A Manager's Perspective on Combining Expert and Model-based Forecasts

Philip Hans Franses and Rianne Legerstee

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2007-083-MKT
Publication	December 2007
Number of pages	19
Persistent paper URL	http://hdl.handle.net/1765/10769
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

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	We study the performance of sales forecasts which linearly combine model-based forecasts and expert forecasts. Using a unique and very large database containing monthly model-based forecasts for many pharmaceutical products and forecasts given by thirty-seven different experts, we document that a combination almost always is most accurate. When correlating the specific weights in these "best" linear combinations with experts' experience and behaviour, we find that more experience is beneficial for forecasts for nearby horizons. And, when the rate of bracketing increases the relative weights converge to a 50%-50% distribution, when there is some slight variation across forecasts horizons.
Free Keywords	Model-based forecasts, experts forecast, combining forecasts
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

A manager's perspective
on combining expert and model-based forecasts

Philip Hans Franses
Rianne Legerstee

Econometric Institute
Erasmus University Rotterdam

December 5 2007

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

We are very grateful to Sander Demouge and his colleagues from Organon BV (Oss, the Netherlands) for allowing us to work with their huge database and to communicate our research findings to the academic community.

A manager's perspective
on combining expert and model-based forecasts

Abstract

We study the performance of sales forecasts which linearly combine model-based forecasts and expert forecasts. Using a unique and very large database containing monthly model-based forecasts for many pharmaceutical products and forecasts given by thirty-seven different experts, we document that a combination almost always is most accurate. When correlating the specific weights in these "best" linear combinations with experts' experience and behaviour, we find that more experience is beneficial for forecasts for nearby horizons. And, when the rate of bracketing increases the relative weights converge to a 50%-50% distribution, where there is some slight variation across forecast horizons.

Key words: Model-based forecasts, Expert forecasts, Combining forecasts

1. Introduction

There is abundant literature on the relative performance of model-based forecasts, experts' forecasts and their combination, see Lawrence et al. (2006) and for example the early work of Blattberg and Hoch (1990). The most common findings are that expert forecasts (before or after seeing the model-based forecast) can improve on model-based forecasts and that a combination of the model-based forecasts with an expert forecast is often even better. The literature so far mainly considers a few single-product, single-horizon and single-expert cases. In our present paper we aim to expand on the currently available literature by considering various products in various product categories, twelve different forecast horizons and thirty-seven experts. A main feature of our analysis is that we know a few characteristics of these experts and we also observe their behaviour. This allows us to correlate the optimal balance between the model and the experts with their characteristics and their behaviour, which in turn gives guidelines from a managerial perspective.

In this paper we empirically analyze a unique and very large database with model-based forecasts, expert forecasts and realizations concerning monthly sales of a range of pharmaceutical products for a large Netherlands-based firm. At the headquarters office, the model-based forecasts are automatically created by a statistical package, where the program each month allows for a re-specification of the model and it also re-estimates all parameters each time. The experts, located in local offices in thirty-seven countries, receive these forecasts and, after that, create their forecasts using their own expertise. We will see that expert forecasts often differ from the model-based forecasts, which is perhaps not unexpected given the fact that the automatic program includes as input only lagged monthly sales values, and that this fact is known to the experts.

The question the firm faces is whether the model-based forecasts are better, or the final expert forecasts, or perhaps a weighted sum of these two. The literature on combining forecasts in for example Clemen (1989) and Timmermann (2006) suggests that linear combinations of forecasts may improve on each of its contributors. So the first question we consider in this paper is whether there is an optimal weight for each of the thirty-seven experts. And, if so, is that robust across forecast horizons and does it differ across experts?

The second question that we try to answer is whether in case of such optimal weights there are any characteristics of the experts that can explain their variation, if there is any. This question is very relevant from a managerial perspective as it facilitates training of experts and also their selection prior to their appointment. Blattberg and Hoch (1990) claim that a 50%-50% rule would be best but this claim corresponds with unconditional weights as it is not correlated with experts' characteristics. Lamont (2002) demonstrated that age (experience) has a positive effect on the quality of an expert, but also that this effect is parabolic. There are also studies like Barber and Odean (2001) and Beyer and Bowden (1997) which find gender differences in (over-)confidence levels, so perhaps there are also such differences across the relative weights of the experts in the combined forecasts.

