Do we make better forecasts these days?

A survey amongst academics

Philip Hans Franses
Econometric Institute Rotterdam

Econometric Institute Report 2003-06

February 17 2003

Acknowledgements:

I am grateful for the helpful comments of Richard Paap, Dick van Dijk, Christiaan Heij, Teun Kloek, Herman van Dijk and Peter Verhoef in various stages of this project. In particular, I thank Rene Segers for his excellent work on the website, which facilitated the collection of the data. Thanks are also due to Scott Armstrong, Herman Stekler, and Lars-Erik Oeller for helpful suggestions.

Address for correspondence:

Econometric Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands, franses@few.eur.nl

Do we make better forecasts these days?
A survey amongst academics

Abstract

This paper presents the results of a survey held amongst all editorial board members of six journals. These journals in part focus on the developments of models and methods for forecasting. The key question was whether one believes that the forecasting discipline has made progress in the last three decades. Amongst various results, the most important one is that modest progress has been made, although the profession seems far from satisfied. This progress appears to be mainly due to the increase in computing power and the fact that we are better able to incorporate important data features in our models. Additionally, progress could have been faster if we somehow were to include the opinions of experts. These last two findings define two important topics on the research agenda.

Introduction

The developments in econometrics and statistics proceed at a rapid pace. Since the advent of the increased use of time series models, initiated by Box and Jenkins' seminal book in 1970, once could notice an explosion of new models and methods, which could be relevant for forecasting. Topics as unit roots, cointegration, switching regime models, neural networks, to mention just a few, easily come to mind to the present-day forecaster. Next to new models and methods, we have also seen an explosion in opportunities to collect economic data. In finance and marketing one can collect data at the individual transaction level. In macroeconomics one witnesses cross-sectional disaggregating, in the sense that, for example, one can now retrieve monthly city-specific unemployment figures. Furthermore, while it took weeks to estimate parameters in a simultaneous equations model some thirty years ago, we can now use simulation-based methods for complicated unobserved components models, which deliver estimates overnight. Hence, increasing computer power has facilitated the analysis of elegant and relevant econometric models. For further reference, Diebold (1998) gives a lucid survey of such developments in macroeconomic forecasting.

Despite these rapid developments, one may wonder whether these also paved the way to generate better out-of-sample forecasts. So, we do have more models to choose from, we can estimate parameters in complicated models, we do have data at a detailed level, but what does it bring us? Headlines in newspapers telling us "the forecasts were again wrong" are quite common, and at first sight this seems in potential contrast with the above-mentioned achievements. Granger (1996) addresses the issue of how we can improve the perceived quality of forecasts, that is, perceived by for example the business press.

To see if we[1] can forecast better these days, one can opt for a large-scale analysis of models, data and forecasts. This would involve decisions on each of these three components, and, for the moment, to me this sounds like an impossible enterprise.

[1] Being a member of the editorial board of three of the considered journals below, I take the liberty of also assigning my research activities to the forecasting profession. Hence,

Indeed, one should separately consider disciplines as finance and macro, forecasts as one-step and multi-step, and models ranging from simple first order autoregressions to systems with 300 equations. And then still, it is quite likely that something is overlooked. A recent attempt to provide details on the quality of forecasts in Fildes and Stekler (2002) focuses on the predictions of a few quarterly macroeconomic series. To replicate their study for other series and other disciplines would require ages of work. Further, I should mention a range of studies, (co-) authored by Scott Armstrong on forecasting principles, see for example Armstrong (1986).

In this paper I want to address the issue of whether forecasts are perceived to have improved by those who develop the models and techniques themselves. It is fashionable these days to examine the relevance of one's own contribution to science and practice, and this study aims to contribute to this self-reflection. Therefore, I decided to follow a different strategy than those mentioned above, and I focused on the individuals, who are partly responsible for the various developments. I discerned six academic journals, which regularly publish articles on models and methods for forecasting, and decided to ask all their editorial board members a few questions. Basically, this is a novelty of this study. The main question concerns the individually perceived progress in the ability to obtain better forecasts, in between roughly the publication of Box and Jenkins' book (1970) and now (2002-2003). Deliberately, various questions are quite general (two respondents called them "vague"), and they should mainly reflect what people feel is the contribution of, say, academic progress to practice. Thus, note that this paper deals with opinions and not with measurable facts, where the latter could indicate that we really make better forecasts these days.

