Experts' Stated Behavior

Youssef Boulaksil and Philip Hans Franses

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2008-001-MKT
Publication	January 2008
Number of pages	17
Persistent paper URL	http://hdl.handle.net/1765/10900
Email address corresponding author	franses@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	We ask various experts, who produce sales forecasts that can differ from earlier received model-based forecasts, what they do and why they do so. A questionnaire with a range of questions was completed by no less than forty-two such experts who are located in twenty different countries. We correlate the answers to these questions with actual behavior of the experts. Our main findings are that experts have a tendency to double count and to react strongly to recent volatility in sales data. Also, experts who feel more confident give forecasts that differ most from model-based forecasts.
Free Keywords	model forecasts, expert forecasts, decision making, stated behavior
Availability	The ERIM Report Series is distributed through the following platforms: Academic Repository at Erasmus University (DEAR), DEAR ERIM Series Portal Social Science Research Network (SSRN), SSRN ERIM Series Webpage Research Papers in Economics (REPEC), REPEC ERIM Series Webpage
Classifications	The electronic versions of the papers in the ERIM report Series contain bibliographic metadata by the following classification systems: Library of Congress Classification, (LCC) LCC Webpage Journal of Economic Literature, (JEL), JEL Webpage ACM Computing Classification System CCS Webpage Inspec Classification scheme (ICS), ICS Webpage

Experts' stated behavior

Youssef Boulaksil

Eindhoven University of Technology

Philip Hans Franses

Erasmus University Rotterdam

Abstract

We ask various experts, who produce sales forecasts that can differ from earlier received model-based forecasts, what they do and why they do so. A questionnaire with a range of questions was completed by no less than forty-two such experts who are located in twenty different countries. We correlate the answers to these questions with actual behavior of the experts. Our main findings are that experts have a tendency to double count and to react strongly to recent volatility in sales data. Also, experts who feel more confident give forecasts that differ most from model-based forecasts.

Key words: Model forecasts, expert forecasts, decision making, stated behavior

This version: January 9 2008

Authors' notes: The address for correspondence is Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, franses@few.eur.nl.
We are very thankful to the headquarters' office of Organon NV (Oss, The Netherlands) for allowing us to send the list of questions to the experts in the local sales offices in various countries. We also thank Paul Goodwin for many excellent suggestions.

1. Introduction

There is a substantial amount of literature on the adjustment of statistical forecasts by knowledgeable experts. Sanders and Manrodt (1994) document that many forecasters claim to adjust model-based forecasts, and also Klassen and Flores (2001) show that experts exercise a substantial impact on the final forecasts. Goodwin (2002) summarizes the reasons for experts to adjust model-based forecasts, and Mathews and Diamantopoulos (1986) and Syntetos et al. (2008) show that such adjustment can be beneficial for the final forecast quality. Indeed, statistical models cannot capture all possible relevant variables, and also experts may know future level shifts or breaks in trends based on specific knowledge. Franses (2004) reports that model-builders themselves believe that integrating expert knowledge with model-based forecasts is beneficial, see also Franses (2008).

Despite the evidence suggesting that merging models with experts would be a sound strategy, the remarks made in Armstrong and Collopy (1989) still hold value and that is that much further research is needed on how integration of models and experts can be improved. In the present paper we intend to contribute to this research issue by basically asking a large group of experts who are in a position to adjust model-based forecasts what they actually do and why they do so. Our study is unique of its kind as we have access to a large variety of experts in various countries located all over the world, and many of them quickly returned the questionnaires.

The outline of our paper is as follows. In Section 2 we review the answers to our survey questions given by only those experts who say that they do rely on the model-based forecasts before deciding on their adjustments. In Section 3 we discuss the answers of all experts, so now also including those who do not look at the model-based forecasts before they make their own. In Section 4 we correlate the actual behavior of the experts with their stated opinions. In Section 5 we summarize the conclusions of these three sections, and give some suggestions for further research.

2. Experts who use the model-based forecast to create their own

We have access to experts who are associated with the local (national) sales offices of a large pharmaceutical company. The headquarters' office of this multinational company uses an automated statistical package that each month generates sales forecasts for one-month to twenty-four months ahead. Sales data concern a range of products within seven product categories. The model-based forecasts are sent to the local sales offices, and the experts quote their own forecasts. Both sets of forecasts are stored. The model-based forecasts are based on lagged sales data only and the experts are very much aware of that. Hence, the expert forecasts are usually different from the model-based forecasts. The models used in the package are regression-type models and this means that on average they do not deliver biased forecasts. Model choice and parameter updating is carried out each month. So, recent exceptional sales figures are incorporated in the model by allowing model and parameters to change.

