University of Pennsylvania
ScholarlyCommons

Marketing Papers

Wharton School

2001

Expert Systems for Forecasting

Fred Collopy
Case Western Reserve University

Monica Adya
University of Maryland - Baltimore

J. Scott Armstrong
University of Pennsylvania, armstrong@wharton.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/marketing_papers

Part of the Marketing Commons

Recommended Citation

Collopy, F., Adya, M., & Armstrong, J. S. (2001). Expert Systems for Forecasting. Retrieved from http://repository.upenn.edu/marketing_papers/151

Suggested Citation:
Collopy, F., Adya, M. and Armstrong, J.S. Expert Systems for Forecasting. In *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Ed. J. Scott Armstrong). Kluwer, 2001.

Publisher URL: http://www.springer.com/business+%26+management/business+for+professionals/book/978-0-7923-7930-0

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/marketing_papers/151
For more information, please contact libraryrepository@pobox.upenn.edu.

Expert Systems for Forecasting

Abstract
Expert systems use rules to represent experts' reasoning in solving problems. The rules are based on knowledge about methods and the problem domain. To acquire knowledge for an expert system, one should rely on a variety of sources, such as textbooks, research papers, interviews, surveys, and protocol analysis. Protocol analysis is especially useful if the area to be modeled is complex or if experts lack an awareness of their processes. Expert systems should be easy to use, incorporate the best available knowledge, and reveal the reasoning behind the recommendations they make. In forecasting, the most promising applications of expert systems are to replace unaided judgment in cases requiring many forecasts, to model complex problems where data on the dependent variable are of poor quality, and to handle semi-structured problems. We found 15 comparisons of forecast validity involving expert systems. As expected, expert systems were more accurate than unaided judgment, six comparisons to one, with one tie. Expert systems were less accurate than judgmental bootstrapping in two comparisons with two ties. There was little evidence with which to compare expert systems and econometric models; expert systems were better in one study and tied in two.

Disciplines
Business | Marketing

Comments
Suggested Citation:
Collopy, F., Adya, M. and Armstrong, J.S. Expert Systems for Forecasting. In *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Ed. J. Scott Armstrong). Kluwer, 2001.

Publisher URL: http://www.springer.com/business+%26+management/business+for+professionals/book/978-0-7923-7930-0

This book chapter is available at ScholarlyCommons: http://repository.upenn.edu/marketing_papers/151

Principles of Forecasting: A Handbook for Researchers and Practitioners,
J. Scott Armstrong (ed.): Norwell, MA: Kluwer Academic Publishers, 2001

Expert Systems for Forecasting

Fred Collopy
The Weatherhead School of Management
Case Western Reserve University

Monica Adya
Department of Information Systems
University of Maryland - Baltimore

J. Scott Armstrong
The Wharton School
University of Pennsylvania

ABSTRACT

Expert systems use rules to represent experts' reasoning in solving problems. The rules are based on knowledge about methods and the problem domain. To acquire knowledge for an expert system, one should rely on a variety of sources, such as textbooks, research papers, interviews, surveys, and protocol analysis. Protocol analysis is especially useful if the area to be modeled is complex or if experts lack an awareness of their processes. Expert systems should be easy to use, incorporate the best available knowledge, and reveal the reasoning behind the recommendations they make. In forecasting, the most promising applications of expert systems are to replace unaided judgment in cases requiring many forecasts, to model complex problems where data on the dependent variable are of poor quality, and to handle semi-structured problems. We found 15 comparisons of forecast validity involving expert systems. As expected, expert systems were more accurate than unaided judgment, six comparisons to one, with one tie. Expert systems were less accurate than judgmental bootstrapping in two comparisons with two ties. There was little evidence with which to compare expert systems and econometric models; expert systems were better in one study and tied in two.

Keywords: inductive techniques, judgmental bootstrapping, knowledge acquisition, production systems, protocol analysis, retrospective process tracing.

Imagine trying to predict the proper medical treatment for a patient. If you have a good measure of success and a substantial record of results from prior cases, you could develop an econometric model to predict which treatment would produce the best results. Based on extensive research over the past half century, the econometric approach would almost always be more accurate than unaided judgment, and as shown in an extensive review by Grove and Meehl (1996), it is difficult to find exceptions to this conclusion. But what if the treatment options include new procedures? You could ask experts to make forecasts about the chances of success for each new procedure. When many forecasts are needed, that approach can be expensive and inconsistent. In such situations, it is useful to develop a model of an expert, then use the model to make forecasts.

One way to develop a model of an expert is to *infer* the rules that experts are using. This approach, called judgmental bootstrapping, is nearly always more accurate than an unaided expert (Armstrong 2001b). Alternatively, you could study how an expert makes predictions and develop an expert system to represent the process. In other words, knowledge is elicited directly from the expert. The focus of this paper is on expert systems for forecasting.

Expert systems have been used to forecast time series and to predict outcomes on the basis of cross-sectional data. Armstrong, Adya and Collopy (2001) discuss their application to time series forecasting. In this paper, we examine the use of expert systems for cross-sectional prediction.

