MAGISTERARBEIT

Titel der Magisterarbeit

Forecast Evaluation and Forecast Combination

Verfasser

Sebastian Paul Koch

Angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften (Mag.rer.soc.oec.)

Wien, im April 2010

Studienkennzahl lt. Studienblatt:	A 066 913
Studienrichtung:	Magisterstudium Volkswirtschaftslehre
Betreuer:	Univ.-Prof. Dipl.-Ing. Dr. Robert Kunst

Abstract

This work is divided in two parts. A first part concentrates on the evaluation of the forecast accuracy of around 50 agencies respective the forecasts of UK GDP growth rates in the last eleven years. For this goal, around 130 publications of "Forecasts for the UK economy" from the HM treasury were assembled in order to construct a usable data set. The individual forecasts are then used in a second part to generate combined forecasts according to the forecast combination literature. Different combination schemes are applied and compared with each other. The findings mainly confirm the evidence of comparable empirical studies.

JEL Classification: C23 and C53

Keywords: UK GDP forecasts, forecast evaluations, accuracy criterion, forecast combinations, simple and advanced combination methods

Contents

Chapter 1

Introduction

In late 2008 the year-ahead forecasts for the UK economic growth rate started to drop as in most other globalised countries. As a matter of fact, in summer 2008 the forecasts for the year-ahead GDP growth were still around two percent. Already in October growth forecasts turned negative and at year-end some forecasting agencies published figures of -2.5%. This trend continued in 2009[1] to forecasts of almost -5%. This includes only serious forecasting agencies. In the media however, figures of -6% and even -7% were published. For people who do not follow forecasts on a regular basis this crisis-caused media attention may have sounded more like a competition of "who's got the worse forecast". By witnessing this sequence of events the idea for this work was born.

The aim of this study is to have a closer look on the forecast figures and behaviour of the different agencies who engage in macro economic forecasting. The fact that there is almost never a comparison between these published figures and the true outturn underlines the need for a more systematic approach. This work shall contribute to the satisfaction of this need.

The importance of this work becomes even clearer when realising that the forecast of this main economic variable[2] provides the basis for economic

[1] for the current-year forecast, i.e. done in 2009 for 2009

[2] This work only looks at GDP. Other main indicators are the inflation rate and employment. Further analysis on the other variables may follow.

decision-making in almost any field. For instance, a company has to decide on a certain investment. The present value analysis which discounts future flows of funds offers major help for the decision finding. It is the economic outlook, mostly in form of a growth forecast, that heavily influences the underlying discount rates. Another quite instructive example is that the state estimates its future tax revenues on the basis of the growth indicators. Thus the economic forecasts severely influence the government budget planning.

Similar work[3] on this field was done before but never in this detail and mostly as a yearly "snapshot", in which a very good accuracy performance of a so-called "one-timer" and an agency who simply has a superior model cannot be distinguished. Also it combines different economic variables into one performance index. This influences the conclusiveness.

This work analyses real GDP growth forecasts from 1997 until 2008 provided by the UK government. UK was chosen as the country of analysis for the simple reason of data availability. The interesting year of 2009 is not yet analysable since reliable GDP data for 2009 still have to be confirmed.

Chapter 2 introduces the terminology of the forecast error which is the main unit for the evaluation of the forecast accuracy. It explains more on data collection and data processing as well as how forecasters produce their figures. It also tells about the different strategies forecasters might follow and what kind of target function could be in use. Chapter 3 analyses the data on an aggregate level and provides a summary statistics. In chapter 4 the different possibilities to measures forecast errors are explained and applied to the dataset. The detailed evaluation of individual forecasters is also part of this chapter. The combination of forecasts with its aim of deriving a more precise and more reliable forecast is discussed in chapters 5 and 6.

[3]An economic journalist of the Sunday Times publishes a yearly compilation of the best forecasters. See figure 9.1 on page 67 in the appendix of this work.

Chapter 2

The data

Starting with a definition of the forecast error in the first section, the second section deals with the collection of the forecast data. Finally, section three explains where the eventuated figures, i.e. the realised data, was taken from and how it was processed. As a reminder, this work takes only the growth rate of the real UK gross domestic product into account. Further analyses of other variables are possible but would blast the limits of this work.

2.1 Definition of forecast error

The forecast error is defined as the difference between the actual realisation of the variable in question and the forecast of that variable, as shown in equation 2.1.

$$\text{Forecast Error} = \text{Actuals - Forecasts} \tag{2.1}$$

The forecast error is the basic unit with which the forecasters' accuracy and therefore their performance can be measured. In some publications you may find that the forecast error is defined the other way around. But this should not bother too much since the main error measures either square the value or take the absolute value from it.

Whereas the collection of the forecasts in the next section is laborious but straight forward, the not so trivial point to be discussed in section 2.3 is which actuals to use. But more on that later on.

2.2 The forecasts

2.2.1 Source

Her Majesty's Treasury (HMT), the UK public finance department, provides at its homepage [1] a monthly publication called "Forecasts for the UK economy - a comparison of independent forecasts", where it publishes monthly forecasts from around 40 different agencies who engage in forecasting. The publication's main part consists of six tables. The first three report the forecasters' current-year forecast, whereas the tables 4, 5 and 6 display the year-ahead forecast. In other words, for every year - with the exception of 1998 - 24 monthly forecasts are published for a broad range of variables.

The forecasts are restricted to the UK economy and can be divided into three main parts: forecasts of GDP and its components, forecasts of prices and forecasts of "other selected variables" like employment. Furthermore, it divides the forecasters into two groups. The first group is called "City forecasters"and the other one "Non-City forecasters". City forecasters are mainly international banks (e.g. Citigroup or UBS) and investment firms (e.g. Goldman Sachs or Morgan Stanley). Non-City forecasters can be described as mostly UK-based research institutes (e.g. Cambridge Econometrics or IHS Global Insight) and international organisations (e.g. EC, IMF or OECD). The HMT sends out a monthly questionnaire to collect the data from the different agencies. A full list of agencies (see figure 9.9) who take part in today's publication is listed in the appendix.

This work considers only the forecasts for the growth rates of real UK GDP, which can be found in the first column of table 1 and 4 of each publication. In total there are 135 monthly publications, which translate into

[1] http://www.hm-treasury.gov.uk/data_forecasts_index.htm

an observed time period of more than 11 years, starting in September 1997 until December 2008.

It should be noted that HMT publishes a medium-term forecast every three months. This forecast comprises two, three and four year-ahead forecasts. Those mid-term estimations are not considered in this study, since it is common opinion that only this-year's and the next-year´s forecasts are of public interest. The forecasts that go beyond that time frame are way too imprecise to rely on.

2.2.2 Data collection

The collection of the data is organized by the HMT. It described the data collection process in the following way. A questionnaire is sent out to forecasters on the first working day of every month. Then on the second Wednesday of each month the forecasters are reminded to fill in their questionnaire, since the deadline is always the second Thursday of each month. If some of the forecasters do not return their questionnaire on time, then the previous month's figures are taken again, but noted that this particular forecast was produced in a previous month. For instance, if a certain forecaster does not compile the questionnaire by the second Thursday in November, then the October values become published again and marked by a " * ". If there was no forecast in October either, then the September forecast is used again and so on. In the case that a forecaster did not update for more than 5 months[2], it will be temporarily taken out of the list until a new forecast is received.

2.2.3 Forecast updating

It is quite unrealistic to believe that every forecaster is updating their data every month or that they always return the survey questionnaire to the HMT back on time. For this reason every data point does not have the same quality. For example, HMT indicates that a forecast is not from a certain month by always displaying the month the figures were originally produced.

[2]with the exception of OECD, who publishes its numbers only twice a year

However, in this work no difference was made between a new forecast and one that is already a couple of months old. There is a reason at hand: simplicity. A further reason is that the dataset would shrink to about two thirds of its original size. But since the superior aim of this work is to find out which of those forecasters are good ones, the inclusion of all data points can be justified in the following way. Compared to those who update their figures just a couple of times per year, the ones who update their data more often should have an advantage for the simple reason that they are more up-to-date.[3] Thus, it comes naturally to leave all data points in the dataset, since by doing so the recentness of the forecast is taken into account when the performance of an agency is calculated.

This fact can be seen by looking at figure 3.2. It shows that versus the end of almost every 24-month forecasting period the average of the forecasts converge to the actual outcome. Hence, by updating only from time to time a six months-ahead forecast might for instance be taken also as the two months-ahead which on average should decrease the forecaster's performance. In a way, this mechanism works as a "punisher" for not updating on a regular, i.e. monthly basis.

2.2.4 Forecasters

In these 12 years of observation, 87 different agencies were listed as forecasters. This number was finally reduced to 51 forecasters who remained in the panel. The reason for these reductions is that a couple of agencies changed their name, or were taken over by another firm who previously did not engage in forecasting. As long as it could be assured that a change in name was just due to mergers or acquisitions, the two series were merged to one while keeping the latter name. Problems could hide here, since it cannot be assured that the forecasts were done by the very same people or very same models run by those people. Finally, a total of 43 forecasters was merged to 19, who then remained in the sample. Another 12 were cancelled because they showed up only for a very short period of time.

[3] Or don't you look up the most recent publication if you have five to decide from?

2.2.5 City vs. non-city forecasters

Batchelor [5] refers to the point that at least governments and multinational agencies should be more accurate than private sector banks or business corporations, just because they have a timelier and complete knowledge of the official statistics. Moreover they have "some insight into their own intentions and likely reactions to future events". But on the other hand, those are subject to political pressure.

In a similar consideration, the difference between research institutes and private sector agencies is a priori hard to tell. On the one hand, researchers might be more accurate just because one could think that is one of their their core competencies. But on the other hand, private companies do much more depend on their in-house analysts for the simple reason that a lot of money depends on their solid assessment.

2.2.6 Forecasters' target function

What should also be kept in mind is that it is not every agency's aim to forecast the true value. For instance, some forecasters could try to engage in the so-called strategic forecasting. This means that a forecaster just may want to maximize publicity by publishing extreme forecasts. So the centre of attention is to get quoted in the media instead of doing a "good job". This can often happen just because very few people are actually checking their statements about the future developments of the economy. It is actually this work's aim to contribute to a reduction of this deficiency. So, knowing that, the forecasting agencies can - without any control mechanisms - announce the figures they think are suiting them best. Expressed in a very insinuating way, it could be argued that announcing extreme forecasts is a low-budget possibility of advertising.

Having said that, their behaviour can also be understood, in a sense that if they hide in the crowd, nobody is listening to them. Thus on the other hand, it can be argued that they are engaging in just plain product

differentiation[4].

Another reason for announcing "extreme" forecasts could be that some forecasters want to provide an upper or lower bound to the variables in question. In some cases, this is also a very useful and important information. Just imagine, some company is deciding about an investment and the question is whether or not to carry it out. By running worst-case and best-case scenarios which may base on aforementioned upper or lower bounds of the forecast distribution the people in charge have additional information at hand for their decision-making.

A further reason why forecasters may diverge from their true belief is that they are closely related to a governmental institution or even a political party. In this case they might follow a strategy of backing policies which are favoured by their political ties.

Another way to think about this issue is that some forecaster, presumably mainly investment banks, do indeed have the "close-to-perfection" model. But obviously they do not want to inform the public about it since they are going to use this information advantage and try to turn it into real money; like in the old saying "If you're so smart, why ain't you rich?"

A last reason why some forecasters either underestimate or overestimate repeatedly the true outcome may just be traced back to a very pessimistic or optimistic person in charge of the forecast.

As it will be shown later, at least one agency in this panel is very likely to engage in one of those strategies, just because it really must be very hard to constantly deliver bad results.

Finally, a last point must not be missing. Even if the target function is not expanded by some of the before-mentioned strategies, other influences can alter the outcome of a forecast. For instance, the forecasting agencies may influence each other. The extent of this factor might actually be quite strong since the interval of updating their beliefs is with one month quite short. Put differently, everybody knows the beliefs[5] of the others. This fact

[4] see Ashiya [4]

[5] although not the underlying target function

has quite strong implications for the possibility of a systemic failure. The more a forecaster is influenced no matter with or without its awareness, the higher the possibility of collectively drifting away from a best possible guess.

In addition to that, there is a free-rider problem which may increase the above effect. Just imagine the following situation. The person or the group of people who are running the forecasts are for some reason, e.g. heavy workload, not able to do a monthly "honest" update. They could be tempted to take their last month's update and just "adapt" it to their competitors' opinion of last month plus some recent information. Even the word "honest" here, may be interpreted differently.

2.3 The actuals

2.3.1 The problematic nature of actuals

In order to check the forecasters' accuracy we need to match the individual forecasts with the actual outcome of the variable. But already at this point the first problem arises, which - if not taken proper care of - may alter the results considerably or even make them useless.

As Batchelor [5] points out, a "problem to be resolved is what data to use to measure the actual or realized value of the forecast variables. This is less trivial than it might seem. Initial estimates of real GDP growth, for example, are often revised several times within the following year, and further revisions may occur years later as the GDP index is re-based." And then he refers to the work of Zarnowitz and Brown [23] from 1992. They suggest taking the value of the variables which are published in August of the following year as actuals. They motivate their view with the reason that on the one hand very early numbers are "too unreliable", whereas on the other hand later data may be subject to re-basing and redefinitions which should not at all enter the judgement of a forecaster's performance. So the conventional view evolved that "forecasters are judged by their ability to predict the values of variables as they appear in the middle of the following

year"[6].

2.3.2 The actual actuals

The data, from which the actuals were derived, was taken from three different sources. The first is the IMF spring publication of its series "World Economic Outlook" [1], which is usually published in April or May of each year. As more data become confirmed or revised during the course of the year, the September edition of IMF's "World Economic Outlook" provides also usable actuals. According to a widespread opinion in the literature, see subsection 2.3.1 above, the actuals published in the second part of the year may reflect the best counterpart to the forecasts since on the one hand already six months of ongoing data revision past, and on the other hand these revisions are still close in a timely meaning to the forecasted year.

UK real GDP Growth															
Issue of the IMF World Economic Outlook	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Spring Actuals	Autumn Actuals
May 1998	2.7	2.2	3.3												3.3	
October 1998	2.7	2.2	3.4													3.4
May 1999	2.8	2.6	3.5	2.1											2.1	
October 1999	2.8	2.6	3.5	2.2												2.2
May 2000	2.8	2.6	3.5	2.2	2										2	
October 2000	2.8	2.6	3.5	2.6	2.1											2.1
May 2001	2.8	2.6	3.5	2.6	2.3	3									3	
October 2001	2.8	2.6	3.5	2.6	2.3	3.1										3.1
April 2002	2.9	2.6	3.4	3	2.1	3	2.2								2.2	
September 2002	2.9	2.6	3.4	2.9	2.4	3.1	1.9									1.9
April 2003	2.9	2.6	3.4	2.9	2.4	3.1	2	1.6							1.6	
September 2003	2.9	2.6	3.4	2.9	2.4	3.1	2.1	1.9								1.9
April 2004		2.7	3.3	3.1	2.8	3.8	2.1	1.7	2.3						2.3	
September 2004		2.8	3.3	3.1	2.9	3.9	2.3	1.8	2.2							2.2
April 2005			3.3	3.1	2.9	3.9	2.3	1.8	2.2	3.1					3.1	
September 2005			3.2	3.2	3	4	2.2	2	2.5	3.2						3.2
April 2006				3.2	3	4	2.2	2	2.5	3.1	1.8				1.8	
September 2006				3.3	3	3.8	2.4	2.1	2.7	3.3	1.9					1.9
April 2007					3	3.8	2.4	2.1	2.7	3.3	1.9	2.7			2.7	
October 2007					3	3.8	2.4	2.1	2.8	3.3	1.8	2.8				2.8
April 2008						3.8	2.4	2.1	2.8	3.3	1.8	2.9	3.1		3.1	
October 2008						3.9	2.5	2.1	2.8	2.8	2.1	2.8	3			3
April 2009							2.5	2.1	2.8	2.8	2.1	2.8	3	0.7	0.7	
October 2009							2.5	2.1	2.8	3	2.2	2.9	2.6	0.7		0.7

Figure 2.1: UK real GDP growth: the actuals

Figure 2.1 shows in a direct comparison the difference of the actuals between the April/May publication of the IMF's "World Economic Outlook" and the September/October issue. It is easy to see that just within those

[6]see Batchelor [5]

six months the actuals differ in most years around one tenth of a percentage point. The biggest difference amounts to full 1.2 percentage points over a period of eight years. But here, keep in mind that this must not necessarily be due to revisions or rebasings of the GDP data, but it might also be the case that the GDP was measured in a slightly different way.