The key question in this survey is whether there is a perceived change in quality, and if so, whether this depends on certain antecedents or personal traits. For the last, the survey also contains questions on individual opinions about the profession and the world, additional to factors like for example age and research interests. This paper reports on the results of this survey amongst academic professionals who develop, implement and evaluate models, which are used for forecasting economic data. I focus on whether

---

the "we" means all academic forecasters. In fact, I am also one of the 76 respondents in the survey.

people believe there are noticeable improvements in forecasting. Next, it is studied which factors can be viewed as antecedents for improvements, if there are any. Finally, are there any opportunities for further improvement of forecasts?

The outline of this paper is as follows. In the next section, I discuss the way the data were collected. To get an impression of the individual characteristics of the respondents, I report on a few descriptive statistics. The subsequent section continues with descriptive statistics concerning the answers to the questions, and with a model, which aims to correlate the overall quality with potential antecedents. The final section contains a discussion of the main results and puts forward a tentative research agenda for the forecasting discipline.

Collecting the data

A four-page survey was designed and made available through a website. Before publication on the web, it was checked for wording and phrasing, and also for the relevance of the answer categories. Next, it was tested out on six colleagues who suggested even better wording and an occasional additional question.

Between December 24 and 31, 2002, e-mail messages were sent out to editorial board members of six journals. These journals are the *International Journal of Forecasting*, *Journal of Forecasting*, *Journal of Econometrics*, *Journal of Applied Econometrics*, *Studies in Nonlinear Dynamics and Econometrics* and the *Journal of Business and Economic Statistics*. It took a while to obtain all the correct e-mail addresses, but in the end the message was well received by 37, 50, 30, 39, 28 and 36 individuals, respectively, giving a total of 220 possible respondents. Responses were recorded between December 24, 2002 and January 13, 2003. The a priori estimate of the time it would take was 10 minutes, but an analysis of the responses indicated that nobody took more than 6 minutes to respond.

I had to delete 4 uninformative surveys. These entailed that no answers were given, although one respondent added a statement like “this is a stupid waste of time”.

The final amount of useful questionnaires is 76, which amounts to a rather acceptable response rate of 34.5 per cent.

As said, respondents could answer the questions through a website. This was fully anonymous. Indeed, one respondent forgot about whether he had already filled it in, but I could not help him by retrieving his questionnaire. Due to the anonymity, most people also turned out to be willing to answer the more personal questions to be discussed below. These questions were included in order to allow for correction for responses styles in the final model.

Table 1 shows that the average age of the respondents is 52 years, and that they on average graduated 23 years ago. The respondents currently spend about 26 hours per week for research, which is approximately 3 days a week. Most respondents (88 percent) did or do publish in academic journals about forecasting matters. Finally, respondents seem to be interested in macroeconomics, finance or both.

As I did ask about their opinions on the quality of the academic profession, it might be relevant to correct for response styles. Indeed, people who have a pessimistic view on the profession might also feel that the quality of forecasts has deteriorated. Additionally, if people are generally optimistic, not so much about the profession but about life in general, they might overrate any change in quality of forecasts. Hence, in econometric models one might want to correct for these personal response styles.

Table 2 gives the results on how we look at our profession. In general, we seem to have agreement on that the economic forecasting profession did make progress in the last three decades, that economic theory is important, that we often go to forecasting talks and that forecasts are important to make good decisions. In sum, it seems that the respondents have a positive attitude towards their profession and actively take part in scientific discussion. With respect to a possible gap between academics and practitioners, there is little consensus, as most respondents nor agree nor disagree.

Table 3 gives some insights into possible pessimistic or optimistic personal views. The respondents believe that today's world is about equally good or bad as before, that a worldwide war is unlikely, that global warming is a serious issue and almost all of us feel safe when we are on our way home. Finally, about all respondents are very interested in

economic matters. Overall, it seems that the academic forecasting community is optimistic, but also conscious of the environment and the economy.

Results

Now it is time to see what the answers to the questions about the quality of forecasts look like. This section first deals with some descriptive statistics. Next, I will turn to some models.

Descriptive statistics

Table 4 reports on the evaluation of the quality of present-day forecasts relative to those in 1970. A few respondents did not find all questions precise enough. A summary of the additional comments is given in Table 7.

