A PUBLICATION OF THE INTERNATIONAL INSTITUTE OF FORECASTERS

FORESIGHT

The International Journal of Applied Forecasting

Issue 8 ■ Fall 2007

THE ESSENTIAL READ FOR THE PRACTICING FORECASTER

www.forecasters.org/foresight

$40 per issue

GOOD AND BAD JUDGMENT IN FORECASTING
Lessons from Four Companies

METHODS TO ELICIT FORECASTS FROM GROUPS
The Delphi Method and Prediction Markets

NEW PERSPECTIVES ON THE COST OF FORECAST ERROR

PHARMACEUTICAL FORECASTING
How to Project Patient Persistency

THE KEYS TO THE WHITE HOUSE
Forecast for 2008

BAYESIAN MODELS FOR SHORT TIME SERIES

The IIF, now in its 27th year, is the leading non-profit clearinghouse of forecasting theory, research and practice.

A GUIDE TO DELPHI

Gene Rowe

PREVIEW

When we rely on the judgment of experts to help produce our forecasts, the key issues are how to get the appropriate information out of our experts and how to get a forecast if we are using multiple experts. Gene Rowe describes the Delphi method, what it offers to the forecaster, and what the pitfalls are in its implementation.

KEY POINTS

- Rather than assembling experts in one place at one time, the Delphi method deliberately separates them to reduce the negative aspects of group interaction. These problems include sucking up to the boss, the most outspoken people not being the most knowledgeable, social loafing (where some people remain anonymous and do not contribute to the debate), and the tendency of people within groups to discuss shared information rather than uniquely held information.
- Delphi uses multiple experts, who are polled via questionnaires on a number of occasions. The experts receive feedback on the anonymous views of the other experts. The forecast is usually taken as the average (e.g. mean or median) of panelists' forecasts on the final round.
- The result of a Delphi process is sometimes described as a *consensual* forecast. True *consensus* is not actually forced but greater consensus is usually achieved, in that the variance in responses decreases over the rounds, and the bases of agreement or disagreement among the panelists are clarified.
- The key for the practitioner is to enact the Delphi process in such a way that *information quality* becomes the key influencing variable.

INTRODUCTION

Articles on Delphi usually start (quite predictably!) by discussing the Oracle in Delphi in ancient Greece and its resident priestess (the Pythia), who dispensed forecasts from the gods. So here you go: I wouldn't wish to disappoint. However, the Delphi technique has little to do with ancient history, being a method developed at the RAND Corporation in the middle of the last century (Dalkey & Helmer, 1963), initially for predicting bombing requirements in a cold-war conflict with the USSR (the modern-day Persians?). In principle, the method is quite simple. The devil, however, is in the details – that is, in *how* and *when* to enact it precisely. In this article I will give a concise overview of the method. I will discuss how it is done, what it offers to the forecaster, and what the potential pitfalls are in its application.

THE DELPHI RATIONALE

In many forecasting situations, there is little relevant information from which a forecast might be extrapolated or a mathematical model produced. In other situations, there may be information, but it may not be of the right type or in an appropriate format, or it may be incomplete. Even if there is sufficient appropriate information for a statistical forecasting procedure, one might suspect that some dramatic change is imminent (e.g. the advent of a novel technology) that would make it dangerous to assume that past trends are a good basis for predicting future trends. In all of these informationally messy scenarios, there is little option but to rely upon the knowledge and abilities of experts – that is, to resort to *judgmental forecasting*.

Gene Rowe is Head of Consumer Science at the Institute of Food Research, Norwich, UK. His interests largely concern how individuals and groups make judgments and decisions and how to improve these. Aside from Delphi, his current research is focused on how people perceive risks and how and when to involve the public in policy-making decisions.

Experts are anticipated to have access to significant knowledge that enables them to predict the future – or at least, to do so better than non-experts. The use of experts does, however, raise various questions. First, how do we select them? How do we know that those we have chosen are the best and most appropriate? Second, how many should we ask? Should we attempt to find the one best expert and ask him/her, or should we seek the views of multiple experts? Third, how do we get the appropriate information out of our experts? Is there some best method to do this? And how do we get a forecast if we are using multiple experts?

The main rationale for Delphi is essentially to retain the merits associated with groups but attempt, through structuring the process more carefully, to remove the possibility of the negative aspects of group interaction.