Finally, the degree of bracketing shall be important for the quality of the combined forecast. Larrick and Soll (2006, p. 112) state that when the rate of bracketing increases, so does the power of averaging forecasts. Their findings were based on experiments, and in the present study we shall seek empirical evidence for this statement based on factual data.

The outline of our paper is as follows. In Section 2 we outline the main features of our unique database. Section 3 deals with the methodology and gives the details of our empirical findings. Section 4 concludes with various implications for managers.

2. Data

Our data concern a firm that creates model-based forecasts and which has thirty-seven experts[1] (in an equal amount of countries) who are allowed to replace these model-based forecasts with their own expert forecasts. Characteristics of these experts are available. The question the firm has is whether specific combinations of these two sets of forecasts are better in terms of point-forecast accuracy, and whether such combinations can be associated with observable characteristics and recent behaviour of these experts.

To start, we have data on $MF_{i,j,t+h|t}$, denoting a model-based forecast created at time t for horizon $t+h$ for sales in country i for product j. The forecasts concern monthly sales of pharmaceutical products, and the sample covers October 2004 to and including October 2006. The countries range from the US, UK, Korea, Austria, Thailand, to Malaysia and Mexico. In our analysis below we label the countries as $i = 1, 2, .., I = 37$ for confidentiality reasons. The index j runs from 1, 2, to J_i, which means that for each country we have a different set of products that belongs to the responsibility of the local expert. The products are associated with seven different product categories. The smallest number of J_i is 10, the largest is 85.

[1] In some countries there are two experts, but for the sake of notation we will label these as a single expert.

Finally, we have forecasts for horizons 1, 2, 3, to 10, 11 and 12 (a year ahead). The headquarters' office uses an automated statistical package to create these forecasts, where the input contains lagged sales data. The model selection process is rerun each month and also parameter estimation is redone each month.

Additional to the evidently enormous amount of model-based forecasts we have access to expert forecasts, to be denoted as $EF_{i,j,t+h|t}$. The experts make these forecasts upon receipt of the model-based forecasts, and they are aware of the fact that the automated program only includes lagged sales. Typically, as we will see bellow, these expert forecasts differ from the model-based forecasts, which is not unexpected given that experts may see various reasons to adjust pure own-history-based projections. Unfortunately we have no information on what exactly drives the expert to make different forecasts and also not on which factors they look at that are included in their final expert forecasts. We therefore simply take these expert forecasts as a second source of future outlook. Finally, we have actual sales denoted by $S_{i,j,t+h}$ corresponding with the two sets of forecasts.

In this paper we aim to draw generalizing conclusions on combining model-based forecasts with expert forecasts, and we decide to analyze the data in the dimensions *i* and *h*, and thus to aggregate across the products for each expert. Unreported experimentation with the data at a more detailed level indicated that differences across the products are not relevant, and so we can safely aggregate along that dimension.

Insert Table 1 about here

To get a first impression of the type of data that we have, we report on some basic statistics in Table 1. For the twelve horizons, we compute the fractions where the model-based forecasts and expert forecasts exceed or do not exceed each other, and also where they differ from the actual sales data. The first two columns of Table 1 correspond with what is called bracketing, meaning that the expert and model forecasts are on both sides of the realizations. The last two columns of Table 1 concern the cases where the expert in a sense adjusts the model-based forecast into the wrong direction.

If we consider the first few forecast horizons, we observe that the fraction that model-based forecasts are modified by the expert in the wrong direction is highest (close to 0.350 by summing the last two columns), where it most often happens that forecasts of experts are too high. Looking at the first two columns we see that bracketing occurs in around 0.270 of the cases. When we compare the various horizons, we see a slight increase in the fraction of bracketing cases, and a decrease in the fraction of wrong direction cases. The middle two columns seem to be rather constant around 0.330 across the forecast horizons.