In this work, the conventional view of Zarnowitz and Brown [23] is followed, which means that the IMF autumn value of each year is used as the actuals the individual forecasts are compared with.

Further actuals were taken from the today's version of the UK's National Statistical Office. These actuals are used in the lower panel of figure 3.1 but just to show the effect of using "false" actuals.

Chapter 3

Data analysis

3.1 Overview of the data - "the blur"

The figure 3.1 on page 17 was given the name "the blur". A revealing overview of the data used in this study is provided with the help of this kind of presentation, it just needs some explanation[1].

The figure 3.1 measures the forecast errors of all forecasters for real GDP growth in the UK economy. Every panel consists of 11 years of covered forecast errors. 1998 only shows 15 months of forecast errors, whereas all the other years reflect the usual 24 months of forecasts: 12 for the year-ahead and other 12 for the current-year forecasts. The 51 individual agencies are listed in the rows.

A dark red dot corresponds to a forecast error of -3 percentage points. A dark blue dot corresponds to a forecast error of $+4$ percentage points. These two values are in fact the lower and the upper bound of all forecast errors available in the observed period of time. A bright yellow dot stands for a forecast error of 0 what would be a perfect match. White dots mark missing data. It consists of 13005 pixels where each pixel represents a single forecast error.[2]

[1] In fact the figure was produced by zooming out of the data containing Excel sheet which was conditionally formatted before.

[2] 51 forecasters times ten years with 24 month and one year with 15 month yields 13005

The figure consists of three parts. Every part reflects a different set of underlying actuals. As mentioned above already, three different actuals were taken into account: the IMF spring revision is in the upper panel, the IMF autumn revision in the middle panel and in the lower panel are the forecast errors which were derived with the data from the UK's National Statistical Office.

3.2 Analysing the data

For now, let us consider only the panel in the middle, which corresponds to the above explained convention on actuals. The very big advantage of this kind of presentation is that certain findings can be found at a glance.

- In the years 1999 and 2000 there is the biggest accumulation of blue points. This means that the forecast errors were around its upper bound of 4 percentage points in this period. Thus, the actuals were much higher than foreseen by almost all of the forecasters. In other words, the strength of the economy was underestimated. A possible explanation for that could be that most analysts expected a recession due to the burst of the dotcom bubble. Since "only" the IT sector and the stock markets were affected badly, but did not influence the real economy whose pace is measured in the growth of real GDP, the actuals in this period did not really drop. So a recession was expected but the real economy did actually quite good, hence a high positive forecast error.

- The largest accumulation of red dots can be found - not really surprisingly - in the year of 2008. That means that especially in the year-ahead forecast for 2008 the forecast errors were around -3. This corresponds to the fact that the actuals were much lower than most forecasters did predict. This is straight forward since "nobody saw it coming". Hence, the expected outcome of the real economy was quite severely overrated.

colour-coded data points.

Figure 3.1: The data at a glance: the blur

- The more abrupt a colour change, the more the forecasters were surprised by the development of the economy. Again, this can best be seen between the years 2007 and 2008.

- A vertical comparison of the forecast errors shows that in almost all of the times almost all forecasters predict values that are very similar to each other. (There is one severe outlier, but more on that later.) This "herding effect" can be easily explained by the fact that all forecasters use the "same" set of information. Put differently, the systematic risk of all forecasters being wrong is quite substantial.

- Versus the end of all 24-month periods, the dots turn yellow. This corresponds to the fact that versus the end of each year a lot of information is available. Thus, surprises may affect the yearly actual growth rate only by a minor extent, and the forecast errors converge against zero.

These findings are in line with that what was expected. Two other findings can not be easily explained, or better, not be explained at all.

- There is a severe outlier. This is indicated by the mostly blue colour-coded horizontal line which can be found between the first and the second third of each panel. Two things are striking. Firstly, this forecaster is always around two to three percentage points below "the herd". This makes this forecaster a pessimist; always betting that the economy is doing worse than it actually turns out. Or put differently, they are in constant fear of a bad surprise. And secondly, how can this forecaster constantly be below the average? There might be two explanations for this. It could be that this forecaster never learns from the mistakes done before. But on the other hand these two characteristics suggest that the forecaster is not more or less informed than the rest, just that they engage in a certain strategy as we pointed out during the introduction of the forecasters earlier in this section.

- From the year 2002 onwards, a missing data points from horizontal white lines. This has to do with the fact that mostly in January and

February the HMT is reporting the year-behind forecast. In other words they report for instance in January 2003 again forecasts for 2000 instead of forecasting the year 2003. A reason might be, that a first publication of the actuals for 2002 still were not produced in January 2003.

Until here, the focus was on the panel in the middle. Let us now compare the three panels to each other. Again, the difference between the three is the underlying actuals used to calculate the forecast errors.

- Between the first and the second panel no visual differences can be found. As a matter of fact, those two just differ at maximum from each other for a tenth of a percentage point, which can be seen again in table 2.1.
- A visual difference exists between the first two and the third panel. From 1998 until 2003 the area is in terms of the colour-code more bluish. Or, put differently, it has a higher positive forecast error than the rest of the years. This could plausibly reflect the above mentioned fact that from today's view on the UK NSO statistic, the 1998 to 2003 forecasts were re-based or redefined in order to make them more comparable to the ones of today. It would be foolish to think that the forecasts on average were biased downwards.

  As a matter of fact, the actuals taken from the UK NSO can actually not be applied at all, since one would neglect that every year has undergone a different amount of revisions. That means that the year 2008 has undergone - let us assume - four revisions till today which may or may not have been substantial (smaller then 0.1 percentage points). But the year 2000 was - again let us assume - subject to 12 revisions and its value may be changed for more than a full percentage point, not to forget possible rebasings or redefinitions.

So from now on, the study only uses the IMF Autumn actuals to measure forecast errors.

3.3 Statistical summary: mean and standard deviation

Figure 3.2 will provide more detailed information about the panel of forecasters in aggregate. The 11 graphs in this figure show for each year from 1998 to 2008 the month-to-month evolution of the most important summary measures: the sample mean and the sample standard deviation. The sample median was not shown since it was always closely overlapping with the mean.

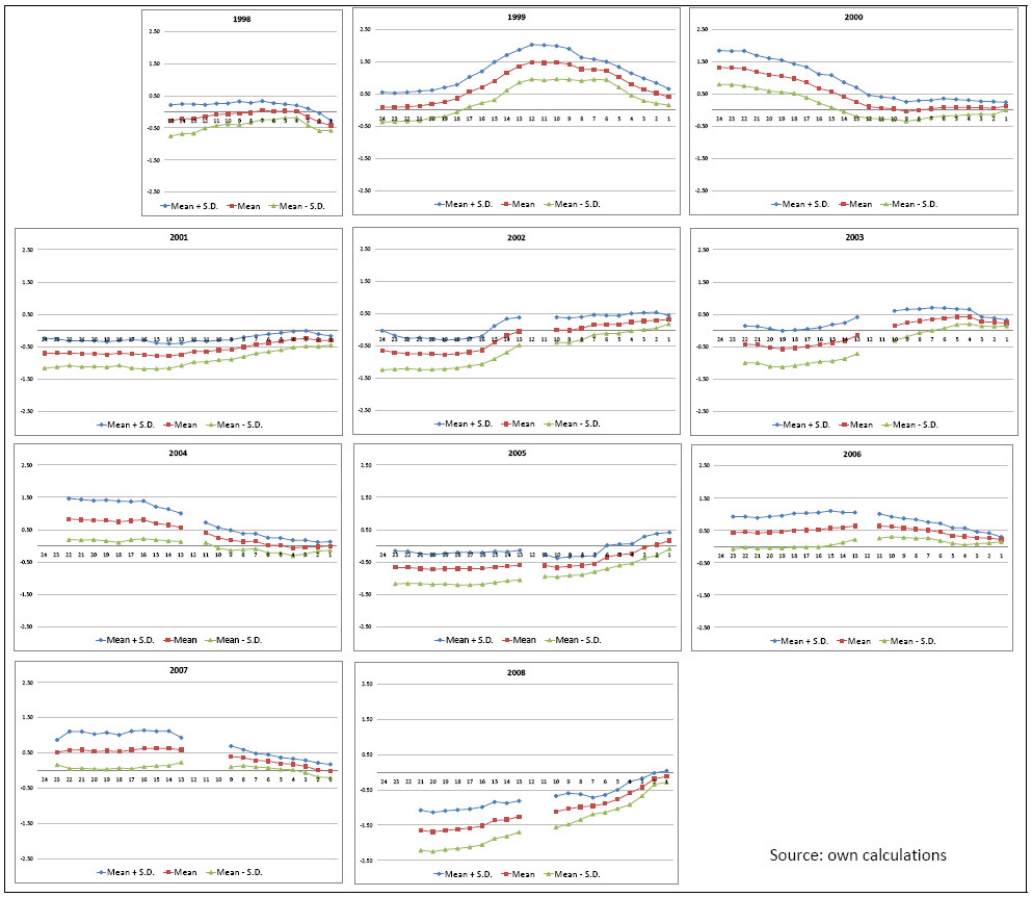


Figure 3.2: Mean ± one standard deviation

The horizontal axis in these graphs represents the chronological progression of the 12 months of year-ahead forecasts as well as the 12 months of current-year forecasts. The axis is labelled in reverse order since the month

number 24 stands for 24 months to go - more precisely 23 and half a months to go - until the year which is forecast has past. For instance, month 24 in the 2007 graph represents January 2006, the month in which the first 2007 forecast was produced. In the same sense month 1 stands for December 2007. The vertical axis measures the forecast error.

As already stated, the graphs describe each year of the panel by three lines. The upper line corresponds to the mean plus one standard deviation, and the lower line to the mean minus one standard deviation. The one in the middle is the mean itself.

3.4 Analysing the statistics summary

From the graphs in figure 3.2, a number of findings can be drawn as well.

- In general, the forecast errors converge versus the end of the 24 month period against zero. This is not very surprising, as with time more and more information about the year to forecast becomes available. But it is far from correct to say that the mean is very close to zero in the end of each two years. Sometimes the mean forecast errors in December amount to around half a percentage point. But the extent to which the forecasters' mean misses the actual outcome seems to be quite arbitrary.

- Also very expected is the fact that the standard deviation decreases over the two years. This comes along with the consideration that the less information is available at the beginning of the 24- month, the more uncertain the forecasters are. This yields a higher variability of their forecasts, and thus the standard deviation is higher. In other words, over the course of two years of information gathering, the standard deviation decreases, as seen in the yearly graphs. The mean standard deviations is shown in figure 3.2 further below in the text.

3.5 The benchmark

According to a broad accordance in the literature the mean over all forecasts - the so-called consensus forecast - outperforms in many cases the individual ones; especially when analysing more than one year. For this reason, this average forecast naturally serves as the benchmark all later trials of forecast combinations have to compete with.

As Batchelor [5] points it out in his 2001 study, "Consensus forecasts are known to be hard to beat. Just as spreading investments over many assets reduces risk, so averaging forecasts across different forecasters reduces the size of the expected error, a point first formalized by Bates and Granger [6] around 40 years ago." And already 1984 Zarnowitz [22] found that, "the group mean forecasts from a series of surveys are on average over time more accurate than most of the individual projections. This is a strong conclusion, which applies to all variables and predictive horizons covered and is consistent with evidence for different periods from other studies."

This section analysed the data in a general way. The upcoming section will take a closer look on the individual forecasters.

Chapter 4

Evaluation of forecasters

4.1 The obvious question: who is the best forecaster?

The question of particular interest in this section is, "is there a forecaster or a group of forecasters who are better than others?" This is still a very vague formulation for two reasons. What is the definition of "better", and what kind of time horizon are we talking about? The former question addresses the topic of how to measure accuracy. A look into the literature provides us with a lot of possible measurements. Among those, three error measurements found broad acceptance in the literature, which we will explain in the next paragraph. The latter question is concerned with whether the success of a forecaster was just arbitrary, or a one-time thing, or whether it performed constantly well over at least a couple of years.

4.2 Error measures

The first conventional error measure is the mean forecast error (MFE), which is a measure of the overall bias of forecasts.

$$MFE(i) = \frac{1}{M} \sum_{t=1}^{M} y_t - f_{i,t} \tag{4.1}$$

where y obviously is the actual outturn of the GDP growth rate. f is here the forecast for the year in question. t indicates the month, whereas i represents the individual agency. M is usually 24 because one forecasting period contains 24 month. In the last section we already introduced the forecast error and its meaning of over- or underestimating the pace of the economy. Here, a positive average forecast error may indicate that over the whole period of observed forecasts the actual values were underestimated, which made the forecasts too low. On the other hand, a negative one indicates overestimation. As already said the average forecast error measures the overall bias and is in this sense no measure of accuracy, since a forecaster could produce large but alternating forecast errors. Although its constant over- and underestimation cancel each other out in a way that the bias is zero or close to it, this forecaster is still not a very good one.

The second commonly used error measure is the mean absolute forecast error (MAFE). It is calculated by averaging all differences between the actual and the forecast, disregarding the sign of the forecast error. So it does not matter whether the forecasts were over- or underestimated, but just the amount of how much the forecast was under- or overestimated.

$$MAFE(i) = \frac{1}{M} \sum_{t=1}^{M} |y_t - f_{i,t}| \tag{4.2}$$

The implication here is that the mean absolute error evaluates a forecast error of 2 percentage points twice as seriously as a forecast error of only 1 percentage point.

The third measure used in this study is the mean squared forecast error (MSFE), which is computed by squaring each forecast error and then taking the average of them.

$$MSFE(i) = \frac{1}{M}\sum_{t=1}^{M}[y_t - f_{i,t}]^2 \tag{4.3}$$

The implicit assumption here is that a forecast error of 2 percentage points is evaluated as 4 times as serious as a forecast error of only 1 percentage point. This in turn means, that the MSFE punishes a couple of heavy forecast errors much more than a lot of small forecast errors. This also means that constantly "good" forecasters become better ranked than the so-called one-timers.

Other error measures which are not used in this work include the scale-independent mean absolute percentage error (MAPE) which is often deployed when the underlying data is denominated in different units. The root mean squared error (RMSE) is like the mean squared forecast error with the difference that it takes the root of the MSFE. Also sometimes found in the literature is the Theil's inequality coefficient[1] which evaluates the forecasting performance relative to a benchmark.

