Experts' adjustment to model-based forecasts: Does the forecast horizon matter?

Philip Hans Franses
Rianne Legerstee

Econometric Institute
Erasmus University Rotterdam

Econometric Institute Report 2007-51

Abstract

Experts may have domain-specific knowledge that is not included in a statistical model and that can improve forecasts. While one-step-ahead forecasts address the conditional mean of the variable, model-based forecasts for longer horizons have a tendency to convert to the unconditional mean of a time series variable. This suggests that added expertise could be most beneficial to forecast quality for immediate horizons (as very recent events are not in the model) and for further away horizons (as they may miss foreseen trends), and less so for intermediate horizons. Relying on a huge database concerning pharmaceutical sales forecasts for various products and adjusted by a range of experts, we examine and verify this and other conjectures. We also document that the forecast horizon that is the most important for supply chain management here entails the heaviest adjustment by the experts. Unfortunately, that additional adjustment harms forecast accuracy.

Key words: Forecasting, Performance Measurement, Human Judgment and Decision-Making

This version: December 3 2007

Address for correspondence: Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, franses@few.eur.nl. Part of the material in this paper has been presented at a seminar in Leuven. We thank the participants for helpful comments.

1. Introduction

In supply chain management and elsewhere statistical models are used to create out-of-sample forecasts. One-step-ahead forecasts are useful for various reasons, but for management it may be more relevant to have forecasts for further away horizons, although not infinitely away. The parameters in statistical models for time series data are typically estimated or updated using some least squares criterion, where usually the squares of the one-step-ahead forecasts are minimized. Hence, it is conceivable that such models would do well for short horizons and less so for further away horizons.

Even though statistical models seek to capture trends and other patterns in the data and to extrapolate these into the future, it is quite likely that these models do not cover recent events that can be relevant for forecasting too. Experts with domain-specific knowledge can have information on these events and hence can add something to the model-based forecast.

In this paper we study whether experts treat model-based forecasts for various horizons in a different way. Given that model-based forecasts for further away horizons convert towards the unconditional mean, it might be that experts have a tendency to change those forecasts more than nearer forecast horizons. We shall study these and other conjectures using a large database with sales forecasts for pharmaceutical products.

The outline of our paper is as follows. To shape our empirical analysis which we believe is the first of its kind we put forward a few hypotheses in Section 2. Next, we outline the empirical methodology in Section 3. Section 4 deals with the empirical results and Section 5 sketches a few areas for further work.

2. Hypotheses

This paper builds and extends on the results in Franses and Legerstee (2007) where the question whether added expertise yields better forecasts was answered. The main conclusion in that study is that experts have a tendency to add too much, and this leads to less accurate forecasts. Indeed, would they downplay their added value a little bit then this already would lead to much better expert forecasts. In that study the forecast horizon was fixed at 1 and 6, and in the present paper we focus on all horizons 1 to 12 and specifically address the link between accuracy and added value of the expert with the horizon.

Variables of interest

The framework is as follows. We consider the following variables

$MF_{t+h|t}$: model-based forecast for horizon *h* (made from origin *t*)

$EF_{t+h|t}$: expert forecast for horizon *h* (made from origin *t*)

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

where *S* denotes (monthly) sales.

The model-based forecasts are linear functions of past sales and where the weights are updated each month. The forecasts are created recursively which means that the parameters are estimated for R in-sample data, and then one-step-ahead to h-step-ahead forecasts are made. Next, the sample is enlarged to R+1, parameters are updated and again the h forecasts are made. The number of forecasts thus obtained is denoted as P, P-1 to P-(h-1).

The experts receive the statistical model-based forecasts and almost always make an adjustment. This is of course partly due to the fact that the experts know that the model-based forecasts only include lagged sales. Part of the adjustment depends on domain-specific variables. Franses and Legerstee (2007) document that the expert forecasts depend on past sales again, which entails some form of double counting (as the model-based forecasts are unbiased), and on those other variables, which makes the expert forecast to nest the model-based forecast. This notion is important when evaluating the relative merits of both forecasts later on.

