Convergence of Expert Opinion via the Wideband Delphi Method: An Application in Cost Estimation Models

Ricardo Valerdi
Massachusetts Institute of Technology
77 Massachusetts Avenue E38-602
Cambridge, MA 02139 USA
rvalerdi@mit.edu

Abstract. This paper discusses the notion of collective intelligence through the application of the Wideband Delphi method as a way to obtain convergence among a group of experts. The specific application is the definition and calibration of cost estimation models that use data collected from experts as part of their calibration. Convergence is important in this case because companies need to commit to cost estimates early in the planning cycle since so many other decisions are dependent on it. Our results demonstrate that, in most cases, convergence among experts can be achieved after three rounds of the Wideband Delphi.

Introduction

The behavior of crowds is not often associated with intelligence or rational choice. People often behave irrationally and the effect can be multiplied when many individuals base judgments off each other.   Examples of this "madness of crowds" include the Mississippi Scheme's manipulation of the French Economy in 1719, the artificial inflation of the tulip trade in the Netherlands in the 16th and 17th centuries (Mackay 1980) and even the stock market volatility of the 1980s (Lewis 1990). However, studies have shown that under the right circumstances, groups can be remarkably intelligent, and are often smarter as a whole than even the smartest individuals in the group (Surowiecki 2005; Griffiths & Tenenbaum 2006).

Groups do not need to be dominated by exceptionally intelligent people in order to be smart. During a weight-judging competition at the West England Fat Stock and Poultry Exhibition in 1906 (Galton 1907), local farmers and townspeople gathered to appraise the quality of each other's cattle, sheep, chickens, horses, and pigs. As part of the weight-judging competition, a fat ox had been selected and placed on display and members of the crowd were able to place wagers on the weight of the ox. Eight hundred people tried their luck, many of them butchers and farmers but also quite a few "non-experts". People with radically different experiences and knowledge about stock or cattle participated in the exercise. The assumption was that by mixing experts with non-experts the average answer would be off the mark. However, when the estimates provided by the contestants were analyzed it was discovered that the crowd's judgment

was nearly perfect. Once the estimates were averaged, the crowd had guessed that the ox would weigh 1,197 pounds; its actual weight was 1,198 pounds. This result suggests that, in the presence of limited information, the collective inputs of a crowd of people can be much more accurate than the input of a single individual.

One of the difficulties in harnessing the wisdom of crowds is the process for capturing their collective intelligence. One approach is the use of prediction markets which provide an environment for traders to buy and sell contracts whose value is tied to an uncertain future event, such as the cost and duration of a product. The first problem with prediction markets is that they do not provide a forum for discussion between experts. Secondly, the selection of traders in prediction markets is not limited to experts which introduces limitations of expertise. That is, there is no rigorous screening process for prediction market participants. In fact, the more people that participate in a prediction market the better it is for trading volume (Wolfers & Zitzewitz 2004).

An alternative approach to harnessing the wisdom of crowds is the Wideband Delphi methodology which, if used correctly, can serve as a way to survey expert opinion and achieve convergence on a topic. While the primary purpose of the Delphi method is to structure group communication (Linstone & Turoff 1975), it can still be used as a mechanism to achieve convergence among experts. To this end, this paper is focused on answering the following hypothesis:

> *The Wideband Delphi method enables convergence of opinion between experts after 3 rounds.*

To test this hypothesis, we provide a specific example of the development and validation of a parametric cost model that demonstrates the Wideband Delphi process in action and quantitative results from three survey rounds. We begin exploring the idea of predictions among large groups of people and discuss the criteria for selecting experts in decision making pursuits. Then we focus on the Wideband Delphi survey method and guidance for planning, executing and analyzing its results. We also discuss the strengths and limitations of the Wideband Delphi method and conclude with a discussion on the convergence of expert opinion.

The Wideband Delphi Survey Technique

The Oracle at Delphi is a female character from Greek mythology believed to sit in a cave upon a tripod from which she recited riddling answers to questions put to her by attending priests. Such oracles were considered to be in close contact with a god and thereby able to act as mediums to the god. In this fashion, Greeks would pilgrimage to the oracle in order to ask important questions of the gods and receive their answers (Anderson 2008). The oracle symbolized a medium who communicated with the gods on a different plane, bringing back to a corrupted language words that could not communicate a logical meaning, only an implied one that a deserving hearer could decipher. The Delphic Oracle was echoed in Roman culture by the figure of the Sybil, who Aeneas encounters in his journeys, directing him to enter the underworld before he can continue his travels (House 2008). The Delphic Sybil was painted by Michelangelo on the ceiling of the Sistine Chapel in the 16th century (see Figure 1) after much

controversy with Pope Julius II (King 2003).