We sent a questionnaire by e-mail in November 2007 to experts in forty countries, who are located in national sales offices. Each expert is responsible for the forecasts of particular product category within a certain country. Out of the around one hundred surveys sent, we received forty-two valid responses, where some countries had more than one expert fill in the questionnaire. There are some missing values for some questions but not that many. The experts were asked to answer on a seven-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). The experts are allocated in Norway (1 expert), Belgium (4 experts), South Korea (4), Russia (4), Spain (3), Austria (2), Finland (3), Poland (2), UK (1), Switzerland (1), Denmark (2), Colombia (1), Thailand (3), France (1), Brazil (1), Turkey (1), Australia (1), Mexico (3), Czech Republic (1), and Chili (1). The distribution of age of these experts is close to uniform with a minimum age of 20 and a maximum age of 65, and there are also about as many men as women.

The first question is whether the expert relies on the model-based forecast in order to create his or her own forecast. This is important as the experts know that the model-based forecasts are created using a regression-type model, with as input only lagged sales data. So, it is quite likely that an expert modifies the model-based forecast if the expert believes that other variables are important too. On the other hand, as the model parameters are updated each time, strictly speaking the experts do not need to again consider lagged sales data when creating their adjustment. In fact, using past sales data again amounts to double counting.

Twenty of the forty-two experts indicated that they relied on the model-based forecasts. It is this group of experts that were asked to answer the first set of questions, while the others were directed towards the second set of questions.

Insert Table 1 about here

In Table 1 we give the fourteen questions that were to be answered by the twenty experts who first take a look at the model-based forecasts before they create their own forecasts. We give in boldface those questions and scores where the mean value differs significantly from 4, which is the neutral category. For this, we use a simple binomial test for the null hypothesis that the probability is 0.5, where we count the number of answers larger (smaller) than 4 and compare this with the answers smaller than and equal to (larger and equal) 4. For a sample size of around 20, this implies that 15 or more (5 or less) answers larger (smaller) than 4 entails significance at the 5% level.

Table 1 shows that only four of the fourteen questions amount to significant outcomes, which is perhaps partly due to the not-so-large samples. From this table we can conclude that the experts, who first consider the model-based forecasts before they create their own final forecasts (i) believe that this model-based forecast is important for their own decision to adjust, (ii) do not believe that the model creates forecasts that divert away from the trend, (iii) are of the opinion that they do about equally well as the model does, but (iv) are convinced that the model does not capture recent country-specific events. These outcomes all seem as expected, given the responses of these twenty experts to the starting question.

3. All experts

The second set of questions was answered by all forty-two experts. We test whether experts who said they use the regression model before creating their own forecast differ from those who say they do not. We find evidence of such differences for just three of twenty-one of the second set of questions.

Insert Table 2

The results appear in Table 2. Again we use a binomial test to examine significance. For a sample size of around 40, this implies that 28 or more (12 or less) answers larger (smaller) than 4 is significant at the 5% level. This table gives a few rather interesting insights. The first is to be observed from questions 2.1 and 2.3 and that is that experts who do not use the model-based forecasts as input for their adjustment state that they can do just as well without those forecasts indeed and also that they use other variables than the model does.

On the other hand, those experts who do use the model forecasts see no reason (question 2.20) to have the model changed.

The second insight is that all experts seem to include recent sales data as input for their adjustment, see question 2.5. As all experts are aware of the fact that the model-based forecasts are based on these lagged sales data, this result suggests that the experts count double. Indeed, as the model parameters are updated each month, and as recent sales figures are always included, there strictly is no need to look again at recent sales figures for adjustment. This result seems to concur with lab findings in the study by Goodwin and Fildes (1999). There, people did not adjust the time series model forecast in case of a special event. In fact, they ignored the model-based forecast, and created their own forecast based on the data and incorporated the effects of the special event.

The third insight (from questions 2.6 and 2.9) is that experts take account of changes in legislation or the behaviour of competitors in their specific countries, while they look less at the country's economy nor at the world economy (see the insignificant questions 2.10 and 2.11). Hence, they do not only look at recent sales but also at other features when they create their own forecasts.