4.3 The measures used in this study

In the following part the forecasters are described individually by the following criteria:

- Counts, which shows the number of forecasts published. (COUNTS)
- Mean forecast error, which reflects a natural measure of bias. (BIAS)
- Mean squared forecast error, as a first accuracy criterion. (MSFE)
- Mean absolute forecast error, as a second accuracy criterion. (MAFE)

The four next figures are structured in exactly the same way for reasons of simplicity and comparability. An aggregated view of the measures for

[1] See Batchelor [5] on Theil's inequality coefficient

COUNTS													
FORECASTER	1998 - 2008	1998 - 2007	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Benchmark Median	233	214	15	24	24	24	22	20	21	22	22	20	19
Benchmark Average	233	214	15	24	24	24	22	20	21	22	22	20	19
Barclays Capital	233	214	15	24	24	24	22	20	21	22	22	20	19
Experian Business Str.	233	214	15	24	24	24	22	20	21	22	22	20	19
Economic Perspectives	232	213	15	24	24	24	22	20	21	22	22	19	19
Hermes	232	213	15	24	24	24	22	20	21	22	21	20	19
UBS	232	213	15	24	23	24	22	20	21	22	22	20	19
Daiwa Institute of Res.	231	212	15	24	23	23	22	20	21	22	22	20	19
HSBC	231	212	15	24	24	24	22	20	21	22	22	18	19
Lehman Brothers	231	213	15	23	24	24	22	20	21	22	22	20	18
Liverpool Macro Res.	231	212	15	24	24	23	22	20	21	21	22	20	19
Morgan Stanley	231	212	15	24	24	24	22	18	21	22	22	20	19
NIESR	231	212	15	23	24	23	22	20	21	22	22	20	19
OEF	231	212	15	22	24	24	22	20	21	22	22	20	19
Cambridge Econometrics	230	212	15	24	23	24	22	20	21	21	22	20	18
CBI	230	211	15	23	24	24	20	20	21	22	22	20	19
Global Insight	230	211	15	24	22	24	22	20	21	21	22	20	19
Citigroup	229	211	15	24	23	24	22	20	21	22	22	18	18
Goldman Sachs	229	210	15	24	24	24	20	19	20	22	22	20	19
ITEM Club	229	211	15	24	23	23	21	20	21	22	22	20	18
Lombard Street	228	209	15	24	24	24	22	20	21	18	21	20	19
CEBR	227	208	15	24	24	24	22	20	21	19	19	20	19
EC	225	207	15	24	24	21	21	20	21	20	21	20	18
RBS Global B&M	225	206	15	24	24	24	22	20	19	18	20	20	19
Abn Amro	222	205	14	23	24	24	21	17	19	22	21	20	17
Credit Suisse	222	203	13	21	24	24	22	20	21	22	22	14	19
J P Morgan	221	210	15	24	24	24	22	20	21	22	22	16	11
OECD	221	202	15	24	24	24	18	18	19	22	18	20	19
EIU	207	193	15	24	24	24	22	20	21	20	11	12	14
WestLB	207	207	15	24	24	24	22	20	21	22	21	14	0
Deutsche Bank	206	187	6	18	24	24	22	20	21	20	21	11	19
IMF	201	183	15	14	20	21	18	17	20	19	20	19	18
Dresdner Kleinwort W.	198	186	15	24	22	24	22	12	13	22	14	18	12
Schroders	198	198	15	24	24	23	18	12	21	22	22	17	0
Williams de Broe	197	197	15	24	24	24	22	20	21	22	18	7	0
Standard Chartered	176	157	0	0	10	22	22	20	21	22	21	19	19
Capital Economics	168	149	0	0	5	17	22	20	21	22	22	20	19
Credit Lyonnais	148	148	15	24	24	24	22	20	19	0	0	0	0
ING Financial Markets	130	111	0	0	0	0	8	18	21	22	22	20	19
Bank of America	127	108	0	0	0	0	7	17	21	21	22	20	19
Fortis Bank	118	111	0	0	0	8	18	19	18	19	22	7	7
Merrill Lynch	117	117	15	24	24	24	22	8	0	0	0	0	0
Societe Generale	111	111	15	22	24	24	19	7	0	0	0	0	0
Moody's Economy	110	92	0	0	0	0	0	9	20	21	22	20	18
Charterhouse	103	103	15	24	24	24	14	2	0	0	0	0	0
Henley	98	98	15	23	24	24	12	0	0	0	0	0	0
Barclays Bank	96	96	15	24	24	22	11	0	0	0	0	0	0
Primark WEFA	91	91	15	24	24	20	8	0	0	0	0	0	0
Bridgewell	88	88	0	0	0	0	3	13	21	22	20	9	0
Chase Manhattan	78	78	15	23	24	16	0	0	0	0	0	0	0
MacroEcon.com	74	58	0	0	0	0	0	0	3	14	22	19	16
Greenwich Natwest	53	53	8	20	19	6	0	0	0	0	0	0	0
Norwich Union IM	49	49	15	23	11	0	0	0	0	0	0	0	0

Figure 4.1: Overview on how much information was gathered

the 12 observed years are given in columns 2 and 3. In the second column the criterion is displayed with respect to all 11 years. Column 3 reports the same as column 2 with the little but maybe insightful difference that the crisis year 2008 was left out of the consideration for a moment. This has the advantage to see to which extent the bias or forecasting accuracy criteria are influenced by a rather extreme year (an extreme year which is supposed to happen every other decade). In the rest of the columns every single year is depicted.

Finally the tables are again colour-coded in order to find certain patterns or outliers easier. Please note that the average forecast was included as "Benchmark Average". The "Benchmark Median" was also included in order to show a possible differences to the the average.

4.3.1 Counts

First, an overview of the forecast counts is given. This has - especially for the general evaluation and interpretation - quite important implications. As we will show later, a forecaster may improve its performance also by not handing in any forecasts. At this point keep in mind what was said in the beginning about the collection of forecasts: to enter the forecast publication it is not necessary to have produced updated forecasts. Again, if a forecaster does not send back any updated figures, the HMT publishes those of the last month.

When looking at the figure 4.1, remember that at first all forecasters who had less than 50 forecast were not included in the panel. There are at least 20 agencies which have almost always published their forecasts. On the other hand the lower two fifths can not offer more than 200 counts. Twelve out of those 51 forecasters did not come up with a forecast in 2008 which most probably means that they ceased to produce forecasts. The table is colour-coded in the following way: green corresponds to the highest number of 233 counts; red on the other hand corresponds to 0 counts.

MEAN FORECAST ERROR

FORECASTER	1998 - 2008	1998 - 2007	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
MacroEcon.com	-0.409	-0.065							-0.150	-0.783	0.137	0.244	-1.657
Charterhouse	-0.312	-0.312	-0.693	-0.308	0.533	-0.767	-0.550	-0.500					
Fortis Bank	-0.267	-0.195				-0.450	-0.356	-0.353	-0.022	-0.695	0.295	0.300	-1.400
Merrill Lynch	-0.201	-0.201	-0.547	0.692	0.238	-0.617	-0.673	-1.000					
EC	-0.194	-0.099	0.093	0.454	0.275	-0.967	-0.514	-0.385	0.352	-0.790	0.181	0.200	-1.289
Barclays Capital	-0.179	-0.081	-0.127	0.700	0.488	-0.754	-0.377	-0.206	0.156	-0.931	0.095	0.090	-1.294
Citigroup	-0.163	-0.076	0.340	0.796	0.178	-0.829	-0.355	-0.345	-0.024	-0.855	0.282	0.182	-1.183
Lombard Street	-0.138	-0.078	-0.500	0.413	0.125	-0.763	-0.318	-0.140	0.145	-0.850	0.245	0.670	-0.795
Credit Suisse	-0.126	-0.013	-0.108	0.805	0.554	-0.550	-0.450	-0.260	0.238	-0.804	0.325	0.175	-1.342
NIESR	-0.121	-0.014	0.113	0.678	0.304	-0.804	-0.448	-0.185	0.362	-0.650	0.277	0.250	-1.316
J P Morgan	-0.109	-0.043	-0.393	1.355	0.125	-0.754	-0.417	-0.402	0.019	-0.591	0.309	0.150	-1.358
Standard Chartered	-0.098	0.011			-0.250	-0.545	-0.059	0.135	0.448	-0.636	0.586	0.374	-1.000
Williams de Broe	-0.087	-0.087	-0.287	0.563	0.121	-0.567	-0.114	-0.120	0.224	-0.650	-0.028	-0.086	
OEF	-0.060	0.041	-0.067	0.845	0.383	-0.563	-0.191	-0.170	0.081	-0.580	0.401	0.263	-1.194
Daiwa Institute of Res.	-0.060	0.033	0.007	0.642	0.370	-0.548	-0.132	-0.195	0.186	-0.609	0.291	0.280	-1.105
Global Insight	-0.051	0.038	0.013	0.517	0.636	-0.729	-0.236	-0.210	0.219	-0.567	0.436	0.300	-1.042
IMF	-0.037	0.097	-0.253	1.286	0.550	-0.310	-0.406	-0.035	0.345	-0.500	0.295	0.195	-1.399
Goldman Sachs	-0.034	0.060	-0.140	1.017	0.433	-0.450	-0.345	-0.247	0.100	-0.523	0.259	0.305	-1.079
Morgan Stanley	-0.030	0.059	-0.500	0.821	0.263	-0.521	-0.314	-0.151	0.398	-0.356	0.313	0.434	-1.028
Deutsche Bank	-0.028	0.080	-0.417	1.172	0.271	-0.713	-0.145	-0.175	0.281	-0.540	0.743	0.264	-1.089
CBI	-0.028	0.076	-0.153	0.800	0.538	-0.538	-0.205	-0.150	0.371	-0.682	0.391	0.285	-1.184
Benchmark Median	-0.022	0.077	-0.150	0.773	0.477	-0.571	-0.277	-0.130	0.362	-0.556	0.411	0.339	-1.138
RBS Global B&M	-0.017	0.087	-0.120	0.854	0.550	-0.513	-0.245	-0.179	0.165	-0.477	0.337	0.304	-1.145
Liverpool Macro Res.	-0.014	0.087	-0.127	0.113	0.346	-0.574	-0.723	-0.035	0.871	-0.343	0.695	0.640	-1.147
Lehman Brothers	-0.011	0.078	-0.420	1.004	0.400	-0.796	-0.536	-0.254	0.389	-0.338	0.750	0.443	-1.054
ITEM Club	-0.007	0.100	-0.167	0.921	0.504	-0.409	-0.019	-0.070	0.243	-0.741	0.386	0.185	-1.261
Abn Amro	-0.004	0.104	-0.179	0.648	0.300	-0.563	-0.252	0.012	0.332	-0.302	0.695	0.299	-1.298
Experian Business Str.	-0.001	0.098	0.120	0.921	0.483	-0.550	-0.245	-0.140	0.457	-0.650	0.216	0.338	-1.116
ING Financial Markets	0.005	0.144					-0.038	-0.167	0.233	-0.568	0.750	0.520	-0.811
Primark WEFA	0.010	0.010	-0.213	0.575	0.433	-0.725	-0.700						
Benchmark Average	0.012	0.110	-0.126	0.782	0.482	-0.587	-0.244	-0.064	0.403	-0.512	0.470	0.412	-1.092
UBS	0.013	0.114	0.120	0.908	0.387	-0.525	-0.264	-0.003	0.495	-0.579	0.331	0.267	-1.117
Moody's Economy	0.013	0.208						0.378	0.130	-0.622	0.607	0.645	-0.983
Hermes	0.014	0.105	-0.213	0.571	0.329	-0.825	-0.391	-0.105	0.633	-0.059	0.433	0.670	-1.011
OECD	0.021	0.151	0.147	0.808	0.913	-0.450	-0.167	-0.211	0.363	-0.632	0.389	0.235	-1.363
Capital Economics	0.026	0.167			0.300	-0.247	-0.051	0.020	0.500	-0.223	0.609	0.465	-1.079
WestLB	0.028	0.028	-0.260	0.908	0.012	-1.142	-0.295	-0.125	0.329	-0.414	0.648	0.900	
Cambridge Econometrics	0.056	0.168	-0.160	0.946	0.822	-0.533	0.123	-0.195	0.390	-0.824	0.591	0.325	-1.261
Dresdner Kleinwort W.	0.087	0.158	0.167	0.858	0.650	-0.708	-0.359	-0.267	0.138	-0.241	0.757	0.722	-1.017
Henley	0.098	0.098	-0.113	1.091	0.625	-0.750	-0.900						
Societe Generale	0.098	0.098	-0.220	0.977	0.613	-0.525	-0.458	-0.100					
Bank of America	0.135	0.367					0.414	0.174	0.727	-0.166	0.643	0.393	-1.186
EIU	0.170	0.270	0.107	0.867	0.654	-0.379	-0.155	0.225	0.624	-0.315	0.809	0.533	-1.214
CEBR	0.184	0.286	0.340	1.046	0.900	-0.580	-0.044	0.282	0.834	-0.744	0.222	0.465	-0.926
HSBC	0.194	0.286	-0.207	0.825	0.504	-0.317	-0.136	-0.050	0.824	-0.095	0.836	0.550	-0.832
Credit Lyonnais	0.218	0.218	-0.240	0.875	0.742	-0.350	-0.114	-0.030	0.453				
Bridgewell	0.230	0.230					-0.033	0.638	0.705	-0.418	0.160	0.356	
Schroders	0.243	0.243	-0.020	1.100	0.308	-0.661	-0.094	0.265	0.563	-0.566	0.855	0.600	
Greenwich Natwest	0.302	0.302	-0.150	0.630	0.374	-0.417							
Chase Manhattan	0.327	0.327	0.280	0.865	0.733	-1.013							
Norwich Union IM	0.333	0.333	-0.360	0.704	0.500								
Barclays Bank	0.335	0.335	-0.327	1.171	0.754	-0.214	-0.400						
Economic Perspectives	1.262	1.323	0.287	1.333	1.458	0.267	1.391	1.860	1.857	0.923	1.886	1.874	0.568

Figure 4.2: Mean forecast error

4.3.2 Mean forecast error

The mean forecast error (MFE) is colour-coded in the same way as "the blur" from section 2. Just that the lower bound of -2 is colour red and the upper bound of $+2$ is coloured blue as the two upper right cells indicate.

For around two thirds of forecasters the forecasts are higher than the actuals, which we can see with the help of the negative sign of the mean forecast errors. This corresponds to an upward bias. The other third shows a downward bias.

With the exception of the two values which show the largest upward respectively downward bias, all forecasters lie within a range of -0.3 to $+0.3$, which we think is quite remarkable. Even further, 28 forecasters lie in the range from -0.1 to $+0.1$ MFEs, which actually can be interpreted as having no bias at all. The highest upward bias by quite a margin is produced by *MacroEcon.com* with bias of almost half a percentage point. On the other end with a downward bias of full 1.2 percentage points, Economic Perspectives gets the title of the "heavy outlier", which is defended in almost all observed years remarkably well. Looking back to figure 3.1 "the blur" at the beginning of this work the observable blue stripe corresponds to that agency. Also note, that many of those who have a relatively high or a relatively low MFE, often have a very small number of underlying counts.

If we compare the second and the third column, we must come to the conclusion that the year of the crisis 2008 on average decreased the general upward bias by 0.15 percentage points. Thus, excluding the crisis year 2008, the pace of the economy was in general underestimated, which makes the collectivity of forecasters a slight pessimist.

As also mentioned above, the vertical columns for the single years show the systematic risk of all forecasters having the same underlying information set. This results in a yearly upward or downward bias.