As the expert forecasts and the model-based forecasts partly rely on the same set of information, that is past sales, the added contribution of the expert to the model-based forecast is not simply $EF_{t+h|t}$ - $MF_{t+h|t}$ as this assumes that both forecasts are independent. To correct for the dependence is to compute

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

where λ gets estimated from a linear regression (see Blattberg and Hoch (1990)), that is,

(2) $$EF_{t+h1} = \lambda MF_{t+h|t} + A_{t+h|t}$$

Where, after OLS estimation, $\lambda MF_{t+h|t}$ and $A_{t+h|t}$ are independent. So, only when $\lambda = 1$, then the adjustment can be computed as $EF_{t+h|t}$ - $MF_{t+h|t}$. Franses and Legerstee (2007) document that the average estimate of λ from (2) is around 0.4.

Hypotheses

Before we outline the methodology, we formulate a few hypotheses to sharpen the discussion below. The main issue in this paper is whether the forecast horizon h matters for the behaviour of experts and for their forecast accuracy relative to model-based forecasts.

The first hypothesis that we wish to put forward is

H1: The added contribution of the expert is about equal for all forecast horizons

This means that model-based forecasts for all horizons experience adjustment, and there is no particular reason to believe that some horizons see less inclination by experts to adjust.

Whether experts are equally successful across horizons is another matter. In fact, one might expect that experts may have access to recent information that never could have been included in the past sales-based model, so the quality of the added contribution of the expert shall be better for very short forecast horizons. On the other end, as regression-based models have a tendency to give h-step-ahead forecasts that converge to the unconditional mean, and hence may miss recently established trends that can be noticed by the expert, the quality of the added value of the expert may be larger for further-away horizons. In sum we conjecture that

H2: The quality of the added contribution of the expert is larger for immediate and further-away horizons, and is smaller for intermediate horizons.

In our empirical analysis below we would roughly label 1, 2 and 3 months ahead as immediate, 10, 11 and 12 months ahead as further away and around 4 to 9 months as intermediate horizons.

Based on our discussions with the managers at the headquarters' office, where they create the model-based forecasts, we know that (around) the 6-month-ahead forecasts are most relevant for supply chain management reasons. So, for these horizons the experts do want to perform particularly well, and in fact part of their bonus payments depends on their success rates here. Given this information, we conjecture

H3: The forecast horizon that is most important for the expert receives most attention, that is, expert's added contribution is largest.

Finally, and this basically reiterates one of the findings in Franses and Legerstee (2007), we propose

H4: When the weight of the contribution of the expert relative to the model is larger, the quality of the expert forecast becomes smaller

In the next section we shall outline the methodology which we will consider to empirically examine the validity of these four hypotheses.

3. Methodology

This section gives a discussion of the database we have and outlines the basic statistics that we compute to examine the hypotheses. We have data concerning product *i* within category *j* for country *c*. We consider 37 countries and there are 7 product categories. We do not have all data for all categories for all countries and within a category there are different numbers of products. The data concern monthly sales for October 2004 to October 2006 of pharmaceutical products. The headquarters' office creates model-based forecasts and sends these to the experts in each of the countries. An expert is allowed to modify the model-based forecasts in a way he or she sees fit.

The added value of the expert

Hypotheses H1 and H3 concern the added contribution of the expert. To answer this question we consider for all products within a country-category combination the following multiple-equation model for each horizon h, that is,

(3)
$$S_{1,t+h} = \alpha_1 + \beta MF_{1,t+h|t} + \gamma A_{1,t+h|t} + u_{1,t+h}$$
$$S_{2,t+h} = \alpha_2 + \beta MF_{2,t+h|t} + \gamma A_{2,t+h|t} + u_{2,t+1}$$
.....
$$S_{n,t+1} = \alpha_n + \beta MF_{n,t+h|t} + \gamma A_{n,t+h|t} + u_{n,t+1},$$

where n denotes the number of products within each such combination. Earlier experience with these data suggest that the error process is best modeled as

$$(1-\rho_1 L-\rho_2 L^2)u_{t+1}=\varepsilon_{t+1}$$

We allow the α parameter to differs per product, but the β and γ parameters are assumed as common across products within a country-category combination. The model in (3) assumes independent equations with cross-equation parameter restrictions. For the sake of computational simplicity, we assume the errors as independent.

