19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011
http://mssanz.org.au/modsim2011

What drives the quality of expert SKU-level sales forecasts relative to model forecasts?

P. H. Franses [a] **and R. Legerstee**[a]

[a] *Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam*
Email: franses@ese.eur.nl; legerstee@ese.eur.nl

Abstract: We examine the quality of expert forecasts for SKU-level sales data (stock keeping unit), and correlate these forecast errors with the difference between expert forecasts and model forecasts. Recent literature suggests that there are at least two types of experts, that is, those who take the model forecast as input and those who ignore the model forecast altogether. This has consequences for analyzing the correlation between accuracy and behaviour. Analyzing a large database with SKU-level sales forecasts for pharmaceutical products, we document that small differences between expert forecasts and model forecasts can for example be associated with experts who do incorporate the model forecasts. Next, we document that these smaller differences are correlated with smaller expert forecast errors.

Keywords: *Model forecasts, expert forecasts, forecasting accuracy*

1. INTRODUCTION

We consider the situation where an analyst has access to statistical model forecasts for SKU-level sales data, to expert forecasts for the same data, and to the actual sales data themselves. The question the analyst faces is whether the potential behaviour of experts, that is, whether deviating from the model forecasts, leads to better forecasts.

There are various reasons why experts may quote different forecasts than the model forecasts, see Goodwin (2000, 2002). One reason is that the expert, upon receipt of model forecasts, believes that somehow the expected forecast error can be reduced by including information that is not in the model. Foreseeable institutional changes, new tax laws, and current incidents may then be taken into account. Another reason for experts to deviate from model forecasts is that experts feel that the model does not incorporate the correct information set or that model parameters are estimated incorrectly due to missing variables, and hence the expert may then ignore the model altogether and use the expert's own model. Goodwin (2000, 2002) also suggests various other reasons, some of which are of a more psychological nature.

It thus seems that there are at least two types of experts who give forecasts that can differ from model forecasts. Intriguingly, as is also documented in Armstrong and Collopy (1998) and various other studies, it is rarely exactly known what it is that experts do. Typically there are no logbooks, no written documentation and no descriptions of formal expert models (if there would be any). This information could be relevant as it may be that the various types of experts may also have difference in performance. Indeed, it would be interesting to know whether large expert forecast errors are caused by (too) large adjustments to model forecasts or by ignoring these model forecasts and by relying on improper expert models. In a recent study, Boulaksil and Franses (2009) survey more than forty such experts and these authors report that close to half of the experts claim not to include the model forecasts as input to their expert forecasts, which seems not a negligible fraction.

In this paper we examine the quality of expert forecasts for SKU-level sales data, and we correlate this quality with the difference between expert forecasts and model forecasts, while we take explicitly account of the possibility that there are at least two types of experts by fitting mixtures of distributions. We first analyze the behaviour of the experts and once we have discerned distinct clusters, we look at the quality of expert forecasts. In Section 2 we first review the relevant literature. Section 3 describes the database and the variables. Section 4 deals with the results and Section 5 summarizes the main findings and gives options for further research.

2. REVIEW OF THE LITERATURE

There is abundant evidence that experts quote different forecasts for SKU-level sales data than statistical model-based forecasts, see Fildes and Goodwin (2007), Bunn and Salo (1996), Sanders and Manrodt (1994), Nikolopoulos et al. (2005) and Syntetos et al. (2009), to mention just a few relevant studies. Typically, SKU-level time series forecasts are created from simple extrapolation techniques like exponential smoothing and trend fitting, and this may motivate managers to adjust, see Goodwin (2000) who also summarizes other reasons to adjust statistical forecasts.

There are studies that suggest that such managerial intervention leads to improved forecasts, see Mathews and Diamantopoulos (1986) and Diamantopoulos and Mathews (1989), but on the other hand Fildes and forecast may outperform model forecasts in terms of forecast accuracy, little seems to be known about what Goodwin (2007) suggest that perhaps model-based forecasts are adjusted too often, consequently leading to a decrease in accuracy.

A recent extensive study of Fildes et al. (2009) concludes that expert forecasts tend to be biased and that over-optimism (expert forecasts exceeding model forecasts) leads to lesser accuracy. Interestingly, these authors also document that large differences between model forecasts and expert forecasts lead to better expert forecasts.