Figure 1. The Delphic Sybil by Michelangelo Buonarroti
Capella Sistina, Il Vaticano (1508-1512)
(source: `www.lib-art.com`[1])

The original Delphi Method for collecting expert opinion originated as part of a post-WWII movement towards forecasting the possible effects of technology development in relation to economic and social re-generation. The technology forecasting studies were initiated by the Douglas Aircraft Company, which established a RAND project in 1946 to study the "broad subject of inter-continental warfare" (Fowles 1978). The theoretical and methodological basis for forecasting was elaborated in a subsequent series of papers produced by the project. These argued that, in the absence of an established evidence base, emergent fields of enquiry could begin to develop such an evidence base through capturing and synthesizing the opinions of domain experts.

RAND research team members, Olaf Helmer, Norman Dalkey, Nicholas Rescher, and others, were given the task of deciding the best way to approach the issue of forecasting future events. They decided that the opinions of a panel of experts should be sought, but were not satisfied with conventional means of gathering group opinions. Dalkey had concluded that prior work by statisticians concerning the statistical properties of group judgment was going in the wrong direction, and would not be useful (Cabaniss 2001). Results of experiments at RAND indicated that the responses of a group were not as accurate as the median of individual estimates without discussion (Dalkey & Helmer 1963). As a result, Olaf Helmer proposed a new approach to the group decision-making process (Cabaniss 2001) that would overcome some of the undesirable aspects of conventional group conferences involving face-to-face contact, specifically:
- Swaying of group decisions by dominant, more vocal personalities
- Semantic noise resulting from individual or group interests

---

[1] `http://www.lib-art.com/artgallery/14144-the-delphic-sibyl-michelangelo-buonarroti.html`

- Reluctance to change prior expressed opinions to avoid loss of credibility
- Reluctance to express opinions that differ from the perceived group consensus
- Fear of reprisal or criticism among junior members for expressing conflicting opinions to senior, more powerful group members

Since its inception over forty years ago, the Delphi method has had various offspring such as the Policy Delphi, Trend Delphi, Delphi II, Problem Solving Delphi, Imen-Delphi, etc. Despite its numerous variations, the central premise behind the Delphi remains the same: to help structure group communication to enable a group of individuals to deal with a complex problem. Group consensus is one byproduct that could result from these discussions, but it is not the only objective (Linstone & Turoff 2010). The utility of the Delphi method is evident in the various applications in fields like policy formulation (Turoff 1970), science and technology planning (Waissbluth & De Gortari 1990), tourism potential (Kaynak, Bloom & Leibold 1994), software engineering training needs (Geier 1985), futures thinking (Passig 1997), educational counseling (Cabaniss 2001), pharmaceutical process validation (Helle, Reijonwn & Mannermaa 2003), and nursing (Cabaniss 2001).

In the late 1970s, as the need for predictive software cost models grew, the Delphi method was applied to the development of parametric cost models. One particular variant was the Wideband Delphi (Boehm 1981) which differed from the conventional Delphi in that it encouraged discussion among participants. The conventional Delphi discouraged discussion between rounds, under the assumption that discussion would introduce biases toward the most vocal participants. When it was applied to cost estimation, Boehm (1981) found that enabling discussion and a broader communications channel produced more accurate results and aided in shared experiences and teambuilding. Compared to the narrowband-communications Rand Delphi method, this revised approach became known as the Wideband Delphi. Recent research in software project estimation has shown that estimates that benefit from group discussion tend to be more accurate (Hoest & Wohlin 1988; Moløkken & Jørgensen 2004; Cohn 2005). Given the greater interaction and communication between the participating experts, we provide a more detailed description of the preparation, execution, and analysis phases of the Wideband Delphi methodology.

Use of Wideband Delphi in Cost Model Development and Validation. Individuals make decisions every day through the use of their experience and intuition. However, these decisions are often suboptimal compared to decisions made by groups of people especially when it comes to the cost of future products. Instead of depending on "estimation by consensus" which often results in large standard deviations, organizations can embed expert opinion in parametric cost models during their development.