MEAN SQUARED FORECAST ERROR (MSFE)

FORECASTER	1998-2008	1998 - 2007	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Williams de Broe	0.231	0.231	0.102	0.440	0.115	0.351	0.160	0.144	0.144	0.535	0.049	0.023	
Greenwich Natwest	0.302	0.302	0.105	0.492	0.225	0.175							
Fortis Bank	0.356	0.255				0.240	0.299	0.411	0.031	0.529	0.088	0.090	1.960
Capital Economics	0.386	0.255			0.090	0.113	0.214	0.152	0.422	0.170	0.414	0.306	1.416
Lombard Street	0.398	0.369	0.459	0.254	0.105	0.694	0.212	0.203	0.185	0.797	0.101	0.793	0.714
Primark WEFA	0.414	0.414	0.055	0.679	0.254	0.528	0.490						
Bridgewell	0.420	0.420					0.010	0.922	0.834	0.255	0.033	0.129	
Daiwa Institute of Res.	0.424	0.323	0.046	0.893	0.356	0.373	0.220	0.240	0.122	0.504	0.161	0.136	1.552
Benchmark Average	0.427	0.339	0.034	0.862	0.478	0.377	0.233	0.148	0.282	0.330	0.236	0.214	1.427
Benchmark Median	0.437	0.335	0.043	0.866	0.446	0.361	0.273	0.173	0.235	0.426	0.184	0.153	1.590
RBS Global B&M	0.440	0.331	0.057	0.892	0.549	0.330	0.355	0.223	0.031	0.371	0.134	0.126	1.628
Credit Lyonnais	0.441	0.441	0.081	1.025	0.913	0.140	0.258	0.166	0.273				
Moody's Economy	0.452	0.350						0.151	0.129	0.491	0.392	0.468	0.971
OEF	0.456	0.338	0.068	1.038	0.462	0.366	0.240	0.233	0.042	0.478	0.190	0.119	1.771
Abn Amro	0.460	0.329	0.036	0.699	0.249	0.370	0.275	0.304	0.267	0.146	0.579	0.233	2.036
Hermes	0.465	0.395	0.075	0.445	0.256	0.841	0.405	0.119	0.677	0.087	0.281	0.640	1.249
Standard Chartered	0.467	0.322			0.085	0.466	0.319	0.036	0.430	0.545	0.367	0.162	1.664
Experian Business Str.	0.471	0.378	0.024	1.075	0.425	0.361	0.392	0.172	0.288	0.570	0.085	0.169	1.514
HSBC	0.481	0.447	0.109	0.992	0.415	0.144	0.104	0.108	1.187	0.029	0.810	0.447	0.867
Merrill Lynch	0.488	0.488	0.464	0.606	0.176	0.388	0.639	1.000					
CBI	0.488	0.381	0.061	1.027	0.564	0.358	0.262	0.142	0.280	0.590	0.187	0.133	1.681
Goldman Sachs	0.488	0.394	0.058	1.393	0.337	0.239	0.635	0.312	0.048	0.454	0.088	0.154	1.534
UBS	0.493	0.392	0.133	1.182	0.265	0.333	0.445	0.123	0.539	0.498	0.121	0.098	1.624
EIU	0.494	0.396	0.020	1.048	0.721	0.168	0.160	0.080	0.471	0.129	0.752	0.313	1.847
Global Insight	0.495	0.412	0.036	0.619	0.881	0.725	0.318	0.326	0.144	0.513	0.219	0.128	1.417
Norwich Union IM	0.498	0.498	0.224	0.796	0.250								
Credit Suisse	0.500	0.338	0.038	0.801	0.400	0.310	0.437	0.162	0.081	0.739	0.123	0.053	2.228
NIESR	0.504	0.369	0.097	0.587	0.190	0.956	0.600	0.154	0.198	0.569	0.103	0.082	2.008
EC	0.513	0.391	0.063	0.480	0.367	0.970	0.590	0.306	0.171	0.734	0.071	0.055	1.918
Morgan Stanley	0.522	0.453	0.477	1.246	0.361	0.451	0.581	0.508	0.285	0.200	0.122	0.227	1.292
ITEM Club	0.523	0.407	0.061	1.233	0.789	0.252	0.045	0.205	0.182	0.799	0.166	0.069	1.877
IMF	0.528	0.349	0.081	1.683	0.469	0.185	0.335	0.184	0.378	0.318	0.106	0.052	2.341
ING Financial Markets	0.531	0.485					0.029	0.596	0.180	0.707	0.693	0.416	0.797
Dresdner Kleinwort W.	0.539	0.496	0.169	0.815	0.788	0.612	0.536	0.142	0.095	0.109	0.661	0.654	1.202
Liverpool Macro Res.	0.546	0.460	0.038	0.557	0.208	0.529	0.808	0.166	0.947	0.174	0.606	0.428	1.499
Barclays Capital	0.558	0.424	0.046	1.137	0.461	0.610	0.360	0.211	0.058	1.040	0.017	0.024	2.059
OECD	0.575	0.426	0.135	1.017	1.235	0.210	0.181	0.159	0.189	0.502	0.153	0.072	2.163
Schroders	0.595	0.595	0.046	1.591	0.753	0.684	0.145	0.149	0.463	0.347	0.782	0.360	
Citigroup	0.599	0.491	0.243	0.983	0.232	0.781	0.396	0.576	0.118	1.117	0.140	0.103	1.872
Cambridge Econometrics	0.605	0.498	0.072	1.152	1.306	0.377	0.065	0.243	0.216	0.776	0.373	0.121	1.874
Bank of America	0.618	0.419					0.190	0.101	1.170	0.083	0.502	0.244	1.745
Deutsche Bank	0.619	0.544	0.175	1.943	0.534	0.685	0.203	0.268	0.219	0.429	0.583	0.119	1.357
Charterhouse	0.651	0.651	0.513	0.435	1.139	0.689	0.326	0.250					
Societe Generale	0.688	0.688	0.098	1.172	1.186	0.383	0.602	0.010					
J P Morgan	0.697	0.592	0.225	2.411	0.331	0.675	0.506	0.417	0.154	0.575	0.145	0.029	2.697
Lehman Brothers	0.705	0.647	0.279	1.518	0.836	0.717	0.840	0.487	0.315	0.231	0.683	0.323	1.392
CEBR	0.721	0.691	0.331	1.538	1.658	0.483	0.174	0.177	1.107	0.667	0.062	0.296	1.056
Barclays Bank	0.739	0.739	0.171	1.645	1.074	0.060	0.160						
WestLB	0.756	0.756	0.147	2.049	0.560	1.852	0.334	0.326	0.227	0.299	0.486	0.841	
Henley	0.837	0.837	0.038	1.651	0.685	0.665	0.925						
Chase Manhattan	0.860	0.860	0.219	1.114	0.903	1.031							
MacroEcon.com	0.932	0.296							0.047	0.876	0.122	0.109	3.236
Economic Perspectives	2.533	2.711	0.106	2.403	2.702	0.097	1.977	3.915	5.450	1.160	4.325	4.956	0.538

Figure 4.3: Mean squared forecast error

4.3.3 Mean squared forecast error

The mean squared forecast error is one of the two accuracy criterions used in this work and is presented in equation 4.3. Figure 4.3 states the calculated values of the MSFE for the different agencies in the according years starting with the best forecaster. Figure 9.2 shows almost the same with the small difference that the value of the MSFE is shown but its corresponding rank in the table. Since the ranks refer only to the very same column in which they are noted, it is easy to tell for each of the stated periods of time who according to the MSFE criterion was the best forecaster. Thus, figure 4.3 and 9.2 provide a first answer to the initial question. The corresponding colour-code was assigned in the way that a MSFE of 0 corresponds to green and a MSFE of 6 corresponds to red.

Williams de Broe leads the tables as best forecaster in terms of the MSFE. They remain best even when we look at the subset which excludes the year 2008. In this 10-year period the other forecasters are closer to the MSFE value of *Williams de Broe*. In other words, they could improve their lead by simply not forecasting in 2008. That reflects the above foretold point that a forecaster can improve their performance by not forecasting in difficult years like in 2008. Although *Williams de Broe* illustrates the argument quite clearly, this fact does not imply the imputation that they actually did so in order to achieve an advantage. Moreover, it shall show that the analysis is quite sensitive to the years included. So, when compressing information in a single index like the MSFE, other mechanisms might be at work as well. This is exactly why figure 4.3 also presents the forecasters' performance per year.

Among the next seven best there are five who did not engage in forecasting for all years. Only *Lombard Street* and *Daiwa Institute of Research* did so over the whole period of time. In this sense, these two forecasters are the only ones who beat as an individual forecasting agency the benchmark, i.e. the simple average over all forecasters. This is quite striking, especially because in the literature they refer to a "hard to beat" benchmark. On the other side one could say that among 51 forecasting agencies there might be

MEAN ABSOLUTE FORECAST ERROR (MAFE)

FORECASTER	1998-2008	1998 - 2007	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Williams de Broe	0.389	0.389	0.287	0.571	0.204	0.567	0.368	0.320	0.300	0.677	0.194	0.143	
Greenwich Natwest	0.457	0.457	0.250	0.630	0.374	0.417							
Fortis Bank	0.472	0.414				0.450	0.389	0.542	0.156	0.716	0.295	0.300	1.400
Bridgewell	0.491	0.491					0.100	0.885	0.714	0.455	0.160	0.356	
Daiwa Institute of Res.	0.497	0.442	0.140	0.783	0.430	0.548	0.405	0.415	0.243	0.627	0.373	0.300	1.105
Lombard Street	0.501	0.475	0.500	0.413	0.217	0.763	0.391	0.400	0.355	0.850	0.245	0.690	0.795
Capital Economics	0.506	0.433			0.300	0.247	0.440	0.380	0.538	0.341	0.609	0.465	1.079
RBS Global B&M	0.510	0.452	0.160	0.854	0.550	0.513	0.482	0.411	0.165	0.535	0.337	0.317	1.145
OEF	0.512	0.451	0.187	0.845	0.483	0.563	0.382	0.430	0.176	0.625	0.401	0.291	1.194
Abn Amro	0.513	0.448	0.179	0.648	0.342	0.563	0.376	0.494	0.426	0.302	0.695	0.351	1.298
Credit Lyonnais	0.521	0.521	0.240	0.875	0.742	0.350	0.450	0.390	0.453				
Goldman Sachs	0.523	0.472	0.153	1.025	0.450	0.450	0.615	0.500	0.160	0.605	0.259	0.335	1.079
HSBC	0.524	0.497	0.313	0.825	0.504	0.317	0.255	0.300	0.862	0.132	0.836	0.561	0.832
Primark WEFA	0.533	0.533	0.213	0.617	0.433	0.725	0.700						
ITEM Club	0.534	0.472	0.167	0.921	0.661	0.435	0.190	0.410	0.357	0.795	0.386	0.205	1.261
Standard Chartered	0.537	0.478			0.290	0.545	0.541	0.135	0.524	0.691	0.586	0.374	1.021
NIESR	0.540	0.471	0.273	0.678	0.346	0.804	0.566	0.315	0.381	0.686	0.277	0.270	1.316
UBS	0.542	0.490	0.293	0.917	0.387	0.525	0.564	0.310	0.544	0.622	0.331	0.277	1.117
CBI	0.542	0.484	0.193	0.826	0.538	0.538	0.455	0.320	0.429	0.700	0.391	0.295	1.184
Experian Business Str.	0.542	0.491	0.120	0.921	0.508	0.550	0.527	0.340	0.457	0.687	0.247	0.362	1.116
EIU	0.544	0.495	0.133	0.867	0.654	0.379	0.364	0.225	0.624	0.355	0.809	0.533	1.214
Hermes	0.545	0.503	0.213	0.571	0.388	0.825	0.555	0.315	0.652	0.277	0.433	0.690	1.011
IMF	0.548	0.464	0.253	1.286	0.550	0.310	0.472	0.388	0.565	0.500	0.295	0.205	1.399
Barclays Capital	0.550	0.484	0.180	0.925	0.504	0.754	0.505	0.379	0.195	0.956	0.106	0.115	1.294
Benchmark Median	0.551	0.499	0.243	0.810	0.531	0.571	0.473	0.408	0.422	0.624	0.411	0.357	1.138
Global Insight	0.555	0.511	0.160	0.700	0.655	0.763	0.464	0.490	0.314	0.633	0.436	0.320	1.042
Credit Suisse	0.556	0.483	0.169	0.805	0.554	0.550	0.559	0.330	0.257	0.833	0.325	0.189	1.342
Norwich Union IM	0.561	0.561	0.360	0.722	0.500								
EC	0.562	0.499	0.227	0.571	0.508	0.967	0.600	0.425	0.371	0.830	0.181	0.220	1.289
Morgan Stanley	0.572	0.531	0.540	0.938	0.446	0.521	0.632	0.638	0.419	0.406	0.313	0.441	1.028
OECD	0.574	0.500	0.280	0.808	0.913	0.450	0.300	0.322	0.363	0.668	0.389	0.245	1.363
Benchmark Average	0.588	0.541	0.286	0.847	0.572	0.602	0.505	0.456	0.474	0.610	0.484	0.432	1.126
Bank of America	0.589	0.484					0.414	0.292	0.792	0.265	0.643	0.401	1.186
Merrill Lynch	0.596	0.596	0.547	0.692	0.304	0.617	0.673	1.000					
Citigroup	0.606	0.556	0.460	0.829	0.404	0.829	0.509	0.615	0.271	0.936	0.318	0.254	1.183
Cambridge Econometrics	0.609	0.554	0.240	0.946	0.822	0.533	0.214	0.435	0.410	0.852	0.591	0.325	1.261
ING Financial Markets	0.609	0.575					0.138	0.689	0.329	0.714	0.750	0.560	0.811
J P Morgan	0.612	0.573	0.393	1.355	0.392	0.754	0.565	0.562	0.343	0.655	0.309	0.150	1.358
Moody's Economy	0.613	0.540						0.378	0.330	0.641	0.607	0.645	0.983
Societe Generale	0.614	0.614	0.220	0.977	0.713	0.525	0.679	0.100					
Dresdner Kleinwort W.	0.616	0.590	0.380	0.858	0.650	0.708	0.659	0.333	0.231	0.323	0.757	0.722	1.017
Liverpool Macro Res.	0.620	0.573	0.180	0.696	0.396	0.574	0.768	0.385	0.890	0.381	0.695	0.640	1.147
Barclays Bank	0.627	0.627	0.327	1.171	0.754	0.214	0.400						
Schroders	0.631	0.631	0.167	1.100	0.708	0.661	0.339	0.365	0.563	0.566	0.855	0.600	
Deutsche Bank	0.636	0.590	0.417	1.172	0.521	0.713	0.400	0.455	0.386	0.630	0.743	0.264	1.089
CEBR	0.651	0.625	0.433	1.046	0.942	0.580	0.377	0.312	0.856	0.801	0.233	0.485	0.937
Lehman Brothers	0.668	0.636	0.420	1.004	0.625	0.796	0.700	0.576	0.444	0.439	0.750	0.474	1.054
Charterhouse	0.669	0.669	0.693	0.467	0.825	0.783	0.550	0.500					
WestLB	0.680	0.680	0.367	1.133	0.613	1.142	0.473	0.525	0.367	0.505	0.648	0.900	
MacroEcon.com	0.683	0.414							0.197	0.811	0.284	0.307	1.657
Henley	0.720	0.720	0.113	1.091	0.625	0.750	0.900						
Chase Manhattan	0.768	0.768	0.413	0.865	0.733	1.013							
Economic Perspectives	1.269	1.329	0.313	1.333	1.458	0.292	1.391	1.860	1.857	0.923	1.886	1.884	0.589

Figure 4.4: Mean absolute forecast error

one or two models in action who beat the average forecast. But of course the next 20 forecasters are in very short range to the benchmark. This becomes even clearer when we compare the benchmark's MSFE with its MAFE.