Interestingly, when recent sales figures fluctuate much, and so there is an increased volatility in recent data, the experts feel the need to adjust more (question 2.8), that is, their forecasts tend to deviate more from the model-based forecasts. This is in accordance with the results from lab experiments in for example Harvey (1995).

Finally, the experts feel quite confident with what they do, as question 2.17 shows that they feel that their adjustments (whether or not they take the model-based forecasts as their starting point) make the final forecasts better. Whether this is true or not is left for further research and is for now outside the scope of this paper. Although, it seems to match with the evidence obtained in a recent survey in Fildes and Goodwin (2007) which is that experts state not to bother too much about whether their adjustments actually improve the final forecasts.

Even though they feel confident, the experts do seem to want to know more about how the model-based forecasts are created (question 2.21). An interesting implication of this finding is that it substantiates the claim that it is needed for forecasting systems to explain how they work, see also Gönül, Önkal and Lawrence (2008) for a recent discussion, see also Yaniv and Kleinberger (2000).

4. Correlation with actual behaviour

Now we have read about what experts say they do and what they perceive as relevant and useful, it is interesting to see if their stated behaviour matches with the observed behaviour. For this purpose we collect data on the size of the adjustment and of its sign, all concerning the one-month-ahead forecasts. We assign the country-specific data to each of the experts in that country, in case we have more than one expert responding. An expert i takes care of n_i products in each of the countries mentioned. This number can vary from 10 to 86. We measure the adjustment $A_{i,j}$ for the j-th product made by the expert as the final expert forecast minus the model-based forecast $M_{i,j}$. To measure the average size of adjustments made by each of the experts we compute

$$A_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{abs(A_{i,j})}{M_{i,j}}$$

where we take absolute values to equally capture both upward and downward adjustments.

The second variable of interest is the amount of upward adjustments as a percentage of all adjustments. We calculate this fraction per country as

$$positiveA_i = \frac{1}{n_i}\sum_{j=1}^{n_i} I(A_{i,j} > 0)$$

where *I(..)* denotes a one-zero indicator variable. As we stated earlier, when there is more than one expert in a country we simply take the country value for each of them.

Before we see if these two variables can be explained by the two sets of questions, we examine if we can reduce the number of regressors by applying principal components analysis (PCA). For the first set of fourteen questions in Table 1 we obtain a cumulative percentage of 77.2% of the variance for as many as five principal components, which does not amount to a serious reduction of variables. For the second set we need no less than eight principal components to capture 75.5% of the variance. Together this suggests that PCA does not lead to much gain in reducing variables, so we decide to simply include all questions in the subsequent regression models.

Insert Table 3

In Table 3 we report on the regression results for the set of questions that appeared in Table 1. We start with including all variables and then subsequently delete the variable with the least significant parameter until all parameters are significant at the 5% level. In the end, the models include only three and two relevant variables, while the fit is 0.274 and 0.448, respectively. From the second column we learn that experts who feel there is always a reason to adjust, who look at their own past performance but who do not believe that they can create the model forecasts have a tendency to have a larger-sized adjustment. From the last column in Table 3 we see that experts who look at past model forecast errors and who believe the model often has the trend wrong have a tendency to quote forecasts in excess of those of the model. These findings suggest that experts' stated behaviour seems to match with observed behaviour, although a substantial amount of variance in behaviour cannot be explained by the included variables.

Insert Table 4

In Table 4 we give the estimation results when the models have as regressors the questions in Table 2. The sample size increases to 39 data points. The fit of the two models is not high, just around 0.250, so quite some fraction of observed behaviour cannot be captured by stated behaviour. From the second column of Table 4 we can learn that some mean reverting behaviour (question 2.13) as well as rather steady adjustment behaviour (no reason to change size of adjustment depending on size of sales values) leads to larger-sized adjustment. The final column of that table indicates that experts who feel they should be rewarded and believe they add to the model on a 50-50 basis do not have a tendency to adjust upwards.

5. Conclusion

We have asked a range of experts who give forecasts that can deviate from model-based forecasts various questions about their decision making, beliefs and opinions. These experts are active in no less that twenty different countries and otherwise age and gender are about equally distributed. The not so unexpected outcomes are that experts feel that the model misses country-specific events and therefore adjustment is needed and that experts believe that their contribution to overall forecast quality is on a 50-50 basis with the model. Whether the experts discount here for the possibility that they use (about) the same information set as the model does, is left for further research.