Parametric cost models are used in industry and government to estimate the cost of future products. The technique involves the characterization of the technical product itself as well as the team that will be working on its development. One of the most commonly used models is the Constructive Cost Model (COCOMO) which was created to help determine how many software engineers would be needed to build certain software systems (Boehm, et al. 2000). The development of parametrics models like COCOMO is accomplished through a series of deliberate steps that help optimize the model's accuracy. This is done by selecting the most

statistically relevant parameters in the model (the independent variables) and validating their relationship to cost (the dependent variable). The modeling methodology is comprised of the following eight steps:

1. Determine modeling needs
2. Analyze existing literature
3. Perform behavioral analysis
4. Define relative significance
5. Perform expert judgment Delphi to determine *a priori* model
6. Gather and analyze empirical data
7. Perform Bayesian calibration to create *a posteori* model
8. Gather more data, refine model

The first four steps in the modeling methodology are designed to define and down select the most relevant drivers of cost, naturally these should also be independent of each other. The resulting model includes only the most significant parameters that drive the cost of products. In step 5, the Wideband Delphi technique serves as a mechanism for guiding a group of experts to a consensus of opinion on each parameter's outcome-influence value. Data definitions and rating scales are also established for the most significant parameters. This exercise is repeated multiple times to give individuals the opportunity to re-estimate their values after group discussion. This process often uncovers overlaps and changes in outcome drivers and ultimately the *a priori* model, which is represents a model based solely on the opinion of experts.

Experts are an essential ingredient to this step in the model development process because they bring subjective knowledge about the sources of cost for previous products. Eventually, their opinion is compared to historical project data through the Bayesian approximation technique. This calibration approach ensures that the cost model is influenced both by expert opinion and historical data to yield a calibrated *a posteori* model (Chulani, Boehm & Steece 1999). Parametric models such as this need to be updated periodically so that their calibration is representative of the latest trends in technology, processes, and personnel experience (Kemerer 1987). The next section provides more information about the strengths and limitations of the Wideband Delphi method.

Strengths and Limitations of the Wideband Delphi. A number of issues have been discussed about the necessary elements for a successful Wideband Delphi. Beyond these critical success factors, there are a number of inherent strengths and limitations that warrant careful consideration. The purpose of the Wideband Delphi is to gather a collective estimate from experts without the risk that certain vocal or influential individuals will skew the results. The method allows estimates to be developed when little physical exists. The method improves upon other methods for the collection of expert opinion, such as interviews and surveys. Interviews can be difficult to control and even more difficult to quantify. They are also extremely time consuming when large sample sizes are needed. Surveys can handle larger sample sizes and provide quantitative results, but they fail to consider the subjective reasons behind expert judgments. The Wideband Delphi method asks experts to give reasons for their estimates and then to debate those reasons, quantitatively capturing knowledge and experience not explicitly asked for on the survey instruments themselves.

The method aims to derive the benefit from the opinions of a group of experts while avoiding the disadvantages of 'group think' and group dynamics where certain individuals can dominate the discussion and instill undue influence. With the help of careful coordination by the facilitator, experts can freely express their opinions and critique ideas since anonymity is maintained throughout the process. If discussions are not dominated by one participant's authority or personality the exercise can generate a much wider set of ideas than what each individual participants would generate by responding to the questionnaire on their own. The facilitator's ability to determine when it is necessary to move on to the next topic is especially important. If discussions drag on for too long then there is a risk of jeopardizing the entire exercise.

It is important to note when a Wideband Delphi is not appropriate. To begin with, the method is not appropriate when one wants to build relationships and ongoing communication between the participants or when one wants to make decisions directly. The key problems reported include: poor internal consistency and reliability of judgments among experts, and therefore difficulty reproducing forecasts based on the results elicited; sensitivity of results to ambiguity and respondent reactivity in the questionnaires used for data collection; difficulty in assessing the degree of expertise held by participating experts. The validity of the results and expertise of the participants is very much dependent on who participates in the workshop. There can also be a high barrier to entry for experts to participate if the topic under discussion is too specialized. This requires extensive training to ensure consistent interpretation of the concepts across the experts. The selection of experts is further discussed in the next section.

The coordination involved with a Wideband Delphi also makes it a time consuming and expensive method. It can sometimes be difficult to coordinate and motivate a group of experts with diverse interests and busy schedules. A major problem has also been the tendency for experts to over-simplify particular issues, and treat them as isolated events. This is particularly the case in forecasting, where experts tend to think in terms of linear sequential events rather than applying a systems view that involves complex chains and associations. Design of survey instruments can help mitigate these effects. A similar question could be presented several times with slightly different wording to measure the consistency of experts' opinions.

Achieving Convergence Among Experts:
Results from Wideband Delphi Rounds

The intended outcome of the Wideband Delphi is either a consensus amongst the participants on likely and possible future developments or a wide range of possible developments and their relative strengths and weaknesses. There are some topics that do not bode well with this method, especially polarizing issues that challenge a person's values or situations that would be better suited for conflict resolution. Other topics such as rare events are also not suitable for Delphi techniques (Goodwin & Wright 2010). If the intent is to achieve convergence, the ultimate measure of effectiveness of a Wideband Delphi is whether there is a central tendency demonstrated by the group. This can be quantified by the mean, median, and mode. In cost model development, convergence is an important goal because it represents the statistical predictive ability of a parametric model. The priority of product development organizations is to make accurate forecasts of future products so that decisions such as make/buy and bid/no-bid can be made.