The same table lists at its very bottom - not very surprisingly - *Economic Perspectives* again . Their MSFE is by far very bad when compared over the entire time of 11 years. As already indicated above, we assume some sort of reason for that behaviour, like they try to niche the forecasting market or try to maximize publicity since it must be quite hard constantly not to learn from earlier mistakes.[2]

On the other hand, *Economic Perspectives* does provide something like a lower bound to the realised outcome. For both years of strong general upward bias, 2001 and 2008, *Economic Perspectives* still expects, though only by a bit, a weaker economy than they predicted by their figures. And that is still true in the case of 2008, which marks the worst economic crisis in the last eight decades. So we could argue that Economic Perspectives provides figures for the worst-case scenario. This is also very valuable information in today's times. But then this aim should be proclaimed as well.

MacroEcon.com is a good example for how a forecaster's performance is ruined by forecasting badly just once, like they did for 2008. But this must also be mitigated in parts: their one-time mistake can not be absorbed very well by other years, because they started forecasting as recently as 2004.

1998 and 2007 are those years which are colour-coded the most in green, indicating that those were the best forecasted years. This is actually only true for the year 2007, since for 1998 the first nine months of the 24-month period, which by far are those that contribute most "mass" to the MSFE, are not included in the data set.

And of course, the worst year - obviously not only in terms of forecasting - was 2008 when the crisis took us by surprise.

Table 9.2 shows the MSFE as well, just in rankings of it. The only thing to mention now is that another three columns were included. The col-

[2]Apparently they use a two-scenario composite model in which one scenario captures a credit market crash.

umn labelled "Min" shows the particular forecaster's best reached position. "Max" on the other hand reports the worst position taken by that forecaster. Quite interesting is also the third column "Spread", because it indicates the stability or instability of good as well as bad forecasts.

Further information can be drawn from this kind of presentation. For instance, *Daiwa Institute of Research* is among the best forecasters because they produce the most stable forecasts. Over the whole period they only rank between the 10th and the 29th position. *CBI's* spread ranks 2nd best, thus making it not only a good forecaster but also a very stable one: only positions from 9 until 32. On the other hand *Chase Manhattan* provides (on a four year basis) stable but comparatively bad forecasts (positions between 27 and 45).

The next section analyse the mean absolute forecast error (MAFE). The structure of the corresponding tables is kept exactly the same.

4.3.4 Mean absolute forecast error

The mean absolute forecast error is the second applied accuracy criterion. The set-up in the two following table 4.4 as well as the colour-coding is all the same, except that the highest value of 2 is coloured in red. In figure 9.3 the forecasters' rankings according to the MAPE criterion are displayed.

As already mentioned in section two, MAFE and MSFE are obviously different, but show almost the same results. So instead of repeating most of the findings for the MSFE, we will go on with the comparison of the two accuracy criteria. For this, see figure 4.5.

4.4 Comparison of MSFE and MAFE

Please note that the second column reports MSFE and the third one MAFE. Also remember that in the introduction to this section we explained that the MSFE punishes a few higher forecast errors much more than a forecaster who produced a lot of small forecasts.

With that in mind we can analyse table 4.5. There are three different topics to address.

Firstly, the forecasters who do average under the MSFE criterion but do much better in the ranking when analysed by MAFE are *Goldman Sachs*, *NIESR*, *ITEM Club*, and *IMF*. A possible explanation is that the last three of them did quite badly in the 2008 forecasts. And since outliers get over proportionally punished under the MSFE, it is them who win in positions.

Secondly, *Merrill Lynch* and *Moody's Economy* do much better in the MSFE ranking then in the MAFE. Also here a possible explanation might be that they did very well in the years when most others did collectively bad. So while the others did over proportionally collected "mass" to their MSFE ranking, they did not.

Thirdly, the benchmark drops quite badly from the 9th rank under MSFE to a 32nd rank under MAFE. This finding, unpleasant as it is, should actually be expected: since the benchmark is an average over 51 forecasters, there are no very high or very low values of forecast errors in the data left, just because they became neutralized by the fact that there are so many (51) who help bearing the burden. Hence, extreme values which become more punished under MSFE are, if at all, quite rare. The benchmark is a very good example to illustrate that many but small mistakes are rewarded more under the MSFE criterion than a few but huge mistakes.

4.5 Comparison of quarterly MSFE

Figure 4.5below shows the three-month average MSFE by forecaster over all eight quarters of the two years. This is in contrast to the above tables which take the average MSFE by years. It is very clear that *Williams de Broe* in general predicts very well the true outcome already early in the forecasting period. In other words, their prediction of the true value of GDP growth is already in the first quarter quite precise on an average basis. The other green colour-coded forecasters *Bridgewell, Greenwich Natwest, Norwich Union IM* and *Primark WEFA* only appear to be good forecasters. When seeing the

FORECASTER	MSFE	MAFE	MSFE	MAFE	Δ
Williams de Broe	1	1	0.231	0.389	0
Greenwich Natwest	2	2	0.302	0.457	0
Fortis Bank	3	3	0.356	0.472	0
Capital Economics	4	7	0.386	0.506	-3
Lombard Street	5	6	0.398	0.501	-1
Primark WEFA	6	14	0.414	0.533	-8
Bridgewell	7	4	0.420	0.491	3
Daiwa Institute of Res.	8	5	0.424	0.497	3
Benchmark Average	9	32	0.427	0.588	-23
Benchmark Median	10	25	0.437	0.551	-15
RBS Global B&M	11	8	0.440	0.510	3
Credit Lyonnais	12	11	0.441	0.521	1
Moody's Economy	13	39	0.452	0.613	-26
OEF	14	9	0.456	0.512	5
Abn Amro	15	10	0.460	0.513	5
Hermes	16	22	0.465	0.545	-6
Standard Chartered	17	16	0.467	0.537	1
Experian Business Str.	18	20	0.471	0.542	-2
HSBC	19	13	0.481	0.524	6
Merrill Lynch	20	34	0.488	0.596	-14
CBI	21	19	0.488	0.542	2
Goldman Sachs	22	12	0.488	0.523	10
UBS	23	18	0.493	0.542	5
EIU	24	21	0.494	0.544	3
Global Insight	25	26	0.495	0.555	-1
Norwich Union IM	26	28	0.498	0.561	-2
Credit Suisse	27	27	0.500	0.556	0

FORECASTER	MSFE	MAFE	MSFE	MAFE	Δ
[continued]					
NIESR	28	17	0.504	0.540	11
EC	29	29	0.513	0.562	0
Morgan Stanley	30	30	0.522	0.572	0
ITEM Club	31	15	0.523	0.534	16
IMF	32	23	0.528	0.548	9
ING Financial Markets	33	37	0.531	0.609	-4
Dresdner Kleinwort W.	34	41	0.539	0.616	-7
Liverpool Macro Res.	35	42	0.546	0.620	-7
Barclays Capital	36	24	0.558	0.550	12
OECD	37	31	0.575	0.574	6
Schroders	38	44	0.595	0.631	-6
Citigroup	39	35	0.599	0.606	4
Cambridge Econometrics	40	36	0.605	0.609	4
Bank of America	41	33	0.618	0.589	8
Deutsche Bank	42	45	0.619	0.636	-3
Charterhouse	43	48	0.651	0.669	-5
Societe Generale	44	40	0.688	0.614	4
J P Morgan	45	38	0.697	0.612	7
Lehman Brothers	46	47	0.705	0.668	-1
CEBR	47	46	0.721	0.651	1
Barclays Bank	48	43	0.739	0.627	5
WestLB	49	49	0.756	0.680	0
Henley	50	51	0.837	0.720	-1
Chase Manhattan	51	52	0.860	0.768	-1
MacroEcon.com	52	50	0.932	0.683	2
Economic Perspectives	53	53	2.533	1.269	0

Figure 4.5: Comparison of MSFE and MAFE

low number of their forecasts especially in January to March of the year-ahead forecast, see figure 4.1 "counts", one has to take into account two other possibilities of their seemingly good performance. First, they could have been just lucky or they have presented their forecasts in a very stable economic environment. This does not mean that they were not good forecasters, but that other influences may have pushed their rank in this analysis.

	Average MSFE by quarter of years								Rankings of Average MSFE by quarter of years							
From month	24	21	18	15	12	9	6	3	24	21	18	15	12	9	6	3
To month	22	19	16	13	10	7	4	1	22	19	16	13	10	7	4	1
Williams de Broe	0.24	0.15	0.30	0.30	0.43	0.24	0.18	0.13	2	2	3	3	3	2	13	39
Greenwich Natwest	0.41	0.29	0.20	0.28	0.35	0.40	0.35	0.19	5	3	1	2	1	28	43	46
Fortis Bank	1.41	0.48	0.52	0.21	1.12	0.39	0.39	0.10	48	7	8	1	48	22	49	27
Capital Economics	0.89	0.72	0.68	0.57	0.96	0.25	0.15	0.07	18	14	20	16	35	3	3	8
Lombard Street	0.65	0.46	0.42	0.74	0.82	0.37	0.21	0.09	7	6	4	41	16	18	22	24
Primark WEFA	0.34	0.34	0.44	0.42	0.44	0.57	0.59	0.14	3	4	5	5	4	45	52	42
Bridgewell	0.38	0.36	0.66	0.51	0.78	0.54	0.28	0.07	4	5	16	9	13	42	34	13
Daiwa Institute of Research	0.78	0.77	0.67	0.52	0.87	0.34	0.18	0.07	11	21	18	12	24	9	14	6
Benchmark Mean	**0.80**	**0.70**	**0.64**	**0.51**	**0.82**	**0.35**	**0.20**	**0.07**	**16**	**13**	**13**	**10**	**22**	**13**	**18**	**12**
Benchmark Median	**0.81**	**0.73**	**0.69**	**0.53**	**0.85**	**0.37**	**0.19**	**0.06**	**18**	**19**	**24**	**14**	**27**	**17**	**17**	**3**
RBS Global Banking & Markets	1.04	0.74	0.65	0.44	0.85	0.40	0.18	0.07	32	19	14	7	21	27	12	11
Credit Lyonnais	0.86	0.70	0.48	0.36	0.51	0.36	0.23	0.14	16	12	6	4	5	15	24	41
Moody's Economy	0.99	0.64	0.61	0.65	0.96	0.40	0.29	0.22	28	10	9	27	37	25	35	48
OEF	0.86	0.74	0.66	0.58	0.98	0.41	0.17	0.09	15	18	17	18	39	31	9	19
Abn Amro	0.90	0.97	0.70	0.43	0.87	0.36	0.25	0.06	21	35	22	6	23	14	29	2
Hermes	0.72	0.76	0.76	0.67	0.79	0.39	0.15	0.07	8	20	27	33	14	24	5	7
Standard Chartered	1.08	0.99	0.90	0.68	0.77	0.17	0.17	0.09	37	38	43	35	11	1	8	20
Experian Business Strategies	0.76	0.73	0.74	0.61	0.88	0.38	0.24	0.08	10	16	25	21	25	19	27	14
HSBC	0.72	0.80	0.74	0.74	0.94	0.25	0.21	0.08	9	24	24	42	31	4	20	17
Merrill Lynch	0.53	0.56	0.64	0.72	0.56	0.35	0.24	0.17	6	9	11	40	6	10	25	45
CBI	0.93	0.79	0.73	0.61	0.89	0.41	0.24	0.10	25	22	23	20	28	30	26	29
Goldman Sachs	0.90	0.81	0.94	0.62	0.94	0.35	0.17	0.07	20	25	45	24	32	11	7	10
UBS	0.86	0.91	0.87	0.63	0.80	0.29	0.22	0.09	18	33	40	26	16	7	23	18
EIU	0.91	0.54	0.65	0.66	0.84	0.56	0.34	0.16	23	8	15	30	19	43	39	43
Global Insight	0.98	1.04	0.88	0.49	0.76	0.27	0.19	0.13	27	43	40	8	8	5	16	38
Norwich Union IM	0.13	0.13	0.21	0.63	0.85	0.81	0.72	0.53	1	1	2	26	20	50	53	54
Credit Suisse	0.93	0.92	0.85	0.59	0.82	0.38	0.32	0.11	26	34	38	19	17	20	37	30
NIESR	1.00	0.68	0.83	0.75	0.91	0.35	0.21	0.12	29	11	35	45	30	12	19	34
EC	0.89	0.72	0.82	0.66	0.88	0.41	0.34	0.19	19	15	33	31	27	29	40	47
Morgan Stanley	1.06	0.80	0.64	0.70	0.96	0.40	0.21	0.09	36	23	12	39	36	26	21	22
ITEM Club	1.06	0.86	0.82	0.62	0.95	0.45	0.25	0.11	35	29	34	23	33	33	28	32
IMF	1.82	0.82	0.80	0.65	0.98	0.39	0.36	0.24	52	27	32	28	41	23	45	51
ING Financial Markets	1.41	1.10	1.03	0.66	1.08	0.30	0.12	0.04	47	47	47	32	45	8	1	1
Dresdner Kleinwort Wasserstein	1.02	0.95	0.89	0.75	0.78	0.36	0.17	0.13	30	35	41	44	12	16	11	36
Liverpool Macro Research	1.02	0.98	0.78	0.57	0.77	0.46	0.29	0.14	31	37	30	17	10	34	36	40
Barclays Capital	0.91	0.88	0.90	0.70	0.95	0.49	0.26	0.08	24	30	42	36	34	38	31	16
OECD	1.04	0.99	0.78	0.67	0.88	0.43	0.39	0.29	34	40	28	34	26	32	50	53
Schroders	1.10	0.82	0.90	0.61	0.77	0.46	0.28	0.10	40	26	44	22	9	35	32	26
Citigroup	1.08	0.99	1.09	0.76	0.98	0.49	0.17	0.07	38	41	49	47	40	37	10	4
Cambridge Econometrics	0.90	0.92	0.75	0.76	0.97	0.52	0.37	0.23	22	33	26	46	38	39	47	49
Bank of America	1.52	1.48	1.22	0.77	1.07	0.29	0.15	0.07	48	52	50	47	43	6	4	5
Deutsche Bank	1.17	0.99	0.78	0.75	1.17	0.54	0.34	0.09	42	39	29	43	51	41	38	21
Charterhouse	0.78	1.22	1.09	0.52	0.38	0.53	0.35	0.24	12	51	48	11	2	40	41	50
Societe Generale	1.65	1.10	0.51	0.53	0.68	0.57	0.36	0.11	50	46	7	14	7	47	44	33
J P Morgan	1.26	1.07	1.12	0.91	1.00	0.39	0.12	0.07	44	45	50	50	42	21	2	9
Lehman Brothers	1.30	1.14	1.01	0.84	1.12	0.56	0.16	0.08	45	49	46	49	49	44	6	15
CEBR	1.93	1.04	0.78	0.66	1.12	0.60	0.38	0.10	53	44	31	29	47	48	48	25
Barclays Bank	1.11	0.84	0.67	0.70	1.18	0.85	0.37	0.16	41	28	19	38	52	52	46	44
WestLB	1.25	1.02	0.84	0.99	1.11	0.57	0.19	0.12	45	43	38	54	48	48	15	35
Henley	1.34	1.10	0.63	0.56	1.15	1.06	0.49	0.10	46	48	10	15	50	53	51	28
Chase Manhattan	1.10	1.31	1.43	0.98	1.03	0.77	0.26	0.09	39	52	52	51	43	49	30	23
MacroEcon.com	1.69	1.21	1.49	1.81	1.99	0.83	0.35	0.11	51	50	53	53	54	51	42	31
Economic Perspectives	3.77	4.51	4.64	3.70	1.56	1.49	0.94	0.28	54	54	54	54	53	54	54	52
Average	1.04	0.89	0.84	0.70	0.90	0.47	0.28	0.13								

Figure 4.6: Mean squared forecast error by quarters of a year

Chapter 5

Introduction to forecast combination

5.1 What is forecast combination?

Most of the times finding out who is the best forecaster is not straight forward. Experience over time is one possibility to determine which forecaster is the more reliable one. But if the pool of agencies to choose from grows bigger in number or the performances of individual forecasters are not stable enough over time to support the decision-making process, experience very fast turns out to be no possibility any more.