Indeed, we see that experts admit to use past sales data again when constructing their own forecasts, and strictly speaking this is not needed, and hence there are indications of double counting. We also see that heavier fluctuations in past sales make experts adjust more.

Upon correlating experts' statements with actual behaviour we see that confidence in own quality leads to larger-sized adjustment. Moreover, less trust in the model, in fact, stating that the model often has the trend wrong makes experts to forecast higher than the model does. We also noticed that experts' behaviour can be explained by stated behaviour with only low values of fit. So, this might have a practical implication, which is that it seems even more relevant for experts to record the reasons for their adjustments. Even better, they should also periodically review whether those reasons were valid or not.

In sum, our survey results could have an impact on how we should evaluate the quality of the contribution of experts to model-based forecasts. If experts rely on past behaviour and on past sales data, then forecast evaluation criteria need to account for that. Also, research into the (over-) confidence levels of experts would be interesting. Finally, a more detailed look at what experts actually do, perhaps by keeping logbooks or the like would be even more useful.

Table 1:

Scores on questions only answered by experts who answer "yes" to the question: "Do you use the model-based forecast from the headquarters' office to determine your adjusted forecast?"

Questions with mean scores that are significantly different from 4 (using a binomial test) are indicated in boldface

		Sample size	Mean	Median	Standard deviation
1.1	**The model-based forecast is an important input for my decision to adjust**				
		18	**4.94**	**5.00**	**1.11**
1.2	I look at past differences (errors) between *actual sales* and the *model-based forecast*				
		18	4.56	5.00	1.15
1.3	There is always a reason to adjust the model-based forecasts				
		18	4.22	4.00	1.26
1.4	I compare the model-based forecast errors and my forecast errors				
		18	4.67	5.00	1.37
1.5	The model-based forecast has a tendency to be off track for long periods of time				
		19	3.68	4.00	1.16
1.6	I would be able to create the model-based forecast myself				
		19	4.26	4.00	1.48
1.7	I prefer to make small adjustments to the model-based forecasts				
		20	4.85	5.00	0.99
1.8	The model-based forecast almost always needs upward adjustment				
		20	3.95	4.00	1.19
1.9	The model-based forecast is not good enough on its own				
		20	3.65	3.50	1.63

1.10 My adjustments make the forecasts better, because my forecasts exceed those of the model

	20	3.80	4.00	1.11

1.11 The model-based forecast typically has the trend wrong

	20	**3.25**	**3.00**	**1.52**

1.12 Sometimes my forecast is better and sometimes the model-based forecast is better

	20	**5.25**	**5.00**	**0.97**

1.13 For the short horizon, the model-based forecast is better, but for the longer horizon, my forecasts are better

	20	3.90	4.00	1.52

1.14 The model-based forecasts miss recent country-specific events

	20	**5.50**	**6.00**	**1.40**

Table 2:

Scores on questions answered by all experts

Questions with mean scores that are significantly different from 4 (using a binomial test) are indicated in boldface

		Sample size	Mean	Median	Standard
2.1	**I could do equally well without the model-based forecast**				
	Yes	18	3.56		
	No	**22**	**5.36**		
2.2	I look at past differences (errors) between actual sales and my own forecasts.				
		40	4.43	5.00	1.57
2.3	**To create the adjusted forecast I use other variables than the model does**				
	Yes	19	4.58		
	No	**22**	**5.36**		
2.4	My recent forecast error is more important than model-based forecast errors				
		40	4.45	5.00	1.50
2.5	**The most recent actual sales data are input for my adjustment**				
		42	**5.40**	**5.50**	**1.23**
2.6	**Some events like changes in legislation or behaviour of competitors are so important that the model-based forecasts have to be adjusted for the next few months**				
		42	**5.74**	**6.00**	**1.27**
2.7	If the recent actual sales are large, then a downward adjustment is required				
		41	4.46	4.00	0.95
2.8	**When the actual sales fluctuate much, I adjust more**				
		42	**4.90**	**5.00**	**1.28**

