CA: RAND Corporation.

Goldenson, D. R., & Stoddard, R.W. (2013). *Quantifying uncertainty in expert judgment: Initial results*. Pittsburgh, PA: Carnegie Mellon Software Engineering Institute.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19–30. doi:10.1037/1040-3590.12.1.19

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind. *Psychological Science, 20*(2), 231–237.

Jorgensen, M. (2007). Forecasting of software development work effort: Evidence on expert judgement and formal models. *International Journal of Forecasting, 23*(3), 449–462. doi:10.1016/j.ijforecast.2007.05.008

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.

Kitchenham, B., Pfleeger, S. L., McColl, B., & Eagan, B. (2002). An empirical study of maintenance and development estimation accuracy. *Journal of Systems and Software*, *64*(1), 57−77.

Levy, R. I. (1973). *Tahitians: Mind and experience in the Society Islands*. Chicago, IL: University of Chicago Press.

Li, Q., Wang, Q., Yang, Y., & Li, M. (2008). Reducing biases in individual software effort estimations. In *Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM 08)*. doi:10.1145/1414004.1414041

McKenzie, C. R. M., Liersch, M. J. & Yaniv, I. (2008) Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Process*, *107*, 179–191.

Morgan, M., Henrion, M., & Small, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge, England: Cambridge University Press.

Nashef, S., Roques, F., Michel, P., Gauducheau, E., Lemeshow, S., & Salamon, R. (1999). European system for cardiac operative risk evaluation (EuroSCORE). *European Journal of Cardio-Thoracic Surgery, 16*(1), 9–13.

Naval Center for Cost Analysis. (2015). *Joint agency cost schedule risk and uncertainty handbook (JA CSRUH)*. Washington, DC: Author.

Nichols, T. (2017, March & April). How America lost faith in expertise. *Foreign Affairs*, *96*(2), 60–73.

Patterson, R. E., Pierce, B. J., Bell, H. H., & Klein, G. (2010). Implicit learning, tacit knowledge, expertise development, and naturalistic decision making. *Journal of Cognitive Engineering and Decision Making, 4*(4), 289–303.

Russo, J. E., & Schoemaker, P. (1992). Managing overconfidence. *Sloan Management Review*, *33*(2), 7–17.

Sanders, N. R., & Ritzman, L. P. (1991). On knowing when to switch from quantitative to judgemental forecasts. *International Journal of Operations & Production Management, 11*(6), 27–37. doi:10.1108/01443579110005523

Shen, L., Chen, X., Gu, J., & Xue, S. (2018). Validation of EuroSCORE II in Chinese patients undergoing coronary artery bypass surgery. *The Heart Surgery Forum, 21*(1), 036.

Soll, J., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299–314.

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgements of experts. *Risk Analysis*, *30*(3), 512–523.

Spetzler, C. S., & Staël Von Holstein, C. S. E. (1975). Probability encoding in decision analysis. *Management Science, 22*(3), 340–358.

Teigen, K. H., & Jorgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology*, *19*(4), 455–475.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Watern, K. (2016). Cost Assessment Data Enterprise (CADE). *Air Force Comptroller Magazine*, *49*(1), 25.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects. *Journal of Experimental Psychology: General*, *125*(4), 387–402.

Winkler, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, *62*(319), 776–800.

Appendix A: Meta-Analysis Results

Author (Year)	Interval Requested	Intervals Containing Truth	Number of Predictions	Expert	Estimating Task
Russo & Schoemaker (1992)					
Advertising 1	90%	39%	750	Corporate Managers	Advertising Industry Knowledge
Advertising 2	50%	22%	750	Corporate Managers	Advertising Industry Knowledge
Computers 1	95%	20%	1,290	Corporate Managers	Computer Industry Knowledge
Computers 2	95%	42%	1,290	Corporate Managers	Computer Firm Knowledge
Data Processing 1	90%	58%	252	Corporate Managers	Data Processing Industry Knowledge
Data Processing 2	90%	38%	261	Corporate Managers	General Business Knowledge
Money Management 1	90%	50%	480	Corporate Managers	Financial Industry Knowledge
Petroleum 1	90%	50%	850	Corporate Managers	Petroleum Industry and Firm Knowledge
Petroleum 2	50%	21%	850	Corporate Managers	Petroleum Industry and Firm Knowledge
Pharmaceutical 1	90%	51%	390	Corporate Managers	Pharmaceutical Firm Knowledge
Security Analysis 1	90%	36%	497	Corporate Managers	Security Industry Knowledge
McKenzie et al. (2008)	90%	44%	1,720	IT Professionals	IT Industry Knowledge
Ben-David et al. (2013)	80%	33%	11,600	Chief Financial Officers	Stock Market Return (S&P 500)
Goldenson & Stoddard (2013)					
Battery 1	100%	10%	140	Graduate Students	Software Development Effort
Battery 2	100%	50%	80	Graduate Students	Software Development Effort
Battery 3	100%	70%	70	Graduate Students	Software Development Effort
Bar-Yosef & Venezia (2014)	95%	7%	30	Brokerage Analysts	Financial Forecasts