DEVELOPING EXPERT SYSTEMS FOR FORECASTING

Building expert forecasting systems consists of three tasks: acquiring relevant knowledge, structuring and applying the knowledge, and testing the system. We discuss the first two of these tasks here. The third task, testing expert systems, is mentioned only briefly because, in general, it requires the same testing procedures as testing other types of forecasting methods. Those procedures are discussed in Armstrong (2001a).

ACQUIRE RELEVANT KNOWLEDGE

The first task in developing an expert system is to acquire relevant knowledge. This is considered the most difficult aspect of developing expert systems, and is often referred to as the knowledge acquisition bottleneck. Acquiring knowledge can be difficult because performing a task is often second nature to experts and they may find it difficult to describe how they do the task.

Some fields lend themselves more readily to knowledge acquisition than others. In rare situations, such as in law cases, detailed documentation might be available on how decisions or predictions were made. Kort (1957) used such knowledge to develop an expert system for the U. S. Supreme Court's decisions on right-to-counsel cases. More generally, you will need to elicit knowledge from experts directly by using questionnaires, in-depth interviews, or retrospective process tracing.

- **Ask experts to describe the rules they use.**

Developers of expert systems should seek knowledge from expert forecasters. There are a number of ways to do this, as listed in Exhibit 1. To find out how this is typically done, Doukidis and Paul (1990) surveyed members of the Operational Research Society, receiving replies from 26% of their sample. Eighty-five percent of the 265 respondents indicated that their organizations had developed expert systems.

Exhibit 1
Knowledge Acquisition Procedures for Developing Expert Systems

	Percentage of developers using*
Interviewing experts	100
Questionnaires	—
Taking experts through case studies	28
Use of induction techniques [judgmental bootstrapping]	18
Retrospective process tracing	—
Recording experts at work [protocols]	16
Literature reviews	—

* The percentages were adapted from Doukidis and Paul (1990)

All of the developers Doukidis and Paul (1990) surveyed said that they had used *interviews* of experts. Such interviews can provide information about the variables experts use to make predictions and the relative weights they assign to the variables. Interviews are most useful when the rules are clear and the expert is aware of the problem-solving process. Even when the process is not easy to explain, however, interviews may help to initially identify important variables.

Collopy and Armstrong (1989) found that when asked about extrapolation methods, experts spoke for about 30 minutes. These interviews were related to general extrapolation strategies. One would expect that interviews directed to a specific problem would provide a richer source of information. This was the case when McClain (1972) asked doctors to describe how they made decisions about using medical resources in specific situations and how they assigned weights to various factors. The task took about three hours per doctor. In addition, McClain spent much time analyzing the recordings.

Although not mentioned by Doukidis and Paul (1990), *questionnaires* are useful when problems are well defined. They provide a cost-effective way to obtain information from many domain experts. Open-ended questionnaires can provide information about general strategies used by expert forecasters. Structured questionnaires are more efficient in that analysts can obtain information about strategies used under various conditions.

In r*etrospective process tracing*, experts are asked to reconstruct the thought processes they used when making judgments. This was not specifically examined by Doukidis and Paul (1990). In some cases, it might yield better descriptions than could be obtained by inferring the process, as shown by Larcker and Lessig (1983). In their study of decisions to purchase stocks, they compared an expert system based on retrospective process tracing to a judgmental bootstrapping model. Thirty-one subjects provided decisions for 45 stocks, based on six information cues (average price, change in earnings per share, average change in earnings per share, dividend yield, debt/equity, and beta). The names of the stocks were not revealed to the subjects. Subjects needed about 30 minutes to provide the retrospective process descriptions. The average number of cases for which the expert systems reproduced the decision was 84.8 %, while for the bootstrapped models it was 73.0 %.

For complex tasks, retrospective process tracing may be ineffective because it relies on memory. Also, experts may remember themselves as being more rational than they were. For example, what personnel managers would remember taking into account height, weight, gender, looks, dress, and accent when describing how they selected the most promising candidates for an office job?

- **When experts lack awareness of their thought processes or when the process is complex, use protocols.**

Sometimes experts have difficulty explaining how they make their judgments. They often refer to their most important judgments as intuitive. Consequently, they may be unable to reveal what knowledge they use and how they use it. Cocozza and Steadman (1978) found that psychiatrists acting as expert witnesses did not have a good understanding of how they made predictions about the potential dangerousness of defendants in court cases. Protocols can be useful in such situations.

In protocol analysis, an expert is asked to think aloud while engaged in a diagnostic process, such as making forecasts. Protocol sessions with experts yield more detailed and specific information about rules than can generally be obtained from interviews. In the Doukidis and Paul (1990) survey, only about 16% of the developers used protocols.

Protocol analysis requires more time than interviewing. In one protocol study, Kleinmuntz (1968) reported tape recording for 60 hours to construct a model of a single decision maker. Clarkson (1962) devoted his entire PhD dissertation to the process a single investment trust officer used to select stocks for an investment portfolio. Because protocols can be expensive, one should use them only if necessary.

- **Incorporate knowledge from empirical literature.**

Research on decision making suggests that experts are good at identifying characteristics of a decision situation. However, they are not able to keep complex relationships straight when there are many variables and they have difficulty assessing the magnitude of relationships. Also, they may see what they expect to see. Chapman and Chapman (1969) asked 32 experts to examine data from homosexual and heterosexual subjects. They contrived the data so that relationships the experts expected did not exist. The clinicians had great difficulty in seeing valid relationships in the data even though their effects were large. Instead, they saw the relationships they expected.