Benefits of Expert Opinion in Cost Model Development. As explained earlier in this paper, the contribution of expert opinion in cost model development provides a reference point for cost model development and validation. In the modeling development methodology, experts are asked to provide their opinion on the relative cost impacts of certain drivers. The aggregate opinion of experts, obtained through the Wideband Delphi, is combined with historical project data through the Bayesian approximation technique. This ensures that the cost model is influenced by experts as well as historical data. The benefit of using experts is that they carry a broad understanding of the issues that drive product costs. They communicate this experience through the Wideband Delphi process, especially during the inter-round discussions. The drawback about using experts is that they can be influenced by a variety of biases that may cloud their judgment.

The benefit of using historical data is that they represent what actually happened on projects. One may wonder why parametric cost models do not depend on historical data alone since past performance is a good predictor of future performance. There are two reasons for this. First, historical project data are not always collected in a consistent fashion. This leads to noise in the data which lead to poor statistical results. Second, there are other aspects of project costs that are often unexplained in the data. If these phenomena are not in the data collection form then they are not captured in the analysis. Expert opinion helps alleviate these two concerns with historical data. When combined, expert opinion and historical data result in a more robust forecasting approach.

Sample Recruitment and Characteristics. Forty participants were recruited through the University of Southern California Center for Systems and Software Engineering (CSSE), an industry consortium comprised mostly of high-tech companies mostly in the aerospace & defense industry. Participant selection could not be controlled since participants volunteered; however, descriptive data were collected from attendees to assess their qualifications to participate in the exercise. The average software systems engineering work experience among participants was 18 years while the average years of cost modeling experience was 6. The range of domain experience among the participants is shown in Figure 2. This demonstrates broad expertise across several technical areas, which is a desired attribute of cost models that are calibrated to estimate complex products in these industries.
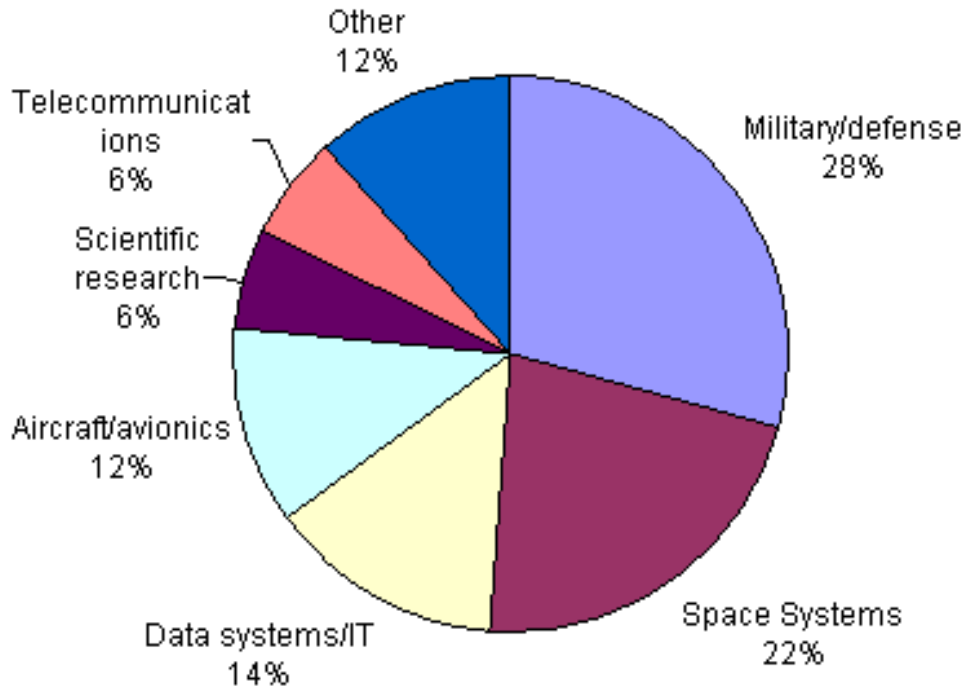
Figure 2. Application domain experience of Delphi participants (n = 40)

