1 Introduction

This monograph deals with the situation where an analyst evaluates expert forecasts and model forecasts, and where it is known that the expert has seen the model forecast and thus that the expert forecast potentially amounts to an adjustment of the model forecast. More precisely, the analyst assumes that

$$\text{Expert-Adjusted Forecast} = \alpha \text{ times Model Forecast} + \text{Adjustment}. \tag{1.1}$$

This additive expression is chosen for analytical convenience, as will become clear in Chapter 2, and also to easily allow for the possibility that the model forecast and the expert forecast have opposite signs.

It is important to stress that the analyst only observes the model forecast and the final expert forecast, and of course also the realized observation, but that the analyst does not observe the value of α, nor the size of the adjustment. In many practical settings, the analyst is usually not the same individual as the expert who adjusts the model forecast, nor is the analyst the same person as the model-builder. In fact, the analyst may have to report to management or to policymakers on the usefulness and relevance of the final expert forecasts, perhaps relative to the model forecast. Such expert forecasts can concern business and economic variables: for example, sales of durable products, earnings of companies or macroeconomic variables like gross domestic product (GDP) or inflation. The forecasts may have to be generated very frequently, for example, hourly, or they may also be quoted just once every half-year.

There is one particular feature that is very important here and that is that the experts are assumed to quote their forecasts *given* that they have received model-based forecasts. It is, however, uncertain if and how they actually incorporate the model forecasts into their decision process, as usually there is no written documentation. So it may be that an expert sets α equal to 0, and fully bases the final expert forecast on his or her own judgement, but this is unknown to the analyst. Indeed, typically, experts do not document how they decide on the values of α and the adjustment.

It is irrelevant whether the model forecasts originate from multi-equation macroeconomic models or from simple extrapolation tools, or anything in between. It may very well be that an expert does not know what the model or forecast algorithm looks like, and in most practical cases it also holds that the expert cannot exercise any influence on how the model forecasts are created. It is usually the case that the expert is not the same person who designs the model, but no specific assumptions on this feature have to be made. The models and their parameters can be updated every single hour, or they may be taken as constant for a long period of time. The statistical tools with which the models are calibrated are largely irrelevant, and it may well be that the expert in fact does not have a clear-cut idea of how the model forecasts were created. In the end, the situation is that the analyst observes an expert forecast, a model forecast and a realization, and the analyst has to evaluate the expert forecast using some criterion.

A key premise of the analysis in this book is that the analyst does not know α or the size of the adjustment and that the analyst is also unaware of *how* the expert has chosen a value of α and the adjustment. The size of these two features, that is, α, and the adjustment, can be set by the expert using his or her own intuition or model, but how that is done is usually unknown. The definition of 'intuition' is obtained from the *Oxford Dictionary* and is 'the ability to understand something instinctively, without the need for conscious reasoning'. At the same time, it may also be that the expert uses knowledge that can be documented and evaluated. For example, a known future regime shift may not be incorporated in the model and thus not in the model forecast, and the expert may use this knowledge to assign a value to the adjustment.

It seems easiest to presume that the expert is a single individual who makes his or her own decisions, but it cannot be excluded that the empirical cases to be analysed below concern cases where a group of experts have jointly decided on α and the adjustment. It can happen that forecasts are adjusted during a group meeting, and that perhaps colleagues interfere, but that is usually unknown in many practical settings. Indeed, if expert forecasts are the outcome of a group process, then all sorts of potential biases may occur, but these are not addressed in this book.

As I have said, it is assumed that the expert is not the same person as the one who is responsible for the model forecast. In other words, experts are assumed to modify a final model forecast, and not particular elements of the econometric model or the statistical algorithm. Of course, model-builders also use their judgement to create their model and subsequent model forecasts, and in the modelling process they can make the adjustments to intercepts, parameters for important variables, and, for example, recent values of explanatory variables. Indeed, much judgement is usually also involved when building a model. One needs to select variables, choose model selection criteria, think of the choice of measurement, and perhaps rely on summarizing techniques like principal components analysis (PCA). One may also have to choose between various parameter estimation methods. All these aspects are assumed to be incorporated in the model forecast that arrives on the expert's desk. So, it is the judgement of the expert that is at stake here, and not the decisions of the model-builder.