To overcome these shortcomings in experts' perceptions, the developer of an expert system can draw upon information from econometric studies. Econometric relationships are typically more valid than those surmised by an expert. They have the advantages of being organized and tested, and they are also more likely to be free of biases. Allen and Fildes (2001) describe econometric procedures.

Meta-analyses of findings from econometric studies are especially valuable. For example, assume that an expert system is needed to forecast how price changes would affect the sales of a technical book. An expert's judgment could be supplemented by a meta-analysis such as the one by Tellis (1988), who reported price elasticities for a variety of products and conditions.

- **Use multiple sources of knowledge.**

A single expert can serve as a starting point, but where possible, you should consult additional experts, perhaps as many as five. Leonard's (1995) system to detect bank fraud was based on interviews with twelve experts. Often, however, researchers do not use multiple sources. Abramson et al. (1996), Clarkson (1962), Moss, Artis and Ormerod (1994), and Stewart et al. (1989) all based their systems on a single expert's input. The use of a single expert may provide an incomplete representation of the knowledge.

Knowledge from prior research can be combined with the knowledge from experts. Greater weight should be placed on empirical results to the extent they are reliable and valid. On the other hand, more weight should be placed on experts' knowledge to the extent that they receive good feedback about their forecasts.

In developing expert systems, one can also draw upon judgmental bootstrapping for knowledge about relationships. This can be helpful when experts make good forecasts but lack awareness of how they are making them. Reagan-Cirincione (1994) followed this procedure. She used structured group procedures to help experts compare the models they described with the estimates from their own bootstrapping models. By focusing on the differences, she was able to make revisions that improved the expert system's accuracy.

Conjoint analysis is still another source of knowledge about relationships (see Wittink and Bergestuen 2001). It is useful when participants can provide good assessments of how they would respond to changes. It can be especially useful in cases where experts are not able to describe how participants might react to changes.

STRUCTURING AND APPLYING THE KNOWLEDGE

Once you gather knowledge from experts, you should represent it so that it can be easily used. The most common way to represent knowledge and expertise in expert systems is as production rules. Production rules are condition-action statements, such as "IF credit history is poor, THEN do not approve loan." In their survey of operations researchers, Doukidis and Paul (1990) found that 62% of their respondents used such rules to represent knowledge in their expert systems.

- **Strive for simplicity.**

In attempting to realistically represent what experts do, there is a danger that the system might become too complex. The interaction effects of many simple rules can be difficult to understand unless the rules are well organized. To avoid overly complex systems, design production systems to make it easy to examine existing rules and to revise them or add new rules.

An expert system should not impose cognitive strain on its users. In an expert system developed by one of the authors to predict the effectiveness of advertisements, an analyst had to rate as many as 235 features of an advertisement. One way to reduce strain is to structure the system so that its organization is intuitive. Another is to make reasonable assumptions about defaults that apply to most common situations and alter these only as needed.

- **Strive for completeness.**

The knowledge encoded in an expert system should represent all key aspects of the problem because users of the system are likely to assume that it is comprehensive. Fischhoff, Slovic and Lichtenstein (1978) studied the use of a fault tree for the maintenance of automobiles. A fault tree describes the paths one can follow to diagnose a problem. For example, one could use a fault tree to diagnose why a car does not start properly (e.g., check battery). They found that subjects, including experts, tended to ignore things left out of the tree. They overlooked omitted conditions even when they probably would not have overlooked them had they relied on unaided judgment. The authors concluded that once a decision aid is adopted, “out of sight was out of mind.”

Dijkstra (1995) conducted an experiment to determine whether experts can be easily misled when an expert system is incomplete. He constructed two expert systems for judging whether a defendant was guilty of a criminal attempt. While both expert systems were logically correct, each omitted critical information; one focused on the act while the other was based on intent. Dijkstra presented nine cases to 30 law students and 33 law professors. Typical case: “Mr. Smith has been increasingly upset at the noise from his neighbor’s motorcycle. One afternoon, he sees it unattended, so he steals it and dumps it into a river. Unknown to him, Mr. Smith’s wife had purchased the bike for him earlier that day.” The two expert systems gave opposite advice on each case. Decisions by the subjects, especially the lawyers, were highly influenced by the expert systems that they were given.

- **Fully disclose the knowledge in the system.**

If users know what research and knowledge are included in the system, they should be better able to judge when it can be used. In addition, other researchers can build upon disclosed expert systems, rather than starting from scratch. Full disclosure also makes it easier to resolve inconsistencies in the rules and allows users to learn from knowledge encoded in the system. Finally, full disclosure of the knowledge used in an expert system allows for judging the face validity of the system.

- **Explanations should be provided by the expert system.**

Expert systems should explain why they make particular recommendations. Complete and well supported explanations provide a way of examining the face validity of the knowledge in the expert system. They may also increase the user’s confidence in the system. In addition, explanations can help the analyst to learn about the process. Finally, explanations may help forecasters to spot situations in which the expert system is not relevant. While it is desirable to provide explanations, the designer should recognize that it is difficult to get people to use them.