In the empirical part below we will report on the fraction of cases that γ is significantly different from zero and that β is equal to γ. In this last case, this would mean that the added contribution of the expert is equal to the contribution of the model. Note that we deal here with the added contribution and not the expert forecast. This can be understood from writing for example

$$S_{1,t+h}=\alpha_1+\beta MF_{1,t+h|t}+\gamma A_{1,t+h|t}+u_{1,t+h} \tag{4}$$

as

$$S_{1,t+h}=\alpha_1+\beta MF_{1,t+h|t}+\gamma(EF_{1,t+h|t}-\lambda MF_{1,t+h|t})+u_{1,t+h}$$

or

$$S_{1,t+h}=\alpha_1+(\beta-\gamma\lambda)MF_{1,t+h|t}+\gamma EF_{1,t+h|t}+u_{1,t+h} \tag{5}$$

In case β is equal to γ, and, as said, $\lambda = 0.4$, then $\beta - \gamma\lambda = 0.6\gamma$, and thus the relative weights of the model-based forecast and the expert forecast are 0.6 and 1, or, scaled to sum to unity, equal to 0.375 and 0.625.

The quality of the expert

To examine the quality of the added contribution of the expert, one can of course just see if the root mean squared prediction errors (RMSPEs) are different, but one may also want to test whether these differences are significantly different. Such a test can then be computed for each product within each country-category combination.

To test the null hypothesis that the root mean squared prediction error of the expert is equal to that of the model against the alternative hypothesis that the expert is better, we need to take account of the fact that expert forecast is based on a model that nests the model used for the statistical forecast. We thus follow the recommendation in Clark and McCracken (2001) and we use the following procedure. We have R in-sample data, where in our case local management informs us that R is around 60. We have P recursively created one-step-ahead forecasts, and here P = 25. Hence, the fraction of forecasts over in-sample data is $\pi = \frac{P}{R} = 0.4$. We need this value of π for the non-standard critical values of the upcoming test. We assume that the experts each time include 2 regressors in X_t. Clark and McCracken recommend using the so-called ENC_NEW test, defined by

(6) $$ENC\_NEW = P \frac{\frac{1}{P}\sum (u_{1,t+1|t}^2 - u_{1,t+1|t}u_{2,t+1|t})}{\frac{1}{P}\sum u_{2,t+1|t}^2}$$

The summation runs for the P one-step-ahead forecasts, and $u_{1,t+1|t}$ denotes the forecast errors for model-based forecasts, and $u_{2,t+1|t}$ concerns the expert forecasts, which nest the model-based forecasts. The 5% critical values are given in Table 1 of Clark-McCracken (2001, page 92). For $\pi = 0.4$ and $k_2 = 2$ it is 1.481. This test is a one-sided test of the null hypothesis that the expert forecast is equally good as the model-based forecast against the alternative hypothesis that the expert forecast is better. So, the outcome of the test is whether the expert yields better forecasts or not.

For multi-step-ahead forecasts we can also compute the test statistic as in (6), but now a complication arises in terms of the asymptotic distribution of that test. For multi-step-ahead forecasts it is well known that the forecast errors are correlated, and this correlation needs to be included in the distribution. Clark and McCracken (2005) outline in detail how to do this in case the variables in X_t are known. One can then use bootstrap techniques to compute critical values for each particular situation at hand. In our case we face the problem that these variables, which are the additional variables used by the expert, are unknown. Fortunately, Clark and McCracken (2005, page 390) note that standard normal critical values would lead to reliable inference, provided that the forecast horizon is relatively short and π is also rather small. Our empirical work below seems to meet these requirements, so we will compute

$$(7) \qquad ENC\_NEW_h = P \frac{\frac{1}{P}\sum(u_{1,t+h|t}^2 - u_{1,t+h|t}u_{2,t+h|t})}{\frac{1}{P}\sum u_{2,t+h|t}^2}$$

and we consider a one-sided test with a 5% critical value equal to 1.645.

4. Results

In this section we report on the empirical results obtained using the methodology outlined in the previous section, where the specific focus is on variation of the results across horizons.

4.1 Added contribution of experts

In total we have data for 203 country-category combinations concerning one-step-ahead forecasts, and within each we have from 1 to as many as 30 products. We will consider the test regression in (3) for 203 cases. Table 1 gives in its second column how many cases we have for each of the 12 different horizons. We observe that we have quite a number of cases, and hence, we can safely draw generalizing conclusions.

Insert Table 1 about here

The third column of Table 1 gives the number of case with significant values for γ. The fourth column shows that the fraction of cases with significant γ seems to decrease with the forecast horizon. However, we must bear in mind that the test for this significance is based on smaller samples each time the horizon increases. In fact, for forecast horizon each equation in (3) gets estimated for 25 observations, while for horizon 12 it is only 14. To provide some form of correction for this power loss, we compute a sample-size corrected fraction of cases with significant γ by multiplying the numbers in the penultimate column of Table 1 with

$$\sqrt{\frac{T}{T-h+1}}$$

where T equals 25 here. Upon doing so, we see from the last column that the fraction of cases with significant is rather constant across the forecast horizons, and this fraction is around 50% of the cases.