Insert Table 2 about here

That experts have a tendency to create forecasts that exceed model-based forecasts is made more explicit in the second column of Table 2. Whether this is due to the very nature of supply chain management which entails perhaps that having too much stock is not as bad as being out of stock is an open issue, but clearly the experts adjust in a positive way more often. This tendency clearly decreases with increasing forecast horizon. At the same time, and as expected, the model-based forecasts are around 50% above and 50% below the realizations. This of course corresponds with the mean-reversion tendency of regression models with symmetric error assumptions, as they are implemented in the automatic program used by the headquarters' office.

Insert Table 3 about here

Finally, and before we turn to the main methodology, in Table 3 we give the available information we have on each of the thirty-seven experts. In some countries there are two experts, and then the data are averaged across these two. For each expert we know the (average) age, gender and the (average) number of years that experts occupy their (forecasting) position within the firm. We see that the average age is close to 40, that there are about as many men as women and that the experts are in office for on average 9 years.

3. Methodology and results

To address the managerial questions of the firm, which are typical questions any firm would have with access to a range of experts, we aim to compute the optimal value of the weights in a combined forecast. This combined forecast for each expert *i* given a horizon *h* is given by

(1) $$a_i MF_{i,j,t+h|t} + (1 - a_i) EF_{i,j,t+h|t}$$

where we compute the value of a_i across all products within an expert-horizon combination. To achieve this aim, we compute the root mean squared prediction error (RMSPE) as

$$\sqrt{\frac{1}{J_i}\sum_{j=1}^{J_i}[a_i MF_{i,j,t+h|t} + (1-a_i)EF_{i,j,t+h|t}) - S_{i,j,t+h}]^2} \quad (2)$$

for a_i = 0.00, 0.05 (with steps of size 0.05), .. 1.00. This gives 21 RMSPE values for each horizon, and we choose the value of a_i with the smallest value of RMSPE. Of course, when a_i = 0.00, the RMSPE for the pure expert forecast is lowest across all 21 cases and when a_i = 1, the RMSPE for the pure model-based forecast is lowest.

Insert Table 4 about here

The results of this first exercise appear in Table 4, and they are already quite interesting. Details of these computations can be obtained from the authors, but for now it suffices to say that almost all sequences of 21 RMSPE values show a (slight) parabolic curve, meaning that an optimum most often is reached within the range of considered a_i values. A closer look at the optimal weights in Table 4 shows that in only 4.73% of the 444 (37 times 12) expert-horizon cases, the value of a_i equals 0.00, and that in only 5.86% of the cases it equals 1.00. This strongly confirms the common finding that combined forecasts are more accurate than their individual components. Here, in 89.41% of the cases yield model-based forecasts when combined with expert forecasts improvement.

When we compute the average of the optimal weights, we get values around 0.50 (see the last row of Table 4), with a slight tendency to increase with increasing forecast horizon. This suggests that the relative weight of the model increases with the horizon. Hence, the unconditional weights 50%-50%, as suggested by Blattberg and Hoch (1990), seems to be a good choice indeed.

Optimal weights and experts

A further impression from the numbers in Table 4 is that there is substantial variation across experts, and hence it seems worthwhile to examine whether the optimal weights can be explained by experts' characteristics and his or her adjustment behaviour. The conditional model that we use for this purpose is

$$
\begin{aligned}
(3) \qquad & optimal\_a_i = \beta_0 + \beta_1 position_i + \beta_2 position_i^{\,2} + \beta_3 age_i + \beta_4 age_i^{\,2} \\
& + \beta_5 number\_of\_products_i + \beta_6 female_i + \beta_7 wrong\_sign_i \\
& + \beta_8 no\_adjustment_i + \varepsilon_i
\end{aligned}
$$

where we estimate the parameters using OLS. As the sample size is only 37, we rely on a 10% significance level. The optimal a_i value per expert follows from Table 4. The variables *wrong sign* and *no adjustment* are fractions of the total amount of forecasts which are adjusted in the wrong way or which are not adjusted, respectively. In the first round we estimate all parameters, and then subsequently delete the least significant ones, until we have at least 10% significant parameters.