TESTING EXPERT SYSTEMS

For the most part, testing an expert system is like testing any forecasting method (Armstrong 2001a). However, a test of face validity, the Turing test (Turing 1950), has been used to compare outputs.

- **Use the Turing test to assess expert systems that replace judgment.**

The Turing test examines whether a panel of experts can distinguish differences in outputs from an expert system and an expert. The panel could present problems and request forecasts, along with explanations, from the expert system and from unaided experts. Based on the responses, the experts on the panel are asked to identify which forecasts come from the expert and which come from the expert system.

Conducting a Turing test is appropriate when comparative accuracy of different methods is difficult to assess, the problem involves much uncertainty, and the prediction problem is complex. For example, when doctors must predict what types of treatment are most appropriate for a patient and they have no prior outcome measures, it is useful to know whether the expert system can produce predictions that are similar to those of the best experts.

CONDITIONS FAVORING USE OF EXPERT SYSTEMS

Expert systems are expensive to develop, so it is important to identify the conditions under which they will be useful.

X **Relative to other forecasting approaches, expert systems are best suited to situations in which:**

. . . experts make repetitive forecasts.

Because expert systems are costly to develop, their use makes most sense when many forecasts are needed. This occurs for many problems, such as: Which drilling sites are most likely to yield oil at reasonable cost? What products are most likely to be profitable in the current market? Which drug treatments will be the most successful for particular patients?

. . . the problems are semi-structured.

The kinds of problems that are most likely to benefit from the use of expert systems are those that are semi-structured. In contrast, for problems that are well structured, statistical techniques (such as regression) can provide good forecasts, while problems that are highly unstructured cannot be translated into rules.

. . . historical data on the dependent variable are unavailable or of poor quality.

When there is not much historical data on the dependent variable or when these data are of poor quality, expert systems may help. They are also expected to be applicable where the underlying processes are subject to changes that are apparent to the experts.

. . . cooperative experts are available.

The development of expert systems depends upon having willing and cooperative experts. It may require extensive time with the expert to develop rules that cover all conditions.

EVIDENCE ON THE EFFECTIVENESS OF EXPERT SYSTEMS

Our search for evidence relied heavily upon citations in papers and books. Some references were provided by researchers specializing in expert systems. Computer searches of the *Social Science Citation Index* and *Social Science Index* were made through early 2000. Using the term "expert systems and forecasting," we located 51 studies. Of these, only two were relevant to our paper, and they were only tangentially related. Requests for help were posted on the forecasting principles website and were sent to e-mail lists, but these produced no additional studies. Given the massive literature on expert systems, it is interesting that we found only about 35 studies directly relevant to the use of expert systems in forecasting.

Our search was difficult because many researchers claiming to use expert systems did not use systems that fit our definition. For example, Moore (1998) claimed to use an expert system to predict the performance of MBA candidates based on information available in their applications. He induced rules using a statistical procedure that related attributes to prior decisions. We would consider this to be a type of econometric method. On the other hand, some who used systems that conformed to our definition did not refer to them as expert systems.

One reason for the small number of papers is that forecasting is only one of many uses for expert systems. Wong and Monaco (1995), in their review of the literature between 1977 and 1993, concluded that out of ten uses mentioned, prediction was the fifth most important use of expert systems. It was well behind planning and monitoring, and about equal with design and interpretation. Eom (1996), in his review of 440 papers on expert systems published between 1980 and 1983, found that only 17 (4%) were applied to forecasting problems.

Of the papers that used expert systems for forecasting, few directly examined comparative forecast validity. The small number of validation studies that we found is consistent with Santhanam and Elam's (1998) survey of knowledge-based systems research in decisions sciences. They found only ten validation studies among the 430 studies of expert systems published in major management journals between 1980 and 1995.

Overall, we found 15 comparisons on the predictive validity of expert systems. This lack of validation testing is perhaps the major conclusion from our search. Even if one were using expert systems for other purposes, such as design or planning, it would be useful to show that they had predictive validity.

We anticipated that researchers would find that expert systems were more accurate than unaided judgment, if only because they use structured knowledge. In comparison with judgmental bootstrapping, we had no prior hypothesis because we could formulate hypotheses favoring either approach. We anticipated that econometric models would be more accurate than expert systems in well-structured situations. This is because they make better use of information. Exhibit 2 summarizes the comparisons. It starts with situations where expert systems were expected to be more accurate.

