# What drives the relevance and quality of experts' adjustment to model-based forecasts?

Philip Hans Franses
Rianne Legerstee
*Econometric Institute*
*Erasmus University Rotterdam*

Econometric Institute Report 2007-43

## Abstract

Experts frequently adjust statistical model-based forecasts. Sometimes this leads to higher forecast accuracy, but expert forecasts can also be dramatically worse. We explore the potential drivers of the relevance and quality of experts' added knowledge. For that purpose, we examine a very large database covering monthly forecasts for pharmaceutical products in seven categories concerning thirty-five countries. The extensive results lead to two main outcomes which are (1) that more balance between model and expert leads to more relevance of the added value of the expert and (2) that smaller-sized adjustments lead to higher quality, although sometimes very large adjustments can be beneficial too. In general, too much input of the expert leads to a deterioration of the quality of the final forecast.

Keywords: Judgemental adjustment; expert forecasts

October 16 2007
We are grateful to Sander Demouge and his colleagues at NV Organon (Oss, The Netherlands) for compiling the database and for allowing us to communicate our findings to the academic community.

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

## 1. Introduction

Experts frequently adjust model-based forecasts. See Lawrence et al. (2006) for a recent survey of the relevant literature. These studies, and there are not that many, show that expert adjustment can improve the final forecast quality, see Mathews and Diamantopoulos (1986). All empirical results available are limited to one or a few cases. More so, there is no study that rigorously examines what are the drivers of better forecast quality due to the expert[1]. In the present paper we aim to provide a substantial amount of empirical evidence on the success factors of experts, and, as a consequence, we formulate some generalizing statements on the interaction between experts and models.

We analyze model-based forecasts, the expert-adjusted forecasts and the realizations, all concerning sales of pharmaceutical products. The company's head office uses a statistical model to generate forecasts for a range of horizons, and sends these to local offices in thirty-five countries. The managers in these countries are allowed to change these model outcomes and should report their final forecasts to the head quarters[2]. We consider various products within seven product categories, and the sample covers twenty-five months of data.

Before we can analyze such a huge database we need to propose a careful and sensible methodology. For example, the managers see the model-based forecasts before they modify them, and this should be taken into account.[3] Next, we should allow for the possibility that managers count double, which means that they overlook the fact that for example exceptional past sales values are already incorporated in the model, and that it is not necessary to correct for these exceptional values twice[4]. And, we should also consider more than one forecast horizon. In the present paper we outline this methodology in detail.

Section 2 deals with a few hypotheses that to some extent should guide our empirical analysis. Then, in Section 3, we discuss the data and the methodology. Section 4 contains the main results, and Section 5 concludes with suggestions for further work.

---

[1] There are two exceptions, which are Fildes and Goodwin (2007) and Nikolopoulos et al. (2005) who study the link between past expert behaviour and current forecast quality, but the methodology used in these studies suffers from various drawbacks, as we will describe below.
[2] Part of the bonus payments of these managers concerns the success rate of their forecasts.
[3] A simple regression based on the recommendations of Blattberg and Hoch (1990) will do.
[4] Bunn and Salo (1996) give a clear description of double counting.

## 2. Hypotheses

There are two phenomena that we wish to explain. The first is the relevance of the added value of the expert. It is important to note that we consider the added value and not the expert forecast itself, because the expert receives the model-based forecast prior to his or her decision to adjust[5]. Hence, we examine if the model-based forecast would on average be equally good with or without the expert. When the expert has a significant added contribution, we are interested in the reasons why. The second variable to be explained is the quality of the expert contribution, which we shall measure in terms of fit.

Now we turn to the features that might explain the relevance and quality of the expert. These features are (1) what he or she does now, (2) what he or she does with the model-based forecast, and (3) the expert's usual behaviour. Concerning (1) we can think of the size and sign of current adjustment. Basically, when the expert is, so to say, overdoing it, this should not be beneficial to the contribution's relevance and quality. So, we postulate that on average

H1: Large-valued adjustment (relative to the model-based forecast) and unidirectional adjustment lead to less relevant and lower quality expert's adjustment to model-based forecasts.