Appendix B: Meta-Analysis Regression Model Outputs

Simple Linear Regression Model Output

```
Call:
lm(formula = Expert.Intervals.Containing.True.Answer ~ Interval.Requested)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.33484 -0.02667  0.00333  0.11333  0.27700 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.05973    0.24994   0.239    0.814
Interval.Requested  0.36328    0.28229   1.287    0.218

Residual standard error: 0.1683 on 15 degrees of freedom
Multiple R-squared:  0.09943,	Adjusted R-squared:  0.03939 
F-statistic: 1.656 on 1 and 15 DF,  p-value: 0.2176
```

Weighted Linear Regression Model Output

```
Call:
lm(formula = Expert.Intervals.Containing.True.Answer ~ Interval.Requested, 
    weights = Sample.Size)

Weighted Residuals:
    Min      1Q  Median      3Q     Max 
-7.1281 -0.4152  0.3115  2.5454  3.5386 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)  
(Intercept)         0.02155    0.14655   0.147   0.8851  
Interval.Requested  0.39675    0.17652   2.248   0.0401 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.819 on 15 degrees of freedom
Multiple R-squared:  0.2519,	Adjusted R-squared:  0.2021 
F-statistic: 5.052 on 1 and 15 DF,  p-value: 0.04007
```

R Code

```
# Import Data
Experts <-
data.frame(Interval.Requested=c(0.90,0.50,0.95,0.95,0.90,0.90,0.90,0.90,0.50,0.90,0.90,0.90,0.80,1.
00,1.00,1.00,0.95),
Expert.Intervals.Containing.True.Answer=c(0.39,0.22,0.20,0.42,0.58,0.38,0.50,0.50,0.21,0.51,0.36,0.
44,0.33,0.10,0.50,0.70,0.07),Sample.Size=c(750,750,1290,1290,252,261,480,850,850,390,497,1720,
11600,140,80,70,30))
attach(Experts)

# Compute simple linear regression and weighted linear regression models
simple.lm <- lm(Expert.Intervals.Containing.True.Answer~Interval.Requested)
weighted.lm <- lm(Expert.Intervals.Containing.True.Answer~Interval.Requested,
weight=Sample.Size)

# Outputs
summary(simple.lm)
summary(weighted.lm)
```

Appendix C: Elicitation Checklist for Cost and Schedule Analysis

Phase 1: Motivating

- ☐ *Analyst*: Familiarize yourself with the requirement needing expert elicitation. Begin formulating questions, and gather data that may be relevant to the expert.
- ☐ *Tell the Expert*: The purpose of this cost estimate is to estimate _____ in support of _____.

Phase 2: Structuring

- ☐ *Ask the Expert*: Should we break down the estimation of the requirement into smaller components?
- ☐ *Ask the Expert*: Would you feel most comfortable estimating the unknown quantity in person-hours, full-time equivalents (FTEs), SLOC, or another unit?
- ☐ *Ask the Expert*: What ground rules and assumptions are you making about the requirement being estimated?

Phase 3: Conditioning

- ☐ *Tell the Expert*: Today I will ask your assistance in constructing the triangular or Beta-PERT distribution that best represents your state of knowledge. I will begin by asking for your low outcome, followed by your high outcome, and lastly I will ask for the most likely outcome.
- ☐ *Tell the Expert*: Next, I will ask you for the probability (or likelihood) that the costs will be lesser/greater than your estimated low and high.
- ☐ *Tell the Expert*: When providing your response for low/high/most likely, please explain the assumptions, rationale, mental model, or analogy used to estimate each outcome. This will help us defend the estimate to decision-makers, and will be useful if the estimate is later revisited.

Phase 4: Encoding

- ☐ *Ask the Expert*: What is the low outcome? Why?
- ☐ *Ask the Expert*: What is the high outcome? Why?
- ☐ *Ask the Expert*: What is the most likely outcome? Why?
- ☐ *Ask the Expert*: Could an outcome be less than your low estimate? If so, what is the probability? What scenario could cause this to happen?
- ☐ *Ask the Expert*: Could an outcome be more than your high estimate? If so, what is the probability? What scenario could cause this to happen?

Phase 5: Verifying

- ☐ *Analyst*: If the expert responded that the interval had a confidence interval of less than 100%, adjust the expert's low and high using JA CSRUH Table 2-8 so that a 100% confidence level is reached. These values are the distribution's absolute min and max.

- ☐ *Ask the Expert*: Are there any conceivable scenarios that could cause the outcome to be less than the minimum? If so, what is the probability? What scenario could cause this to happen?
- ☐ *Ask the Expert*: Are there any conceivable scenarios that could cause the outcome to be more than the maximum? If so, what is the probability? What scenario could cause this to happen?
- ☐ *Ask the Expert*: Does the distribution require any further adjustments? Does it best represent your current state of knowledge?
- ☐ *Analyst*: The elicitation is complete. Thank the expert for their time. Compute the elicited distribution's coefficient of variation (CV), and consider adding 30% additional uncertainty if the CV is low (less than 0.25). Note that the expected CV will vary depending on the requirement being estimated and the milestone that the estimate is supporting.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net