Exhibit 2
Comparative Accuracy of Expert Systems

Expert Systems vs.:	Study	Task	Criteria	Results
Judgment				
ES Better	• Reagan-Cirincione (1994)	Teachers' salaries	Correlation	ES much more accurate
	• Reagan-Cirincione (1994)	Baseball team records	Correlation	ES much more accurate
	• Kleinmuntz (1967)	Student adjustment	Classification errors	ES error less by 16%
	• Smith et al. (1996)	Gas demand	Mean absolute deviation	ES error less by 10%
	• Michael (1971)	Mail order catalog sales	Sales volume	ES error less by 5%
	• Silverman (1992)	Army equipment capability	Bias	ES eliminated bias
Similar	• Stewart et al (1989)	Weather (hail)	Correlation	—
ES Worse	• Leonard (1995)	Credit card fraud	Classification errors	ES detected 71% fraud; experts detected 80%
Bootstrapping				
ES Better	—	—	—	—
Similar	• Yntema & Torgerson (1961)	Artificial task	Correlation	—
	• Einhorn et al. (1979)	Nutrition	Classification errors	—
ES Worse	• Schmitt (1978)	Academic success	Correlation	ES less accurate
	• Einhorn et al. (1979)	Psychological adjustment	Classification errors	ES 73% correct vs. bootstrapping 91%
Econometric Model				
ES Better	• Leonard (1995)	Credit card fraud	Classification errors	ES detected 71% fraud; AID detected 66%
Similar	• Moninger et al. (1991)	Weather	Correlation	—
	• Stewart (1989)	Weather (hail)	Correlation	—
ES Worse	—	—	—	—

Comparisons with Judgment

The overwhelming superiority of judgmental bootstrapping over judgment has been thought to result largely from its greater consistency (Armstrong 2001b). Given that expert systems also provide consistency to the forecasting process, one would expect that they would also be more accurate than unaided expert forecasts. As it turned out, they were more accurate in six comparisons, tied in one, and worse in one.

In a paper that does not use the term *expert system*, Reagan-Cirincione (1994) provides an excellent example of the application of principles for the development of an expert system. She asked judges to describe how they would make predictions for two problems, the first being to predict the average teacher's salary in each of the 50 states, and the second to predict the number of games a baseball team won during a season. Using this information, she developed expert systems. She asked her judges to make predictions for a sample of cases. The expert systems were much more accurate than the judges' direct predictions.

Kleinmuntz (1967) employed protocols to code the rules an expert used to predict how subjects who had sought counseling would adjust to college life. The rules were based on information from a psychological inventory (the MMPI). Kleinmuntz's comparison between clinicians and the expert system used data on 720 students from five colleges. Eight clinicians, all with reputations for their skills at interpreting MMPI results, misclassified 34.4% of the cases. In contrast, the expert system missed on only 28.8% — a reduction of 16.3% in the error.

In a study of gas demand, Smith, Hussein and Leonard (1996) described an expert system that British Gas used to forecast short-term demand. Expert knowledge from 72 geographically scattered shift officers was obtained through structured interviews, questionnaires, and retrospective case descriptions. The expert system proved to be more accurate than forecasts from the shift control officers.

Michael (1971) developed an expert system based on the rules used by an expert who forecasted catalogue sales. He obtained the rules by asking the expert to explain how he had made specific forecasts in the past (i.e., retrospective process tracing). The forecasting task, which involved 42 items, was difficult because it was performed at the beginning of the season before any feedback had been received on sales. In terms of unit sales, the average error for the expert was 28.8%. The expert system was more accurate with an average error of 27.1% — a reduction of almost six percent. He obtained similar results when using sales dollars as the criterion (a 4.3% reduction in error). The biggest improvements were achieved for the "major new articles," but there were only three of these.

In his study of army planners, Silverman (1992) compared an expert system to unaided judgment. Silverman developed a system to help analysts predict how new equipment would perform in various environments. The system was designed to remove biases from these forecasts by identifying the application of irrelevant or overlooked knowledge. Protocol sessions with one expert helped Silverman to identify recurring biases. He found biases in each of the 22 assessments in which subjects did not use the expert system. When nine subjects repeated the task using the expert system, unaware of the biases in their earlier answers, none of their forecasts contained biases.

Stewart et al. (1989) compared an expert system with judgment. The expert system, developed from conversations with only one expert, consisted of 250 rules based on seven cues. Stewart et al. presented seven meteorologists with Doppler radar scans of 75 storms, and each made probability forecasts of hail and severe hail. Forecasts from the expert system were a little less accurate than all but one of the experts for forecasts of hail, and a bit more accurate than all but the best of the experts for forecasts of severe hail. Relative to judgment then, the expert system's performance was mixed.

Leonard (1995) described an expert system for detecting bank fraud. Twelve bank managers were involved in developing the rule base. It made use of a dozen predictor variables, such as the number of authorizations at the same merchant and the current balance as a percent of limit. Examples of the rules used are "If there have been previous purchases within the last 24 hours at the same merchant, then call customer" and "If the time since the last transaction is less than 30 minutes, then investigate further." The resulting expert system had a slightly higher overall accuracy than the classifications of the bankers themselves (92% vs. 90%). However, with respect to the primary function of the model, it identified only 71% of actual frauds, compared with 80% by the experts.

Comparisons with Judgmental Bootstrapping

Yntema and Torgerson (1961), using an artificial task, provided pictures of 180 ellipses to six judges. The ellipses were assigned values based on their size, shape, and color. The worth of an item increased with size, thinness, and brownness. Yntema and Torgerson developed an expert system for each judge by asking judges what weights they placed on each of the three variables. The resulting models were as accurate as those based on judgmental bootstrapping.

Schmitt (1978) asked 112 students to predict the academic success of a group of subjects based on four variables. For this problem, students were expected to have some expertise. The data used were contrived (simulated). After practicing on 20 "applicants," the students made comparisons for 30 new "applicants." Three different approaches for asking questions led to expert systems of comparable accuracy. These expert systems were a bit less accurate than judgmental bootstrapping.