The Delphi technique was developed as a method to specifically address the third question. The answers to the first two questions are not formally included within the Delphi package, although various Delphi proponents have suggested answers that might improve the effectiveness of the technique.

The method has at its heart the assumption that "N heads are better than one" (where N is an integer greater than one). Its proponents assume that experts have different amounts and types of relevant knowledge about the issue to be forecast and that a better forecast may be achieved by somehow acquiring and combining that information.

This strategy differs from a *best-member approach*, which would seek to identify one expert who was more knowledgeable and had better forecasting skills than the others (whatever such skills may be), and simply ask that expert to make a forecast. As such, Delphi is a group technique, but one with a difference. Rather than assembling experts in one place at one time, the method deliberately separates them for theoretically informed reasons, although the process has practical benefits too.

Much human decision making – particularly within organizations – takes place within groups. Logically, groups possess *at least* as much information/expertise as that held by the best individual, but they will usually possess *more* than this. Furthermore, groups have *process-gain* potential: it is possible that the interactions between group members may lead to solutions that are synergistically better than those of the best member, and better than might be achieved by simply averaging the judgments or opinions of all members.

Empirical research has tended to support the effectiveness of groups in judgmental, as opposed to forecasting tasks (e.g. Hill, 1982; Rowe, 1992). However, in practice, groups may suffer from a variety of problems that may undermine their general performance. This can lead to what is known as *process loss*, where performance is worse than might be achieved by simply choosing the best member or worse than the average of the individual members' judgments (these being the two main benchmarks for assessing group performance).

These problems are often social or political and relate to the kind of game playing that is commonly seen in groups (sucking up to the boss, undermining competitors, trying to maintain a positive social face, and so on). They may also be personality related: some people are simply more dogmatic and forceful within groups, while others are shyer, yet it is not clear that the most outspoken people are necessarily the most knowledgeable or have the best arguments.

There are also structural problems that can emerge from having multiple people within one room with limited time (after all, only one person can effectively speak at a time). With larger groups this may lead to processes such as *social loafing* by people who remain anonymous and do not contribute to the debate. In the latter case we might ask, are the available human resources being maximally utilized, when some do not speak, while others speak too much?

Social psychologists have studied group processes and discovered some interesting group trends. For example, they have found that group polarization often takes place, in which the prevailing trend in the group's individual opinions tends to become exaggerated and amplified during the group process (i.e. group views may become more extreme, rather than more considered). Further, research suggests that people within groups tend to discuss shared information rather than uniquely held information – a trend that would appear to undermine the positive potential of groups that emerges from having diverse experts.

The main rationale for Delphi is essentially to retain the merits associated with groups but attempt, through structuring the process more carefully, to remove the possibility of the negative aspects of group interaction.

The following list shows a number of business forecast applications that are described in the literature.

Examples of Practical Uses of Delphi

- Forecasting demand for, and supply of, academic veterinarians (Prince et al., 2006)
- Forecasting market prospects for organic food in Europe (Padel & Midmore, 2005)
- Forecasting the likelihood of changes in the international business environment over the next decade (in the context of globalization) and the impact of such changes on policy and corporate practices (Czinkota & Ronkainen, 2005)
- Forecasting the future development of Taiwan's machinery industry (Chang et al., 2002)
- Forecasting emerging technologies (Hala et al., 1998)
- Forecasting the state of electronics and information technology in India (Chakravati et al., 1998).

SO WHAT DOES DELPHI INVOLVE?

The Delphi method elicits judgments (including forecasts) of experts through questionnaires. These are the key features.

[1] Delphi panel members never meet and may take part in diverse locations. This allows key experts to participate without the expense and inconvenience of having to travel to a single location for a set time.

[2] Those taking part remain anonymous: no one in a Delphi panel knows who the other panel members are. This feature is intended to preempt the process-loss social and political aspects of groups, since the panelists are free to say what they think without any fear of offending their bosses or important peers, and with the payoff reduced for grandstanding and the like.

[3] Delphi allows equal contributions from experts: all those involved have an equal opportunity to have a say and be influential – and not just those with the loudest voices.

[4] The process has multiple rounds: those involved are asked their opinions on the questions of interest a number of times, allowing them to change their minds and thus allowing the judgment or forecast to evolve.