Even though expert forecasts are often available additional to model-based forecasts, and even though expert managers or experts actually do and why they do so. The wish expressed in Armstrong and Collopy (1998), see also Sanders and Ritzman (2001) and Lawrence et al. (2006), which is that experts should keep records of their activities, is still not met. As there are not many, if at all, of these records, we thus need to rely on actual data to derive what it is that experts really do.

Franses and Legerstee (2009), accordingly, provide an extensive analysis of a large database (the same as will be used below) and they study the properties of (ten thousands of) expert forecasts. They show that in about 90% of all cases expert forecasts differ from model forecasts. On average, there is also a slight tendency that this difference is positive. Furthermore, on average, they find that the difference between expert forecasts and model forecasts is predictable. Finally, they document that the size of the differences strongly depends on past adjustment (habit formation, persistence), about three times as much as that it depends on past model-based forecast errors.

All results documented so far in the literature provide averages across all cases. No distinction is made for the behaviour of experts who use the model forecasts as input and those who ignore the model forecast, simply because precise information is not available. That this distinction is important however, can be concluded from the survey results documented in Boulaksil and Franses (2009). Out of the forty-two experts who responded to the survey, twenty indicated that they take the model forecasts as input for their own forecasts. Analyzing the answers to various other questions of those twenty and comparing these with the answers of the other twenty-two, Boulaksil and Franses (2009) conclude that experts who take the model forecast as input to their own expert forecasts (i) believe that this model-based forecast is important for their own decision to adjust, (ii) prefer to make small adjustments, (iii) do not believe that the model typically has the trend wrong, but (iv) are convinced that the model does not capture recent country-specific events (ibid, page 169). In contrast, those experts who ignore the model forecasts tend to have large differences between expert and model forecasts, have a tendency to quote forecasts that very frequently and sequentially differ from the model forecasts, believe the model persistently misses important variables, and also more often quote expert forecasts that are higher than model forecasts (these forecasts must then be biased).

In other words, expert forecasts which do not deviate much from the model forecasts are created by experts who show the behaviour as summarized in the following four hypotheses. That is, small differences are associated with experts who

H1: use the model forecast as input to their own forecast

H2: do not have a tendency to more often quote higher than lower forecasts

H3: give forecasts that do not persistently differ from the model forecasts, and

H4: give forecasts with unpredictable differences from the model forecasts.

The first three hypotheses straightforwardly follow from the survey results in Boulaksil and Franses (2009). The fourth hypothesis may need some more explanation. When experts take the model as starting point, and only once in a while see a reason to differ from the model forecasts, taking aboard information that is not in the model and only rarely is relevant, then it is clear that the difference between expert forecasts and model forecasts is not predictable from past data. However, when the expert makes an own model, which potentially also includes the variables used for the model forecasts, but with different parameters, then a difference between expert and model forecasts can be predictable from past sales data and also from past differences, see Franses and Legerstee (2009) for a more detailed discussion.

3. THE DATA AND THE VARIABLES

In this section we discuss the data that we have available and the construction of the relevant variables.

3.1 Data

Our data concern the SKU-level sales of pharmaceutical products. We have data on sales in 7 product categories, and the sample covers 25 months, running from October 2004 to and including October 2006. The products are from a pharmaceutical company which has its headquarters in the Netherlands, but also has local offices in 35 countries.

The headquarters' office uses an automated professional statistical package to generate model-based forecasts. These forecasts concern one-step ahead to at least twelve-steps-ahead forecasts. In this paper we will analyze the one-step-ahead forecasts and the six-step-ahead forecasts, where the latter choice is guided by advice from the headquarters' managers who indicate that due to supply chain management reasons this six-step-ahead horizon is important.

The forecasts are based on a statistical model which is, each month, selected from a range of models. The input variables are lagged SKU-level sales data, and this is known to the experts in the local offices. The program considers simple extrapolation models, Box-Jenkins' ARIMA type models, Holt-Winters' smoothing models, and various combinations. Each and every month, the most recent sales data are fed in, all models are calibrated again and, based on past but recent forecast performance, a choice is made for one of the models. Therefore, the models used to generate forecasts and estimated parameters can change over time.

The headquarters' forecasts are communicated with the local managers (experts) in 35 countries covering all continents, see the appendix. There are data for the US, the UK, Australia, China, Korea, but also for Peru, Algeria, Sweden, to mention just a few. These experts also give SKU-level sales forecasts, and hence the analyst has access to expert forecasts, model forecasts and the actual sales data. As mentioned however, it is unknown which expert takes the model forecasts as input. However, the survey results in Boulaksil and Franses (2009) suggest that a substantial fraction of experts simply ignores the model forecasts. This is important information as it is thus not true that the difference between the expert forecast and the model forecast is always based on judgemental adjustment. Indeed, this difference may also appear if the expert fully ignores the model forecast and simply gives an own quote.

The products can be captured in seven categories, but not all products are sold in all countries. In fact, we have 194 country-category combinations concerning one-step-ahead forecasts. Due to data limitations we have 190 such combinations for six-step-ahead forecasts. It is not the case that these combinations also associate with single experts. It more often holds that one expert in a certain country is responsible for more than one product category. The total amount of individual experts is about 50.

3.2 The Dependent Variable

The dependent variable is based on the absolute percentage expert forecast error

$$abs\left(\frac{Sales - ExpertForecast}{Sales} x100\%\right) \tag{1}$$

which we compute for every observation in each of the 194 (and 190 for six month horizons) country-category combinations. The dependent variable is then constructed using the average of all absolute percentage forecast errors per the country-category combination, which means that these averages are calculated over different amounts of data points, see the appendix. As the data are rather skewed, we apply the natural log transformation to finally get

$$Error = \log(average(abs\left(\frac{Sales - ExpertForecast}{Sales} x100\%\right))) \tag{2}$$

Notice that our analysis concerns a cross-sectional analysis, as we focus on experts who are responsible for these country-category combinations.

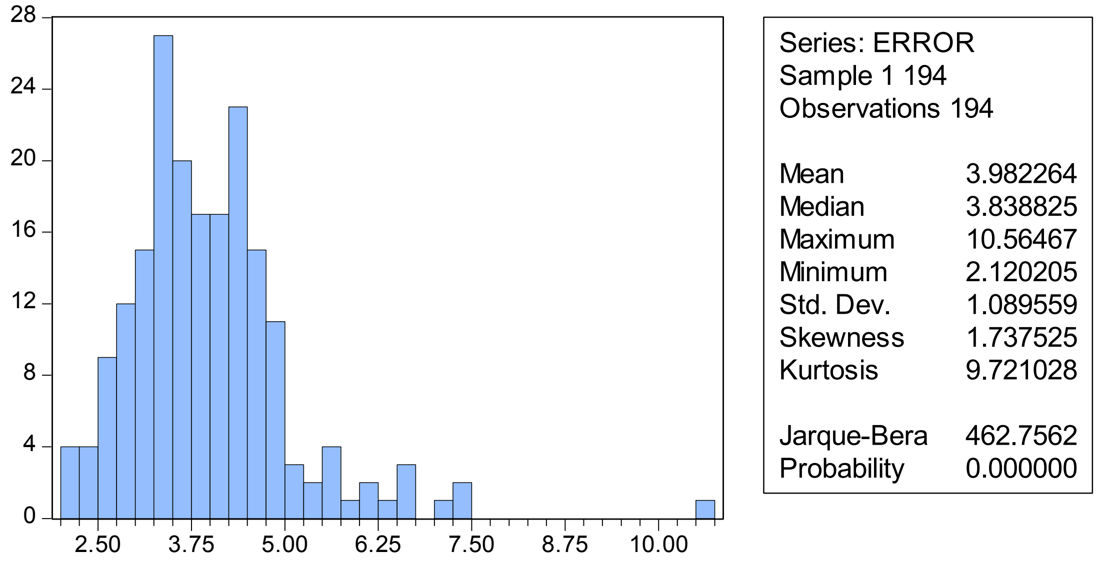

Figure 1. Distribution of *Error* values for 194 country-category combinations, the case of one-step-ahead forecasts.

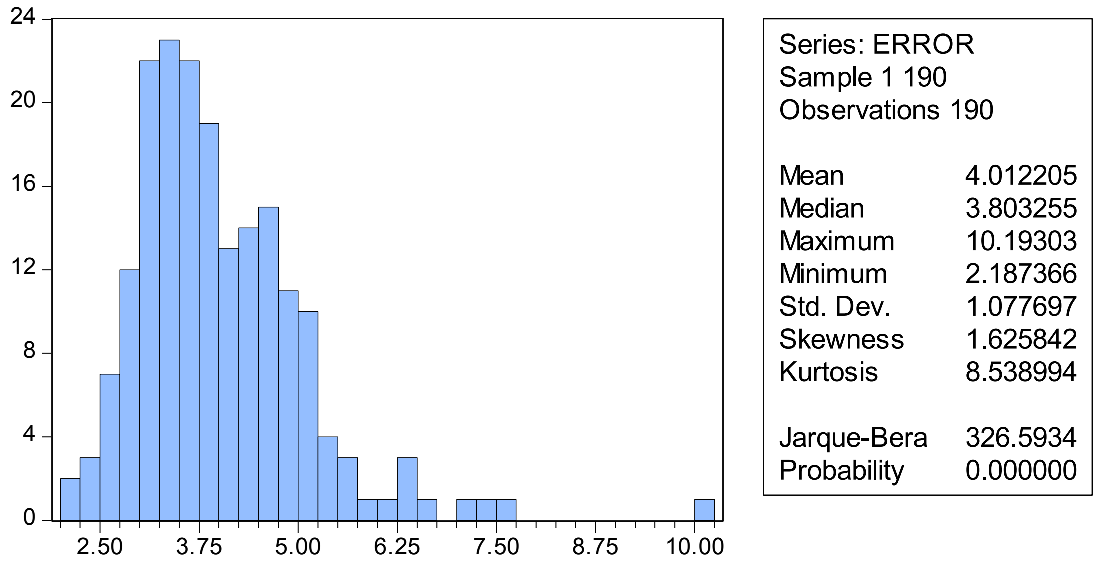

Figure 2. Distribution of *Error* values for 190 country-category combinations, the case of six-step-ahead forecasts

The *Error* variables are depicted in Figures 1 and 2. Clearly, even after the natural log transformation, the data are still skewed to the right, meaning that some expert forecasts are of very poor quality.

3.3 The key explanatory variable

In our analysis we include five explanatory variables. The most important variable is of course the difference between the expert forecast and the model forecast, as we aim to see to what extent this deviation improves the expert forecasts. Hence, we consider

$$\frac{ExpertForecast - ModelForecast}{ModelForecast} x100\% \tag{3}$$

as the explanatory variable, after similar transformation as in (2), that is

$$Difference = \log(average(abs\left(\frac{ExpertForecast - ModelForecast}{ModelForecast} x100\%\right))) \tag{4}$$

where *Difference* denotes the relative difference between the expert forecast and the model forecast and where again averages are computed across the observations within each country-category combination. Note that we do not use the word "adjustment" anymore as it is unknown whether the experts actually use the model forecast as their input for their own forecast.

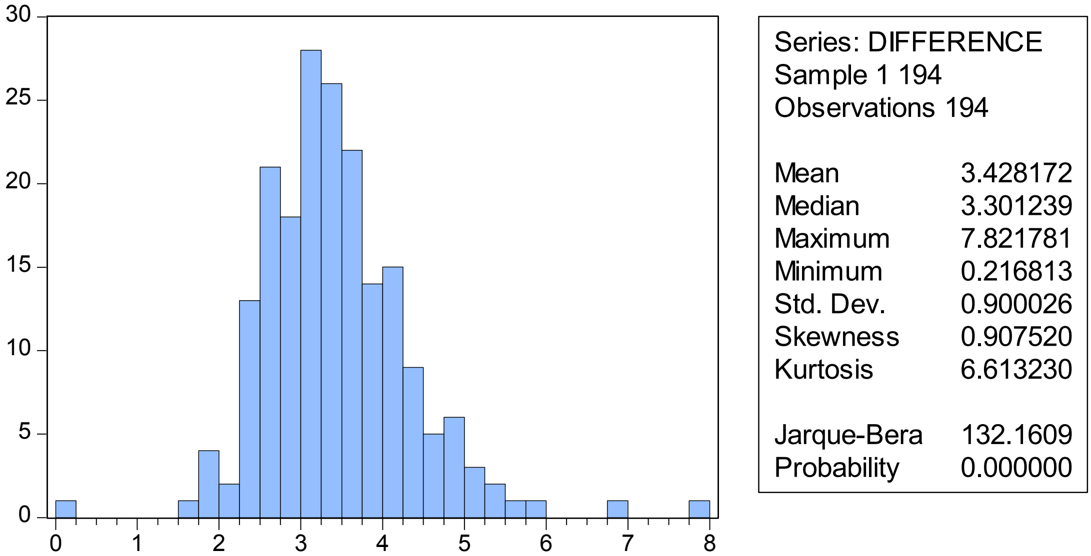

Figure 3. Distribution of *Difference* values for 194 country-category combinations, the case of one-step-ahead forecasts

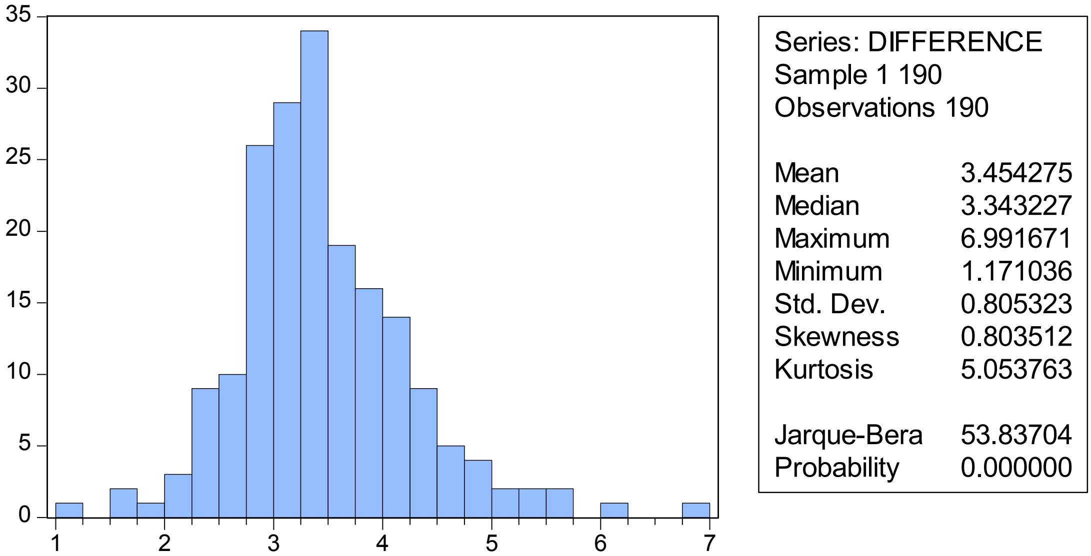

Figure 4. Distribution of *Difference* values for 190 country-category combinations, the case of six-step-ahead forecasts

As with the *Error* variable, also this variable is skewed to the right, meaning that some expert forecasts are very different from the model forecasts.

Fildes et al. (2009) document that large differences lead to more forecast accuracy. In contrast one might also postulate that, with regular updates of the parameter estimation, with recursive model selection (as is done in our case), and with unbiased model forecasts, it may happen that it is just small differences (on average) that can put the forecast on track. Given the variation in arguments, we do not postulate an a priori effect of *Difference* on *Error*, although we are tempted to hypothesize a positive effect that is, large differences give large errors.

3.4 Other explanatory variables

The next four variables that we include in our analysis are those which are associated with the hypotheses in Section 2. The first explanatory variable measures the correlation between the expert forecast and the model forecast. This is computed using the regression model that has also been recommended in Blattberg and Hoch (1990), and it reads as

$$Expert\ Forecast = \mu + \lambda Model\ \ Forecast + \varepsilon \tag{5}$$

with ε a random error term. The parameter of interest is λ, and we compute its average estimate across all regressions within each country-product category. When $\lambda = 1$, one can see from (5) that the difference between the expert forecast and model forecast is fully random, and also that the expert forecast entails the sum of the model forecast and a random term. Hence, in our analysis we include $|\lambda\text{-}1|$ as the first additional explanatory variable, where large values mean lesser relevance of the model forecast for creating the expert forecast. In the context of H1, we expect a positive effect of $|\lambda\text{-}1|$ on the size of the difference between expert forecasts and model forecasts.

The next variable is the fraction of positive differences between expert forecasts and model forecasts. Again we average across all products within categories and we obtain a percentage for each country-category combination. Denote this variable as "% positive". In line with H2 we expect that large values of this variable are associated with large differences between expert forecasts and model forecasts.

The last two variables are obtained from the auxiliary regressions that are also used in Franses and Legerstee (2009). There, the authors regress the differences between expert forecasts and model forecasts on past such differences (lags 1, 2, and 6), on past model forecast errors, and on past expert forecast errors (all at lags 1, 2 and 6). The sum of the estimated parameters for the own past differences (the so-called autoregressive terms) is called persistence, as is also commonly done in advertising studies where the interest lies in the long-run effect of advertising. We denote persistence by ρ, and H3 stipulates that larger persistence is associated with larger differences between expert and model forecasts. Finally, we use the fit of these regressions (R^2) as a measure of predictability, and again we expect a positive effect.

4. RESULTS

In this section we first analyze the one-step-ahead forecasts and then turn to the six-steps-ahead forecasts. Each time, we first look at the behaviour of the experts, and after that we discuss the link between *Error* and *Difference*.

4.1 Modelling the behaviour of experts, one-step-ahead forecasts

Figure 3 shows a histogram of the variable *Difference* for one-step-ahead forecasts, and Figure 1 that of *Error*. Clearly, these variables are not distributed as normal, and perhaps can be better characterized as a mixture of normal distributions. We fit a mixture of two and of three such distributions (with common variance). We compare fit, parameter estimates and standard errors, and conclude that three such distributions match with these data for *Difference* and two for *Error.*

Table 1. Estimation results of fitting mixtures of normal distributions to the *Difference* and *Error* variables, the case of one-step-ahead forecasts

Variable	Estimated mean	(standard error)	Probability
Difference	3.212	(0.129)	0.852
	4.473	(0.453)	0.137
	7.350	(0.735)	0.010
Error	3.830	(0.067)	0.959
	7.141	(0.237)	0.041

Table 1 shows that the distribution of the difference between expert forecasts and model forecasts can be captured by a mixture of three normal distributions (each with standard deviation estimated as 0.676 with standard error 0.046), while the distribution of the forecast errors can be captured by two such distributions (each with standard deviation estimated as 0.836 with standard error 0.049). There is a probability of 0.852 to be a member of the largest cluster of differences, and this cluster shows the smallest difference (3.212), which amounts to around a 25% absolute difference between the expert forecasts and model forecasts. There is a very small probability to be in the smallest cluster (0.010), which shows differences of on average of exp(7.350) which is about 1600% (!). We consider this cluster as a cluster with outliers, and from now the focus is on clusters 1 (25%) and cluster 2 (probability of 0.137, and mean about 90%).

For each of the 194 cross-sectional observations, our estimation routine gives the conditional probability to be in cluster 1 or in cluster 2 (and 3). Call these *cpd1* (small-sized difference) and *cpd2* (large-sized difference). It is now of interest to see which variables that measure the behaviour of experts correlate with these conditional probabilities. As discussed in the previous section, the explanatory variables for the *Difference* variable are X1 ($|\lambda-1|$), X2 (%positive), X3 (ρ) and X4 (R^2). The first regression for the small-differences cluster reads as

$$\log\left(\frac{cpd1}{1-cpd1}\right) = \mu + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \tag{6}$$

Given hypotheses H1 to H4 we expect all β parameters to be negative. The second regression is for the larger-differences cluster and it reads as

$$\log\left(\frac{cpd2}{1-cpd2}\right) = \mu + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \tag{7}$$

Of course, given hypotheses H1 to H4 we expect all β parameters to be positive. Applying Ordinary Least Squares (OLS) to (6) and (7) gives the estimation results as in Table 2.

Table 2. Modelling the probability of being in a cluster of small differences and of larger differences between expert forecasts and model forecasts (White heteroskedasticity-consistent standard errors). The case of one-step-ahead forecasts (mixture of 3 normal distributions)

Variables	Small differences Parameter (standard error)			Larger differences Parameter (standard error)	
Intercept	5.196	(0.989)	*	-6.082	(0.728)*
Correlation ($\lvert\lambda-1\rvert$)	-1.234	(0.398)	*	1.084	(0.382)*
%Positive adjustments	-2.720	(1.324)	*	3.804	(1.034)*
Persistence (ρ)	-0.862	(0.245)	*	0.952	(0.243)*
Predictability (R^2)	0.799	(0.699)		-0.477	(0.661)
P-value F-test		0.004			<0.001

Note: Some of the explanatory variables cannot be measured for all 194 observations due to data shortage. The effective sample size is 171 observations. Significance at 5% is indicated with *.

The results in Table 2 show that more persistent adjustment, a smaller correlation with the model forecast and more often positive adjustment is associated with a large probability to be in the cluster with larger differences with the model forecasts. In sum, we find strong support H1, H2 and H3, but not for H4.

4.2 Modelling quality of expert forecasts (one-step-ahead)

Now we turn to models for the error measure (2). From Table 1 we already learned that there are two clusters, the low quality cluster with very large forecast errors (mean is about 1260% (!)) and a larger cluster with reasonably small forecast errors (with mean around 46%). Again we can compute the conditional probabilities for each of the 194 observations, and we call these *cpq1* (small errors) and *cpq2* (large errors). Next, we consider the model (where we also include the four other explanatory variables as control variables) for the smaller errors

$$\log\left(\frac{cpq1}{1-cpq1}\right) = \mu + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 cpd1 + \beta_6 cpd2 + \varepsilon \tag{8}$$

Note that a model for *cpq2* is not needed as there are only two clusters. Applying OLS to (8) gives the estimation results as presented in the left-hand panel of Table 3.

Table 3. Modelling the probability of being in a cluster of (relatively) small forecast errors (White heteroskedasticity-consistent standard errors)

Variables	One-step-ahead forecasts Parameter	(standard error)	Six-step-ahead forecasts Parameter	(standard error)
Intercept	-3.497	(1.623)	2.270	(2.841)
Correlation (|λ-1|)	0.791	(1.062)	0.154	(0.293)
Positive adjustments	3.483	(2.821)	2.657	(2.487)
Persistence (ρ)	0.010	(0.529)	-.279	(0.407)
Predictability (R^2)	-2.420	(1.734)	0.088	(2.371)
cpd1	13.78	(0.497)*	6.576	(1.538)*
cpd2	2.762	(1.974)		
P-value F-test	<0.001		0.072	

Note: Significance at 5% is indicated with *. The effective sample for one-step-ahead forecasts is 171, and for six-step-ahead forecasts it is 164.

From the left-hand panel of Table 3 we see that the only relevant predictor of being in a cluster with small forecast errors is the probability of being in a cluster with small differences between expert forecasts and model forecasts. Note that this is in contradiction with the findings in Fildes et al. (2009), but supports our prior thoughts.

Table 4. Modelling the quality of expert forecasts in cases of (relatively) small forecast errors (White heteroskedasticity-consistent standard errors)

Variables	One-step-ahead forecasts Parameter	(standard error)	Six-step-ahead forecasts Parameter	(standard error)
Intercept	1.368	(0.336)*	1.698	(0.264)*
Correlation (|λ-1|)	-0.146	(0.123)	-0.038	(0.043)
% Positive adjustments	-0.725	(0.423)	-0.509	(0.367)
Persistence (ρ)	-0.048	(0.078)	0.028	(0.084)
Predictability (R^2)	0.300	(0.210)	0.242	(0.234)
Difference	0.845	(0.089)*	0.664	(0.072)*
P-value F-test		<0.001		<0.001

Note: Significance at 5% is indicated with *. The effective sample for one-step-ahead forecasts is 164, and for six-step-ahead forecasts it is 156.

Finally, we consider the regression

$$Error = \mu + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 Difference + \varepsilon \quad (9)$$

for the cases where forecast errors are small (*cpq1* > 0.5, which amounts to 163 effective cases for regression model (9)) and for the cases where these errors are large (*cpq2* > 0.5, 8 effective cases). The F test for the second cluster has a p-value of 0.411, meaning that no variables can predict very large expert forecast errors. The estimation results for the first cluster are given in the left-hand panel of Table 4. This indicates that the crucial variable is the size of the deviation of expert forecasts from model forecasts.

Table 5. Estimation results of fitting mixtures of normal distributions to the *Error* and *Difference* variables, the case of six-step-ahead forecasts

Variable	Estimated mean	(standard error)	Probability
Difference	3.323	(0.059)	0.929
	5.176	(0.229)	0.071
Error	3.871	(0.069)	0.957
	7.141	(0.248)	0.043

Note: Sample size is 190 observations.

Table 6. Modelling the probability of being in a cluster with small differences between expert forecasts and model forecasts (White heteroskedasticity-consistent standard errors). The case of six-step-ahead forecasts (mixture of two normal distributions)

Variables	Parameter	(standard error)
Intercept	9.109	(1.074)
Correlation (|λ-1|)	-0.967	(0.169)*
Positive adjustments	-4.141	(1.461)*
Persistence (ρ)	-0.456	(0.222)*
Predictability (R^2)	0.146	(0.157)
P-value F-test		<0.001

Note: Some variables could not be measured for all 190 observations due to data shortage, so the effective sample size is 164 observations. Significance at 5% is indicated with a *.