Einhorn, Kleinmuntz and Kleinmuntz (1979) compared judgmental bootstrapping and process-tracing models in two experiments. The first experiment was to assess the degree of adjustment of 96 students based on a psychological assessment (MMPI profiles). Each profile contained 16 variables. The judges sorted the students into 12 categories based on their degree of adjustment. The researchers used process tracing with one judge to develop an expert system. A four-variable judgmental bootstrapping model did a much better job of modeling the actual judgment than the expert system. It had 9 misclassifications of 65 students, versus 26 misclassifications for the expert system. In their second experiment, a single subject rated the nutritional quality of breakfast cereals on the basis of 11 cues. A protocol analysis produced seven rules for the expert system. The resulting judgmental bootstrapping and expert systems had similar accuracy.

Comparisons with Econometric Methods

Expert systems might have advantages over econometric models because they can handle messier problems. But econometric models typically make more effective use of information on the dependent variable for problems that are well structured.

Leonard (1995), in a study of credit-card fraud, examined predictions for 12,132 accounts. Although an econometric model (developed using Automatic Interaction Detector, or AID) was slightly more accurate overall, the expert system was more effective for fraud cases (71% correct versus 66%).

Moninger et al. (1991) evaluated systems based on artificial intelligence to forecast severe storms. Three of these systems were traditional expert systems, another was a hybrid system including a linear model augmented by a small expert system, and two others were based on linear (econometric-type) models. On each day of a three-month test, the systems generated 2 to 9-hour forecasts of the probabilities of occurrence of nonsignificant, significant, and severe weather in four regions of Colorado. The two traditional expert systems appeared best able to discriminate significant from nonsignificant weather events. Both of these systems required the analyst to make sophisticated meteorological judgments. However, one of the expert systems produced forecasts that were biased.

In Stewart et al. (1989), an expert system to forecast hail was better than only one of a number of regression models. When the forecasts were limited to severe hail, the expert system was better than all of the regression models.

IMPLICATIONS FOR PRACTITIONERS

There are benefits and risks associated with expert systems. Given the lack of validation studies, there is also much uncertainty.

Benefits

Expert systems can improve accuracy by making the predictions of the best experts available to anyone who wishes to use them. Thus, users might obtain predictions for medical cases from the top medical specialists or they might get legal advice from the leading lawyers.

Like judgmental bootstrapping and econometric models, expert systems can improve consistency to allow for comparisons among forecasts for alternative policies. Consistency can convey the impression of rationality and this may help to persuade people to use the forecasts. For instance, one of the benefits cited for Texas Instruments' Capital Expert was its ability to enforce consistency in the preparation of capital expenditure proposals across the company (Gill 1995). Consistency can also enhance fairness, which can be important for the allocation of resources by government agencies, schools, hospitals, and other organizations.

Expert systems can improve the persuasiveness of recommendations. Dijkstra, Liebrand & Timminga (1998) presented 85 subjects with four problems concerning dyslexia, law, cardiology, and train tickets. Experts and expert systems provided the same advice, but the subjects believed that the advice from the expert systems was more objective and more rational than that from experts.

By describing the current process, expert systems may provide clues about how to improve the process. Various aspects of the problem can be studied and the findings can be incorporated into the expert system.

Perhaps the most important benefit is that expert systems can reduce the cost of making decisions and forecasts. Based on his survey of publications describing 440 expert systems in business, Eom (1996) concluded that cost savings are the primary motivation for their use.

When cost-saving and consistency are the prime considerations, expert systems can be justified if they merely reproduce the experts' forecasts. For example, Kort (1957) developed an expert system for the U. S. Supreme Court's decisions on right-to-counsel cases. The forecasts by the expert system matched the actual decisions by the Court for all 14 cases in a validation sample in later years. Although one could not replace the Supreme Court, expert systems could be used in many tasks such as in the selection of candidates for programs in higher education.

Risks

Design, implementation, and maintenance of expert systems are expensive. The process of eliciting, reconciling, and validating knowledge from multiple sources is difficult. As the complexity of the problem increases, it becomes more difficult to elicit knowledge and to extract meaningful rules. Many rules may be required to represent a complex problem.

It can be difficult to maintain expert systems when domain knowledge changes. Even more important perhaps is ensuring that the expert systems are acceptable to new decision makers. In 1987, Gill (1995) identified 97 expert systems that had been introduced into organizations during the early and mid-1980s. In 1992, he used phone interviews to determine the status of each system, obtaining information on 73 of the systems. Of this group, only about one-third of the expert systems were still being used. The decline occurred even though the expert systems always improved consistency and 86% of the users thought that they led to better decisions. Explanations that were given for discarding the expert systems involved system-maintenance expenses and such organizational factors as changing priorities and loss of developers. Developers of expert systems often failed to obtain and maintain user commitment to the systems. Fewer than one-quarter of the abandoned systems were criticized for bad performance.

Some expert systems might have failed because the experts viewed them as a threat to their positions. After all, how many of us would be agreeable to our organizations replacing us with an expert system? Armstrong and Yokum (2000) found that potential adopters perceived significant risks associated with the use of expert systems. On the other hand, they viewed expert systems positively with respect to their compatibility with their job, the ability to experiment with parts of the system, and ease of understanding.