Before we estimate the parameters, we formulate some prior thoughts on the possible relevance and sign of the parameters in (3). We measure the experience of an expert by the number of years he or she is in that particular position and by his or her age. The results in Lamont (2002) suggest that the effect of experience is positive for the quality of the expert, which here means that the parameters β_1 and β_3 would have a negative sign (giving smaller values of the optimal a_i and hence less weight to the model). Lamont (2002) also documents that much younger or much older experts perform not as good as medium aged experts, and hence we expect that β_2 and β_4 are positive. Additionally, we include the variable that counts the number of products an expert has to deal with as a measure of experience. We expect β_5 to be negative too. The studies in Barber and Odean (2001) and Beyer and Bowden (1997) suggest that female experts have a lower tendency to be overconfident, and hence might adjust less often, and this may also give the model more weight, so β_6 is expected to be positive. On the other hand, female experts may quote forecasts that differ less from the model-based forecasts, and this in turn may lead to more weight of the model, and then β_6 would be negative.

Concerning the actual behaviour of experts, we include two variables in (3). Adjustments by experts in the wrong direction would of course lead to less weight of the expert in the combined forecast, so we expect β_7 to be positive. Also, more cases with no adjustment would make the model more relevant, and so we expect β_8 to be positive.

As far as forecast horizons are concerned we have no particular prior hypotheses, except perhaps that, based on Table 2, a smaller difference of the expert forecast from the model forecast might be beneficial to the weight of the expert. This would mean that the parameter β_8 becomes more relevant for further away horizons.

Insert Table 5 about here

A first immediate conclusion that can be drawn from Table 5, where we only report on the 10% significant parameters, is that the variables *position*, *position*^{2}, *female* and *number of products* do not matter at all. Further, the results in Table 5 show that the horizon matters to fit the conditional mean of the optimal value of *a*. For the short term horizons like 1 to 5 experience matters while for further away horizons the degree of no adjustment matters. Also, for some horizons, the degree of wrong signed adjustments influences the optimal value of *a*.

We obtain the expected sign for experience, for experts' forecasts on the wrong side of the model-based forecasts and for the degree of no adjustment. Forecasts of older experts have more weight in the optimal combined forecast, and, as the squared variable is significant too, too young or too old gives less weight. This quadratic effect clearly supports the findings in Lamont (2002). The age which gives the minimum values of optimal a_i, and hence gives most weight to the expert, is around 40 years. Interestingly, experience does not seem to matter much for further away horizons.

For horizons 6, 7, 8 and 10, we find that the optimal weight cannot be predicted by the explanatory variables used in this paper and hence the best predictor is the unconditional mean. Looking at the standard errors for the intercept parameter, we see that 0.50 is within the 90% confidence bounds, which supports the claim in Blattberg and Hoch (1990).

Finally, for further away horizons we see that the degree of no adjustment becomes relatively more important when gaining weight for the expert in the combined forecast. This suggests that the degree in which the expert does not change the model-based forecast is indeed relevant.

Given that we find a useful predictive model for the optimal weight for various horizons, we conclude that the unconditional 50%-50% rule can sometimes better be replaced by a conditional rule. This insight adds to the insights given in Blattberg and Hoch (1990).

Bracketing

Theory predicts that bracketing, that is, model-based forecasts and expert forecasts are each on one side of the realization, makes combining forecasts a fruitful exercise. More so, as is argued in Larrick and Soll (2006), if the rate of bracketing increases, the power of simply taking averages increases. This argument rests on the assumption that the location on both sides of the realization of both the model-based forecast and the expert forecast obeys a uniform distribution. So, if *MF* is on one side of *S* and *EF* is on the other side, and the location of *MF* and *EF* is uniformly distributed, meaning that it does not occur that say *MF* is always closer to *S* that *EF* is, then on average *MF* and *EF* are equally close to *S*, and in that case the 50-50 rule should be optimal.