4.3 Results for six-step-ahead forecasts

In Table 5 we report the results of fitting mixtures to the *Difference* and *Error* variables for the six-steps-ahead forecasts. Again we find clusters of observations and they bear similarities with those in Table 1. Table 6 basically gives the same results as Table 2. Further, the right-hand side panels of Table 3 and 4 also do not differ much from the left-had side panels.

All this suggests that our empirical findings are quite robust across forecast horizons.

5. CONCLUSION

This paper has studied the deviation of expert forecasts from model forecasts and its impact on forecast accuracy. Before doing so, we recognized that it is mandatory to acknowledge that there could be at least two types of experts quoting forecasts. One cluster could take the model forecast as input and take it from there, while a second cluster could simply ignore the model forecast and experts would create their own forecast, perhaps using the same variables as were used for the model forecast.

Analyzing a large database concerning SKU-level sales data for pharmaceutical products managed by various different experts, we found clear evidence that experts who rely on the model forecast have a tendency to deviate less from that forecast, show lesser signs of over-optimism, and show lesser predictable behaviour. Further, larger deviations from the model forecasts led to poorer forecast performance. This result turned out to be robust across forecast horizons.

The novelty of our study is the fact that we discern at least two types of experts. We do not know exactly who they are, but based on behaviour we can estimate their latent class membership. We believe that this membership is important for studying the success of the interactions between modes and experts. Further understanding of the key characteristics of the various types of experts is needed.

APPENDIX

Number of products in the categories for each country

	Category						
Country	A	B	C	D	E	F	G
I	3	2	1	4	2		
II	3	4		2	7	1	
III	3	5	6	10	8	4	1
IV	2	1	3	4	2	2	
V	11	6	7	9	5	4	1
VI	2			2			
VII	7	4		1	6	1	1
VIII	9	4	2	6	6	6	
IX	2	1		1		2	
X		1	1		6	3	
XI	10	4		8	7	4	
XII	7	9	4	7	8	1*	
XIII		2	2	2	3	3	
XIV	12	10	2	9	8	5	
XV	23	3		6	18	4	1
XVI	32	20	1	16	10	5	1*
XVII	7	2	4	2	11	3	
XVIII	12	4	2	5	8	4	
XIX	10	5	3	5	8	3	
XX	1		1	2	6		
XXI	6	1	2	4	9	5	
XXII	1		1	2	6		
XXIII	9	5	15	10	12	4	1
XXIV	6	9	2	3	6	1	
XXV	3	3	2*	1	2	2	
XXVI	6	4	2	6	3	4	
XXVII	11	3	7	4	8	3	
XXVIII	7	2		5	3	2	
XXIX	12	7	4	6	10	2	
XXX	15	7	6	8	7	4	1
XXXI	15	12	3	11	9	5	1
XXXII	1				3		
XXXIII	8	8	13	15	12	5	1*
XXXIV	7	8	2		6	2	
XXXV	2	5		2	7	3	

* These cases are only available for the one-step-ahead forecasts but not for the six-step-ahead forecasts. So, the one-step-ahead forecasts concern 194 country-category combinations, while for the six-step-ahead forecasts there are 190 such cases.

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*. Wiley, New York.

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

Boulaksil, Y. and P.H. Franses (2009). Experts' stated behavior. *Interfaces*, 39, 168–171.

Bunn, D.W. and A.A. Salo (1996). Adjustment of forecasts with model consistent expectations. *International Journal of Forecasting*, 12, 163–170.

Diamantopoulos, A. and B.P. Mathews (1989). Factors affecting the nature and effectiveness of subjective revision in sales forecasting: An empirical study. *Managerial and Decision Economics*, 10, 51–59.

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.

Fildes, R., P. Goodwin, M. Lawrence and K. Nikopoulos (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25, 3–23.

Franses, P.H. and R. Legerstee (2009). Properties of expert adjustments on model-based SKU-level forecasts. *International Journal of Forecasting*, 25, 35–47.

Goodwin, P. (2000). Improving the voluntary integration of statistical forecasts and judgement. *International Journal of Forecasting*, 16, 85–99.

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

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.

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

Nikolopoulos, K., R. Fildes, P. Goodwin and M. Lawrence (2005). On the accuracy of judgemental interventions on forecasting support systems, Lancaster University Management School Working paper 2005/022.

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*. Kluwer, New York.

Syntetos, A., K. Nikolopoulos, J.E. Boylan, R. Fildes and P. Goodwin (2009). The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics*, 118, 72–81.