Guimaraes, Yoon and Clevenson (1996), in their study of 1,200 expert systems at E. I. DuPont, concluded that it was important to establish training programs for developers and end-users. Developers must be trained in using appropriate knowledge elicitation techniques and modeling knowledge effectively. Their study found a strong relationship between the impact of expert systems on end-users jobs and the success of such systems.

Expert systems are sometimes used uncritically. For example, over 20 students in one of our classes used an expert system to predict the persuasiveness of some advertisements. As it turned out, a programming error had rendered about 25% of the system inoperable. None of the students recognized that their inputs to that part of the program had no effect on their ratings of an ad's effectiveness.

We suggest that the use of expert systems be restricted to complex situations in which forecasts are made repeatedly, when there is little or poor data on the dependent value, and when the alternative would be unaided judgment. Where possible, expert system forecasts should be supplemented by forecasts from other approaches. Finally, expert systems should be comprehensive as they might be used uncritically.

IMPLICATIONS FOR RESEARCHERS

Despite the extensive literature on expert systems, little of it concerns the development and use of expert systems for forecasting. Research is needed on different approaches to developing expert systems. When is it best to ask people to say directly how they solve a problem, when should protocols be used, and when should a combination of these approaches be used?

Research is also needed on the conditions under which expert systems are superior to alternative procedures. Researchers should examine accuracy and other criteria, such as relative costs and acceptability. We expect that expert systems will prove most appropriate for messy problems for which experts can make fairly accurate predictions.

Expert systems might also help analysts to select the best forecasting method for a particular situation. Weitz (1986) and Nute, Mann and Brewer (1990) developed general models though they did not test them. Ashouri (1993) developed an expert system to decide which of a set of forecasting methods would be most effective in predicting daily gas demand.

CONCLUSIONS

When acquiring knowledge for an expert system, it is desirable to use many techniques. Protocol analyses of experts who are actually engaged in the task can produce usable knowledge in complex situations where the rules are not self-evident.

Knowledge representations should be simple so that users can know what the system is doing. Because users will tend to become dependent on the system, it is important that it be valid and comprehensive.

The most surprising finding was that so little research has been done to examine the predictive validity of expert systems. We found only 15 validation comparisons of expert systems for forecasting. Expert systems were more accurate than unaided expert judgment in six of eight comparisons. In the four comparisons we found with judgmental bootstrapping, expert systems were less accurate in two and tied in two. Expert systems were more accurate than econometric models in one study and tied in two.

Given the high development costs and the meager evidence on improving predictive validity, we see two major uses of expert systems. The first is to develop systems when one needs many forecasts and the problem is too messy for judgmental bootstrapping. In such cases, one can expect some gains in accuracy. Second, and more important, expert systems can produce substantial cost savings by merely matching the accuracy of the best experts in semi-structured problems that do not lend themselves to judgmental bootstrapping.

REFERENCES

Abramson, B., J. Brown, W. Edwards, A. Murphy & R. L. Winkler (1996), “Hailfinder: A Bayesian system for forecasting severe weather,” *International Journal of Forecasting*, 12, 57-71.

Allen, G. & R. Fildes (2001), “Econometric forecasting,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S. (2001a), “Evaluating forecasting methods,” in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S. (2001b), “Judgmental bootstrapping: Inferring experts’ rules for forecasting,” in J. S. Armstrong (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S., M. Adya & F. Collopy (2001), “Rule-based forecasting: Using judgment in time-series extrapolation,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S. & J. T. Yokum (2000), "Potential diffusion of expert systems in forecasting," *Technological Forecasting and Social Change* (forthcoming).

Ashouri, F. (1993), "An expert system for predicting gas demand: A case study," *Omega*, 21, 307-317.

Chapman, L. J. & J. P. Chapman (1969), "Illusory correlations as an obstacle to the use of valid psychodiagnostic observations," *Journal of Abnormal Psychology*, 74, 271-280.

Clarkson, G. P. E. (1962), *Portfolio Selection*. Englewood Cliffs, NJ: Prentice-Hall.

Cocozza, J. J. & H. J. Steadman (1978), "Prediction in psychiatry: An example of misplaced confidence in experts," *Social Problems*, 25, 265-276.

Collopy, F. & J. S. Armstrong (1989), "Toward computer-aided forecasting systems," in G. R. Widemeyer (ed.), *DSS 89 Transactions.* Providence, R.I.: TIMS College on Information Systems, 103-119. Full text is provided at hops.wharton.upenn.edu/forecast.

Dijkstra, J. J. (1995), "The influence of an expert system on the user's view: How to fool a lawyer," *New Review of Applied Expert Systems*, 1, 123-138.

Dijkstra, J. J., W. B. G. Liebrand & E. Timminga (1998), "Persuasiveness of expert systems," *Behavior & Information Technology*, 17, 3, 155-163.

Doukidis, G. I. & R. J. Paul (1990), "A survey of the application of artificial intelligence techniques within the OR Society," *Journal of the Operational Research Society*, 41, 363-375.