Instead of relying on a single opinion about the future, forecast combination - as the name says - combines multiple opinions into one which, as the empiric evidence often shows, provides a very good support to the decision makers compared to just relying on individual forecasts. Thus, forecast combination is a tool which helps decision makers assess multiple forecasts of the same variable.

A first combination method was already introduced in the first part of this study. The "Benchmark Average" in the tables of the last chapter is per se a forecast combination since it was calculated by combining all forecasts with equal weight; here the emphasis lies on the words *all* and *equal*.

However, forecast combination is more than just averaging over all possibilities. It rather tries to exploit the underlying information of the individual forecasters in a best way. The main idea behind that is the use of diversification gains.

Forecast combination originates from the field of financial economics. Predicting financial data as stock prices, indices or commodity prices better than potential competitors might gain a company a very powerful tool in its daily business. For this reason the literature on this topic is quite extensive. Nevertheless, the next couple of sections will provide insight into the theory and empirical evidence on forecast combinations. It follows to some extent the paper of Timmermann [20] which has in contrast to this work a rather theoretical approach to the topic.

5.2 The idea of forecast combination

As Timmermann [20] and Clements and Harvey [8], as well as many others point out, the combination of forecasts is motivated by a simple portfolio diversification argument which was first discussed in this context by Bates and Granger [6] already in 1969. "The idea is simply that the individual forecasts are each based on partial, and incompletely overlapping, information sets, as might be the case if they reflect private information"[1]. Thus, "it is not feasible to pool the underlying information sets and construct a 'super' model that nests each of the underlying forecasting models."[2]. In this sense, the underlying idea of forecast combination is to exploit the extra information of each forecast in a similar way as the portfolio diversification tries to minimize risk by spreading the investments over a number of securities.

Another commonly found reason in the literature for the strength of forecast combination methods is best summarized by Timmermann [20]: "individual forecasts may be very differently affected by structural breaks caused, for example, by institutional change or technological developments. Some

[1] Clements and Harvey [8]
[2] Timmermann [20]

models may adapt quickly and will only temporarily be affected by structural breaks, while others have parameters that only adjust very slowly to new post-break data. The more data that is available after the most recent break, the better one might expect stable, slowly adapting models to perform relative to fast adapting ones as the parameters of the former are more precisely estimated. Conversely, if the data window since the most recent break is short, the faster adapting models can be expected to produce the best forecasting performance. Since it is typically difficult to detect structural breaks in 'real time', it is plausible that on average, i.e., across periods with varying degrees of stability, combinations of forecasts from models with different degrees of adaptability will outperform forecasts from individual models. This intuition is confirmed in Pesaran and Timmermann (2005)."

A third advantage of forecast combination is that it is more robust against model misspecification and possible measurement errors in the underlying data set. By combining forecasts which were produced by different models, the risk that the chosen forecast model is misspecified is reduced. For further detail, see again the article of Timmermann [20] in which he refers to a couple of studies focusing on this issue.

So far, the theoretical considerations about the success of the combination of forecasts were discussed. But what about the empirical evidence on its success? The next paragraphs shed some light on this issue. Since there are different methods of combining forecasts (see chapter 6 for some of them), the evidence presented right below mostly refers to combination method of the simple mean. The expressions "simple mean forecaster", "average forecast" or "consensus forecast" are used here synonymously.

The evidence on the success of forecast combination is overwhelming as the following will show. The first paper to look at is Clemen [7] who provides the first summary on the literature of forecast combination. As he states in his introduction, "The results have been virtually unanimous: combining multiple forecasts leads to increased forecast accuracy. This has been the result whether the forecasts are judgemental or statistical, econometric or extrapolation. Furthermore, in many cases one can make dramatic performance improvements by simply averaging the forecasts." In his annotated

bibliography he lists more than 200 articles dealing in a broader sense with forecast combination. Among those, a lot of articles - without listing them right here - are in strong favour of an accuracy improvement achieved by combining forecasts.

Another summarizing study with a much narrower focus on empirical evidence of forecast combination was conducted by Armstrong [3] who surveyed 30 empirical studies. He states that " in [those] 30 empirical comparisons, the reduction in ex ante errors for equally weighted combined forecasts averaged about 12.5% and ranged from 3 to 24 percent. Under ideal conditions, combined forecasts were sometimes more accurate than their most accurate components." Further down in his study he rephrases that "Combined forecasts are more accurate than the typical component forecast in almost all situations studied to date. Sometimes the combined forecast will surpass the best method." He also gives an overview of the underlying factors which might modify the extent of the accuracy improvement.

As a last piece of evidence we would like to refer to section 4.3.3 of this study. In table 4.3 the benchmark which is nothing else than the simple average over all forecast is among the best. When ruling out all the forecasters who did not take part over the entire time horizon only *Lombard Street* and the *Daiwa Institute of Research* did slightly better when applying the MSFE accuracy criterion.

Most of the above mentioned improvements were achieved through simple combination methods. But there are so many different possibilities to combine multiple forecasts of the same variable with each other. Thus, a more detailed differentiation of the methods of combination is required. But this raises immediately a new question: may there be even further accuracy improvements when using more complex combination methods?

The next chapter presents different combination methods. While the first discusses methods which - strictly speaking - do not combine forecasts, section 6.3 deals with simple ways of combining forecasts and section 6.4 tells about more advanced possibilities of forecast combinations. Since almost all the studies and articles - to the knowledge of the author - focus on a particular

detail of the combination problem and since a general discussion of the topic - like in a book - is missing, it shall be noted that the discussion in the next sections is mainly rooted in the article of Timmermann [20].

Chapter 6

Forecast combination methods

The essential idea behind the different methods to be introduced now is that "they" all try to be the most accurate forecasting scheme. They differentiate from each other basically in two aspects. First, a loss function has to be specified to make the individual forecasts comparable. In chapter 4 two different loss functions were presented. The first bases on the MAFE whereas the second one bases on the MSFE. For this study, the MSFE was chosen as the criterion of accuracy. This rather arbitrary choice can be justified by the idea that "large errors are proportionally more serious than small".[1] And the second aspect is that a certain rule has to be chosen which tells about the weight of each forecast in the combination forecast. From now on this study is all about possible rules of combining. It also shall be noted right away that different rules or methods can jointly be applied at the same time. Or a certain rule can be applied first and then followed by a second one.

Before getting started with a more detailed discussion of the different schemes, some general notations should be introduced. ω represents a weight. Thus, ω_i stands for the specific weight assigned to forecaster i whereas i stands for all different forecasters, $i = \{1, 2, \ldots, N\}$ and N indicates the

[1] see Clements and Hendry [9]. Next to that, Clements and Hendry [9] justify the use of MSFE by the fact that over- and underestimation have similar costs. Moreover, they say that the "quadratic function is largely due to mathematical tractability and reflects the pervasiveness of the least-square principle in the econometrics literature."

number of forecasters in the sample. The number of months is indicated by M which mostly is equal to 24, since for every year to be forecasted there are 12 year-ahead forecasts and 12 current-year forecast.

Further the reader shall remember that in order to correctly compare the different methods of combination, the data set is usually split into two parts. The first part is used to collect information, ω_i, about the performance of the individual forecasters. Then in the second part this information is applied by producing a combined forecast, $\widehat{y}_t$, from the individual forecasts of the year in question, $f_{i,t}$, with the help of the collected information, ω_i.

For example, we want to produce a combined forecast for the year 2005. Since we know the performances of the forecasters in 2004, we apply this information by weighting the new January 2005 forecasts with the year-2004-based weights. Then for February 2005, and so on, the same is done until a complete new series of forecasts is created. Done that, the accuracy of this new combined forecast is again measured by an accuracy criterion. Since a forecast year contains of 12 year-ahead and 12 current-year ahead creating the weights and applying them is a bit confusing, a more detailed discussion is foregone.

Another important information on how the data was processed is that unlike in many other studies the data was - except for the trimming competition - analysed by years and not on a monthly basis. This means that the MSFE of a combined forecast is based on 24 forecasts which all have a different time horizon: a 24-month forecast, a 23-month forecast, ..., and finally, a one-month forecast. Since all forecasts were treated in the same way, this approach can be justified. Thus, one has to keep in mind, that months with a longer forecasting horizon will add more "mass" to the MSFE than months with a shorter forecasting horizon.

To perfectly confuse the reader we start now with a non-combination scheme which earned its entry into this comparison through its high degree of popularity.

6.1 Non-combination scheme: the naive forecast

In the literature it is often referred to the so-called naive forecast. The main difference to all the other methods here is that its next year's prediction is not based on the performance of earlier forecasters, but simply on this year's realization of the variable in question. The naive forecast basically adopts the idea of the random walk: the best guess for tomorrow is today's value plus some uncertainty about it. So basically this method looks at this year's realisation of the GDP growth rate and assumes that next year's growth rate will stay the same as equation 6.1 shows.[2] Note that no individual forecasts, $f_{i,t}$, are used to form the prediction.

$$\widehat{y}_{t+1} = y_t \tag{6.1}$$

where $\widehat{y}_{t+1}$ indicates the "combined" forecast and t represents the year. Thus, next year's forecasts is this year's actual outturn.

6.2 Quasi combination schemes

In this text quasi combination schemes are combinations of forecasts which assign to only one forecaster a weight of one and to all the other forecasters a weight of zero. Since this can be translated into simply choosing the best model or the median model for instance, and no "real" combination happens, they are called quasi combinations.

6.2.1 Last year's best

A first combination scheme is the following: the best forecaster of last year is identified and then this agency's current-year forecast is taken as the best guess. This does not sound like a very solid idea, but due to the lack of a

[2] empiric evidence and references still missing

publicly available track record we could imagine that this method of forecasting is quite common in practice. Especially since agencies might want to profit from their last year's good performance and advertise with their supposedly good forecasting skills. As the example of *Economic Perspectives* shows, the underlying reason might just be luck. It is left to the reader to compare the ranking shown in figure 9.1 with the ranking of this study in figure 9.2.[3]

6.2.2 Last years' best

A similar method is identifying the forecaster with the best track record in the last couple of years. The only difference to the before mentioned strategy is that not only last year's performance of the different forecasters is under evaluation but all years or a rolling window over the last three, five or eight years is analysed. The forecaster with the best guesses over this enlarged horizon is identified and then the forecast of this specific forecaster is used to predict next year's realisation of the GDP growth rate.

But as Timmermann puts it "choosing the single forecast with the best track record is often a bad idea". He refers to a couple of studies which support this statement. Among them are Makridakis et al [14] as early as 1982, Makridakis and Winkler [15] in 1983, Gupta and Wilton [12]in 1987. More recent studies reporting the same results are Stock and Watson [18] in 1998 as well as Hendry and Clements [13] in 2002.

6.2.3 Median forecaster

This method identifies the median forecaster and takes its forecast as the best prediction for the realization of the values in question. Stock and Watson [18] report a huge success of this group of forecast combination methods. The median forecasters performance is much better than the previous two

[3]Obviously the rankings are not constructed in the same way. However, the ranking by the Sunday Times is the only publicly available information about the performance of forecasters in the UK.

explained methods and as good as the average forecaster which is explained next, but was already used several times in this study. In the literature the method of the median forecaster is mostly labelled as a simple combination method. Since in the literature there is not yet a formal characterisation and classification established, the word "simple" here describes more the fact that it is easy to calculate rather than the affiliation to a certain group or class of methods. It was classified as a quasi-combination method since it is based unlike the average forecaster only on one individual forecast, namely the median forecaster.

6.3 Simple combination schemes

6.3.1 Average forecast

The explanation of this method is straight-forward. Take the average over all forecasts of the different agencies published in a certain month and you get the average forecast. Again, this averaging method can also be seen as a special case of weighting in which each forecaster in the sample is assigned the weight of one or when the sum of all weights shall result one, the weight of each forecaster is $1/N$ where N is the number of forecasters in the sample.

"Simple combination schemes are hard to beat"[4]. This means that this actually quite simple method of combining forecasts does often better than "more sophisticated rules relying on estimating optimal weights that [for instance] depend on the full variance-covariance matrix of forecast errors". Timmermann refers to Palm and Zellner [17] who summarize in general the advantages of a simple average forecast in the following way. First, the weights do not have to be estimated. According to their findings, this is a huge advantage in cases where little evidence on the performance of the forecasting agencies is known or in cases with time-varying parameters. Second, the average reduces the variance and the bias by cancelling out each others bias. And thirdly, "it will often dominate, in terms of MSE[5], forecasts based

[4] again Timmermann [20]

[5] which in this study is labelled MSFE

on optimal weighting." For a deeper discussion and a theoretical justification of the advantages of simple averaging methods see Timmermann 2005 and other studies mentioned in his work.

Due to this method's outstanding performance in the relevant literature and its easy way of calculation, we decided to declare this simple averaging method as the benchmark for this study. The benchmark is then taken to analyse all further combination schemes for further accuracy improvements.

An explanation why the average forecaster performs so well despite its simplicity when compared to other more complex models is best summarised by Stock and Watson's[18] " 'This forecast combination puzzle' - the repeated finding that simple combination forecasts outperform sophisticated adaptive combination methods in empirical applications - is, [as Stock and Watson think], more likely to be understood in the context of a model in which there is widespread instability in the performance of individual forecast, but the instability is sufficiently idiosyncratic that the combination of these individually unstably performing forecasts can itself be stable."

6.3.2 Trimming

Trimming is a technique similar to a outlier-reduction rule where forecasters who show a lower-ranking track record measured by an accuracy criterion like for instance the MSFE are assigned a weight of zero. Hereby, it does not matter in which way the other weights are calculated. This means for instance that before the average is calculated or before the median forecaster is identified, forecasters are trimmed off, i.e. dropped, from the sample. Obviously, the rule on which these cancellations base has to be determined, which hides a certain extent of arbitrariness. For this study, a MSFE-based ranking was generated and then an increasing amount of forecasters were dropped. First only the worst two, then as many as are necessary that remain the best 30, the best 20, the best 10 and finally the best 5 in the sample.

Empirical evidence is offered by a couple of authors. For instance "Winkler and Makridakis [21] find that including very poor models in an equal-

weighted combination can substantially worsen forecasting performance. Further, Stock and Watson[6] find that the simplest forecast combination methods such as trimmed equal weights and slowly moving weights tend to perform well" as Timmermann [20] describes. He concludes with "trimming of the worst models often improves performance".

The extent to which the lower-ranking forecasters are trimmed off has also been analysed. Granger and Jeon [11] suggest up to ten percent of the worst models, whereas Aiolfi and Favero [2] favour a much more aggressive trimming. They found that trimming off up to 80% of the worst models yield the best performance results.

6.4 Advanced combination schemes

Until now, the weights in the above sections had only two different values: zero or $1/n$; where again n is the number of forecasts to be combined out of all forecasting agencies N.