From Table 4 we can see that generally everything is perceived to be better than in 1970, that is, the answer category "not better than in 1970" is rarely selected. The accuracy of point forecasts, density forecasts and longer horizon forecasts are perceived as only a little better, while all other issues suggest a substantial improvement.

These results in Table 4 are to be used to see if the overall judgment in Table 5 matches with specific components. Table 5 shows that the average perceived quality of today's forecasts, which is 5.91, compares favorably with the score for 1970, which is 4.30. As an absolute number, 5.91 is not high, also knowing that on a scale of 1 to 10, people usually tend to give scores around 7 in other types of surveys. Most respondents, that is 60 of the 76, perceived an improvement, while 14 respondents did not experience an improvement, and 2 even feel things have gone worse.

The correlations of the differences between the scores for today and for 1970 with the statements in Table 4 range from 0.22 to 0.55. A factor analysis of the statements in Table 4 gives a 48.9 per cent of the variance for the first factor, where this factor is roughly the average of the 10 statements in Table 4. Hence, it seems that the overall opinion in Table 5 is representative for the various more detailed opinions in Table 4.

Finally, Table 6 gives the descriptive statistics for possible antecedents of the perceived quality. We seem to agree on the notion that we can better incorporate data

features in the models and that we would be better off with combining models with experts. Structural breaks are seen as important causes for forecast failure, substantiating the theoretical discussion in for example Clements and Hendry (1998). Handling unit roots does not seem to matter much. On the other hand, the increase in computing power helped us to get better forecasts.

Model

The key dependent variable in my survey is the change in the score for today's forecasts and for those in 1970, as given in Table 4. The multiple correlation coefficient of this change with the variables as given in Table 5 is 0.42, which again indicates that this key variable approximately summarizes various underlying factors, like the quality of density forecasts and of evaluation criteria.

The next step is to see if this key variable can be correlated with some of the antecedents in Table 6, when corrected for possible influences of the personal traits, as summarized in Tables 1 to 3. To reduce the number of variables, I take the first principal components of the variables in Table 4 and of the variables in Table 5. Both components cover about 33 per cent of the variances. The weights of the five underlying variables in the first principal components have the expected sign, at least, given the wording of the questions.

What can we expect about the possible effects of the antecedents in Table 6? I think that we can expect a positive effect from the first antecedent, that is, that we can make better forecasts these days as we can better incorporate the relevant data features in the models. Positive effects can also be expected from the fifth antecedent, which says that handling unit roots would have a positive effect, and from the sixth antecedent, which concerns the notion of having more computing power these days. At first sight, the unit roots issue might seems to be redundant, but I included it, as the unit root literature is the most cited in applied econometrics. Key articles in this area receive hundreds and sometimes thousands of citations, which is quite a lot for our discipline.

Possible negative effects are the following. The second and third antecedents in Table 6 state that we did not make enough use of external information, perhaps from an expert, so more agreement with these statements should lead to a smaller value of

perceived progress. The fourth antecedent says that we might have experienced structural breaks and therefore progress is slower than was hoped for, and hence a negative sign is expected as well. Finally, if economic data have become less predictable, than progress is experienced to be smaller, and hence I expect also a minus sign here.

For the control variables, which deal with various personal traits, I expect that more positive views on the profession would give higher values of the increase in quality. Next, for the variables and time for research, one might contemplate that more experience might lead to less positive scores. Finally, I would not have any firm thoughts about whether people in finance, marketing or macro believe things have become much better.

The dependent variable is the change between the scores for now and then (1970). The full standard regression model contains 16 explanatory variables for 66 effective observations. This last number is due to the fact that 10 respondents with one or more "no answers" are not included. The adjusted R-squared is 0.17. Deleting 10 redundant explanatory variables makes this measure to increase to 0.28, now for 70 effective observations.

The smaller model results are as follows. The notion that we can better incorporate salient features of the data in our model has an expected positive effect (t-ratio is 2.26), and also more computing power has a positive effect (t-ratio is 2.31). That we rely too much on a single model has the expected negative effect (t-ratio is -2.20). Next, more years since graduation and more publications in forecasting journals have a positive (t-ratio is 1.91) and negative (t-ratio is -1.97) effect, respectively. Finally, if the interest lies in macro, one tends to be more positive (t-ratio is 2.19).