A further important premise in this book is that the analyst actually observes the model forecast. Naturally, this facilitates the evaluation of the expert forecasts. This is not always the case, however. Think of the forecasts generated by the IMF, the OECD or the World Bank, where the underlying model is not usually presented, nor are the associated model forecasts (if there are any) displayed in their reports. From an econometric perspective, this lack of available model forecasts can be accommodated by assuming that the analyst has his or her own econometric skills and can create a model using the information available. That is, the analyst can somehow approximate the unavailable model forecasts by designing his or her own econometric model based on publicly available data, and use these as the pseudo-model forecasts, but again that can only be viewed as an approximation. As will become clear in later chapters, these approximations can also be quite informative when evaluating final expert forecasts that could have been based on model forecasts.

Another important stance in this book is that the analyst can only sensibly evaluate the expert-adjusted forecasts if the analyst can approximate what the expert did when he or she received the model forecasts. In other words, to properly analyse the usefulness and accuracy of the final expert forecasts, the analyst somehow needs to infer values for α and the adjustment. Indeed, experts may decide to fully incorporate the model forecasts (meaning that they set α at 1) and just add or subtract a little bit, but they may also wholly ignore the input from an econometric model or statistical algorithm altogether (meaning that they set α at 0). As will be argued in Chapter 2, it will be quite relevant to approximate what the expert does in order to properly evaluate the forecasts. That same chapter will indicate the optimal values of α and the adjustment to make the expert-adjusted forecasts most useful, at least from an econometric perspective.

In a nutshell, the critical questions in this book are the following. If it is assumed that:

$$\text{Expert-Adjusted Forecast} = \alpha \text{ times Model Forecast} \\ + \text{Adjustment},$$

what then are the optimal values and properties of α and the adjustment, at least from an econometric perspective? As will become clear from Chapter 2, among other insights, is that the adjustment can be large or small: that is, the size of the adjustment does not matter, at least in theory, but the adjustment better not be based on the same information that is used to create the model forecast. The next question is, given the optimality results, how close to the optimal setting are empirical values of α and the adjustment in a range of actual practical cases? And when prevailing practice does not match theoretical optimality, as it seems not to do, how does this effect forecast accuracy? Finally, if there seems to be a gap between theory and practice and it does hurt accuracy, are there any sensible strategies for improvement?

INSIGHTS FROM THE LITERATURE ON DECISION-MAKING

Before the focus in this book becomes an econometric one, it seems relevant to consult the literature on decision-making to learn about potential insights concerning α and the adjustment. The decision-making literature is very large and is still growing, but much of the relevant material for the present book is summarized in Kahneman's *Thinking, Fast and Slow* (2012).

One angle on forecasting could be that the expert forecast is not based on a model or algorithm, but that it holds true that

$$\text{Expert Forecast} = \text{Intuition}.$$

In the first part of Kahneman (2012) it is convincingly argued that when forecasts are based not on statistical algorithms but only on intuition, it is quite likely that all kinds of biases are in play, and that these biases negatively impact on forecast accuracy. For example, individuals have a tendency to ignore the phenomenon called 'regression to the mean', which entails that when exceptional events occurred, say, yesterday, it is quite likely that such events will not occur again today. In fact and in contrast, individuals seem ready to believe that recent exceptional events mark the start of a series of such events, and hence a trend will be spotted where there effectively is no such trend. The fact that there is a focus on only a single exceptional event also masks the notion that other events could have occurred too, and that basically the sample size is equal to 1. In his Chapter 18, Kahneman (2012) thus argues that intuition-based forecasts are often based on too much confidence and are often too extreme, and that they ignore the regression-to-the-mean tendency (see also Shanteau, 1992). Building on the influential work of Taleb (2007), who addresses the bias called 'the illusion of understanding', individuals have a tendency to be confident in their interpretation of past events, and they seem to ignore the fact that matters could have been different. Additionally, due to hindsight bias, individuals also have difficulties in reconstructing how they relied on their intuition the last time they created a forecast.

The various biases that can hamper the quality of expert forecasts are convincingly illustrated in the analysis of Tetlock (2005). Political forecasts created by experts turned out not to be so good, and the suggestion is therefore that: 'Another reason for the inferiority of expert judgement is that humans are incorrigibly inconsistent in making summary judgements of complex information' (Kahneman, 2012: 224). Hence, it does not seem wise to ignore a model forecast, if indeed there is one, in favour of a forecast based wholly on the intuition of an expert. Dawes (1979) recommends the use of simple algorithms instead of complicated regression models, but, as mentioned, for the experts adjusting model forecasts the type of model is not very important. And Simon (1992) proposes relying on intuition when it is based on pattern recognition – that is, a set of rules that can be understood and replicated – and this recommendation comes close to what will be reported in Chapter 2 below. That is, the values of α and the adjustment should best be based on the replicable knowledge of an expert.

Kahneman (2012) also convincingly argues in favour of relying on a model or algorithm: 'Because you have little direct knowledge of what goes on in your mind, you will never know that you might have made a different judgement or reached a different decision under very slightly different circumstances. Formulas do not suffer from such problems. Given the same input, they always return the same answer' (Kahneman, 2012: 225). In fact, he concludes that: 'The research suggests a surprising conclusion: to maximize predictive accuracy, final decisions should be left to formulas, especially in low-validity environments' (Kahneman, 2012: 225). Based on this, one may now wonder if Expert Forecast = Intuition should be replaced by

Expert Forecast = Model Forecast,

implying that experts should not adjust model forecasts at all. The decision-making literature suggests not, and in fact the suggestion is that 'intuition adds value . . . but only after a disciplined collection of objective information and disciplined scoring of separate traits' (Kahneman, 2012: 231–2). Hence, from the decision-making literature one seems to conclude with Kahneman (2012) that α is perhaps best set at 1, and thus that

Expert-adjusted Forecast = Model Forecast + Adjustment.

Interestingly, as becomes very clear in Chapter 2, this conclusion closely matches the outcome of an econometric perspective on experts adjusting model forecasts, where the optimal adjustment has a few particular properties.

Naturally, the question now is how experts arrive at a numerical value of the adjustment. It can be expected that similar biases to those mentioned above can be at stake when assigning a value to the adjustment. This is true, but perhaps the potential problematic effects of biases can be alleviated by making explicit what is, in an econometric sense, the sign and size of the adjustment. Chapter 2 will start with this issue by proposing that the adjustment should be equal to the expert's knowledge about the future forecast error associated with the model. For example, when the model forecast is equal to, say, 4, and the expert believes that the associated realization will be 1 higher than what might be expected due to, for example, a known future change of policy, then the expert knows part of the future forecast error and can modify the model forecast of 4 to a final expert-adjusted forecast equal to 5. Chapter 2 will also discuss how decisions on the value of that amount of 1 can be made explicit, so that hindsight bias can be alleviated in the future.

EARLY RESULTS

There were a few studies in the late 1980s and the beginning of the 1990s where the authors examined empirical cases where they had expert forecasts, model forecasts and realizations. In a range of studies, Mathews and Diamantopolous (1986, 1989) investigated how experts performed relative to models in terms of out-of-sample forecast accuracy. Their data concerned sales of repeat-purchase products from a manufacturing company in the UK healthcare industry and their main findings were that expert-adjusted forecasts can be better in terms of out-of-sample root mean squared prediction error (RMSPE).

Bunn (1992) provided an overview of the body of knowledge on the synthesis of expert judgement and statistical forecasting, especially in the light of the then emerging concept of decision support systems (DSS) (see also Belsley, 1988 and Fischhoff, 1988). Huss (1986), Edmundson *et al*. (1988) and Willemain (1989) considered cases where forecasts from simple models were subjected to substantial managerial adjustment, apparently with a successful forecast track record. Bunn (1992: 253) concluded that: ‘It seems that, while well-specified time series models can be most effective in filtering out noise and projecting past patterns in the data, expert intervention will pay off in practice when there is extra information about new untypical circumstances.’ Bunn (1992) also provided a range of reasons why model forecasts might need the adjustment, like low data quality, a change in parameters, omitted variables and the like. There was also an allusion to the notion of somehow combining model forecasts and expert forecasts, which is a strategy that will be analysed below in Chapters 4 and 5. Finally, we may also consider the alternative situation where an initial expert forecast is modified using the information from a model forecast. This interesting situation is, however, beyond the scope of this monograph.

At the beginning of the 1990s, there was also an interest in analysing judgement exercised for macroeconomic forecasts. Drawing on early insights in Howrey *et al*. (1974) and Haitovsky and Treyz (1972), the studies of McNees (1990), Turner (1990) and Donihue (1993) address final expert forecasts for consumer expenditures, gross national product (GNP), exports and inventory investment, to mention a few, when large-scale macroeconomic models delivered the model forecasts.

Donihue (1993: 83) observes that 'virtually none of the macroeconomic forecasting activities in this country [USA] are entirely model-based'. These studies all concern the notion that the model-builders exercise substantial judgement before they arrive at their final model forecast, and it is found that, 'The adjusted forecasts tend to be more accurate overall, although important exceptions are found' (McNees, 1990: 287).

Another important early study is Blattberg and Hoch (1990), who examined expert and model forecasts for catalogue sales of fashion merchandise. They documented that expert-adjusted forecasts can be a little better than model forecasts. In addition, Blattberg and Hoch (1990) showed that when model forecasts and expert forecasts are taken together – that is, somehow combined – their weighted forecast is more accurate. This latter study was for a long time the one that set the agenda, as since then (until recently) almost no studies have appeared where researchers have considered and compared expert-adjusted forecasts with model forecasts. Their 50 per cent model and 50 per cent manager quote (part of the title of their page) was echoed in many later studies.

Quite interestingly, the finding that the 50 per cent/50 per cent rule would work, as a balance of expert and model forecasts, has rarely been disputed. This is particularly relevant as it does matter whether experts have consciously modified model-based forecasts or whether they have ignored them. In other words, the value of α in (1.1) is important before the 50 per cent/50 per cent rule can be recommended. If experts had wholly ignored the model forecast, the 50 per cent/50 per cent rule would indeed seem to be a useful way of combining two independent forecasts based on a model and on pure intuition, respectively. In the case where α is not equal to 0, however, the 50 per cent/50 per cent rule would very much overweigh the expert input by over-emphasizing the model forecast. Indeed, if the expert simply adds a small number to the model forecast, then the newly combined forecast with a 50 per cent/50 per cent balance would count the model forecast twice. In other words, it very much matters that we know what an expert does, before we can make a claim about a seemingly beneficial 50 per cent/50 per cent rule, as will be discussed in Chapter 5.

A REVIVED INTEREST

Recently, large databases with model forecasts, expert(-adjusted) forecasts and realizations have become available in the areas of macroeconomic forecasting and business forecasting, and this has spurred a revived interest in analysing expert-adjusted forecasts. Franses, Kranendonk and Lanser (2011) document that the forecasts from the 1945-founded Netherlands Bureau for Economic Policy Analysis (CPB), which are based on an econometric model of 2,000+ equations, are *all* manually adjusted by domain-specific experts. In sales forecasting, where typically large numbers of forecasts need to be created very frequently, there is, as mentioned, a long tradition of an interaction between forecasting tools and experts. And also, in this latter area, large databases have recently become available (see Fildes *et al*. 2009, and Franses and Legerstee, 2009, 2010), where the first impression is that typically over 95 per cent of all statistical model forecasts are manually modified.

At the same time, research on the evaluation of economic forecasts has intensified. For a long time, researchers usually reported just some statistics on forecast accuracy, but rarely did people bother about which method was more statistically significant than another method. Exceptions are some of the contributions of Clive Granger and co-authors (such as Granger and Newbold, 1986: Chapter 8), but it seems fair to say that a revived focus on forecast evaluation was initiated by Diebold and Mariano (1995). Since then, many studies have appeared on the proper evaluation criteria, how accuracy should be evaluated for rolling-window samples versus recursive samples, and on whether the models are nested or not (Clark and McCracken, 2001, West, 1996, to mention only a few). There are also studies where just plain forecasts are compared, without making assumptions on how they were created (Patton and Timmermann, 2007a, 2007b), but then various specific assumptions – for example, on loss functions – have to