Insert Table 2 about here

In Table 2 we report on the fraction of cases where β equals γ, first across all cases and second across the cases where γ is significantly different from 0. Again, we correct for the decreasing sample size with increasing horizon, but as it now concerns the size of the test we compute the sample-size adjusted fractions by

$$1-\sqrt{\frac{T}{T-h+1}}(1-fraction)$$

When we consider the relevant columns in Table 2 we observe that there is a slow decrease in the fraction of cases with β equal to γ, and this certainly holds for the cases with significant γ parameters. Hence, from Table 1 we saw that the expert contribution is stable, but here we see that this contribution is less on a 50-50 basis when the horizon increases. So, we find support for hypothesis H1, and there are some indications that the relative weights of model and expert are not stable across the forecast horizon.

Insert Tables 3 and 4 about here

To have a closer look at these relative weights, we compute the mean, median, minimum and maximum values of the estimated parameters in (3) across the cases mentioned in the second column of Table 1. In Table 3 we first look at the estimated β parameters, and in Table 4 we consider the γ parameters.

The first impression from both tables is that the empirical distribution of the estimated parameters can have serious outliers, and typically to the left. The non-normality is further substantiated by the sometimes large differences between mean and median. Therefore we only look at the medians. From Table 2 we noticed that with further away horizons, there is less evidence of a 50-50 model-added expert link and Tables 3 and 4 might suggest which factor is causing this. The median of the estimated β parameters suggests a parabolic shape, with a dip around horizons 6 or 7. On the other hand, we see from Table 4 that the estimated γ parameters seem to get smaller with increasing horizon.

Insert Table 5 and 6 about here

Due to this inconclusiveness, it is perhaps better to look at the estimated values of β – γ as in Table 5, or even better, at β – γ in case the γ is significant (the cases in the third column of Table 1) as in Table 6. The median values in Table 5 are already suggestive, but most clear-cut evidence is obtained from the median values in Table 6. The parabolic shape that is visually obvious is even further substantiated when we consider the regression

$$(8) \qquad Median_{\beta-\gamma} = \mu + \alpha_1 h + \alpha_2 h^2 + \varepsilon$$

for h = 1,2,…,12. The α_1 parameter gets estimated as -0.071 (with standard error 0.020) and α_2 becomes 0.006 (0.001) and the R^2 is 0.602. The fit of the model is depicted in Figure 1 and it is clear that the minimum value of β – γ is obtained for h = 6, which interestingly is the most important forecast horizon.

Insert Figure 1 about here

Let us have a look at what this means for the relative contribution of the model and the expert. Earlier we noticed that when λ = 0.4 in (5), and when of β – γ is 0, that then β – γλ = 0.6γ, and thus the relative weights of the model-based forecast and the expert forecast are 6 to 10. When however at h = 6 the difference is -0.2, so for example that β = 0.2 and γ = 0.4, then plugging these values in (5) we see that the model-expert relative contribution is 1 to 10. So, at the most important horizon for the experts, they exercise 10 times as much weight than the model forecast does. This finding strongly supports our hypothesis H3.

4.2 Quality of added contribution

Now we have seen that experts add most expertise to the model-based forecasts for horizon 6, the next obvious question is whether this is successful or not.

Insert Table 7 about here

In Table 7 we present the mean, median, minimum and maximum values of the empirical distribution of the percentage of improvement in root mean squared prediction error (RMSPE) of the expert-adjusted forecasts over the model-based forecasts for all cases (as in Table 1, column 2). Note that this thus concerns all products within the expert-category cases. The results in this Table 7 are pretty clear. The empirical distribution is strongly skewed to the left, meaning that when experts do better they do less good in absolute sense than when they do worse. For all horizons we see that minimum values can be around -1000 while on average a maximum improvement is around 60%. So, here again we have a closer look at the medians instead of the means. At first sight we again see a parabolic shape, with a dip around horizons 6 or 7. This is confirmed when we estimate the parameters of

(9) $$Median\_\%improvement\_in\_RMSPE = \mu + \alpha_1 h + \alpha_2 h^2 + \varepsilon$$

for $h = 1,2,\ldots,12$. The α_1 parameter gets estimated as -0.713 (with standard error 0.347) and α_2 becomes 0.060 (0.026) and the R^2 is 0.398. The fit of this regression against horizon h is depicted in Figure 2.