Convergence Results. Several studies indicate that three rounds are sufficient to collect the necessary information and reach a consensus among experts (Cyphert & Gant 1971; Brooks 1979; Ludwig 1994; Custer, Scarcella & Stewart 1999). Unfortunately, these studies do not provide quantitative evidence for these claims. The measure of convergence is the reduction in standard deviation throughout the rounds. Participants were asked to assess the relative impact of the following fourteen systems engineering parameters that form the basis of a cost model (Valerdi 2008):

1. Requirements understanding
2. Architecture Understanding
3. Level of Service Requirements
4. Migration Complexity
5. Technology Risk
6. Documentation to Match Lifecycle Needs
7. Number and Diversity of Installations/Platforms
8. Number of Recursive Levels in the Design
9. Stakeholder Team Cohesion
10. Personnel/Team Capability
11. Personnel Experience/Continuity
12. Process Capability
13. Multisite Coordination
14. Tool Support

The results of three Wideband Delphi rounds are provided in Figure 3. The triangles indicate the average score of the respondents in that round while the squares indicate the default values provided at the beginning of that round. In most cases, the Wideband Delphi led to a convergence in opinions among experts as evident by the box plot which shows five-number summaries: the smallest observation (sample minimum), lower quartile (Q1), median (Q2), upper quartile (Q3), and largest observation (sample maximum).

In addition to a visual representation of convergence, we can determine the coefficient of variation for each round. This is calculated as the ratio of the standard deviation, σ, to the mean, μ, as shown in Equation 1.

$$c_v = \sigma/\mu \quad \text{(Equation 1)}$$

The coefficient of variation for each of the three rounds is shown in Table 1. Results show that c_v for the 14 cost drivers decreased in 8 of them, remained relatively constant in 3 of them, and increased in 3 of them.

Table 1. Coefficient of variation for three Wideband Delphi Rounds

Cost Driver	Round 1 c_v	Round 2 c_v	Round 3 c_v
Requirements Understanding	0.44	0.41	0.34
Architecture Understanding	0.34	0.32	0.34
Level of Service Requirements	0.22	0.20	0.21
Migration Complexity	0.12	0.11	0.20
Technology Risk	0.20	0.18	0.15
Documentation to match lifecycle needs	0.08	0.07	0.06
# and diversity of installations/platforms	0.19	0.18	0.09
# of recursive levels in the design	0.27	0.25	0.07
Stakeholder team cohesion	0.25	0.23	0.24
Personnel/team capability	0.24	0.22	0.16
Personnel experience/continuity	0.23	0.22	0.07
Process capability	0.20	0.18	0.17
Multisite coordination	0.07	0.07	0.17
Tool support	0.11	0.10	0.26

It appears that the Wideband Delphi provided three types of results. The first type is the group of cost drivers that benefitted from the exercise as demonstrated by a significant reduction in c_v (58% of the cases). The second type is the group that had no relative change in c_v as a result of the exercise (21% of the cases). The third type is the group that exhibited an increase in c_v (21% of the cases). This means that the hypothesis that three rounds of Wideband Delphi are needed to achieve convergence is partially supported. What is evident is that two Wideband Delphi rounds were not sufficient for these types of questions. A majority of the reduction in c_v values came in the third round.
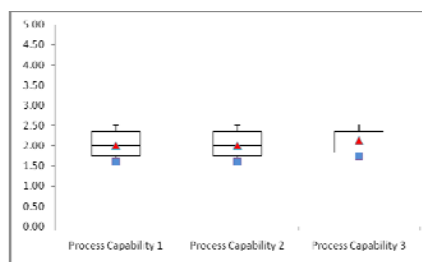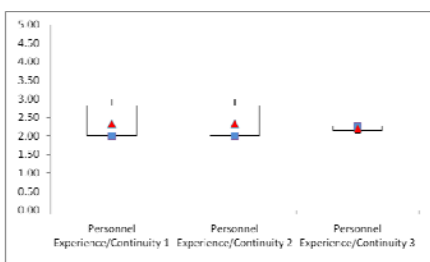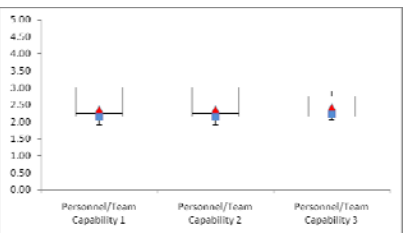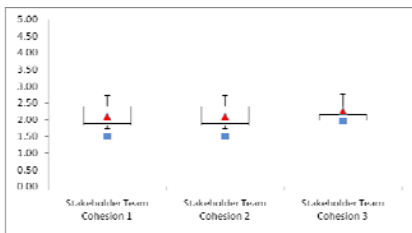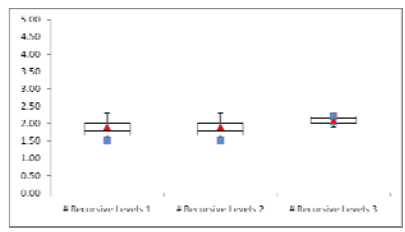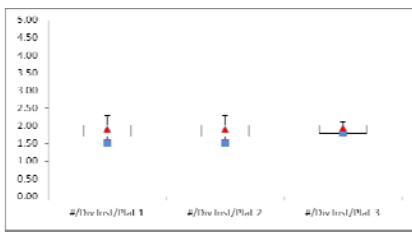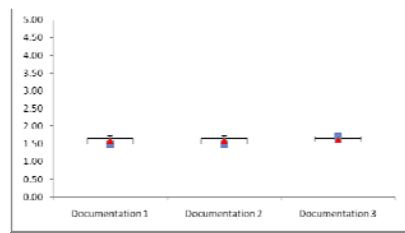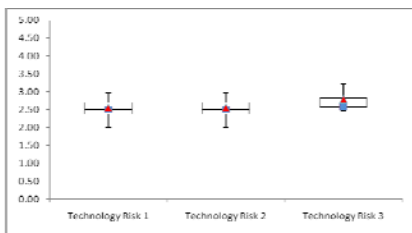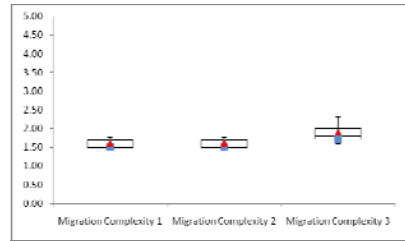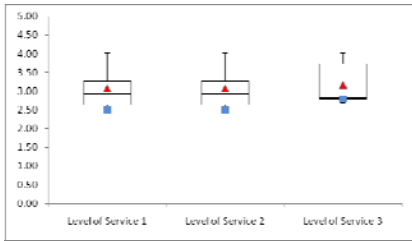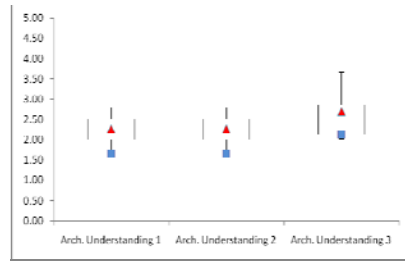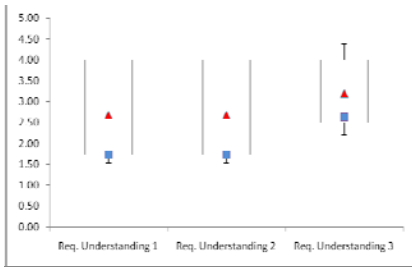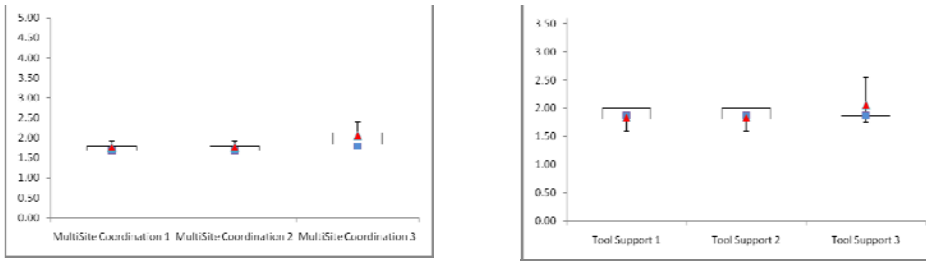
Figure 3. Wideband Delphi Results for Three Rounds (n = 40)

Threats to Validity and Generalizability. The Wideband Delphi method applied to the development and calibration of cost estimation models does not suffer from the same limitations as other forecasting techniques since the application is limited to estimating parameter values. Similarly, the convergence results illustrated in this paper may not apply to situations in which the primary objective is to forecast future events. Nevertheless, it is important to discuss threats to validity of this study and explore areas in which these results can be generalized.

One of the major threats to internal validity was sample selection. Even though significant effort was invested in making sure the experts were well qualified for the Wideband Delphi, there is a chance some people did not have the necessary experience. We did measure year of experience in software systems engineering and cost modeling but it is possible that years of experience do not necessarily make someone an expert especially if the lack breadth of experience.