To examine this conjecture for actual data and not experiments as in Larrick and Soll (2006), we run the following simple regression, that is,

$$(4) \qquad (optimal\_a_i - 0.5)^2 = \beta_0 + \beta_1 bracketing_i + \varepsilon_i$$

where the explanatory variable is the fraction of forecasts that bracket the realization. We argue that when the conjecture is valid, that then β_1 in (4) is significant and negative. The relevant estimation results are displayed in Table 6.

Insert Table 6 about here

Similar to the results in Table 5, we observe that for intermediate horizons the distribution of the optimal value of a_i is hard to predict (see the low fit values for the horizons 6 to 10 in the last column of Table 6). On the other hand, for horizons 1 to 5 and for 11 and 12, we obtain strong evidence that the difference between the optimal a_i and 0.5 gets smaller for higher rates of bracketing, meaning that indeed bracketing makes simple averaging more powerful.

4. Discussion

Our paper analyzed a very large database with model-based forecasts and expert forecasts to see if combining these forecasts would be beneficial. Blattberg and Hoch (1990) predicted that unconditional weights of 50%-50% would be best. One of the novelties of our study is that we examined if these weights could be predicted by experts' characteristics and actual behaviour or performance, that is, whether there are perhaps conditional weights.

Main findings

We first confirmed that the model-based forecasts are unbiased, at least on average. This is very important as if that would not be the case, all subsequent analysis should have been modified. Evidence indicated that model forecasts are indeed unbiased, and also that experts have a tendency to deliver forecasts that exceed model-based forecasts in particular for nearby forecast horizons. Additionally, we documented that the fraction of bracketing is substantially smaller than the fraction of expert forecasts being on the wrong side of the model and of reality. In fact, bracketing occurs in only around one-fourth of all cases.

When we computed the optimal weights of combining model forecasts and expert forecasts, we found that the unconditional weights (across horizons and on average) are indeed close to a 50%-50%, but that there also is a strong variation across the experts. In fact, we showed that combined forecasts improve on the component forecasts in about 90% of the (large amount of) cases. Next, the optimal weights were shown to depend on experience (age), the degree of wrong-signed expert forecasts, and the degree of no adjustment in the hypothesized way for various forecast horizons. Hence the unconditional 50%-50% rule can be improved by including experts' characteristics and actual behaviour. Finally, we found that more bracketing leads to more indication that the 50%-50% rule is optimal.

Managerial implications

Our findings have various managerial implications. The first is that it is almost always best to combine model-based forecasts and expert forecasts. When the manager has no information on what the expert does or who he or she is, then the unconditional weights 50%-50% seem to be the best choice. However, when experts more often take extreme positions (on the wrong side of the model forecast) or do not adjust or have less experience the weights of the model-based forecast should be higher. On the other hand, when experts' and model forecasts would bracket the realizations, then the 50%-50% becomes more useful again.

When training new experts it is important to inform them that bracketing makes their contribution more relevant and that taking extreme positions (that is, the wrong side of the model-based forecast) does not. What also would help is to demonstrate to these experts that the model-forecasts are in general unbiased, and hence that most often quoting above or below a model forecast simply cannot be appropriate.

When hiring new experts it makes sense to ask for their past credentials in terms of their forecasts relative to the model forecasts. Note that simply choosing for the expert or for the model because the associated RMSPE is smaller than that of the other is not the best strategy, as we have seen that combined forecasts are almost always better anyway. So, their degree of bracketing matters, and as we saw, their experience does too. Literature suggests that too novice or too established experts have a tendency to take more extreme positions, and our findings suggest that this makes their relative contribution in a joint forecast smaller, at least for nearby horizons.