Insert Figure 2 about here

Clearly, the lowest point in this graph is attained when h is around 6. This result supports our hypothesis H4.

Insert Table 8 about here

The conclusion drawn from Table 7 was based on the analysis of all forecasts, and the next question is whether this finding would also hold when we only look at the significant contributions of the experts, which now cover the averages within categories for the products with significant test values. When we look at the percentage improvement of experts in case the ENC_NEW tests are significant, as in Table 8, we observe that when the expert has a significant and positive added value to the final forecast, this value becomes larger for further away horizons.

Insert Table 9 about here

When the expert is not adding anything significant and positive, at least according to the ENC_NEW tests, we see from Table 9 that the median values do seem to have a parabolic shape. This is confirmed by the regression

(10) $$Median\_\%RMSPEwhenENC\_NEWinsignificant = \mu + \alpha_1 h + \alpha_2 h^2 + \varepsilon$$

for $h = 1,2,\ldots,12$. The α_1 parameter gets estimated as -1.590 (with standard error 0.509) and α_2 becomes 0.138 (0.038) and the R^2 is 0.643. The fit of this regression against horizon h is depicted in Figure 3.

Insert Figure 3 about here

The minimum value is clearly attained around $h = 6$. Taking the results in the last two tables and figures together, we see that when experts do significantly better they do even more so when the forecast horizon increases. Moreover, when they do not perform significantly better, their forecasts are worst around the horizon that is most important for their business. This matches with the findings in our earlier work that when experts exercise too much value in the final forecast, notably here at horizon 6 where the model-expert ratio is 1 to 10, the quality of this forecast quickly deteriorates.

Insert Table 10 about here

To demonstrate that less weight of the expert leads to improvement for all horizons, we report on the mean, median, minimum and maximum values of the percentage of improvement in RMSPE when we would take the average of the model-based forecast added to the expert forecast, versus the pure model-based forecast. Clearly, the median is positive across all cases and the minimum values are not that low anymore. Most improvement is seen for horizons 1 and 12, and least for 6 and 7. This suggests that one might wish to assign the weights to the model and the expert depending on the horizon.

5. Conclusions

Upon analysis of the model-based forecasts and expert-adjusted forecasts for a large range of products within various categories and the experts being located in countries spread all over the world, we arrive at the following four conclusions concerning the impact of the forecast horizon on relevance and quality of experts' expertise. The first is that the model-based forecasts for all horizons experience adjustment, and there are no horizons for which experts see less inclination to adjust. The second conclusion is that the quality of the added contribution of the expert is larger for immediate and further-away horizons, and is smaller for intermediate horizons. This intermediate horizon in our case happens to be close to the forecast horizon that is most relevant for supply chain management. In fact, our third conclusion is that the forecast horizon that is most important for the expert receives most attention, that is, expert's added contribution is largest. But, and that is the fourth conclusion, when the weight of the contribution of the expert relative to the model is larger, the quality of the expert forecast becomes smaller.

So to answer the question in the title of our paper: Yes, the forecast horizon matters. This finding has consequences for management, in particular when training the experts. It is seen that experts exercise too much impact when the horizons matters most to them and then the quality of the added expertise quickly decreases. Apparently, experts should be trained not to over-adjust model-based forecasts for some horizons.

TABLES

Table 1:
Testing for the relevance of the added value of the expert by testing whether γ is significant in the system of equations in (3)

			Fraction	
Horizon correction	Cases	Cases with $\gamma \neq 0$	No correction	Corrected for power loss
1	203	105	0.517	0.517
2	201	109	0.542	0.553
3	201	103	0.512	0.534
4	199	96	0.482	0.514
5	199	90	0.452	0.493
6	196	87	0.444	0.496
7	194	88	0.454	0.521
8	189	78	0.413	0.487
9	189	75	0.397	0.481
10	190	76	0.400	0.500
11	186	65	0.349	0.451
12	185	69	0.373	0.498

Table 2:
Is the added contribution of the expert equal to that of the model, that is, is β equal to γ in (3)?
(Fractions are in parentheses)

Horizon	Cases	Cases with $\beta = \gamma$		Cases with $\beta = \gamma$ and $\gamma \neq 0$
1	203	124	(0.611) (0.611)	54 of 105 (0.514)
2	201	118	(0.587) (0.578)	53 of 109 (0.486)
3	201	129	(0.642) (0.627)	53 of 103 (0.515)
4	199	126	(0.633) (0.609)	47 of 96 (0.490)
5	199	125	(0.628) (0.594)	41 of 90 (0.456)
6	196	119	(0.607) (0.561)	36 of 87 (0.414)
7	194	125	(0.644) (0.592)	40 of 88 (0.455)
8	189	123	(0.651) (0.589)	40 of 78 (0.513)
9	189	118	(0.624) (0.544)	33 of 75 (0.440)
10	190	111	(0.584) (0.480)	36 of 76 (0.474)
11	186	106	(0.570) (0.445)	22 of 65 (0.338)
12	185	111	(0.600) (0.465)	21 of 69 (0.304)

Table 3:
Empirical distribution of the estimated value of β in (3) (number of cases is given in Table 1, second column)

Horizon	Mean	Median	Minimum	Maximum
1	0.250	0.338	-3.509	1.843
2	0.267	0.350	-6.486	1.961
3	0.249	0.289	-2.152	2.416
4	0.310	0.272	-4.341	5.035
5	0.291	0.270	-2.060	2.643
6	0.286	0.218	-3.765	6.079
7	0.216	0.207	-18.711	4.335
8	0.372	0.259	-4.356	6.229
9	0.298	0.237	-2.343	2.679
10	0.446	0.323	-6.302	8.046
11	0.332	0.329	-6.695	6.281
12	0.076	0.169	-15.025	6.255

Table 4:
Empirical distribution of the estimated value of γ in (3) (number of cases is given in Table 1, second column)

Horizon	Mean	Median	Minimum	Maximum
1	0.168	0.240	-12.606	1.992
2	0.282	0.290	-3.467	2.276
3	0.250	0.239	-4.916	1.852
4	0.230	0.229	-5.230	1.662
5	0.262	0.204	-2.474	5.377
6	0.249	0.167	-4.801	4.368
7	0.236	0.154	-2.983	4.836
8	0.229	0.178	-1.605	2.131
9	0.223	0.163	-1.556	2.829
10	0.232	0.173	-4.513	5.167
11	0.146	0.099	-15.347	5.202
12	0.170	0.121	-17.639	8.409

Table 5:
Empirical distribution of the estimated value of β-γ in (3) (number of cases is given in Table 1, second column)

Horizon	Mean	Median	Minimum	Maximum
1	0.082	0.074	-3.326	9.097
2	-0.015	0.044	-6.139	4.037
3	-0.001	-0.020	-2.703	3.032
4	0.080	0.039	-2.748	5.187
5	0.029	0.101	-7.352	2.793
6	0.037	-0.006	-6.726	5.377
7	-0.019	0.002	-18.290	5.149
8	0.143	0.056	-4.501	5.614
9	0.075	0.066	-2.525	2.518
10	0.214	0.102	-4.802	7.507
11	0.185	0.134	-5.105	15.274
12	-0.093	0.077	-12.915	17.286

Table 6:
Empirical distribution of the estimated value of β-γ in (3), given that $\gamma \neq 0$ (number of cases is given in Table 1, third column)

Horizon	Mean	Median	Minimum	Maximum
1	0.064	0.041	-3.326	9.097
2	-0.158	-0.113	-6.139	1.704
3	-0.193	-0.156	-2.703	3.013
4	-0.133	-0.067	-2.748	1.283
5	-0.091	-0.128	-2.509	1.854
6	-0.182	-0.235	-6.726	5.377
7	-0.165	-0.161	-2.718	3.471
8	-0.135	-0.166	-1.675	1.052
9	-0.093	-0.098	-2.525	2.191
10	0.062	-0.114	-3.302	4.008
11	0.163	0.015	-4.742	15.274
12	-0.148	-0.049	-12.915	17.286

Table 7:
Empirical distribution of the percentage of improvement in RMSPE of expert-adjusted forecast over the model-based forecast (all products in all cases given in Table 1)

Horizon	Mean	Median	Minimum	Maximum
1	-19.892	-2.290	-553.00	47.730
2	-23.950	-2.600	-515.62	49.540
3	-26.747	-3.130	-478.07	55.190
4	-30.641	-2.840	-999.72	51.900
5	-25.799	-2.850	-629.45	56.330
6	-24.912	-3.045	-661.45	56.280
7	-23.309	-3.735	-573.03	55.440
8	-17.998	-4.740	-481.40	54.490
9	-17.121	-1.580	-497.12	54.430
10	-304.43	-2.760	<-1000	56.260
11	-18.288	-3.345	-478.59	54.840
12	-12.357	-0.230	-445.50	60.690

Table 8:
Distribution of the percentage of improvement in RMSPE of expert-adjusted forecast over the model-based forecast, in case the ENC-NEW test is positive and significant and improvement is not zero (test is performed for each product in all categories and only for those products where the test is significant, the numbers in the table have been computed)

Horizon	Mean	Median	Minimum	Maximum
1	17.266	15.480	0.050	84.920
2	17.419	15.400	0.580	78.100
3	17.752	15.120	1.700	79.820
4	18.382	15.870	0.460	86.050
5	20.401	17.460	1.210	81.480
6	20.413	17.795	0.880	91.170
7	19.943	17.390	0.690	91.120
8	21.840	18.180	0.960	91.450
9	21.022	18.330	0.015	73.800
10	21.802	19.510	0.980	73.170
11	21.989	20.060	1.430	73.940
12	24.412	22.515	1.430	60.690

Table 9:
Distribution of the percentage of improvement in RMSPE of expert-adjusted forecast over the model-based forecast, in case the ENC-NEW test is NOT positive and significant (the test is performed for each product in all categories and only for those products where the test is NOT significant, the numbers in the table have been computed)

Horizon	Mean	Median	Minimum	Maximum
1	-40.008	-14.910	-859.42	1.020
2	-44.450	-18.040	-753.63	2.720
3	-48.948	-18.570	-635.62	1.350
4	-53.986	-17.230	-999.72	1.950
5	-52.001	-19.440	-632.57	3.670
6	-50.969	-21.110	-661.45	3.240
7	-52.857	-21.605	-685.73	2.440
8	-48.172	-20.640	-743.74	2.880
9	-43.073	-18.950	-497.12	3.120
10	-614.79	-21.175	<-1000	4.400
11	-40.978	-21.530	-478.59	5.780
12	-38.484	-18.820	-445.50	4.200

Table 10:
Empirical distribution of the percentage of improvement in RMSPE of expert-adjusted forecast over the model-based forecast, if the final forecast is 50-50 (all products in all cases given in Table 1)

Horizon	Mean	Median	Minimum	Maximum
1	1.571	7.010	-253.56	53.920
2	-0.113	6.250	-227.63	56.780
3	-1.229	6.400	-199.42	60.540
4	-3.318	6.790	-459.84	60.290
5	-0.577	6.360	-283.85	59.320
6	-0.240	6.220	-297.75	58.410
7	-0.825	5.965	-251.42	67.970
8	2.791	6.130	-163.61	63.120
9	3.175	7.090	-167.52	57.660
10	-140.27	5.915	<-1000	54.870
11	2.867	6.440	-152.95	68.580
12	4.965	8.490	-139.43	54.170

FIGURES

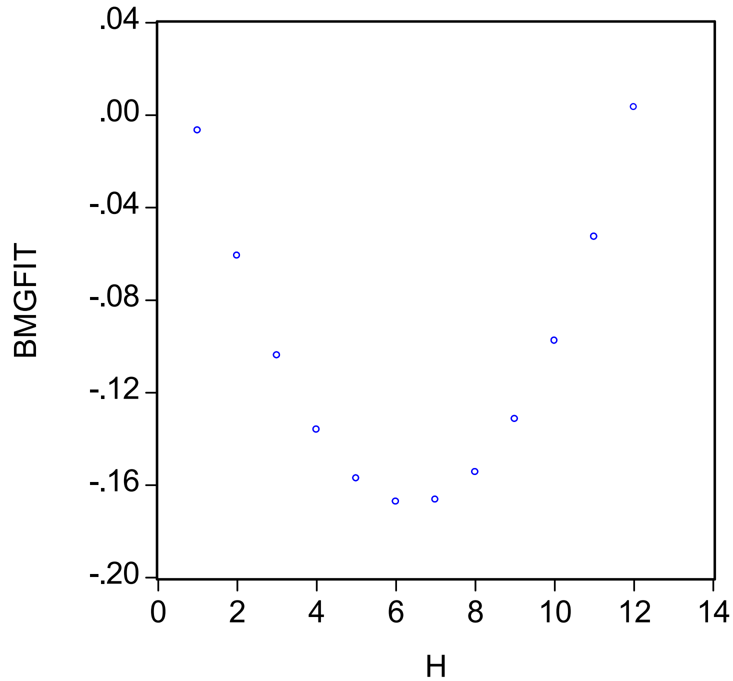

Figure 1: The fit of the regression model versus the variable *h* in model (8)