Quasi-combination schemes assign the value of one only to a single forecaster's weight whereas the remaining forecasters $N-n$, with $n=1$, get the weight of zero. In the case of the average forecast, every forecaster out of all N forecasters get the same weight of $1/n$ with $n=N$, whereas in trimming schemes the weights depend on how many forecasters n are chosen to be combined out of all agencies N. That means the non-zero weights ω_i depend on the degree of trimming: $\omega_i = 1/n$ with $n = 1, 2, ..., N$.

In this section now more than two (0 or $1/n$) different values of weights are allowed. The determination of the value of the individual weight can be based on several methods. A first way was already implicitly introduced, when the ranking-based[7] trimming method was explained. The drawback of using the rank is that it does not use all available information because it only asks whether a forecaster is better than another and not for how much a forecaster is better than the other. A second way which uses all available

[6] Stock and Watson [19] cited in Timmermann

[7] a ranking based on an accuracy criterion like the MSFE or the MAFE

information, is introduced in the next subsection.

Other methods like the minimisation of the summed MSFE, exponential smoothing and a time-discounted method will also be explained in this section.

6.4.1 Inverse MSFE

A first more advanced strategy is that the forecasts are weighted by their inverse MSFE. Since the better forecasting agencies should have a lower mean squared forecast error, weighting them with their inverse MSFE, makes their guess more dominant in a combination of forecasts. Thus the weight of forecaster i is determined by the relation of its inverse MSFE to the sum of the inverse MSFEs of all forecasters, as equation 6.2 points out.

$$\omega_i = \frac{[MSFE(i)]^{-1}}{\sum_{i=1}^{N}[MSFE(i)]^{-1}} \tag{6.2}$$

According to Stock and Watson [18], this combination scheme belongs to those ones, next to the median and the average forecast, which show empirically strong evidence of improved accuracy. Similarly, in a time-series simulation experiment, Winkler and Makridakis [21] find that a "weighted average with weights inversely proportional to the sum of squared errors or a weighted average with weights that depend on the exponentially discounted sum of squared errors perform better than the best individual forecasting model, equal-weighting or methods that require estimation of the full covariance matrix for the forecast errors."[8]

6.4.2 MSFE minimisation

This combination scheme assigns different weights to forecasts by giving those a higher weight which according to an accuracy criterion like the MSFE are most accurate. This method involves solving an optimization problem.

[8] in Timmermann [20]

In short, this method searches for a specific set of weights for which the forecasts produce the smallest MSFE when calculated over the respective time of information gathering. This is done by the following recipe which is demonstrated for the year 2008 in the case of a one-year rolling window. In this case, the weights for 2008 are calculated by looking on the performance of the forecasters in 2007. In the case of a three-year rolling window, obviously we need to look at the years 2005, 2006 and 2007. Thus in the one-year rolling window, the weights are calculated by solving the following minimization problem: for which weights is the MSFE of 2007 minimised? This is done by solving problem 6.3.

$$\min_{\omega_{i,t}} \frac{1}{M} \sum_{t=1}^{M} \left[y_t - \sum_{i=1}^{N} f_{i,t} \omega_{i,t} \right]^2 \tag{6.3}$$

where again ω_i represents the weight respective a forecasting agency $i \in N$, $t \in M$ stands for the month which total up to $M = 24$. Since we just look at a one-year rolling window, in this case the year 2007, there is no need to make the minimisation equation more complicated. Once, a solution for all ω_i is found, the combination of the forecasts to a new "combined" forecast can be done. For this, the weights ω_i are multiplied with their corresponding forecasts from the year 2008. This is done in equation 6.4, where $h = 24$ since we combine the forecasts for 2008 based on the information we gathered for the year 2007.

$$\widehat{y}_{t+h} = \omega_{1,t} f_{1,t+h} + \omega_{2,t} f_{2,t+h} + \ldots + \omega_{N,} f_{N,t+h} \tag{6.4}$$

By doing this for all months $t+h$, we get a series of 24 combined forecasts, $\widehat{y}_{t+h}$, for the year 2008. After taking the MSFE, its accuracy can be compared to all other combination schemes, by taking the MSFE($\widehat{y}_{t+h}$):

$$MSFE(\widehat{y}_{t+h}) = \frac{1}{M} \sum_{t=1}^{M} [y_{t+h} - \widehat{y}_{t+h}]^2 \tag{6.5}$$

Usually, one would have to wait a year until that out-of sample year's true outturn is available. Again, all numbers produced here are constructed

from the point of view of the year in question.

Since these optimisation problems were derived from the regression problems as in Timmermann [20] or Clements and Harvey [8], there is little evidence on the performance of these minimisation methods. Following their logic, there are three different cases to differentiate. Until here the weights ω_i are unconstrained. That means they do not necessarily add up to one and most probably are not positive for all ω_i. Thus, there is the possibility to impose a constraint such that all weights ω_i sum up to one, as it was done in equation 6.6.

$$\min_{\omega_{i,t}} \frac{1}{M} \sum_{t=1}^{M} \left[y_t - \sum_{i=1}^{N} f_{i,t}\omega_{i,t} \right]^2 \tag{6.6}$$
$$\text{s.t. } \sum_{i=1}^{N} \omega_{i,t} = 1$$

Since this still does not imply that all weights ω_i are positive or at least 0, a second constraint can be introduced, as in the following equation 6.7

$$\min_{\omega_{i,t}} \frac{1}{M} \sum_{t=1}^{M} \left[y_t - \sum_{i=1}^{N} f_{i,t}\omega_{i,t} \right]^2 \tag{6.7}$$
$$\text{s.t. } \sum_{i=1}^{N} \omega_{i,t} = 1$$

$$\text{and s.t. } \omega_{i,t} \geq 0 \qquad \forall i \in \{1, 2, \ldots, N\}$$

In the above case a rolling window of one year was assumed. Of course, there can be different specifications as for instance a three-year rolling window or a recursive algorithm, i.e. an ever expanding window, which never "forgets". These kind of considerations will be discussed in the next subsection.

6.4.3 Time-varying combination schemes

In order to construct weights one has consider about which information to use. In trimming schemes certain information was neglected, i.e. certain forecasters were cut out of the sample. Another issue is the time horizon on which a certain decision is made. Do I only consider the forecasters' performance in the last year? Or in the last three years? Or do I base my decision on all the information that is available to me, i.e. on all available years? Apart from this decision a different kind of time-varying combinations can be thought off: they summarize under the term of exponential smoothing. This is a technique which gives - independently on the number of years in consideration - more weight on recent years then on years which passed some more time ago.

Empirical evidence on this topic is best summarized by Timmermann's [20] "Limited time-variation in the combination my be helpful". He mentions a couple of studies, among those are Bates and Granger [6], Winkler and Makridakis [21] and Newbold and Granger [16], which all tend to result in favour of a longer rolling window length. Results are contradicting in the case of exponential smoothing methods. Whereas Bates and Granger prefer a stronger emphasis on recent years, Winkler and Makridakis find evidence for a relatively low value of the discount factor.

Whereas smoothing methods will not be discussed in this study, time-variation is applied for most cases in form of a comparison of a one-year rolling window, a three-year rolling window and a recursive, i.e. expanding, window.

6.4.4 Uncertainty-discounted methods

This method is - at least to the author's knowledge - a novelty to the literature of forecast combination. Figure 6.1 on page 54 shows that over the horizon of 24 month the standard deviation is decreasing from a value of about 0.5 in the January year-ahead forecast to a value of under 0.2 in the current-year December forecast. This is nothing new and not at all different

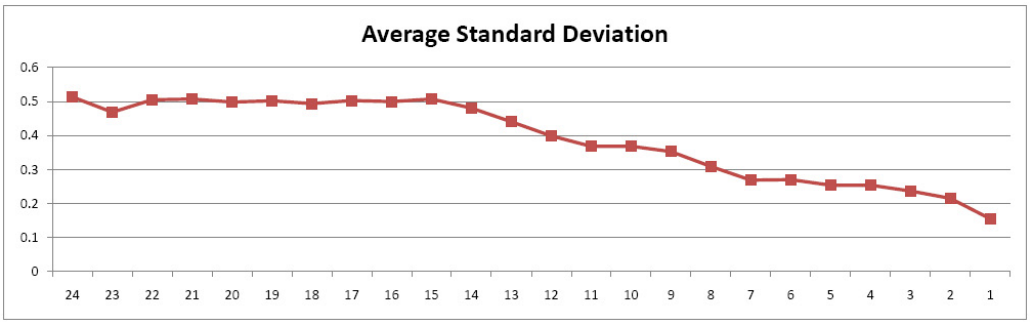


Figure 6.1: Average Standard Deviation (1998-2008)

to what is expected. It simply shows that the more information is available the more certain are the different forecasting agencies about the true outcome of the GDP growth rate. The novelty now is that when calculating the weights of each forecaster the original forecasts are multiplied by a certain factor λ which takes the uncertainty of earlier forecasts into account. This λ for instance could be generated by the following rule:

$$\lambda = \frac{\text{SD of month}}{\text{SD of last month}} \tag{6.8}$$

This would mean that in January of the year-ahead forecast the multiplier λ is around 3 since $\lambda = \frac{0.51}{0.16} \approx 3$. Obviously the λ for the last month is 1. Thus, the multiplier would be between 3 and 1. The effect of this procedure is the following. We already know that the earlier a forecast is published the more uncertainty there is about its correctness. Thus, a forecaster who is very good in predicting more close to the realisation should be rewarded for doing so. On the other hand, a forecaster being far away from the "truth" should be punished. This mechanism is implemented by multiplying the forecasts with their month-specific uncertainty, expressed by λ. In the end, the better "early forecasters" get assigned a much lower MSFE than in the case without a multiplier.

Since a more detailed evaluation of this method would go beyond the scope of this work, it will not further be discussed. A future study will exclusively discuss this issue.

Other combination methods and methods for calculating the weights are

of course existent. For example, estimating the weights via a regression. The literature reports a lot of mixed results on that topic along with some more severe difficulties concerning the application of an estimation. In this context, the combination method of shrinkage is mentioned from time to time. Nevertheless, these two issues are not discussed in this work. The interested reader is, thus, referred to the literature of Timmermann [20] and Clements and Harvey [8].

Also not included in this work is the more recent line of research of "forecast encompassing". For the newest developments on this issue, the reader is referred to Costantini and Kunst [10].

Chapter 7

Applied forecast combination

Is the benchmark really as hard to beat as Timmermann [20] and other studies claim? Are their findings robust enough to be applied also to this particular work?

After having explained the main methods of combining forecasts in the foregone chapter, this final chapter shows the results of the forecast combination which is based on this work's data set.

7.1 The output tables

In figure 7.1 the performance of five different combination schemes are displayed. Three of them were introduced in section 6.2 in which the so-called "quasi"-combinations are explained. A fourth scheme is very similar to the method of the "Last Year's Best", just that it tries to identify the best forecaster based on all past years in the data set and not only on the last year or on the last three years. The fifth method is the average forecast, i.e. the benchmark which is to beat.

In order to understand how to read this as well as the following output tables, figure 7.1 will be explained in more detail. Most important fact to remember is that the numbers for each year were calculated out-of-sample,

Performance of different forecast combination methods measured in MSFE

	Method	Year	2002	2003	2004	2005	2006	2007	2008	2001-2008
Y	Best performer in last 1 year		1.099	3.424	0.736	0.654	0.249	0.403	2.059	1.219
Y	Best performer in last 3 years			0.628	0.790	0.275	0.276	0.647	1.718	0.633
Y	Best performer in all available years		0.214	0.152	0.422	0.267	0.174	0.306	1.416	0.408
Y	Median of forecasts		0.273	0.173	0.235	0.426	0.184	0.153	1.590	0.410
Y	Average of forecasts (Benchmark)		0.233	0.148	0.282	0.330	0.236	0.214	1.427	0.394

Figure 7.1: Performance of quasi- and simple combination schemes

i.e. from the point of view of each particular single year[1]. This means that in the year 2002 we know which forecasting agency was the best in 2001. Thus, the MSFE of 1.099 (first column,first row) was formed by first identifying the best forecaster of 2001, and then taking this particular agency's forecast as *the* forecast for 2002. The number 1.099 then represents this forecaster's performance measured by the MSFE accuracy criterion. The next number for instance is not available. The reason for this is that in the year 2002 no full three years of information gathering was possible since the data for 1998 are not fully available to calculate a MSFE on the basis of a three-year window. In the third row however, the rule on which the information is evaluated is defined by taking all past possible information. Thus, it does not really matter that the year 1998 is not complete.[2]

For the column of 2008, only the information in 2007 was used to identify the last year's best performer. For the best performer in the last three years, obviously the years 2005, 2006 and 2007 were evaluated. And for the best performer since September 1997 - the start date of the data set - all years were used to identify the best forecaster. Then for all the three methods the forecasts from 2008 were taken to be measured by an accuracy criterion.

[1]To be more precise, the numbers represent the point of view of the end of each year.

[2]Remember the data set starts with the year-ahead forecast in September 1997 and not in January 1997

And finally they here got listed in this table.

The remaining two methods use by definition different information than the above explained schemes. In June 2008 for example, the median forecasting method identifies the median of all individual forecasts published in June 2008 and takes this value as *the* forecast for 2008. Then an evaluation of the accuracy follows exactly in the same way as for the other schemes. For the average the respective mechanism is true.

And finally the last column summarizes the accuracy performance of all five methods over the whole period. Thus, the last column tells about which method is performing better than the others in the long run, or, out differently, is not based on luck. In other words, this column tells about how to handle the available information in order to have a good prediction about the next realisation of the real UK GDP growth rate.

7.2 Competition of quasi- and simple combination schemes

Since now it should be clear how to read the table, the initial question actually can be answered. Thus, in a competition of the above five explained (quasi)-combination methods in figure 7.1, the average forecast - the benchmark in this study - is not beaten. The overall[3] MSFE of the averaging method is, although to a small extent, the smallest. The best performer in all available years and the median forecaster share a close, but second place. Now the yearly consideration pays off, since it can be shown that out of the seven years, it was once the best method and another four times missed the first place by just a little. In the other years it performed quite well. Also, it is very clear that by simply relying on the last year's best forecaster is not really a good choice. This method is almost five out of seven years the worst adviser; and this to quite an extent.

These findings are very much in line with the empirical evidence of other

[3]the last column which summarizes the accuracy performance

studies. Another empirical fact can also be confirmed by this work. As mentioned in subsection 6.4.3 time-varying models with a longer rolling window length tend to perform better. This is also true for this competition. Whereas a rolling window of one year performs quite poorly, a window of length three performs better, but not as good as a forecasting method which incorporates all available information.

7.3 Competition of trimming schemes

The produced empirical evidence respective the trimming schemes is extensive as tables 9.4 till 9.7 on page 70 show. In this trimming competition a new dimension of comparison[4] is also introduced and then evaluated in figure 9.8. But more on that later.

Figure 7.2 which is taken from the lower part of figure 9.5 is presented here to serve the purpose of demonstration. For a full comparison, find the tables in the appendix.