Numerous threats to validity come from cognitive biases in Delphi participants (Goodwin & Wright 2010). Although the Delphi method was developed to eliminate biases that emerge during group decision making, the facilitator must remain vigilant against any that do appear. Quantitative consensus must not be mistaken for practical consensus. If an individual chooses an answer choice just to match his or her peers, then the resulting estimate is of little practical use. The facilitator should be aware of the following relevant biases (Tversky & Kahneman 1974):

- Anchoring – when a certain value unduly influences the decision maker and biases their opinion towards that value regardless of its relevance to the question at hand.
- Projection bias – the tendency to unconsciously assume that others share the same or similar thoughts, beliefs, values, or positions.
- False consensus effect – the tendency for people to overestimate the degree to which others agree with them.
- Belief bias – the tendency to base assessments on personal beliefs (also known as belief perseverance and experimenter's regress).
- Availability bias – the tendency to associate more recent event or well-publicized events with the relative importance of that event.
- Confirmation bias – the tendency to search for or interpret information in a way that confirms one's preconceptions.
- Anthropic bias – the tendency for one's evidence to be biased by observation selection effects, or not accurately representing a population through sampling. For example, when an expert has a very narrow range of experience that might represent an outlier from others in the field.

- Self-serving bias – the tendency to claim more responsibility for successes than failures. It may also manifest itself as a tendency for people to evaluate ambiguous information in a way beneficial to their interests (also known as group-serving bias).

These biases are difficult to diagnose but nevertheless important in evaluating progress towards convergence. An important consideration in any Delphi study is the percentage agreement that a researcher would accept. The answer may lie with the importance of the research topic. For instance, if it were a life and death issue such as whether or not to switch off a respirator in an intensive care unit or the retirement of a nuclear plant, a 100% consensus level may be desirable. Alternatively, if the topic was related to the selection of a new nurses' uniform, a consensus of 51% may be acceptable. As with most aspects of the Delphi technique, the literature provides few clear guidelines on what consensus level to set. Loughlin and Moore (1979) suggest that consensus should be equated with 51% agreement amongst respondents. By contrast, Green, et al. (1999) employed an 80% consensus level. Establishing the standard is crucial as the level chosen determines what items are discarded or retained as the rounds unfold. It is necessary for the research team to establish a definition of consensus before data collection begins.

The generalizability of the results in this paper is limited to decisions surround the design and validation of parametric cost models. Furthermore, the application domain is limited to large high-technology companies in aerospace & defense industries. Extreme care should be taken when applying these results, both from a methodological standpoint and practical standpoint, to areas that are dramatically different.

Conclusion

This paper has provided a number of contributions to the body of knowledge. First, it has provided guidance for the administration of the Wideband Delphi method. Second, it has provided quantitative evidence for the need for three rounds of the Wideband Delphi in order to achieve convergence among experts. Reaching convergence is extremely important in the development of parametric cost estimation models because they are used to guide investment decisions for future products. Not reaching consensus could mean one of two things. Either the experts do not agree because of the diversity of their experience or because the topic under discussion is too controversial. Disagreement could also stem from the misinterpretation of the definitions or context of the items being measured. Whatever the reason for divergence, the Wideband Delphi method can be a useful technique for harnessing the wisdom of crowds if used correctly. But convergence is not the only goal of the Wideband Delphi method. It can also be valuable in alerting the participants to the complexity of issues by having them challenge their assumptions Fink, et al. 1984; Hartman 1981).

References

K. Anderson, N. Freund, The last days of socrates, http://socrates.clarke.edu/aplg0219.htm, 2008.

Boehm, B. W. 1981. *Software engineering economics*, Prentice-Hall.

Boehm, B. W., Abts, C., Brown, A. W., Chulani, S., Clark, B., Horowitz, E., Madachy, R., Reifer, D. J. and Steece, B. 2000. *Software cost estimation with COCOMO II*, Prentice Hall,

Upper Saddle River, NJ.

Brooks, K. W. 1979. Delphi technique: expanding applications. *North Central Association Quarterly* 54 (3): 377-385.

Cabaniss, K. 2001. Counseling and computer technology in the new millennium – an internet delphi study, PhD Dissertation, Virginia Polytechnic Institute and State University.

Chulani, S., Boehm, B. W. and Steece, B. 1999. Bayesian analysis of empirical software engineering cost models. *IEEE Trans. Softw. Eng*. 25 (4): 573-583.

Cohn, M. 2005. *Agile estimating and planning.* Addison Wesley.

Custer, R. L., Scarcella, J. A. and Stewart, B. R. 1999. The modified delphi technique: a rotational modification. *Journal of Vocational and Technical Education* 15 (2): 1-10.

Cyphert, F. R. and Gant, W. L. 1971. The Delphi technique: A case study. *Phi Delta Kappan*: 272-273.

Dalkey, N. and Helmer, D. 1963. An experimental application of the delphi method to the use of experts. *Manage. Sci*. 9 (3): 458-467.