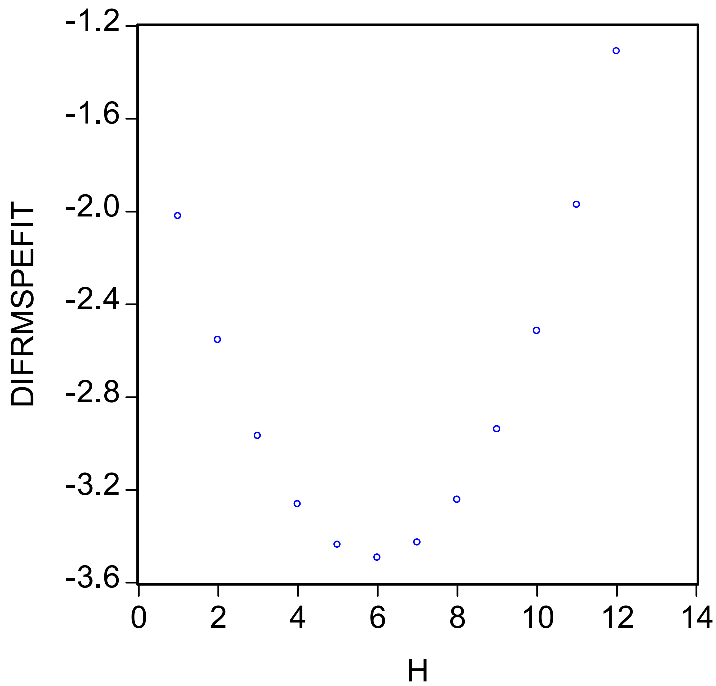

Figure 2: The fit of the regression model versus the variable *h* in model (9)

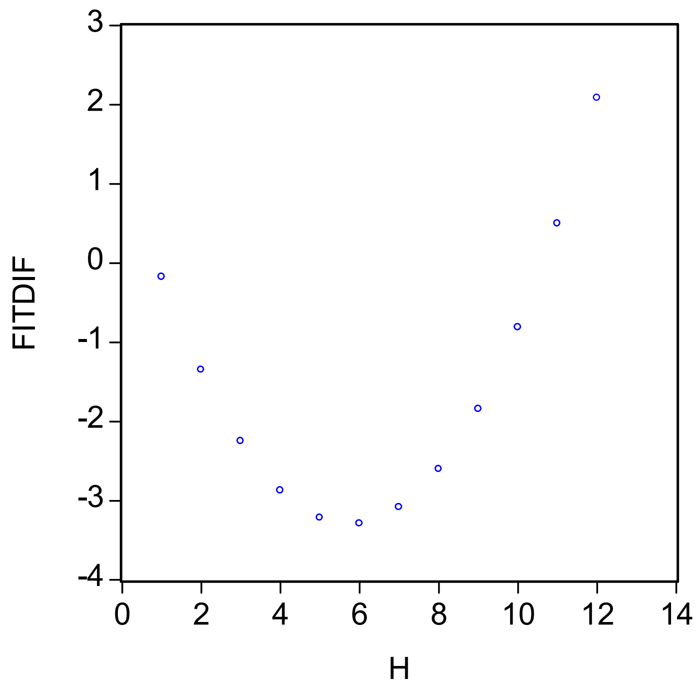

Figure 3: The fit of the regression model versus the variable *h* in model (10)

References

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

Clark, Todd E. and Michael W. McCracken (2001), Tests of equal forecast accuracy and encompassing for nested models, Journal of Econometrics, 105, 85-110

Clark, Todd E. and Michael W. McCracken (2005), Evaluating direct multi-step forecasts, Econometric Reviews, 24, 369-404

Franses, P.H. and R. Legerstee (2007), Does experts' adjustment to model-based forecasts contribute to forecast quality?, Econometric Institute Report 2007-34, Erasmus University Rotterdam.