Table 1:
The differences between the model-based forecasts, the expert forecasts and the corresponding realization, measured in fractions across all cases J_i and I

Horizon	Bracketing				Adjustment in wrong direction	
	MF<S<EF	*MF>S>EF*	*MF<EF<S*	*MF>EF>S*	*EF<MF<S*	*EF>MF>S*
1	0.162	0.100	0.168	0.166	0.133	0.229
2	0.164	0.102	0.162	0.166	0.132	0.232
3	0.164	0.108	0.160	0.168	0.130	0.226
4	0.165	0.108	0.164	0.166	0.133	0.221
5	0.162	0.108	0.159	0.167	0.139	0.220
6	0.160	0.109	0.163	0.171	0.133	0.216
7	0.163	0.115	0.157	0.173	0.136	0.208
8	0.165	0.116	0.152	0.171	0.141	0.205
9	0.160	0.118	0.151	0.174	0.140	0.207
10	0.160	0.122	0.152	0.173	0.140	0.203
11	0.160	0.127	0.153	0.166	0.146	0.196
12	0.154	0.129	0.160	0.173	0.144	0.187

Table 2:
Expert forecasts relative to model-based forecasts and model forecasts relative to realizations, measured in fractions across all cases J_i and I

Horizon	*EF > MF*	*MF > S*
1	0.559	0.495
2	0.557	0.500
3	0.550	0.502
4	0.549	0.495
5	0.541	0.495
6	0.539	0.497
7	0.527	0.496
8	0.521	0.492
9	0.517	0.499
10	0.516	0.498
11	0.509	0.488
12	0.502	0.489

Table 3:
Characteristics of the experts (mean values if there are two experts) (and sample average)

Expert	Age (years)	Gender (1 = female)	Position (years)
1	30	0	10
2	37.5	0	15
3	55	1	15
4	40	0	5
5	35	1	5
6	30	0	5
7	47.5	0	17.5
8	45	0	10
9	60	0	20
10	30	1	10
11	20	1	1
12	45	1	5
13	35	1	5
14	40	1	15
15	45	1	5
16	50	0	15
17	40	1	3
18	40	1	1
19	30	0	5
20	45	0.5	10
21	35	1	5
22	25	1	2
23	50	0	20
24	32.5	0	7.5
25	45	0	10
26	35	0	10
27	45	0	7
28	35	0	3
29	35	1	5
30	30	1	3
31	55	0	10
32	35	0	5
33	60	1	15
34	30	0.5	4
35	55	1	20
36	30	1	5
37	30	0	2
Mean	40.07	0.47	8.61

Table 4: Optimal weights in the combined forecast, that is, $a_i MF_{i,j,t+h|t} + (1-a_i) EF_{i,j,t+h|t}$, when aggregated across products

