

Predicting Public Opinion from Press Coverage

By David P. Fan

Surveys are not ideal tools for analyzing public opinion. To begin with they're expensive. Then, there's the problem of decreasing response rates. The public has become ever more annoyed with the amount of polling, especially telemarketing. Furthermore, surveys can only be taken today: More than one analyst has wished that another question had been asked yesterday. Also, surveys are typically taken only after an interesting event has occurred. Without prior polls, there is insufficient historical context. The analyst will not know what the trend has been.

Given these deficiencies, it would be useful to have a complementary method which is cheap, not plagued by nonresponse and other survey problems, and can give estimates of opinion over time—even for the past.

Such a method does indeed exist and was developed by noting that public opinion is akin to sailboats. Neither will budge in the absence of external forces. With the arrival of new persuasive information, people will change their minds, much like sailboats responding to the wind and currents. If we know the forces behind the change, then we should be able to compute the trajectories of both sailboats and public opinion over time.

The Power of the Press

The strategy for predicting opinion is to quantify the key persuasive forces acting on the public and then use these forces to compute predicted opinion values. The prediction is validated if the measures generated by the chosen causal forces match the actual survey data.

For such predictions, the sine qua non is examining the persuasive forces acting on opinion over time. So what are likely to be these explanatory variables? For opin-

ions evolving over decades and centuries, there can be a wide range of forces. However, the case is different for opinions which change over days or weeks. An example here is the University of Michigan's Index of Consumer Sentiment, a composite of 5 survey questions asked monthly.¹ For these unstable opinions characterized by rapid shifts up and down, there is one prime candidate for the driving force: the print and electronic media. Nothing else can reach so many people in such a short time. Of course, the public also makes its own

It would be useful to have a complementary [survey] method which is cheap [and] not plagued by nonresponse and other survey problems.

observations, but the major influence is likely to be the press—in the same way the population depends on media forecasts rather than glancing out the window to form impressions of the weather. People know that their own windows on the economy only give very narrow, and possibly distorted, views of the world. Therefore, it is reasonable to propose that the press contains the prime persuasive forces to predict the Michigan Index.

Consumer Sentiment

To test this theory, the analysis began with the scoring of stories from a dozen to two dozen newspapers, all available on electronic databases. The advantage of the on-line databases is that the text is already in a form which can be analyzed by computer. Since the goal was to predict consumer sentiment, the machine was programmed to score stories for the number of paragraphs pro and con on the health of the

economy. The computer did much more than simply look for the occurrence of individual words or phrases. A number of words could be combined to give complex thoughts. Thus the computer could exclude discussion of foreign economies, convert high inflation or unemployment to unfavorable data on the economy, and take into account negation so that "no momentum" would also be scored as negative. A total of 9,385 stories were scored from January 1, 1990 to May 29, 1994. The result for each story was a date and the number of paragraphs pro and con. These scores were then entered into a statistical model developed to predict the Index of Consumer Sentiment.²