Conclusion

The results of the survey discussed in this paper suggest various conclusions. First, the survey shows that the respondents represent a part of the academic community, which in general has a positive attitude towards their profession and to various aspects of life. Next, they are interested in their object of research, the economy, and they spend quite some time on research during the week. Progress is also experienced, in terms of being

able to generate better forecasts these days, and this general progress correlates with various factors like densities, evaluation criteria and the quality of models. The stated progress is not due to the personal traits. The positive change in quality is mainly attributed to the increase in computing power and the fact that we are better able to incorporate the salient data features in the models. However, it seems that we could have made even more progress if we would have explicitly included the opinion of experts and not just relied on a single model.

This survey points towards two marked topics on the future research agenda. The first is that we should keep on striving for developing models and methods, which incorporate the relevant features of the data. This aspect was seen as an important underlying factor of the positive progress. The second is that we should somehow try to explicitly incorporate the opinion of an expert into a forecasting model and to use more than just one model. This last issue is perhaps more difficult than one might think. At present there are many studies, which deal with the combination of expert opinions, but there are not many illustrations of how one can precisely incorporate such opinions into an available model.

Finally, several respondents made a few additional remarks, other than that "they liked the survey and wanted to the results", which I would like to share with the reader. One respondent mentions, "As far as I can tell economic forecasting depends largely on short-term stability of purely statistical factors, and as a result its accuracy rapidly deteriorates with horizon. If anything, macroeconomic theory seems to me to have drifted backward since the 1970s." A complaint about the survey was formulated as follows: "Many of the questions were ambiguously stated. Questions on data for example talked of quality, but not of quantity, nor ease of access; the latter two are vastly improved relative to the former. Structural breaks are only a very small part of our difficulties; the real issue is that we are dealing with an evolving system for which the parameters of adjustment are changing." If the latter is true, then we should, even more so, spend efforts summarizing the salient data features.

One respondent feels that "You do not allow for the fact that forecasts may have got worse. I believe that they have (a) because of increased turbulence and (b) an

emphasis on mathematical elegance which, while valuable, is often negatively correlated with contextual immersion in practice."

Another respondent says, "I think that the issue of forecasting errors and related decision making are very closely related. It would be very helpful for future research to investigate the extent to which economic forecasts are used to shape policy precisely. Long-range forecasts are likely to be more important and qualitatively more accurate than the short run forecasts. For example, forecasts based on demographic trends can be potentially very valuable in shaping pension policy, but the value of accurate quarter-to-quarter forecasts of GDP growth is less obvious. Private and academic forecasts also have a very useful role of countering government forecasts, which may be politically contaminated. Useful research might involve comparison of government and central bank forecasts and those from academics and research institutes, assuming that the two groups are reasonably independent." This seems to match with the second topic on the research agenda.

Finally, I do not want to withhold the reader from the intriguing statement of one of the respondents, which is that "I doubt it is worth doing", although I am not sure whether he or she means "forecasting" or "this survey".

Table 1:
Characteristics of the respondents*

	Mean	Median	Maximum	Minimum	Std
Age (74)	51.85	52	76	35	9.17
Years since PhD (74)	22.69	22	46	6	9.42
Hours per week for research. (74)	25.65	22	60	4	13.06

		Yes	No
Publish about forecasting/time series (75)		0.88	0.12
Research interests are (76)			
	macro	0.53	0.47
	finance	0.52	0.48
	marketing	0.18	0.82
	other	0.50	0.50

***** Here in all the following tables I give in parentheses the number of respondents who did answer the particular question.

Table 2:
How do we look at our profession?
The respondents could mark "1" if they strongly disagreed, until a "5" if they strongly agreed, with a "3" implying indifference. The cells contain the percentage of respondents who marked an answer category.

	1	**2**	**3**	**4**	**5**
The economic forecasting profession did not make much progress the last 3 decades (76)					
	0.25	**0.39**	0.21	0.13	0.01
Economic theory is irrelevant for constructing forecasts (75)					
	0.36	**0.37**	0.11	0.13	0.03
I often go to conferences' sessions on forecasting (76)					
	0.16	0.21	0.20	**0.27**	0.17
There is a decreasing gap between academic forecasters and practitioners (74)					
	0.01	0.24	**0.38**	0.30	0.07
Forecasts have shown to be very important for making good decisions in government and in business (75)					
	0.03	0.07	0.21	**0.47**	0.23

Table 3:
How do we look at ourselves?
The respondents could mark "1" if they strongly disagreed, until a "5" if they strongly agreed, with a "3" implying indifference. The cells contain the percentage of respondents who marked an answer category.