2.9	**I keep track of local changes in legislation**	**42**	**5.83**	**6.00**	**1.32**
2.10	The country’s economy should be closely watched	40	4.73	5.00	1.48
2.11	My adjustment depends on what happens in the world economy	40	2.88	3.00	1.36
2.12	I should better forecast too high than too low	40	4.33	5.00	1.33
2.13	When my adjustment in the previous month was large, I now adjust less	40	3.90	4.00	1.26
2.14	Very low recent actual sales do not affect my forecast adjustments	40	3.83	4.00	1.50
2.15	What happened more than a year ago is not relevant	41	3.29	3.00	1.58
2.16	My market knowledge cannot be included in the model-based forecast	41	3.85	3.00	2.08
2.17	**My adjustments make the forecasts better**	**41**	**5.29**	**5.00**	**1.09**
2.18	I should be rewarded for my forecast accuracy	41	4.24	5.00	1.93
2.19	On average, the model-based forecasts and my adjustment are in 50%-50% balance	39	4.18	4.00	1.10

2.20 I would recommend the headquarters' office to change the way the model-based forecasts are created

Yes	**19**	**2.89**		
No	22	4.23		

2.21 I would like to learn more about how the model-based forecasts are created

	41	**5.10**	**5.00**	**1.37**

==

Table 3:

Auxiliary regressions of log(A_i) and positive A_i on answers to the first set of questions. Standard errors appear in parentheses and are White-corrected for heteroskedasticity. Sample size is 18.

Only parameter estimates that are significantly different from 0 are reported.

Question	log(A_i)	positive A_i
1.2		0.049 (0.021)
1.3	0.136 (0.065)	
1.4	0.158 (0.064)	
1.6	-0.217 (0.098)	
1.11		0.073 (0.019)
R^2	0.274	0.448

Table 4:

Auxiliary regressions of log(A_i) and positive A_i on answers to the second set of questions. Standard errors appear in parentheses and are White-corrected for heteroskedasticity. Sample size is 39.

Only parameter estimates that are significantly different from 0 are reported.

Question	log(A_i)	positive A_i
2.13	0.261 (0.074)	
2.16	0.100 (0.039)	
2.18		-0.016 (0.008)
2.19		-0.041 (0.016)
R^2	0.250	0.234

References

Armstrong, J.S. and F. Collopy (1998), Integration of statistical methods and judgement for time series forecasting: Principles for empirical research, in G. Wright and P. Goodwin (eds.) *Forecasting with Judgement*, New York: Wiley.

Fildes, R. and P. Goodwin (2007), Against your better judgement? How organizations can improve their use of management judgement in forecasting, *Interfaces*, 37, 570-576.

Franses, P.H. (2004), Do we think we make better forecasts than in the past?: A survey of academics, *Interfaces*, 34, 466-468.

Franses, P.H. (2008), Merging models and experts, *International Journal of Forecasting*, in print.

Gönül, M.S., D. Önkal, and M. Lawrence (2008), The effects of structural characteristics of explanations on use of a DSS, *Decision Support Systems*, in print.

Goodwin, P. (2002), Integrating management judgement with statistical methods to improve short-term forecasts, *Omega,* 30, 127-135.

Goodwin, P. and R. Fildes (1999), Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy?, *Journal of Behavioral Decision Making*, 12, 37-53.

Harvey, N. (1995), Why are judgements less consistent in less predictable task situations?, *Organizational Behavior and Human Decision Processes*, 63, 247-263.

Klassen, .D. and B.E. Flores (2001), Forecasting practices of Canadian firms: Survey results and comparisons, *International Journal of Production Economics*, 70, 163-174.

Mathews, B.P. and A. Diamantopoulos (1986), Managerial intervention in forecasting: An empirical investigation of forecast manipulation, *International Journal of Research in Marketing*, 3, 3-10.

Sanders, N. and K.B. Manrodt (1994), Forecasting practices in US corporations: Survey results, *Interfaces*, 24, 92-100.

Sanders, N. and L. Ritzman (2001), Judgemental adjustments of statistical forecasts, in J.S. Armstrong (ed.), *Principles of Forecasting*, New York: Kluwer.

Syntetos, A., K. Nikolopoulos, J.E. Boylan, R. Fildes and P. Goodwin (2008), The effects of integrating management judgement into intermittent demand forecasts, *International Journal of Production Economics*, in print.

Yaniv, I. and Kleinberger, E. (2000), Advice taking in decision making: egocentric discounting and reputation formation, *Organizational Behavior and Human Decision Processes,* 83, 260-281.

Publications in the Report Series Research* in Management

ERIM Research Program: “Marketing”

2008

Experts' Stated Behavior
Youssef Boulaksil and Philip Hans Franses
ERS-2008-001-MKT
http://hdl.handle.net/1765/10900

---

* A complete overview of the ERIM Report Series Research in Management:
https://ep.eur.nl/handle/1765/1

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship