On the Optimality of Expert-Adjusted Forecasts

Philip Hans Franses

Econometric Institute

Erasmus University Rotterdam

Henk C. Kranendonk

Debby Lanser

Netherlands Bureau of Economic Policy Analysis

Econometric Institute Report 2007-38

20 September 2007

Notes: We thank Dick van Dijk, Henk Don, Clive Granger, Rocus van Opstal and Johan Verbruggen for helpful comments. The address for correspondence is: Econometric Institute, Erasmus University Rotterdam, P.O.Box 1738, NL-3000 DR Rotterdam, The Netherlands, franses@few.eur.nl

Abstract

Official forecasts of international institutions are never purely model-based. The preliminary results of models are adjusted with expert opinions. What is the impact of these adjustments for the forecasts? Are they necessary to get 'optimal' forecasts? When model-based forecasts are adjusted by experts, the loss function of these forecasts is not a mean squared error loss function. In fact, the overall loss function is unknown. To examine the quality of these forecasts, one can rely on the tests for forecast optimality under unknown loss function as developed in Patton and Timmermann (2007). We apply one of these tests to ten variables for which we have model-based forecasts and expert-adjusted forecasts, all generated by the Netherlands Bureau for Economic Policy Analysis (CPB). We find that for almost all variables the added expertise yields better forecasts in terms of fit. In terms of optimality the effect of adjustments for the forecasts are limited, because for most variables the assumption that the forecast are not optimal can be rejected for both the model-based and the expert-adjusted forecasts.

Key words: Expert-Adjusted Forecasts, Optimality
JEL code: C53, E17

1. Introduction

There is substantial literature on expert-adjusted forecasts, see for example Clements (1995) and the references cited therein. Such forecasts imply that experts change the result of the original preliminary model-based forecast. There are many reasons why an expert may wish to change such a forecast, see Clements and Hendry (1998, Chapter 8). For instance, they adjust to obviate known shortcomings in the economic model or to mimic the effects of economic events outside the model. Expert adjustment implies that while the model-based forecasts could have been obtained under mean-squared error loss[1], the expert-adjusted forecasts are obtained under an unknown loss function. In the present paper we address the issue of examining the quality of such expert-adjusted forecasts.

The forecasts under scrutiny are those created by the Netherlands Bureau for Economic Policy Analysis (CPB) for ten variables for the period 1997-2006. We were able to re-run the original models that delivered the model-based forecasts[2]. These re-runs were possible by using a unique notebook kept at the CPB which documents all adjustments made between the initial model-based forecasts and final publication. The models were used four times a year to generate forecasts up to the next full year.

In this paper we evaluate the quality of a forecast not only by statistics of the forecast errors but also by means of its optimality. A forecast is considered optimal when given all information available at the time of forecasting and including the preferences of the forecaster, the forecaster provides the best possible estimate. For that purpose, the forecaster is regarded as a decision-maker who minimizes a loss function penalizing the disutility associated with deviations between future realizations and the forecast. In case of both the CPB model-based and expert-adjusted forecasts this loss function is unknown.

To test the null hypothesis of optimality of forecasts under unknown loss functions, we use a test proposed in Patton and Timmermann (2007). When a mean squared error loss function can be assumed, one can rely on the test procedures outlined in Nordhaus (1987), Diebold, Gunther and Tay (1998) and Christoffersen (1998), but when the loss function is

---

[1] As will be discussed below, large-scale macroeconomic model-based forecasts are typically not based on mean-squared error loss. This is due to the fact that these models may contain many equations, and this does not allow for least-squares based estimation of all parameters. Indeed, some parameters are simply fixed at levels which the experts deem reasonable. So, the sequel of this paper also concerns the evaluation of large-scale macroeconomic model-based forecasts themselves.

[2] It is perhaps of interest here to note that this re-running of the models actually amounted to most of the empirical work done for this paper.

unknown, other procedures are required. Patton and Timmermann (2007) recommend the use of the test proposed in Engle (1982).

The outline of our paper is as follows. In Section 2 we discuss reasons for intervening in the model outcomes and we illustrate the quantitative impact of adjustments. In Section 3 we outline our methodology and we present the empirical results of the optimality tests. In Section 4 we give the main conclusions.

2. Adjusting model-based forecasts

A recent study of Lanser (2006) analyses the importance of four sources of uncertainty when making forecasts with a large macro-economic model. These sources concern uncertainty in preliminary data, in exogenous data, in model parameters and in residuals. Model users should be aware of these uncertainties and use their models with a critical mind.

In the current paper, we investigate a fifth source of uncertainty, that is, uncertainty in add factors or autonomous terms. Don and Verbruggen (2006) instructively refer to this source, by stating "*When using models, it is critical to bear in mind the limitations and weaknesses of the models, in order to prevent misleading outcomes. This implies that a model may have to be adjusted as necessary for the analysis in question, with messy compromises from the economic theoretical and econometric perspectives often unavoidable*". This can lead to changes in the model itself or in the way the model is used. We quote again: "*There are two ways of taking account of observed imperfections when using models: interfering with the model' structure or adding autonomous terms or 'add factors'*".

An add factor can be applied for several reasons. A first example is when actual time series data not fit well with the estimated behavioural equation, for example because of revisions of the national accounts. Awaiting re-estimation of this equation, a systematic residual for recent years should be extrapolated to the future. A second possibility is to incorporate specific knowledge for the near future about contracts or plans into the model. A third reason can be to adjust for a specific period the effect of economic behaviour of households or firms because of sudden shocks in confidence or announced changes of tax-rates.

In this paper we investigate the add factors applied to forecasts computed with the CPB-models applied in the last ten years, FKSEC, SAFE and SAFFIER.[3] These models contain more than two thousand equations. However, the core of the model, concerning the behaviour of households and firms, consists of only around thirty estimated or calibrated equations. Each of these behavioural equations contains an add factor which can be used by the forecaster to adjust the outcome of the equation. By far most other equations concern identities or detailed descriptions of Dutch institutional relations.

CPB is able to re-run all published forecasts since spring 1996 with the models and databases corresponding to these publications. These databases include all relevant information on add factors. Since 1999 CPB keeps a detailed logbook of these add factors for the most important behavioural equations with information on the reasons, and the quantitative effect, for the adjustments.

Insert Table 1 around here

Table 1 provides the list of the investigated equations, which concern the most relevant demand components of GDP, both in volumes and prices, the wage rate and the demand for labour. For eight equations the average value of the add factor is close to zero, meaning that over the considered ten-year period both positive and negative adjustments have occurred about equally. For three variables, which are Re-exports (volume), Imports investment goods (volume) and Imports consumption goods (volume) the average adjustment differs 0.5% point or more from zero. Note that although the average adjustment for most variables is close to zero, this can be the average of rather large-sized add factors, as the standard deviations in Table 1 show.

2.1 Properties of expert adjustment

With its model, the Netherlands Bureau of Economic Policy Analysis (CPB) creates forecasts for several macroeconomic variables. Of these, we analyze a subset of ten important ones. We focus on forecasts for the next full year. These forecasts are published four times a year (March, June, September and December) indexed by quarters. The relevant variables are GDP (volume), Exports of goods (volume), Imports of goods (volume), Private consumption

[3] FKSEC was used at the CPB in the nineties, afterwards SAFE was in use up to 2004. Since late 2004 SAFFIER is the model for short-term and medium-term forecasts. See CPB(1992), CPB(2003) and CPB(2007).

(volume), Business investment (volume), Employment business sector, GDP (price), Contractual wages market sector, CPI, and Exports of goods (price).

Additional to the expert-adjusted forecasts, we need the model-based forecasts to see if adjustment leads to improvement. These forecasts are obtained by rerunning the original models with an alternative set of inputs. We reproduced the forecasts based only on the model and data available at that specific time t and neglecting expert opinion for the year t+1. In sum, we consider the quarterly forecasts for the years 1997 until 2006, published in the year before, providing ten forecast errors per forecast origin (the quarter) and per variable.

Insert Table 2 around here

For each of the variables, the effects of the expert adjustments are presented in Table 2. The size of the average adjustment for most variables is small, with exceptions for Investment (volume) and Contractual wages. The average effect on GDP growth is only 0.2% point. Higher growth rates for Consumption and Investment are almost compensated by lower growth rates for exports. These macro effects coincide only partly with the add factors in the behavioural equations of these variables. This is most relevant for Contractual wages. Endogenous effects of the add factors lead on average to more productivity, higher wages and export prices and lower export volumes.

2.2 Effect of adjustment for forecast accuracy

What is the effect of expert adjustment on forecast accuracy? That depends not only on the forecasts but also on the 'realisation'. For this paper we apply the preliminary yearly figures published in the national accounts of Statistics Netherlands. Those figures are available when preparing the forecasts and are relevant for the optimality criterion to be defined below.

Table 3 gives two statistical criteria, that is the mean forecast error and the root mean square of the forecast error (RMSFE) for the model-based (M) and the expert-adjusted (A) forecasts. For eight of ten variables the mean error is equal or closer to zero. For nine out of ten variables the RMSFE's are equal of smaller for the adjusted forecasts than for the model-based forecasts. Hence, generally it seems that expert adjustment matters, and in fact in a positive sense.

Insert Table 3 around here

The effects of the adjustments are however small with three noticeable exceptions, that is, GDP (price), Contractual wages and CPI. For these variables the forecast errors for the expert-adjusted forecasts are much smaller than for the model-based forecast. In other words, the experts adequately adjusted the model. On the other hand, the adjustments in the investment equation worsened the quality of the forecast for investment and thereby to a lesser extent the forecasts of other variables like GDP growth.

What remains though is that so far we have compared forecast errors without taking the loss function into account. As we already indicated, for both the model-based and for the expert-adjusted forecasts, this loss function is unlikely to be known. So, a better way to see if expert adjustment matters is to see if such adjustments makes model-based forecasts closer to optimal or not.

3. Optimality

In the present paper we use the information provided by the Netherlands Bureau of Economic Policy Analysis. They concern the forecasts for ten variables up to the next full year as given in their quarterly reports.

3.1 Methodology

For the sake of clarity, we introduce some notation. Let us denote forecast errors that correspond with the officially released forecasts as $\varepsilon_{A,i,q,t}$, where A denotes "expert-adjusted", where i denotes variable i, where i runs from 1 to 10, where q is 1, 2, 3, 4 and where t denotes years, here 1997 to 2006 Hence, these forecast errors concern the published forecasts after applying expert adjustment.

For these years we were able to re-run the CPB's macroeconomic model-versions used for those forecasts, to compute the forecast errors for the same ten variables but then purely based on the model, that is, the forecast errors found without expert adjustment. Let us denote these forecasts errors as $\varepsilon_{M,i,q,t}$, where M denotes "model".

We evaluate the quality of both our model-based and expert-adjusted forecasts by a test of their optimality. A forecast is considered optimal when given all information available at the time of forecasting and including the preferences of the forecaster, the forecaster provides the best possible estimate. For that purpose, the forecaster is regarded as a decision-maker who minimizes a loss function penalizing the disutility associated with deviations between future realizations and the forecast.

For forecasts errors produced under mean-squared error loss, we can use the familiar tests for optimality. If the parameters have been estimated using a mean-squared error loss function, the forecasts based on the conditional mean are optimal in a mean-squared error sense. In case of the considered macroeconomic model this holds true only partly. The model contains more than two thousand equations and variables, assigned to various blocks, and this seriously limits the feasibility of least squares estimation. About twenty equations are behavioural equations with parameters found by estimation. Some of their parameters are fixed by modellers based on extensive domain knowledge. In addition, past forecast errors are used to change these values if needed. So, basically, the model-based forecasts are not constructed using a mean-squared error loss function for all variables, and also here the loss function is unknown.

To evaluate the quality of the forecast errors, in this case both for model-based and expert-adjusted forecasts, we thus need to rely on the methodology recently proposed by Patton and Timmermann (2007). They have shown that under some weak assumptions on the data generating process (DGP) of the forecast realisations an analysis of forecast optimality is still possible.

Following Patton and Timmermann (2007), we consider the class of data generating processes (DGP's) for which the conditional mean may contain a dynamic component, that is, the expected value of the realisation depends on time and its higher order (un)conditional moments do not. Under the assumption of an error-based loss function and this restriction on the DGP they obtain that optimal forecast errors are serially uncorrelated for lags greater than or equal to the forecast horizon and that the variance of the minimal forecast error increases with this horizon.

The optimality property can be tested for by means of an ARCH test as proposed in Engle (1982). First, we need to determine the forecast horizon. The forecast for year t is prepared in year t-1. At that time information on the current year is only partly available and this information is rather preliminary. Therefore we assume that the forecast for year t is made on all information available at t-2. This information is published in the national accounts as 'preliminary' data, which is revised twice afterwards.

In our notation, the ARCH test then concerns testing the significance of $\rho_{2,M,i,q}$ and of $\rho_{2,A,i,q}$ in the equations

$$
\begin{aligned}
\varepsilon^2_{M,i,q,t} &= \mu_{M,i,q} + \rho_{2,M,i,q}\varepsilon^2_{M,i,q,t-2} + \nu_t \\
\varepsilon^2_{A,i,q,t} &= \mu_{A,i,q} + \rho_{2,A,i,q}\varepsilon^2_{A,i,q,t-2} + \eta_t
\end{aligned} \tag{1}
$$

for the model-based and expert-adjusted forecast errors, respectively. To apply this test a time series should be serially uncorrelated for lags two and higher. This can be checked by a simple AR(2) test.

These tests can only be run for the ten forecast errors for each of the variables, and this may limit the power of these tests. To increase this power, we also perform the same tests in panel version, where we assume the ρ parameters to be equal across the quarters (that is, $\rho_{2,M,i,q}= \rho_{2,M,,i}$ and $\rho_{2,A,i,q}= \rho_{2,A,,i}$).

When a model-based forecast is found a not optimal, this would imply that parameters might have been set at the wrong values. When an expert-adjusted forecast is not optimal, this could mean that an expert adds too much perhaps for too long a period.

3.2 Results of the optimality test

As the test for second-order ARCH requires two observations as a starting value, each test regression can only be run for eight effective data points[4]. This is not much, and hence we investigate additional a way of pooling, which seem sensible given the nature of the forecasting approach. We pool across the forecast origins, which means that we create a four-equation model, and we assume that the parameter for lagged squared forecast errors is the same across the horizons. We adopt a significance level of 10%.

Insert Table 4 around here

The results of the tests for each of the ten variables are given in Table 4. We indicate with an "A" the forecast errors obtained from using expert-adjusted forecasts and with an "M" those from the model-based forecasts. The columns with the header "Forecast made in quarter 1, 2, 3 and 4" concern the p-values of test regressions as in (1), and hence each time concern eight effective observations.

For eight of the ten variables we see that for both A and M, the forecasts for the quarterly forecasts the test statistics are insignificant, so we can conclude that for these

[4] It is allowed to apply the ARCH test because the tests on serial correlation of order higher than 2 indicate that there is no such correlation, and this holds for all the variables and quarterly publications

variables the forecasts are optimal in the sense that the null hypothesis of optimality is not rejected. For two variables (Employment and Contractual wages) the P-value in specific situations is lower then the critical value of 0.10. In two situations this relates to the model-based forecasts and in two situations to expert-adjusted.

As said, these results may be deflated by the small sample size and therefore we estimate the parameters in a four-equation panel model (concerning all four forecast origins) using OLS, while restricting the focal test parameters to be equal across equations. The related P-values appear in the last column of Table 4. Now we see that expert adjustment is beneficial to Consumption (volume) as the not optimal model-based forecast is made optimal by the expert. When we match this with the result for this variable in Table 3, we see that here the experts do a very good job.

In contrast, for Employment and GDP (price) we see that the model-based forecasts are made not optimal by the expert. Interestingly, Table 3 shows that expert adjustment is beneficial in terms of fit. This suggests that even further refinement of what the experts do could lead to even more accurate final forecasts.

4. Conclusion

This paper has proposed and applied a simple methodology to evaluate the quality of large-scale macroeconomic model-based forecasts and expert-adjusted forecasts. It is quite unlikely that both sets of forecasts are generated under a mean squared error loss function, and hence a straightforward comparison of root mean squared forecast errors is not exclusively informative. We followed the recommendations of Patton and Timmermann (2007) and used single-equation and pooled tests for ARCH effects in the forecast errors. Our illustration concerned the quarterly forecasts made by the Netherlands Bureau for Economic Policy Analysis, for which we were able to re-run the original models and also for which we had information on how the experts of that Bureau adjusted the model-based forecasts.

Our unique data set, joint with the simple statistical tests, is informative at least in two ways. We see that for some variables the added value of the experts is very substantial because their intervention reduces forecast errors. However for most variables expert adjustment makes almost no difference. Secondly our research suggests that the model-based forecasts are already rather good, in the sense that all the relevant information is probably included in the forecast. At least, the opposite can not be proven from the available dataset. For a few variables we see challenges for further improving the model or the add factors.

Table 1: Mean and standard deviation of the adjustment by experts to the autonomous variables in the macro model, 1997-2006 (in percentages)

Variable	The adjustment (add factor)	
	Mean	Stand.dev.
Exports domestically manufactured goods (volume)	0.0	1.2
Reexports (volume)	1.1	2.5
Imports intermediate goods (volume)	-0.1	0.9
Imports investment goods (volume)	-1.3	1.8
Imports consumption goods (volume)	-1.3	1.2
Consumption (volume)	0.2	0.5
Investment (volume)	0.4	4.2
Employment market sector	0.1	0.7
Contractual wages	0.3	0.7
CPI	0.4	0.6
Exports (price)	-0.1	0.3

Table 2: Effect of add factors for the mean and standard deviation of the adjustment by experts to the model-based forecast, 1997-2006 (in percentages)

Variable	The adjustment	
	Mean	Stand.dev.
GDP (volume)	0.2	0.5
Exports (volume)	-0.3	0.8
Imports (volume)	-0.2	0.5
Consumption (volume)	0.2	0.7
Investment (volume)	1.2	2.7
Employment	0.0	0.6
GDP (price)	0.5	0.7
Contractual wages	0.9	1.1
CPI	0.4	0.7
Exports (price)	0.1	0.3

Table 3: Mean and RMS of the forecast error for the expert-adjusted forecasts (A) and the model-based forecasts (M), forecast origins for quarterly forecasts 1997-2006

Variable	Mean forecast error		RMS forecast error	
	A	M	A	M
GDP (volume)	0.3	0.1	1.5	1.5
Exports (volume)	0.7	1.0	3.9	4.5
Imports (volume)	0.3	0.5	4.4	4.4
Consumption (volume)	0.2	-0.2	2.1	2.4
Investment (volume)	-1.0	-2.2	5.5	4.9
Employment market sector	-0.2	-0.2	1.0	1.3
GDP (price)	-0.5	-0.9	0.9	1.3
Contractual wages	-0.1	-1.0	0.6	1.3
CPI	-0.4	-0.8	0.8	1.2
Exports (price)	-1.7	-1.8	3.3	3.4

Table 4: P-values of tests for second-order ARCH using equation (1) for each of the variables for each of the forecast origins and when pooled across all origins. The forecast errors concern the officially released forecasts, that is, the expert-adjusted forecasts (A) and the model-based forecasts (M). In boldface we indicate the 10% significant values.

Variable	Forecast	Forecasts made in quarter				
		1	2	3	4	All
GDP (volume)	A	0.76	0.93	0.76	0.21	0.96
	M	0.60	0.85	0.92	0.72	0.75
Exports (volume)	A	1.00	0.99	0.89	0.71	0.85
	M	0.98	0.95	0.97	0.80	0.89
Imports (volume)	A	0.73	0.79	0.72	0.82	0.50
	M	0.74	0.87	0.72	0.81	0.54
Consumption (volume)	A	0.51	0.53	0.73	0.64	0.23
	M	0.37	0.43	0.40	0.19	**0.05**
Investment (volume)	A	0.63	0.75	0.61	0.83	0.59
	M	0.49	0.77	0.86	0.62	0.56
Employment	A	**0.09**	**0.02**	0.59	0.85	**0.05**
	M	0.50	0.39	0.41	**0.03**	0.33
GDP (price)	A	0.46	0.14	0.27	0.33	**0.07**
	M	0.70	0.79	0.27	0.23	0.33
Contractual wages	A	0.37	0.76	0.66	0.61	0.13
	M	0.63	**0.01**	0.68	0.49	0.54
CPI	A	0.74	0.54	0.78	0.81	0.43
	M	0.91	0.70	0.88	0.95	0.94
Exports (price)	A	0.64	0.72	0.94	0.94	0.52
	M	0.65	0.77	0.97	0.78	0.63

References

Christoffersen , P.F. (1998), Evaluating Interval Forecasts, *International Economic Review*, 39, 841-862.

Clements, M.P. (1995), Rationality and the Role of Judgement in Macroeconomic Forecasting, *The Economic Journal*, 105, 410-420.

Clements, M.P. and D.F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.

CPB (1992), FKSEC, a macroeconometric model for the Netherlands, Stenfort Kroese Uitgevers, Leiden/Antwerpen.

CPB (2003), *SAFE*, A Quarterly Model of the Dutch Economy for ShortTerm Analyses, CPB document 42.

CPB (2007), *SAFFIER*, A Multi-purpose model of the Dutch Economy for short-term and medium-term analyses, CPB document 144.

Diebold, F.X., T. Guther and A. Tay (1998), Evaluating Density Forecasts, with Applications to Financial Risk Management, *International Economic Review*, 39, 863-883.

Don, H. and J. Verbruggen (2006), Models and Methods for Economic Policy, 60 years of Evolution at CPB, *Statistica Neerlandica*, 60, 145-170.

Engle, R.F. (1982), Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of UK Inflation, *Econometrica*, 50, 987-1008.

Lanser, D. (2006) ....

Nordhaus, W.D. (1987), Forecasting Efficiency: Concepts and Applications, *Review of Economics and Statistics*, 69, 667-674.

Patton, A. and A. Timmermann (2007), Testing Forecast Optimality under Unknown Loss, *Journal of the American Statistical Association*, in print.