Expert	Horizon											
	1	2	3	4	5	6	7	8	9	10	11	12
1	0.00	0.00	0.00	0.00	0.00	0.25	0.60	0.65	0.75	0.65	0.85	0.85
2	0.20	0.20	0.35	0.25	0.30	0.30	0.15	0.45	0.50	0.50	0.55	0.50
3	0.65	0.65	0.70	0.55	0.45	0.40	0.35	0.35	0.30	0.30	0.25	0.25
4	0.25	0.50	0.40	0.20	0.20	0.20	0.70	0.65	0.35	0.45	0.50	0.35
5	0.20	0.20	0.30	0.50	0.75	0.70	0.65	0.70	0.80	0.80	0.95	0.90
6	0.65	0.35	0.30	0.30	0.10	0.20	0.30	0.15	0.25	0.15	0.05	0.00
7	0.00	0.05	0.25	0.05	0.00	0.05	0.25	0.30	0.35	0.25	0.50	0.45
8	0.55	0.60	0.50	0.50	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
9	0.65	0.95	0.95	0.90	0.90	0.85	0.55	0.55	0.50	0.50	0.50	0.35
10	0.55	0.40	0.45	0.40	0.40	0.45	0.45	0.50	0.50	0.50	0.50	0.55
11	0.90	0.90	0.90	0.90	0.90	0.90	0.95	0.95	0.90	0.95	1.00	1.00
12	0.75	0.65	0.85	0.90	0.85	0.65	0.75	0.75	0.70	0.65	0.55	0.40
13	0.10	0.15	0.15	0.25	0.15	0.20	0.40	0.35	0.40	0.35	0.05	0.05
14	0.65	0.65	0.55	0.45	0.40	0.40	0.35	0.35	0.30	0.35	0.40	0.35
15	0.45	0.45	0.45	0.45	0.55	0.55	0.45	0.40	0.65	0.75	0.70	0.70
16	0.35	0.35	0.30	0.30	0.30	0.30	0.35	0.35	0.40	0.45	0.50	0.45
17	0.45	0.50	0.45	0.40	0.45	0.40	0.35	0.35	0.30	0.35	0.35	0.20
18	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.20	0.15	0.20	0.65
19	0.85	1.00	0.60	0.60	0.45	0.05	0.00	0.30	0.15	0.25	0.40	0.75
20	0.65	0.55	0.55	0.45	0.50	0.60	0.65	0.60	0.65	0.70	0.75	0.55
21	0.80	0.85	0.80	0.95	0.95	0.90	0.95	1.00	0.95	1.00	0.80	0.95
22	0.50	0.40	0.35	0.20	0.90	0.10	0.15	0.10	0.20	0.10	0.10	0.10
23	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.70	0.70	0.70	0.70	0.70
24	0.70	0.65	0.70	0.60	0.55	0.60	0.60	0.60	0.70	0.75	0.70	0.55
25	0.55	0.70	0.70	0.60	0.65	0.75	0.75	0.85	0.80	0.80	0.80	0.85
26	0.25	0.50	0.45	0.35	0.25	0.25	0.40	0.45	0.10	0.10	0.05	0.05
27	1.00	1.00	1.00	1.00	1.00	0.80	0.65	0.65	0.45	0.60	0.80	0.75
28	0.45	0.35	0.50	0.45	0.55	0.55	0.60	0.65	0.50	0.50	0.50	0.40
29	0.50	0.40	0.15	0.20	0.45	0.55	0.50	0.70	0.55	0.65	0.85	0.90
30	0.50	0.65	0.75	0.70	0.70	0.65	0.70	0.65	0.60	0.60	0.50	0.55
31	0.40	0.05	0.00	0.00	0.00	0.35	0.50	0.55	0.50	0.10	0.00	0.00
32	0.50	0.50	0.50	0.45	0.50	0.50	0.50	0.55	0.55	0.50	0.50	0.45
33	0.55	0.60	0.65	0.70	0.70	0.75	0.75	0.75	0.75	0.75	0.80	0.70
34	1.00	0.90	0.95	1.00	1.00	0.95	0.60	0.55	0.90	1.00	1.00	1.00
35	0.30	0.40	0.65	0.90	0.00	0.15	0.30	0.65	0.55	1.00	0.25	0.25
36	0.20	0.30	0.50	0.40	0.50	0.60	0.40	0.65	0.60	0.80	0.50	0.80
37	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.90
Mean	0.51	0.52	0.52	0.50	0.50	0.49	0.51	0.55	0.54	0.55	0.53	0.53

Table 5:
Estimation results for model (3), where insignificant parameters are deleted sequentially (using a 10% significance level). Parameters have been estimated by OLS, with a White-type correction for potential heteroskedasticity

Horizon	Variable (parameter and standard error in parentheses)					Fit
	Intercept	Age	Age^2	Wrong Direction	No Adjustment	
1	0.580 (0.574)	-0.044 (0.021)	0.0005 (0.0002)	2.248 (0.846)		0.246
2	0.727 (0.671)	-0.043 (0.026)	0.0005 (0.0003)	1.724 (0.984)		0.158
3	1.617 (0.579)	-0.057 (0.028)	0.0007 (0.0003)			0.079
4	1.668 (0.654)	-0.062 (0.031)	0.0008 (0.0004)			0.082
5	1.115 (0.729)	-0.058 (0.030)	0.0007 (0.0004)	1.609 (0.736)		0.199
6	0.489 (0.046)					0.000
7	0.508 (0.040)					0.000
8	0.547 (0.037)					0.000
9	0.508 (0.040)				0.547 (0.297)	0.063
10	0.553 (0.045)					0.000
11	0.502 (0.051)				0.614 (0.204)	0.062
12	-0.102 (0.304)			1.690 (0.840)	1.398 (0.369)	0.174

Table 6:
Estimation results for model (4). Parameters have been estimated by OLS, with a White-type correction for potential heteroskedasticity. The sample size is 36. The data on expert 11 are not included as they amount to an outlier.