[5] The method involves feedback of group opinion on successive rounds, so that members can appreciate and be informed by what others think about the issue.

[6] Delphi provides an aggregate judgment or forecast at the end. After the final round, the output is determined by equal weighting of the views of all the panelists, thus providing a mean (or median) forecast.

The result of the final round is sometimes described as a *consensual* forecast. True *consensus* is not actually forced but greater consensus is usually achieved, in that the variance in responses decreases over the rounds, and the bases of agreement or disagreement among the panelists are clarified.

There is considerable latitude in how one might enact the Delphi process. To begin with, the number of rounds may be variable. The process sometimes starts with a relatively open first round, in which a moderator asks the expert panel to comment on the suitability of the questions to be asked. Alternatively the panel is asked a number of open questions from which more precise, quantitative questions may emerge for use in subsequent rounds.

Consider an example in which you want to predict the mortality from certain diseases (perhaps as a consequence of increasing incidence caused by climate change). In a first round, the panel could be asked to suggest which are the most relevant diseases. In subsequent rounds, the panel could be asked to make quantitative forecasts about the numbers likely to be affected or to rank the selected diseases as those most likely to be problematic.

There also is no reason why the process cannot be used over successive rounds to generate and clarify *qualitative* scenarios, rather than producing a *quantitative* forecast, although Delphi is most commonly used to produce a numerical output.

The number of quantitative rounds may also vary. There is no hard-and-fast rule as to when to stop: the usual heuristic is to continue polling until some stability in responses is achieved and there is little evidence of panelists changing their minds much further. Attempting to force an absolute consensus does not make sense, since a panelist who fundamentally disagrees with a forecast is not likely to change (especially given the removal of social conformity pressures), regardless of the number of rounds. Most change generally occurs within the first round or two (for example, see Brockhoff, 1975). The process moderator needs to be aware that experts are by their nature busy people and unlikely to want to continue with the process for very long.

Delphi implementations differ in how feedback is used. The form of feedback is limited by the type of responses acquired from panelists: if only numerical figures are requested (e.g. forecast amount, date, rating), this is all that can be used. Although Delphi feedback generally comprises only the mean or median of judgments/forecasts from previous rounds, qualitative information might be used as feedback as well, specifically, the opinions of those whose forecasts fall outside certain limits, like the upper and lower quartiles (Martino, 1983).

Statistics alone are unlikely to provide a particularly rich source of information to panelists and, in fact, may induce opinion change for the *wrong* reasons. Statistics without rationales may provide a compelling social magnet for responses (people converging their response to the norm for social reasons), when what is needed instead is *select* opinion change in response to novel or compelling information. In one of the few studies that have evaluated Delphi as a process, my colleague George Wright and I (Rowe & Wright, 1996) found evidence that feedback of reasons led to higher quality forecasts than provision of statistics alone.

Other sources of variety in Delphi include the number of panelists to use and the medium through which the process is enacted. There is no clear consensus as to what number of panelists is best. The rules that apply to normal interacting groups do not necessarily apply to Delphi groups, since the structural difficulties of hosting large numbers of people – who need to speak consecutively rather than being able to interact concurrently/asynchronously – do not apply to the latter. However, Delphi group size will be limited by the complexity of the topic and the nature of intended feedback. If only statistical feedback is expected to be used, then a larger panel can be accommodated than if feedback of arguments is intended, simply because the latter may require extensive information dispersal at the end of each round, making the process more time-consuming and onerous for the panelists (and moderators).

There also may exist only a limited amount of information on a particular issue. With increasing panel size there is decreasing additional utility from additional information – an important factor when one considers how busy one's panelists may be. It is likely better to use a smaller number of dedicated experts (e.g. a dozen or less) and elicit and dispense greater informational richness from them than to use a large number of experts and take and give simple numerical responses.

Regarding process, Delphi may be enacted as a paper questionnaire sent via post or as an electronically distributed event. There are a number of dedicated Delphi programs that do exist, but the method may still be conducted using e-mail. Each method has its own largely practical advantages and disadvantages related to the computer literacy of the Delphi moderators, but there is no clear theoretical reason for preferring one mode over another.

HOW WELL DOES DELPHI WORK?

If you search the main publications databases, you will find cases in which the technique has been used *practically* to address a certain problem with multiple experts. But you will find few articles that actually address the Delphi process in any *evaluative* or *critical* sense. In other words, there is relatively little research into what goes on during the Delphi process and whether the method is helpful to forecasters and decision makers.