Einhorn, H. J., D. N. Kleinmuntz & B. Kleinmuntz (1979), "Linear regression and process-tracing models of judgment," *Psychological Review*, 86, 465-485.

Eom, S. E. (1996), "A survey of operational expert systems in business," *Interfaces*, 26, (September-October) 50-70.

Fischhoff, B., P. Slovic & S. Lichtenstein (1978), "Fault trees: Sensitivity of estimated failure probabilities to problem representation," *Journal of Experimental Psychology*, 4, 330-344.

Gill, T. G. (1995), "Early expert systems: Where are they now?" *MIS Quarterly*, March, 51-81.

Grove, W. M. & P. E. Meehl (1996), "Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy," *Psychology, Public Policy, and Law*, 2, 293-323.

Guimaraes, T., Y. Yoon & A. Clevenson (1996), "Factors important to expert system success: A field test," *Information & Management*, 30, 119-130.

Kleinmuntz, B. (1968), "The processing of clinical information by man and machine," in B. Kleinmuntz (ed.), *Formal Representation of Human Judgments*. New York: John Wiley.

Kleinmuntz, B. (1967), "Sign and seer: Another example," *Journal of Abnormal Psychology*, 72, 163-165.

Kort, F. (1957), "Predicting Supreme Court decisions mathematically: A quantitative analysis of the 'right to counsel' cases," *The American Political Science Review*, 51, 1-12.

Larcker, D. F. & V. P. Lessig (1983), "An examination of the linear and retrospective process tracing approaches to judgment modeling," *Accounting Review*, 58, 58-77.

Leonard, K. J. (1995), “The development of a rule based expert system model for fraud alert in consumer credit,” *European Journal of Operational Research*, 80, 350-356.

McClain, J. O. (1972), “Decision modeling in case selection for medical utilization review,” *Management Science*, 18, B706-B717.

Michael, G. C. (1971), “A computer simulation model for forecasting catalogue sales,” *Journal of Marketing Research*, 8, 224-229.

Moninger, W. R., J. Bullas, B. de Lorenzis, E. Ellison, J. Flueck, J. C. McLeod, C. Lusk, P. D. Lampru, R. S. Phillips, W. F. Roberts, R. Shaw, T. R. Stewart, J. Weaver, K. C. Young & S. M. Zubrick (1991), “Shootout-89: A comparative evaluation of knowledge-based systems that forecast severe weather,” *Bulletin of the American Meteorological Society*, 72, 9, 1339-1354.

Moore, J. S. (1998), “An expert systems approach to graduate school admission decisions and academic performance prediction,” *Omega*, 26, 670-695.

Moss, S., M. Artis & P. Ormerod (1994), “A smart automated macroeconometric forecasting system,” *Journal of Forecasting*, 13, 299-312.

Nute, D., R. I. Mann & B.F. Brewer (1990), “Controlling expert system recommendations with defeasible logic,” *Decision Support Systems*, 6, 153-164.

Reagan-Cirincione, P. (1994), “Improving the accuracy of group judgment: A process intervention combining group facilitation, social judgment analysis, and information technology,” *Organizational Behavior and Human Performance*, 58, 246-270.

Santhanam, R. & J. Elam (1998), “A survey of knowledge-based system research in decision sciences (1980-1995),” *Journal of the Operational Research Society*, 49, 445-457.

Schmitt, N. (1978), “Comparison of subjective and objective weighting strategies in changing task situations,” *Organizational Behavior and Human Performance*, 21, 171-188.

Silverman, B. G. (1992), “Judgment error and expert critics in forecasting tasks,” *Decision Sciences*, 23, 1199-1219.

Smith, P., S. Hussein & D. T. Leonard (1996), “Forecasting short-term regional gas demand using an expert system,” *Expert Systems with Applications*, 10, 265-273.

Stewart, T. R., W. R. Moninger, J. Grassia, R. H. Brady & F. H. Merrem (1989), “Analysis of expert judgment in a hail forecasting experiment,” *Weather and Forecasting*, 4, 24-34.

Tellis, G. (1988), “The price elasticity of selective demand: A meta-analysis of econometric models of sales,” *Journal of Marketing Research*, 25, 331-341.

Turing, A. M. (1950), “Computing machinery and intelligence,” *Mind*, 59, 443-460.

Weitz, R. R. (1986), “NOSTRADAMUS: A knowledge based forecasting advisor,” *International Journal of Forecasting*, 2, 273-283.

Wittink D. R. & T. Bergestuen (2001), “Forecasting with conjoint analysis,” in J. S. Armstrong (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers.

Wong, B. K. & J. A. Monaco (1995), “Expert system applications in business: A review and analysis of the literature (1977-1993),” *Information and Management*, 29, 141-152.

Yntema, D. B. & W. S. Torgerson (1961), "Man-computer cooperation in decisions requiring common sense," in W. Edwards & A. Tversky (eds.), *Decision Making*, Baltimore: Penguin Books.

Acknowledgements

Dennis A. Ahlburg, Sean B. Eom, Carlos Mate, Mark S. Silver, Barry G. Silverman, Leonard J. Tashman and J. Thomas Yokum provided useful comments. Raphael Austin, Ling Qiu and Mariam Rafi made editorial revisions.

December 21, 2009