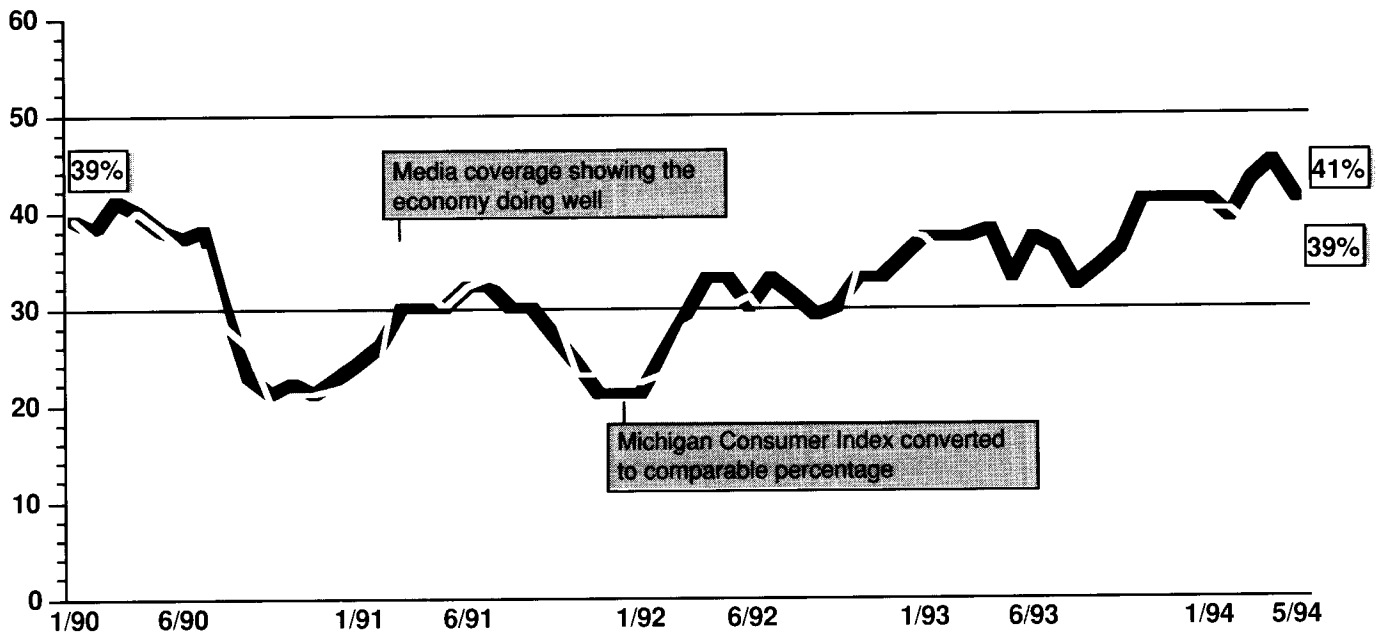
To allow for comparability between the Index and the print media analysis, the Index values were all adjusted to yield numbers equivalent to the percentage of paragraphs portraying the economy positively. Therefore, any significant divergence in the tracking of these two public opinion measures (predicted from the press coverage versus drawn from surveys) would then be attributable to a poor model for predicting public opinion through print media analysis. However, if the two data sources track closely, then the model is performing well.

The final test of this theory was to compare the predictions with the measured index values. As seen in Figure 1, qualitatively, the predicted values and the Index of Consumer Sentiment have moved together for the entire period of more than four years.³ Thus this model has been quite accurate at predicting public sentiment on the economy.

Other Uses of This Technology

Besides these tests for consumer sentiment, the theory was tested for its ability to make other predictions. A good example is

Figure 1
Index of Consumer Sentiment Shifts Follow Closely
Media Picture of How the US Economy is Doing



Source: Index of Consumer Sentiment Survey by the Survey Research Center, University of Michigan.

cocaine use in the last 30 days by high school seniors.⁴ Cocaine use was fairly flat during the beginning of the 1980s. There was a rise from 1983 to 1985 due to the introduction of crack. Then there was a sustained drop from almost 7% use in 1986 down to just over 1% in 1991 (See Figure 2). The entire drop could be attributed to press coverage of the adverse medical and legal consequences of cocaine consumption.

Another successful example modeled the increase in caffeine-free cola purchases after news about the harmful effects of caffeine. It was also possible to predict the measured decreases in high risk sex by gay men after the beginning of the AIDS epidemic. The result of the latter was a drop in the rate of HIV infection which was also modeled satisfactorily.⁵

Beyond these examples, good predictions were made for both the 1988 and 1992 presidential elections. Print and wire ser-

vices could also capture key persuasive information about the 1988 catastrophic health bill, including the mass mailings opposing the bill.⁶ Opposition by senior citizens could be predicted entirely from the Associated Press (AP) or stories from any of three major papers around the country.⁷

All these studies point to the similarity in coverage of national stories by different news organizations. With trends being predictable for a number of opinions and behaviors from persuasive messages in the press, it is likely that the method has general applicability. So how would a public opinion analyst use this new technology?