The model-based forecasts are generated automatically each month, and we know that the parameters are updated each month. This implies that each month, the forecasts are geared towards the mean of the sales data, which is the basic notion behind regression analysis. So, on average the model should do well, and unidirectional adjustment neglects this feature. Basically, one would want that the expert only adds expertise which is clearly not in the model and that is also not always relevant (think of some anticipated institutional changes or country-specific major events). Hence, the model can be expected to summarize past trends rather well and expert adjustment should be adding to that and not replacing it[6]. So, we postulate

H2: More trust in the model, and thus not simply replacing it by one's own opinion, leads to more relevant and higher quality expert's adjustment to model-based forecasts.

---

[5] Here we differ from the approach followed in Fildes and Goodwin (2007), for example.

[6] There is some literature (see Lawrence et al. 2006) that suggests that only using experts' knowledge leads to biased forecasts and that the combination of models and experts (perhaps even with equal weights) is best.

One remark here is that in case when the model-based forecast is obviously poor, one would want that the expert adjusts with a large added value (Fildes and Goodwin 2007).

Additional to what the expert is doing now, it is important to have some impression of what the expert is used to do. Indeed, as Bunn and Salo (1996) indicate, double counting is not beneficial for the final quality, as the model already incorporates recent exceptional events (and in particular, as in our case, when the parameters are updated each time). Note that when the expert counts double, we can predict future experts' adjustment by past sales data, see below. Therefore, the optimal scenario of expert's behaviour is that expert's adjustment is unpredictable. In that case the expert takes the model-based forecasts as given and adds, on top of that, knowledge on events that could not systematically be predicted. Indeed, if they were, the model would need revision as it seems to lack important variables. In sum, we postulate

H3: More predictability of the expert's behaviour (which includes double counting) leads to less relevant and lower quality expert's adjustment to model-based forecasts.

In the next section we shall operationalize these variables such that they become useful for our empirical analysis.

## 3. Methodology

In this section we first describe the database we have, and, based on that, we describe our methodology. This methodology is designed such that we can examine the validity of the three hypotheses in the previous section.

### 3.1 Data

Our data concern the sales of pharmaceutical products. We have data on sales in seven categories, and the sample covers 25 months, running from October 2004 to and including October 2006. A summary appears in the appendix. The headquarters' office uses an automated (professional) statistical package to generate one-step-ahead forecasts until and including twenty-four-step-ahead forecasts. Due to data limitations (we will need to run various regressions over time) we confine the analysis to one-step-ahead (short horizon) and six-step-ahead (long horizon) forecasts. This second choice is guided by advice from the headquarters' managers who indicate that, due to supply chain management reasons this six-step-ahead horizon is an important one. The headquarters' forecasts are communicated with the local managers in thirty-five countries, covering all continents. So, there are data for the US, the UK, Australia, China, Korea, but also for Peru, Algeria, Sweden, to mention just a few. The products can be captured in seven categories, and below we will mostly focus on the data within those categories. In sum, we shall analyze the data for 171 country-category combinations for the one-step-ahead forecasts and 164 such combinations for the six-step-ahead forecasts[7].

In this study we examine whether experts improve model-based forecasts by adding domain knowledge. We thus consider the following variables

$MF_{t+h}$: $h$-step-ahead model-based forecast (made from origin $t$)

$EF_{t+h}$: $h$-step-ahead expert forecast (made from origin $t$)

$S_{t+h}$: realization at time $t+h$

where $h$ will be 1 or 6 in our empirical work. The variable $S$ denotes monthly sales. The model-based forecast is some linear function of past sales, where the weights are updated each month, which entails a so-called recursive forecasting scheme. In short-hand, the model-based forecast can be written as

$$MF_{t+h} = \mu_1 + \rho_1 S_t + \rho_2 S_{t-1} + \rho_3 S_{t-2} + \ldots.. \quad (1)$$

The recursive scheme means that the parameters are estimated (using OLS, minimizing one-step-ahead forecast errors, as usual) for $R$ in-sample data, and then an $h$-step-ahead forecast is (iteratively) made. Next, the sample is enlarged to $R$+1, parameters are re-estimated and again $h$-step-ahead forecasts are made.

The expert receives the statistical model-based forecasts and quite often (in fact, as we will see: almost always) makes an adjustment. It is quite likely that part of that adjustment is based on past sales (again) and part on other domain-specific variables, say $X_t$. Note that the

[7] These numbers are slightly smaller than the amount of available cases mentioned in the appendix, which is due to the fact that sometimes there are not enough data points to calibrate one or more of the explanatory variables.

inclusion of past sales in experts' added value implies double-counting. In sum, the expert-adjusted forecast could be written as

(2) $$EF_{t+h} = \mu_2 + \delta_1 S_t + \delta_2 S_{t-1} + \delta_3 S_{t-2} + ..... + \beta X_t + ...$$

We have no specific information on $X_t$. Also, we do not know what are the values of the parameters and hence whether the $\delta$ parameters in (2) differ from the $\rho$ parameters in (1).

As said, the optimal expert's contribution would relate to the case where

(3) $$EF_{t+h} = MF_{t+h} + \beta X_t$$

with the two components being orthogonal and with $X_t$ being unpredictable. In the latter case, the added contribution of the expert to the model-based forecast would simply be $EF_{t+h}$ - $MF_{t+h}$, which assumes that both forecasts are independent and that the expert takes the model-based forecasts as given and adds his or her expertise. This assumption is made in Fildes and Goodwin (2007), amo0ng others. However, when the components in (3) are not independent, it is wise to correct for common patterns by computing

(4) $$A_{t+h} = EF_{t+h} - \lambda MF_{t+h},$$

where $\lambda$ gets estimated from a linear auxiliary regression of $EF$ on $MF$, as it is recommended in Blattberg and Hoch (1990). The value of $\lambda$ can be seen as a measure of trust in the model, as we will indicate below. Note that when $\lambda < 1$, this corresponds to more positive than negative adjustments.

### 3.2 Constructing the dependent variables

The dependent variables of interest are relevance and quality. To measure relevance, we consider the question whether the added value of the expert actually matters (note that this can be either in a positive or a negative way). This question can be answered by looking at the following auxiliary test regression

(5) $$S_{t+h} = \alpha + \beta MF_{t+h} + \gamma A_{t+h} + u_{t+h}$$

When the expert adds something that is relevant, the contribution of $A_{t+h}$ in (5) should be non-zero. So, a test for $\gamma = 0$ in (5) is important. Some unreported prior analysis reveals that the error term in (5) is not white noise. In fact, it is best to allow for second-order autoregressive dynamics, that is, to consider $(1-\rho_1 L - \rho_2 L^2)u_{t+h} = \varepsilon_{t+h}$ where $L$ is the familiar lag operator.

We have access to a maximum of 25 monthly observations for each product, and this sample becomes 23 observations for the case where $h = 1$, and only 18 for $h = 6$. This reduces the power of the test on the parameter of interest, and therefore we pool the estimates for (5) across all the products within each country-category combination. That is, we consider

$$
\begin{aligned}
(6)\qquad & S_{1,t+h} = \alpha_1 + \beta MF_{1,t+h} + \gamma A_{1,t+h} + u_{1,t+h} \\
& S_{2,t+h} = \alpha_2 + \beta MF_{2,t+h} + \gamma A_{2,t+h} + u_{2,t+h} \\
& \ldots\ldots\ldots \\
& S_{n,t+h} = \alpha_n + \beta MF_{n,t+h} + \gamma A_{n,t+h} + u_{n,t+h},
\end{aligned}
$$

where $n$ denotes the number of products within a category. We assume the $\beta$ and $\gamma$ parameters to be common across products within a country-category combination. For the sake of computational simplicity, we also assume the errors as independent.

The first dependent variable is now created as follows. We run the multiple-equation regression in (6) and document whether the $\gamma$ parameter is statistically significantly different from 0. If it is, it can be negative or positive. We label the outcomes as follows, that is a -1 if $\gamma$ is significant and negative, a 0 if it insignificant and a 1 if $\gamma$ is significant and positive. This amounts to an ordered categorical variable, and we therefore shall resort to an ordered probit model in our empirical analysis below.

The second dependent variable concerns the quality of the expert forecast relative to the model-based forecast. For this, we simply compute the difference between Root Mean Squared Prediction Errors (RMSPE) for the final expert forecasts versus the model-based forecasts. This is a continuous variable, and the more positive it is, the better is the quality of the experts' added knowledge.

### 3.3 Constructing the independent variables

We aim to find supportive or non-confirmatory evidence for three hypotheses, and we shall include one or two explanatory variables in the models below for each of these hypotheses. For Hypothesis H1, we need to construct large-valued adjustment and unidirectional adjustment. To measure the first we compute the average absolute size of adjustment, relative to the model-based forecast, that is, $|A|/MF$ averaged for all products within a country-category combination. To have a measure of the direction of adjustment, while taking account of the fact that the expert forecast and the model-based forecast are correlated, we use the value of $\lambda$. Interestingly, for Hypothesis H2 on trust in the model, we also use the estimated value of $\lambda$ in the auxiliary regression (4). When it is zero, or even negative, the expert expresses not much confidence in the model-based forecast, while when $\lambda$ is positive, the reliability of the model is appreciated. Most confidence in the model is obtained when $\lambda = 1$, and much beyond this value is also a sign of not much trust in the model. Later on, we will therefore include this variable as $|\lambda - 1|$ in the models and this variable relates to H1 and H2.

For Hypothesis H3 we need to spend a bit more effort in constructing the relevant variables. Some unreported prior experimentation indicated that a useful forecast model for current expert adjustment (at least for our data at hand) turns out to be

$$\begin{aligned} A_{t+h} &= \mu + \rho_1 A_t + \beta_1 (MF_{t-1} - S_{t-1}) + \rho_2 A_{t-1} + \beta_2 (MF_{t-2} - S_{t-2}) \\ &+ \rho_6 A_{t-5} + \beta_6 (MF_{t-6} - S_{t-6}) + \omega_1 A_t^2 + \lambda_1 (MF_{t-1} - S_{t-1})^2 \\ &+ \omega_2 A_{t-1}^2 + \lambda_2 (MF_{t-2} - S_{t-2})^2 + \omega_6 A_{t-5}^2 + \lambda_6 (MF_{t-6} - S_{t-6})^2 + \varepsilon_t \end{aligned} \quad (7)$$

We run this regression for all $n$ products with a country-category combination, and we compute the average $R^2$ across these $n$ regressions. This measure of fit indicates the degree of predictability of the expert's behaviour. Finally, to obtain a measure of double-counting, we compute

$$\rho = \rho_1 + \rho_2 + \rho_6 \quad (8)$$

from (7) which measures the total impact of past adjustment, or the persistence of adjustment.

Insert Tables 1 and 2 about here

In Table 1 we summarize the above operationalization of the explanatory variables. And, in Table 2, we give the expected signs of the various variables in the upcoming models that match with the hypotheses in the previous section. In the next section we discuss the data in more detail and we examine whether some hypotheses get supported or not.

## 4. Empirical results

The variables to be explained are the relevance and quality of expert adjustment. Relevance is quantified as an ordered variable with outcomes -1, 0 and 1, while quality is quantified as the difference between RMSPE's. The explanatory variables are (absolute) relative adjustment, the degree of unidirectional forecasts or, similarly, the correlation with the model-based forecasts and, finally, the measures for double-counting and predictability (H3a and H3b). As it might be that the size of adjustment has a parabolic effect, we additionally include relative absolute adjustment squared. This gives a total of five explanatory variables. The linear regression model for quality additionally contains an intercept, whereas the ordered probit model for relevance includes two thresholds (as there are three categories).

Insert Tables 3 and 4 about here

Before we turn to the estimation results, we first have a look at the variables. Table 3 presents some key statistics of the dependent variables. We observe that in 85 of the 171 cases for the one-step-ahead horizon the contribution of the expert is significant and positive, while in only 12 cases it is significant and negative. For the six-step-ahead forecasts the related fractions are roughly similar. For the differences across the RMSPE's we have interesting statistics, see Table 3, bottom panel. The mean value is negative for both horizons, implying that experts do worse than models, at least on average. The median value is much closer to zero, albeit still negative. Clearly, the data are skewed to the left, as can also be seen from the minimum and maximum values. Hence, there are country-category combinations where the expert's added value implies a seriously poor forecast.

In Table 4 we present some key statistics of the explanatory variables, where we notice that these statistics are roughly similar across the two forecast horizons. We observe that the fraction of positive adjustment is high, so most often experts adjust upwards, which is also reflected by the values of λ. We see that adjustment on average is equally large as the model-based forecast itself. And, trust in the model seems not high, at least on average, as the mean estimated value for $\lambda$ ranges from 0.315 (one-step-ahead) to 0.397 (six-steps-ahead). Finally, the sum of the $\rho$ parameters is close to 0.5 in both cases, so there is a clear indication of double-counting. And, experts' behaviour seems largely predictable, with a fit ranging from 0.512 (one-step-ahead) to even 0.616 for the six-steps-ahead forecasts. Overall, these numbers suggest that the experts we study here put substantial weight on their own added knowledge, where they often adjust upwards, with relatively large values and they do so, on a regular basis.

Insert Tables 5 and 6 about here

The estimation results for the ordered probit models[8] for relevance are given in Tables 5 and 6, for the one-step-ahead and six-steps-ahead forecasts, respectively. Each time, and also later on for the regression models, we report on the first round model with all five explanatory variables included and on the final model where only 5% significant parameters are included. Table 5 shows that the final model fit is good, with a p-value of 0.004, and that only one of the five variables matters. The parameter for $R^2$ has the expected sign. Table 6 shows that after deleting insignificant terms, the final model again contains only a single significant parameter with the expected sign while the overall fit of the model is not significant (p-value is 0.097). This suggests the relevance of the contribution of an expert increases when the experts' behaviour is less predictable and when they do have trust in the model.

Insert Tables 7 and 8 about here

The estimation results for the linear regression models for the difference between RMSPE's of expert versus model appear in Tables 7 and 8, for the one-step-ahead and six-steps-ahead forecasts, respectively. The final model in Table 7 has substantial fit (an $R^2$ of 0.264), and it contains three out of the five variables. For relative absolute adjustment we obtain the expected outcome but for trust the outcome is different than expected. Combining the outcomes for squared (relative, absolute) adjustment and λ we observe that there are cases

[8] Estimation is carried out using Eviews, version 5.0. All standard errors are obtained after correction for heteroskedasticity.

where experts need unidirectional and very large adjustments to improve the model-based forecasts. Table 8 gives the same of type of results for the six-steps-ahead forecasts, except now $\lambda$ is not significant. In sum, too large an adjustment does not lead to more accuracy, but sometimes very large positive or very large negative adjustments are beneficial.

Insert Table 9

In Table 9 we summarize the results in Tables 5 to 8, and match these with the three hypotheses. In 11 of the 16 cases we find neither supportive nor opposite results, while in 4 cases we confirm our hypotheses.

Our first main finding is that a lack of trust in the model and predictable behaviour of the expert leads to less relevant added expertise. Our second main conclusion is that smaller sized adjustments lead to more accuracy, although some extremely-valued adjustments can be helpful too.

## 5. Conclusion and implications

This study systematically analyzed the potential drivers of the relevance and quality of added experts' knowledge to model-based forecasts, using a unique and very large database concerning monthly sales of pharmaceutical products in many countries spread over the globe. We developed a useful and reliable methodology, which improves upon standard methods by allowing for experts who show autoregressive adjustment patterns, for experts' behaviour that can be predicted and for the fact that the final expert forecast is most likely correlated with the model-based forecast.

We formulated a few hypotheses, and based on this, we could shape our empirical analysis. The most dominant result is that more trust in the model, and thus not simply replacing it by one's own opinion and, ceteris paribus, having smaller sized adjustment, leads to more relevant and higher quality expert's adjustment to model-based forecasts. More precise, when the expert adds information that is not very predictable (that is, the model is taken as given) and when this addition is not very large relative to the model-based forecast itself, then the expert more likely adds relevant and higher quality knowledge. In sum, a modest addition to the model-based forecast is best although we found that in exceptional cases making the right adjustment is beneficial. Finally, double counting does not seem to play a role as a driver.

We could see that sometimes the added values are very large, and that performance can be much skewed to the left. This means that experts sometimes adjust too much, and that they lack trust in the model. The implications of our results are that experts should be better informed on how the model works and on what the consequences are of heavy adjustment.

**Appendix**

Number of products in the categories for each country

| Country | Category 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. For our analysis we cannot use all data, as is explained in the text.

Table 1: Operationalization of variables

| Variable | Measurement |
| --- | --- |
| Actual adjustment | Average size of absolute adjustment relative to model-based forecast<br>Frequency of positive adjustments (measured by $\lambda$) |
| Trust in model | Estimated value of $\lambda$ (averaged over products within category) |
| Recent behaviour | Estimated value of $R^2$ of (7) (as a measure or predictability)<br>Estimated value of $\rho$ in (8) (as a measure of double counting) |

Table 2: Hypotheses

| Variable (if higher or more) | | Relevance | Quality |
|---|---|---|---|
| (H1a) | Relative absolute adjustment | - | - |
| (H1b) | Unidirectional adjustment | - | - |
| (H2) | Correlation with model-based forecast | + | + |
| (H3a) | Double counting | - | - |
| (H3b) | Predictability | - | - |

Table 3: Some key features of the data, variables to be explained

| Variable | Forecast horizon | | | | |
|---|---|---|---|---|---|
| Relevance | | -1 | 0 | 1 | |
| | One-step-ahead (171) | 12 | 74 | 85 | |
| | Six-steps-ahead (164) | 10 | 79 | 75 | |
| RMSPE of expert minus RMSPE of model | | Mean | Median | Min. | Max. |
| | One-step-ahead (171) | -9.764 | -2.209 | -121.4 | 47.73 |
| | Six-steps-ahead (164) | -10.338 | -1.705 | -320.9 | 53.69 |

Table 4: Some key features of the data, explanatory variables

| Variable | Mean | Median | Minimum | Maximum |
|---|---|---|---|---|
| *One-step-ahead forecasts (171 cases)* | | | | |
| Relative absolute adjustment | 1.041 | 0.794 | 0.137 | 24.319 |
| Positive adjustments | 0.898 | 0.953 | 0.440 | 1.000 |
| Correlation (λ) | 0.397 | 0.401 | -1.015 | 1.709 |
| Double counting (ρ) | 0.459 | 0.465 | -1.036 | 3.273 |
| Predictability ($R^2$) | 0.512 | 0.495 | 0.116 | 0.989 |
| *Six-step-ahead forecasts (164 cases)* | | | | |
| Relative absolute adjustment | 1.029 | 0.844 | 0.212 | 9.829 |
| Positive adjustments | 0.884 | 0.950 | 0.139 | 1.000 |
| Correlation (λ) | 0.315 | 0.368 | -7.212 | 1.621 |
| Double counting (ρ) | 0.484 | 0.504 | -2.956 | 4.201 |
| Predictability ($R^2$) | 0.616 | 0.598 | 0.140 | 0.988 |

Table 5: Estimation results for relevance, one-step-ahead forecasts (ordered probit model with Huber/White standard errors) (171 effective observations)

| Variable | First round Parameter | (Standard error) | Final model Parameter | (Standard error) |
|---|---|---|---|---|
| Relative abs. adjustment | 0.199 | (0.339) | | |
| Relative abs. adjustment squared | -0.002 | (0.013) | | |
| Correlation (\|λ-1\|) | -0.632 | (0.384) | | |
| Double counting (ρ) | 0.118 | (0.170) | | |
| Predictability ($R^2$) | -1.276 | (0.458) | -1.309 | (0.451) |
| Threshold 1 | -2.349 | (0.347) | -2.193 | (0.296) |
| Threshold 2 | -0.791 | (0.308) | -0.662 | (0.254) |
| P-value fit | 0.028 | | 0.004 | |

Table 6: Estimation results for relevance, <u>six-step-ahead forecasts</u> (ordered probit model with Huber/White standard errors) (164 effective observations)

| Variable | First round | | Final model | |
|---|---|---|---|---|
| | Parameter | (Standard error) | Parameter | (Standard error) |
| Relative abs. adjustment | -0.236 | (0.262) | | |
| Relative abs. adjustment squared | 0.022 | (0.026) | | |
| Correlation (\|λ-1\|) | -0.267 | (0.318) | -0.196 | (0.084) |
| Double counting (ρ) | -0.170 | (0.124) | | |
| Predictability ($R^2$) | 0.162 | (0.487) | | |
| Threshold 1 | -1.833 | (0.358) | -1.704 | (0.166) |
| Threshold 2 | -0.141 | (0.343) | -0.026 | (0.116) |
| P-value fit | 0.342 | | 0.097 | |

Table 7: Estimation results for quality, one-step-ahead forecasts (linear regression model with White standard errors, 171 effective observations)

| Variable | First round | | Final model | |
|---|---|---|---|---|
| | Parameter (Standard error) | | Parameter (Standard error) | |
| Intercept | 10.099 | (7.918 ) | 8.212 | (5.107) |
| Relative abs. adjustment | -43.244 | (10.757) | -42.836 | (10.384) |
| Relative abs. adjustment squared | 1.626 | (0.428) | 1.610 | (0.413) |
| Correlation ($\|\lambda$-1$\|$) | 31.516 | (13.464) | 31.564 | (13.564) |
| Double counting ($\rho$) | 1.447 | (5.409) | | |
| Predictability ($R^2$) | -4.237 | (11.153) | | |
| Fit | 0.265 | | 0.264 | |
| P-value fit | <0.001 | | <0.001 | |

Table 8: Estimation results for quality, six-step-ahead forecasts (linear regression model with White standard errors, 164 effective observations)

| Variable | First round | | Final model | |
|---|---|---|---|---|
| | Parameter (Standard error) | | Parameter (Standard error) | |
| Intercept | 13.122 | (12.401) | 32.304 | (5.962) |
| Relative abs. adjustment | -52.580 | (15.987) | -49.321 | (6.645) |
| Relative abs. adjustment squared | 2.624 | (1.464) | 4.030 | (0.783) |
| Correlation (\|λ-1\|) | 20.792 | (12.950) | | |
| Double counting (ρ) | -5.238 | (4.022) | | |
| Predictability ($R^2$) | 21.674 | (12.682) | | |
| Fit | 0.369 | | 0.317 | |
| P-value fit | <0.001 | | <0.001 | |

Table 9: Results for hypotheses (blank cells -- concern insignificant parameters)

| Hypothesis | Horizon | Relevance | Quality |
|---|---|---|---|
| *Actual adjustment* | | | |
| H1 | Short | -- | Confirmed |
| | Long | -- | Confirmed |
| *Trust* | | | |
| H2 | Short | -- | Opposite result |
| | Long | Confirmed | -- |
| *Recent behaviour* | | | |
| H3a | Short | -- | -- |
| | Long | -- | -- |
| H3b | Short | Confirmed | -- |
| | Long | -- | -- |

## References


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

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

Fildes, R.A. and P. Goodwin (2007), Effective forecasting for supply-chain planning: an empirical evaluation and strategies for improvement, Unpublished Manuscript, Lancaster University.

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

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.