There is some evidence, however, that Delphi may be an effective process. In a review of research studies on Delphi, Rowe and Wright (1999) found that the practice does tend to result in improved judgments (including forecasts) against two key benchmarks: comparative interacting groups and the statistical aggregation of experts. That is, accuracy tends to increase from the first, pre-feedback Delphi round to the final round. However, the total number of studies reviewed in this article was small and the studies were typically simplistic, involving groups of students making judgments about the answers to almanac questions, rather than actual experts making complex forecasts. On the other hand, there are reasons to suspect that Delphi is likely to be *more* effective in more-real scenarios than in the rather artificial laboratory ones used in the studies reviewed in the Rowe and Wright article.

The key for the practitioner is to enact the Delphi process in such a way that *information quality* becomes the key influencing variable.

Importantly, what we really need to understand is *why* Delphi should work. Parenté and Anderson-Parenté (1987) attempted to answer this question by recourse to the so-called "Theory of Errors." Their interpretation of this theory is that, during Delphi, the more knowledgeable experts will hold out while the less knowledgeable will swing; that is, they will move their judgments towards the mean of the panel on subsequent rounds. If this happens, then it can be mathematically demonstrated that, over rounds, the judgmental mean will edge from that of the first round towards the true value, and hence accuracy will increase.

There is some empirical support for this theory (Rowe & Wright, 1996), but there are difficulties with its likely validity in the real world. The theory might well explain why judgmental enhancement occurs during Delphi in cases where a clear separation between the more- and less-expert is apparent to panelists (as arguably is the case in many of the artificial laboratory studies). But in situations of large issue uncertainty (such as forecasting scenarios), and in panels totally comprised of high-level experts with diverse areas of expertise, the likelihood that panelists will be able to appropriately identify their relative level of knowledge and then to appropriately swing or hold out seems far from guaranteed. Hence judgmental enhancement may not *necessarily* follow.

Moreover, if you could identify the best experts in advance, why would you wish to include less-good experts in your panel (i.e. those who will subsequently swing, leading to improved judgment in subsequent rounds)? In Delphi enactments, other factors will explain who changes their opinions during the process – such as personality, confidence, and revealed majority opinion (e.g. Rowe et al., 2005). Thus whether judgmental enhancement takes place will depend upon whether such factors correlate with actual expertise or not.

The key for the practitioner is to enact the Delphi process in such a way that information quality becomes the key influencing variable. Unfortunately – and as previously discussed – many Delphis provide feedback that is unlikely to be all that helpful in this regard, but the lack of controlled evaluation of Delphi studies undermines our ability to learn about and advance the method.

PRACTICAL ISSUES OF IMPLEMENTATION

Although Delphi is a potentially valuable method, it is not without its drawbacks. Conducting a Delphi can sometimes be a complex, time-consuming, and frustrating business. Perhaps the main difficulty is in identifying and recruiting suitable experts to take part and in keeping them on board during a process that can take weeks or even months, and one that may require significant commitment. Below are a few hints that may help the practitioner (and see Rowe & Wright, 2001).

- Be clear on the research question: what exactly are you trying to forecast? What are you *not* trying to forecast? Define key issues and parameters.
- Put a lot of effort into identifying *appropriate* experts – people with real expertise (not just high-status people who have acquired a senior position through age, patronage, social skills, force of personality, etc.) who are willing to commit to the process.
- Over-recruit experts: it is highly likely that, irrespective of any initial commitments voiced, there will be a significant dropout rate.
- Get the experts on board before approaching them with the first round questionnaire.
- Follow standard rules for good questionnaire design (as can be found in any good social science text book) – such as avoiding leading questions or confusing or emotive terminology.
- Think carefully about the nature of feedback you are going to provide, as this will have an impact on the information you will need to elicit in your questionnaire. It is good to elicit rationales as well as numerical data – but the more complex and qualitative the information gained, the more time-consuming it will be for panelists to consider the feedback in subsequent rounds, and the smaller the panel you should use.
- Do not use too many rounds unless you wish to alienate your panel. Two structured rounds may suffice if you are also using an unstructured first round. There is probably little point in going beyond three structured rounds.
- If using an electronically implemented Delphi, do a trial run on the questionnaire and process, and then run it again (even if using a dedicated Delphi programme). Things *will* go wrong.