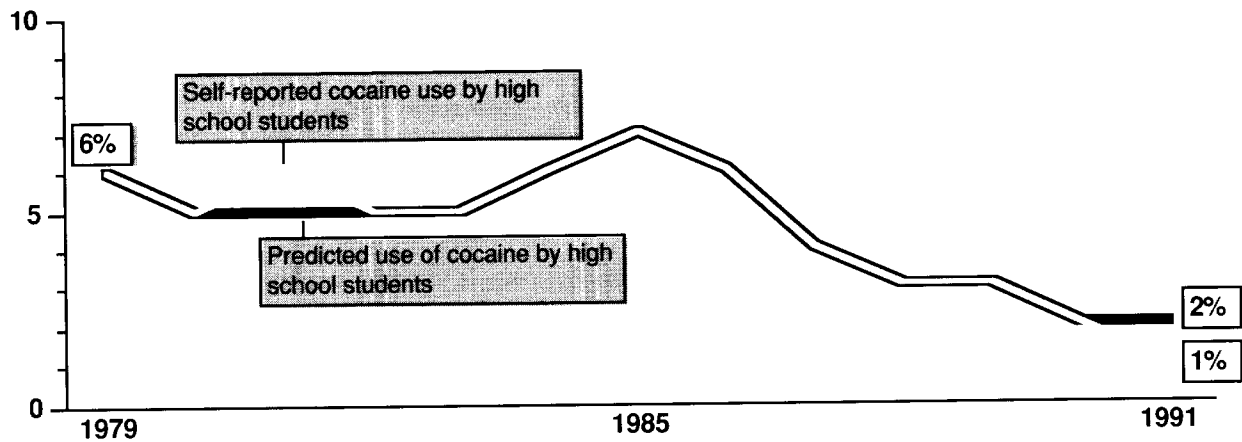
Applications and Advantages

If the predictions match actual survey results over time, as was the case for Figures 1 and 2, it is likely that the key persuasive information can be separated out to give the relative importance of different

influences. For instance, dissection of media messages showed that the most important anti-cocaine information in the late 1980s was descriptions of the medical consequences. This causal conclusion is stronger than those inferred from responses to survey questions, because the prediction model uses persuasive information measured directly and not recalled by respondents to surveys. Therefore, one important benefit of the method is new light on the causes of opinion formation.

Besides the advantage of direct information about correlates of opinion change, the new technique gives a rapid method of tracking expected poll results. The computer can be programmed to extend the analysis by dialing up the electronic database, retrieving text, analyzing it, and doing the modeling—all automatically. Therefore, it is easy to do periodic analyses like the ones for Figure 1 which were replicated every month. The pollster can follow these trends as early warning signals of important

Figure 2
Predicted Use of Cocaine Mirrors Self-Reported Use



Source: David P. Fan and William B. Holway, "Media Coverage of Cocaine and its Impact on Usage Patterns," *International Journal of Public Opinion Research*, Issue 2, forthcoming 1994.

shifts in key persuasive forces and hence expected changes in opinion.

While this method is not prone to survey errors, it does have potential problems in its own right, the principal one being the possible omission of relevant persuasive information for the analysis from the print media databases. Therefore, it is still desirable to do confirmatory surveys. If the estimates from the media are seen to be reasonably accurate based on prior poll results, then polls only need to be commissioned when the persuasive climate has shifted substantially.

With this technology being subject to different kinds of error than survey results, finding agreement with polls strengthens both measures. The method also gives a direct measure of the consequences of persuasive information and is cheap when text can be obtained from electronic databases.

The applicability of the technology is now increasing rapidly as more and more news sources become retrievable electronically. Thus it is possible to perform detailed analyses of persuasive press information back almost 20 years. Also, the persuasive messages in the media can be dissected by computer to determine the prevalence of different arguments at different times. Such analyses are useful for planning new communication strategies and evaluating old ones.

Endnotes:

- ¹ Richard T. Curtin, "Indicators of Consumer Behavior: The University of Michigan Survey of Consumers," *Public Opinion Quarterly*, Vol. 46, 1982, pp. 340-352.
- ² David P. Fan, *Predictions of Public Opinion from the Mass Media: Computer Content Analysis and Mathematical Modeling* (Westport, CT: Greenwood Press, 1988); David P. Fan, Hans-B. Brosius and Hans M. Kepplinger, "Predictions of the Public Agenda

from Television Coverage," *Journal of Broadcasting and Electronic Media*, Vol. 38, 1994, pp. 163-177; and David P. Fan, "Predictions of Consumer Confidence/Sentiment from the Press." Paper delivered at AAPOR Conference, Chicago, 1993.

³ More quantitative statistical tests have confirmed the key role of the media in shaping the Index. As an additional validation test, these predictions have been faxed to a number of interested parties each month since May 1991 on the weekend before the last Tuesday of every month. The predictions were made without the benefit of measured index values since the computations were made well before the appearance of the Index on the Commerce Department's Economic Bulletin Board.

⁴ David P. Fan and William B. Holway, "Media Coverage of Cocaine and its Impact on Usage Patterns," *International Journal of Public Opinion Research*, Issue 2, forthcoming, 1994.

⁵ Contact the author for citations.

⁶ Ibid.

⁷ Ibid.

David P. Fan is professor of Genetics and Cell Biology, University of Minnesota