	1	2	3	4	5
Today's children grow up in a better world than I did. (75)					
	0.04	0.17	**0.37**	0.27	0.15
It is unlikely that we will witness a worldwide war in the next few decades. (74)					
	0.05	0.19	0.27	**0.30**	0.19
Global warming activists do not exaggerate their issue. (74)					
	0.11	0.22	0.22	**0.28**	0.18
Most of the time, I feel safe when I am on my way home. (76)					
	0.01	0.03	0.12	0.39	**0.45**
Personally, I am not much interested in economic matters. (75)					
	0.73	0.17	0.04	0.04	0.01

Table 4:
Comparing today's quality of (aspects of) forecasts with their quality in 1970. The answer category 1 is "not better than in 1970", 2 is "not much better", 3 is "only a little better", 4 is "better" and 5 is "much better". The cells contain the percentage of respondents who marked an answer category.

	1	2	3	4	5
Our forecasts are in general (74)	0.04	0.12	0.36	**0.39**	0.08
The quality of our forecasting models is (74)	0.01	0.11	0.19	**0.41**	0.28
The quality of our data is (75)	0.01	0.09	0.17	**0.45**	0.27
The skills of an average practitioner are (74)	0.05	0.20	0.30	**0.36**	0.08
The accuracy of point forecasts is (73)	0.07	0.21	**0.36**	0.32	0.05
Density forecasts (65)	0.02	0.11	**0.34**	0.29	0.25
Forecasts for longer horizons are (72)	0.10	0.22	**0.36**	0.24	0.08
More variables can be forecasted (74)	0.03	0.15	0.32	**0.46**	0.04
Forecast evaluation criteria are (73)	0.01	0.04	0.27	**0.42**	0.25
One-step ahead forecasts are (73)	0.03	0.14	0.34	**0.36**	0.14

Table 5:
Judgment on the overall quality score,
on a scale from "1" (very bad) to "10"(very good).

	Mean	Median	Maximum	Minimum	Std
Today (76)	5.91	6	8	2	1.42
1970 (76)	4.30	4	8	2	1.35
Today - 1970 (76)	1.61	2	5	-2	1.30

Table 6:
Antecedents of the perceived quality of forecasts.
The respondents could mark "1" if they strongly disagreed, until a "5" if they strongly agreed, with a "3" implying indifference. The cells contain the percentage of respondents who marked an answer category.

	1	2	3	4	5
We can make better forecasts these days as we can better incorporate the relevant data features in our models. (76)	0.01	0.08	0.21	**0.51**	0.18
Our forecasts are still not good as we rely solely on a model. (74)	0.08	0.28	**0.36**	0.20	0.07
We would be better off by frequently including the opinion of an expert. (76)	0	0.22	0.30	**0.34**	0.13
Structural breaks are the main cause of forecast failure. (76)	0.07	0.13	0.22	**0.39**	0.18
Due to the fact that we handle unit roots in economic data much better now we can make better forecasts for longer horizons. (74)	0.16	**0.32**	0.24	0.24	0.03
The increase in computing power helped us much to get better forecasts. (75)	0	0.11	0.15	**0.55**	0.20
Economic data have become less predictable. (76)	0.14	0.22	**0.36**	0.22	0.05

References

Armstrong, J.S. (1986), Research on Forecasting: A Quarter-Century Review, 1960-1984, *Interfaces*, 16, 89-103.

Clements, M.P. and D.F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.

Diebold, F.X. (1998), The Past, Present, and Future of Macroeconomic Forecasting, *Journal of Economic Perspectives*, 12, 175-192.

Fildes, R. and H. Stekler (2002), The State of Macroeconomic Forecasting, *Journal of Macroeconomics*, 24, 435-468.

Granger, C.W.J. (1996), Can We Improve the Perceived Quality of Economic Forecasts?, *Journal of Applied Econometrics*, 11, 455-473.