- Assume that panelists will not respond to deadline, and allow for sending reminders at set times in the process.
- Assume that the process will take much longer than your initial estimates.

CONCLUSIONS

The Delphi technique is a judgmental forecasting process for use when insufficient data is present to construct any form of statistical model. It uses multiple experts, who are polled via questionnaire on a number of occasions, with feedback on the anonymous views of the panel presented on later rounds, and with the forecast taken as the average (e.g. mean or median) of panelists' forecasts on the final round.

Apart from the practical advantage of allowing diverse experts to be consulted without having to bring them together in one place at one time, experimental evidence suggests that the process *does* lead to enhanced judgments with respect to traditional groups. The likely mechanism for Delphi's success is that the structured process pre-empts group process-loss difficulties due to social/political factors and allows panelists to focus on the quality of information instead.

However, many Delphis provide limited feedback (means/medians only), which may undermine the utility of the process. There are also many practical difficulties in the implementation of a Delphi. It is up to the academic community to conduct more and better evaluative research on the method so that we can determine how best to enact the process and clarify when it is most likely to be of use to practitioners. There is certainly scope for further modernization of the process.

REFERENCES

Brockhoff, K. (1975). The performance of forecasting groups in computer dialogue and face-to-face discussions. In H. Linstone & M. Turoff (Eds.), *The Delphi Method: Techniques and Application,* London: Addison-Wesley.

Chakravati, A.K., Vasanta, B., Krishnan, A.S.A. & Dubash, R.K. (1998). Modified Delphi methodology for technology forecasting: Case study of electronics and information technology in India, *Technological Forecasting and Social Change*, 58 (1-2), 155-165.

Chang, P.C., Wang, C.P., Yuan, B.J.C. & Chuang, K.T. (2002). Forecast of development trends in Taiwan's machinery industry, *Technological Forecasting and Social Change*, 69 (8), 781-802.

Czinkota, M.R. & Ronkainen, I.A. (2005). A forecast of globalization, international business and trade: Report from a Delphi study, *Journal of World Business*, 40 (2), 111-123.

Dalkey, N. & Helmer, O. (1963.) An experimental application of the Delphi method to the use of experts, *Management Science,* 9, 458-474.

Hala, W.E., Kull, M.D. & Leffman, A. (1998). The George Washington University forecast of emerging technologies: A continuous assessment of the technology revolution, *Technological Forecasting and Social Change*, 59 (1), 89-110.

Hill, G.W. (1982). Group versus individual performance: Are *N + 1* heads better than one? *Psychological Bulletin,* 91 (3), 517-539.

Martino, J. (1983). *Technological Forecasting for Decision Making* (2nd ed.), New York: American Elsevier.

Padel, S. & Midmore, P. (2005). The development of the European market for organic products: Insights from a Delphi study, *British Food Journal*, 107 (8), 626-647.

Parenté, F.J. & Anderson-Parenté, J.K. (1987). Delphi inquiry systems. In G. Wright & P. Ayton (Eds.), *Judgmental Forecasting,* Chichester: Wiley.

Prince, J.B., Andrus, D.M. & Gwinner, K. (2006). Academic food-supply veterinarians: Future demand and likely shortages, *Journal of Veterinary Medical Education*, 33 (4), 517-524.

Rowe, G. (1992). Perspectives on expertise in the aggregation of judgments. In G. Wright & F. Bolger (Eds.), *Expertise and Decision Support*, London: Plenum, 155-180.

Rowe, G. & Wright, G. (2001). Expert opinions in forecasting: Role of the Delphi technique. In J.S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Norwell, MA.: Kluwer Academic Publishers, 125-144.

Rowe, G. & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis, *International Journal of Forecasting*, 15, 353-375.

Rowe, G. & Wright, G. (1996). The impact of task characteristics on the performance of structured group forecasting techniques, *International Journal of Forecasting*, 12, 73-89.

Rowe, G., Wright, G. & McColl, A. (2005). Judgment change during Delphi-like procedures: The role of majority influence, expertise and confidence, *Technological Forecasting and Social Change*, 72, 377-399.

CONTACT
Gene Rowe
Institute for Food Research, UK
gene.rowe@BBSRC.AC.UK