Horizon	Intercept	Fraction of bracketing	Fit
1	0.191 (0.053)	-0.486 (0.178)	0.141
2	0.229 (0.055)	-0.575 (0.195)	0.164
3	0.192 (0.063)	-0.442 (0.212)	0.104
4	0.216 (0.061)	-0.489 (0.210)	0.115
5	0.246 (0.057)	-0.573 (0.198)	0.149
6	0.105 (0.031)	-0.127 (0.093)	0.014
7	0.035 (0.032)	0.065 (0.100)	0.004
8	0.030 (0.037)	0.069 (0.126)	0.005
9	0.055 (0.031)	-0.015 (0.103)	0.003
10	0.110 (0.046)	-0.138 (0.150)	0.017
11	0.163 (0.046)	-0.317 (0.151)	0.095
12	0.174 (0.044)	-0.334 (0.132)	0.107

References

Barber, B. and T. Odean (2001), Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics*, 116, 261-292.

Beyer, S. and E. Bowden (1997), Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias, *Personality and Social Psychology Bulletin*, 23, 157-172.

Blattberg, Robert C. and Stephen J. Hoch (1990), Database models and managerial intuition: 50% model + 50% manager, *Management Science*, 36, 887-899.

Clemen, Robert T. (1989), Combining forecasts: A review and annotated bibliography (with discussion), *International Journal of Forecasting 5*, 559-583.

Lamont, O.A. (2002), Macroeconomic forecasts and microeconomic forecasters, *Journal of Economic Behavior & Organization*, 48, 265-280.

Larrick, Richard P. and Jack B. Soll (2006), Intuitions about combining opinions: Mis-appreciation of the averaging principle, *Management Science*, 52, 111-127.

Lawrence, M., P. Goodwin, M. O'Connor and D. Onkal (2006), Judgemental forecasting: A review of progress over the last 25 years, *International Journal of Forecasting*, 22, 493-518.

Timmermann, Allan (2006), Forecast combinations, Chapter 4 in Graham Elliott, Clive W.J. Granger and Allan Timmermann (eds.), *Handbook of Economic Forecasting Volume I*, Amsterdam: Elsevier, pp 135-196.

Publications in the Report Series Research* in Management

ERIM Research Program: "Marketing"

2007

Marketing Communication Drivers of Adoption Timing of a New E-Service Among Existing Customers
Remco Prins and Peter C. Verhoef
ERS-2007-018-MKT
http://hdl.handle.net/1765/9405

Indirect Network Effects in New Product Growth
Stefan Stremersch, Gerard J. Tellis, Philip Hans Franses and Jeroen L.G. Binken
ERS-2007-019-MKT
http://hdl.handle.net/1765/9406

Demand-Driven Scheduling of Movies in a Multiplex
Jehoshua Eliashberg, Quintus Hegie, Jason Ho, Dennis Huisman, Steven J. Miller, Sanjeev Swami, Charles B. Weinberg and Berend Wierenga
ERS-2007-033-MKT
http://hdl.handle.net/1765/10069

Identifying Unknown Response Styles: A Latent-Class Bilinear Multinomial Logit Model
Joost van Rosmalen, Hester van Herk and Patrick J.F. Groenen
ERS-2007-045-MKT
http://hdl.handle.net/1765/10463

The Debate on Influencing Doctors' Decisions: Are Drug Characteristics the Missing Link?
Sriram Venkataraman and Stefan Stremersch
ERS-2007-056-MKT
http://hdl.handle.net/1765/10520

A Manager's Perspective on Combining Expert and Model-based Forecasts
Philip Hans Franses and Rianne Legerstee
ERS-2007-083-MKT
http://hdl.handle.net/1765/10769

---

* 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