Fink, A., Kosecoff, J., Chassin, M. and Brook, R. 1984. Consensus methods: characteristics and guidelines for use. *Am. J. Public Health* 74 (9): 979-983.

Fowles, J. 1978. *Handbook of futures research*. Greenwood Press: Connecticut.

Galton, F. 1907. Vox populi. *Nature* (75): 450-451.

Geier, J. D. 1985. The delphi survey methodology: an approach to determine training needs, Proceedings of the 8th Conference on Software Engineering Education, 389-402.

Goodwin, P. and Wright, G. 2010. The limits of forecasting methods in anticipating rare events, Technol. Forecast. Soc. Change 77 (3): 355-368.

Green, B., Jones, M., Hughes, D. and Williams, A. 1999. Applying the delphi technique in a study of GPs information requirements. *Health Soc. Care Community* 7 (3): 198-205.

Griffiths, T. L. and Tenenbaum, J. B. 2006. Optimal predictions in everyday cognition, *Psychological Science* 17 (9): 767-773.

Hartman, A. 1981. Reaching consensus using the delphi technique. *Educ. Leadersh*. 38 (6): 495-497.

Helle, M. R., Reijonwn, M. and Mannermaa, J. P. 2003. Using a delphi survey to assess the value of pharmaceutical process validation part 1: survey methodology. *Pharm. Technol. Europe*.

Hoest, M. and Wohlin, C. 1988. An Experimental Study of Individual Subjective Effort Estimations and Combinations of the Estimates. *Proceedings of the 20th International Conference on Software Engineering*, 332-339.

House, J. 2008. Riddles & Mazes: Transcending the Mundane, `http://www.jeffhouse.addr.com/mythology/riddles.htm.`

Kaynak, E. Bloom, J. and Leibold, M. 1994. Using the Delphi technique to predict future tourism potential, *Mark. Intell. Plann*. 12 (7): 18-29.

Keeney, S. Hasson, F. and McKenna, H. 2006. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *J. Adv. Nurs.* 53 (2): 205-212.

Kemerer. C. F. 1987. An empirical validation of software cost estimation models. *Communications of the ACM* 30 (5): 416-429.

King, R. 2003. *Michelangelo and the Pope's Ceiling*, Walker & Company, New York, NY.

Lewis, M. 1990. Liar's poker: rising through the wreckage on wall street, Penguin.

Linstone, H. A. and Turoff, M. (Eds.) 1975. The delphi method: techniques and applications, Addison-Wesley, Reading, MA.

Linstone, H. A. and Turoff, M. 2010. Delphi: A brief look backward and forward, *Technol. Forecast. Soc. Change* (in press).

Loughlin, K. G. and Moore, L. F. 1979. Using Delphi to achieve congruent objectives and activities in a pediatrics department. *J. Med. Educ*. 54 (2): 101-106.

Ludwig, B. G. 1994. Internationalizing Extension: An exploration of the characteristics evident in a state university Extension system that achieves internationalization. Unpublished doctoral dissertation, The Ohio State University, Columbus.

Mackay, C. 1980. *Extraordinary popular delusions and the madness of crowds*. Harmony Books, New York.

Moløkken, K. and Jørgensen, M. 2004. Group Processes in Software Effort Estimation, *Empir. Softw. Eng*. 9 (4): 315-334.

Passig, D. 1977. Imen-Delphi: a Delphi variant procedure for emergence, *Hum. Organ.* 56 (1): 53-63.

Surowiecki, J. 2005. *The Wisdom of Crowds*, Anchor Books, New York.

Turoff, M. 1970. The design of a policy Delphi, *Technol. Forecast. Soc. Change* 2 (2): 149-171.

Tversky, A. and Kahneman, D. 1974. Judgment Under Uncertainty: Heuristics and Biases, *Science* 185: 1124-1131.

Valerdi, R. 2008. *The Constructive Systems Engineering Cost Model (COSYSMO): Quantifying the Costs of Systems Engineering Effort in Complex Systems*, VDM Verlag.

Waissbluth, M. and De Gortari, A. 1990. A methodology for science and technology planning based upon economic scenarios and Delphi techniques, *Technol. Forecast. Soc. Change* 37 (4): 383-397.

Wolfers, J. and Zitzewitz, E. 2004. Prediction Markets, *J. Econ. Perspect*. 18 (2):107-126.

Biography

Ricardo Valerdi is a Research Associate in the Engineering Systems Division at the Massachusetts Institute of Technology. He is the co-Editor-in-Chief of the Journal of Enterprise Transformation and served on the Board of Directors of the International Council on Systems Engineering.