Performance of different forecast combination methods measured in MSFE

Method		Year	2002	2003	2004	2005	2006	2007	2008	2001-2008
Y	Data Average Benchmark		0.233	0.148	0.282	0.330	0.236	0.214	1.427	0.394
Y	Data Median Benchmark		0.273	0.173	0.235	0.426	0.184	0.153	1.590	0.410
Trimming based on Last Year's Best Yearly Forecasters (= rolling window length of 1 year)										
Y	Data Average Best"-2"		0.273	0.135	0.239	0.381	0.193	0.175	1.580	0.410
Y	Data Median Best"-2"		0.285	0.167	0.235	0.437	0.173	0.149	1.624	0.423
Y	Data Average Best30		0.311	0.109	0.257	0.382	0.196	0.180	1.651	0.425
Y	Data Median Best30		0.323	0.158	0.239	0.472	0.182	0.168	1.674	0.444
Y	Data Average Best20		0.346	0.106	0.320	0.337	0.195	0.236	1.779	0.457
Y	Data Median Best20		0.359	0.166	0.280	0.438	0.197	0.212	1.833	0.480
Y	Data Average Best10		0.309	0.094	0.385	0.293	0.218	0.321	1.601	0.444
Y	Data Median Best10		0.248	0.145	0.359	0.260	0.221	0.329	1.743	0.453
Y	Data Average Best05		0.316	0.155	0.389	0.226	0.252	0.459	1.627	0.472
Y	Data Median Best05		0.219	0.112	0.266	0.180	0.264	0.471	1.763	0.447

Figure 7.2: Trimming competition, exemplarily

The first row shows the average forecaster, the benchmark, accompanied by the median forecaster. These two lines are in fact row five and four,

[4] indicated by the "Y" or the "M" at the beginning of each line

respectively, of the above figure 7.1 which showed the results of the quasi- and simple combination competition.

In figure 7.2, as an example, the case of a rolling window with a length of one year is shown. Thus, the performance of the forecasters were ranked according to the MSFE accuracy criterion always on a year-before-displayed basis. In rows 3 and 4 then, the worst two forecasters were eliminated which shall be indicated by the notation of *best"-2"*. In lines 5 to 10 the sample is further reduced until in the last two lines only the best five forecasters based on last year's information remain. Then their forecasts were averaged (or the median was taken), after that the accuracy was measured and finally displayed according to the years indicated in the table. The last column again summarizes the accuracy performance over the entire evaluated period of time.

As a first finding of figure 7.2 it should be said that the benchmark again was not beaten by any of the trimming methods when evaluated over the whole 7-year period. This is quite impressive. Secondly, the higher the extent of the trimming, the lower its overall accuracy, even though the "Data Median Best05" appears to work not as bad as others. And as a last finding of this table, we can state that their is no clear superiority of neither the average forecast nor the median forecast.

Now, the findings of tables 9.4 and 9.5 are summarized which were placed in the appendix on page 70 for reasons of space. Notice, that these two tables show exactly the same numbers, but are differently colour-coded. The first table 9.4 is colour-coded in a way that it supports a visual comparison of the three blocks of each table which differ by the length of the rolling window: in the upper panel information was gathered only in the last respective year, in the middle panel a three-years rolling window was applied and in the lower panel the underlying method was a expanding window. The second table 9.5 emphasises a visual comparison within of each block which means that it compares the different methods based on the extent of trimming. Here are the findings of the tables 9.4 and 9.5:

- From the first table (figure 9.4) we can see that the methods that

have a longer horizon (lower panel) for determining the ranking - on which the trimming decision is based - perform better since the MSFE for these methods are mostly smaller than in the upper or the middle panel. Thus, these results confirm again the empirical findings of other studies.

- The second table (figure 9.5) leads us to the conclusion that trimming off either works when done only slightly or done quite extensively. In between those extremes, there seems to be evidence, that trimming performs not that good. This finding actually confirms the contrary opinion of other studies on this issue as pointed out in the last paragraph of subsection 6.3.2 on page 48.

- But - and here comes the "but" - the benchmark is beaten by the "Best05" specification in the case of an expanding window.

- There is no clear evidence that the median or the average forecaster is superior. Very often their values do not differ much from each other.

The next two tables 9.6 and 9.7 introduce a new dimension of comparison which is indicated by the "M", instead of the "Y", in front of each line. "Y" stands for the yearly calculation of the MSFE. This means that all forecasts for one year of a specific agency are evaluated by the MSFE. This way may neglect the fact that forecasts with different horizons are evaluated with the same criterion. But as explained in earlier chapters, this does not affect the aim of this study since all forecasting agencies are treated in the same way. Still it was interesting to see what happens when the construction of the rankings are based on a monthly comparison similar to figure 4.6 on page 37. This means for the three-year rolling window that for every forecaster all three January year-ahead forecasts were evaluated and ranked. Then the same is done for all of those in February of the year-ahead, and so on, until all 24 months are finished. After that, it was decided on a monthly basis who are the best 5, 10, 20 and so on forecasters. On this information then the forecasts were combined.

The findings in the tables 9.6 and 9.7 are different to those in the first two tables:

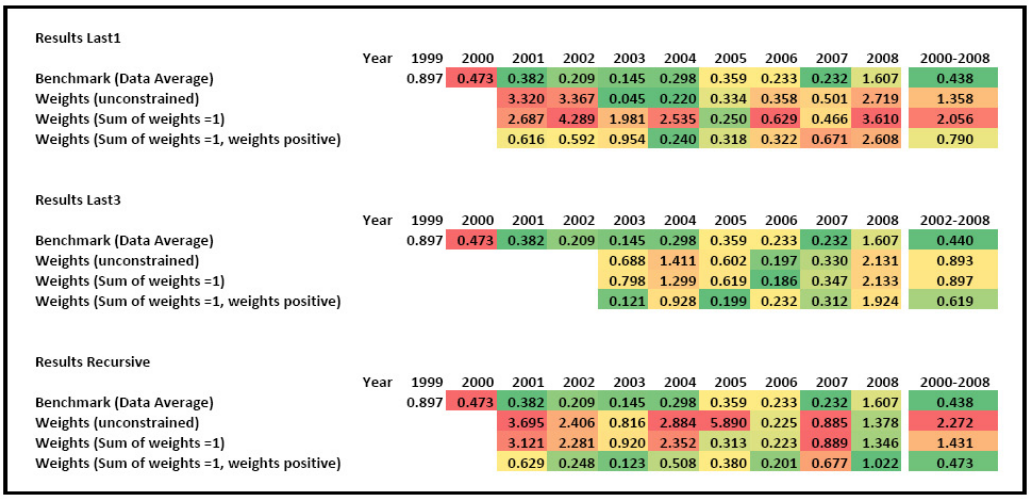
- On the one hand longer horizons seem again to beat shorter horizons.
- But on the other hand extensive trimming performs uniformly worse.
- None of those specifications is able to beat the benchmark.

The question now is only which of those two ways of constructing the rankings is better? Table 9.8 shows a direct comparison of every method involved. The figures in that table were calculated by taking the MSFE of the monthly rankings and subtracting from that the MSFE generated by the yearly rankings which is also indicated by "M-Y" at the beginning of each line. Since the majority of the figures are above zero, we must conclude that the monthly MSFEs are slightly higher than those based on years. In other words, the yearly based rankings produce better combined forecasts.

7.4 MSFE minimisation

A last competition was conducted. As explained in the subsection 6.4.2 on page 50, it was tried to minimize the MSFE via an optimisation process. The weights of the individual agencies were chosen in a way that they minimise the MSFE subject to the entire time period of information gathering. Within the already well known differentiation into a one-year rolling window, a three-year rolling window and ever expanding window, the weights were also formed by constraining them in three different ways. First, there is no constraint applied. In the second type the sum of all weights have to add up to 1 and in the third one the weights are not only forced to add up to 1, but also to be all positive. For further detail recall equations 6.3, 6.6 and 6.7.

The first line of all three blocks show again the average forecaster as the benchmark. Notice that its figures are not exactly the same as in the two competitions before since the data sample was slightly modified: blanks



Results Last1

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2000-2008
Benchmark (Data Average)	0.897	0.473	0.382	0.209	0.145	0.298	0.359	0.233	0.232	1.607	0.438
Weights (unconstrained)			3.320	3.367	0.045	0.220	0.334	0.358	0.501	2.719	1.358
Weights (Sum of weights =1)			2.687	4.289	1.981	2.535	0.250	0.629	0.466	3.610	2.056
Weights (Sum of weights =1, weights positive)			0.616	0.592	0.954	0.240	0.318	0.322	0.671	2.608	0.790

Results Last3

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2002-2008
Benchmark (Data Average)	0.897	0.473	0.382	0.209	0.145	0.298	0.359	0.233	0.232	1.607	0.440
Weights (unconstrained)					0.688	1.411	0.602	0.197	0.330	2.131	0.893
Weights (Sum of weights =1)					0.798	1.299	0.619	0.186	0.347	2.133	0.897
Weights (Sum of weights =1, weights positive)					0.121	0.928	0.199	0.232	0.312	1.924	0.619

Results Recursive

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2000-2008
Benchmark (Data Average)	0.897	0.473	0.382	0.209	0.145	0.298	0.359	0.233	0.232	1.607	0.438
Weights (unconstrained)			3.695	2.406	0.816	2.884	5.890	0.225	0.885	1.378	2.272
Weights (Sum of weights =1)			3.121	2.281	0.920	2.352	0.313	0.223	0.889	1.346	1.431
Weights (Sum of weights =1, weights positive)			0.629	0.248	0.123	0.508	0.380	0.201	0.677	1.022	0.473

Figure 7.3: MSFE minimisation competition

were filled up and agencies with less than 200 out of 240 data points were dropped.

A first obvious finding is that in the case of unconstrained weights and weights which only have to add up to 1, the overall MSFEs are clearly higher. Further, the benchmark is not beaten. When it comes to the comparison of the weights it can be seen: the ones with an expanding window performs best, followed by the three-year rolling window which make the one-year rolling window again the worst performer.

Chapter 8

Conclusion

Forecast evaluation

This study started with an evaluation of the forecasters' ability to predict accurately this year's and next year's UK real GDP growth rate. Only two agencies out of a total of 51 were able to beat the "consensus" forecast which is the simple average of all forecasts. Quite striking, but conforming to the empirical evidence, the average forecaster as well as the median forecaster perform very good.

Forecast combination

The second part of this study looks for methods of combining the individual forecasters in a way that leads to better forecasts than the average forecaster. Finally, a small summary about the findings of combination of forecasts is shown:

- The simple average was just once beaten accuracy-wise by another combination method.
- Relying on the last year's or the last three year's best forecaster is not a good idea.
- The median forecaster seems to be - to a very small extent - not as

good as the average forecaster. But the distance between them is so small that we must say they are equally performing very good.

- Trimming is performing well when done either only slightly or to a very high extent. In only one case, trimming has beaten the benchmark.
- The longer the horizon to gather information the better are the combined forecasts. Thus, an ever expanding window is better than a three-years rolling window which is in turn better than a one-year rolling window.
- Minimising the MSFE by changing the weights of the individual forecasters performs quite poorly except in the situation where the weights are constrained to sum up to one and have to be all positive. But still, that methods is not as good as the benchmark.

Thus, most findings confirm the empirical evidence of other studies.

Chapter 9

Appendix

THE YEAR'S BEST FORECASTERS

		GDP growth %	Inflation Q4 %	Current account £bn	Jobless Q4 m	Bank rate %	Score out of 10
	Outturn	0.8	4.1	-39	1.07	2	10
1	Standard Chartered	1.2	2.0	-50		4.25	5
2	HSBC	1.5	2.0	-50	1.02	4.5	5
3	Lehman Brothers	1.7	2.7	-50	0.87	5.0	4
4	Economic Perspectives	-0.1	2.3	-33	1.05	4.5	4
5	Daiwa Institute	1.8	2.0	-39.7	0.95	5.0	4
6	ING Financial Markets	1.6	1.8	-49	0.92	4.75	4
7	MacroEcon.Com	3.0	3.2	-41.7	0.81	6.25	4
8	Lombard Street	1.4	2.2	-52.3	0.93	4.25	3
9	Ingenious Securities	1.8	2.5	-70.7	0.91	5.25	3
10	BNP Paribas	1.6	2.2	-65	0.88	4.5	3
11	ABN Amro	1.9	2.2	-55.7	*1.89	5.0	3
12	Deutsche Bank	1.8	2.1	-43.7	0.9	5.0	3
13	Hermes	2.0	2.2	-42	0.9	4.75	3
14	Barclays Capital	2.0	2.1	-45.7	0.81	5.0	3
15	Dresdner Kleinwort	1.9	2.0	-39		5.0	3
16	Beacon Econ Forecasting	2.1	2.8	-74.1	0.8	5.2	2
17	Experian	1.7	2.3	-80.1	0.87	5.0	2
18	Goldman Sachs	1.7	2.3	-50.5	*1.64		2
19	CBI	2.0	2.4	-40.2	0.81	5.25	2
20	Citigroup	1.8	2.2	-64.3	0.8	4.25	2
21	CSFB	2.0	2.4	-60		4.75	2
22	UBS	1.8	2.2	-51	0.86	4.5	2
23	Global Insight	1.8	2.1	-52	0.91	4.5	2
24	Morgan Stanley	1.8	2.1			5.25	2
25	Commerzbank	1.8	2.1	-55.1	0.84	4.75	2
26	Ernst & Young Item Club	1.8	2.0	-65	0.89	4.5	2
27	Liverpool Macroeconomics	2.1	**2.3	-50.9	1.23	5.4	2
28	OECD	2.0	2.2				2
29	Bank of America	1.8	1.9	-73.9		4.5	2
30	Capital Economics	2.0	2.1	-75	0.87	4.5	2
31	CEBR	1.7	1.7	-62.4	0.96	4.5	2
32	European Commission	2.2	2.0	-49.9	*1.74		2
33	Moody's Economy	1.7	1.5	-27.7	0.97	5.25	2
34	Oxford Economics	1.9	1.9	-52.1	0.92	5.0	1
35	Cambridge Econometrics	2.1	1.9	-36.9	0.9	5.4	1
36	NIESR	2.2	2.0	-47.7	0.9	5.4	1
37	Treasury	2.25	2.0	-45			1
38	EIU	2.3	2.0	-45.1	0.89	5.0	1
39	RBS Financial Markets	2.1	2.1	-58.9	0.89	5.0	0
40	Fortis Bank	2.1	1.9	-58	0.89	5.0	0
41	Daiwa Securities	2.3	2.0			5.5	0
42	IMF	2.3	2.0	-52.5			0

SCORING SYSTEM **GDP growth:** 3 points for 0.4%-1.2%; 2 points for 0%-1.6%; 1 point for -0.4%-2%. **Inflation:** 3 points for 3.5% or above; 2 points for 3% or above; 1 point for 2.5% or above. **Current account:** 1 point for a deficit between £30 billion and £50 billion. **Unemployment:** 1 point for 0.95m or above. **Bank rate:** 1 point for 3% or below. Bonus point for inflation exceeding growth. Where scores are equal, forecasters are ranked by their proximity to the growth and inflation outturns. Key: * Labour Force Survey unemployment; ** RPIX inflation

Figure 9.1: Best Forecaster 2008, Source: Sunday Times Dezember 2008

RANKING MSFE																
FORECASTER	1998-2008	1998 - 2007	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	MIN	MAX	SPREAD
Williams de Broe	1	1	26	3	4	16	7	10	13	27	3	1		1	27	26
Greenwich Natwest	2	5	27	6	8	7								6	27	21
Fortis Bank	3	3				11	23	38	2	26	7	10	31	2	38	36
Capital Economics	4	2			2	3	15	14	33	7	31	31	12	2	33	31
Lombard Street	5	16	42	1	3	39	14	24	18	38	9	41	2	1	42	41
Primark WEFA	6	27	13	11	12	30	36							11	36	25
Bridgewell	7	29					1	44	39	11	2	19		1	44	43
Daiwa Institute of Res.	8	7	11	19	17	22	16	29	10	24	20	21	18	10	29	19
Benchmark Average	9	13	3	16	27	24	17	11	28	14	26	26	14	3	28	25
Benchmark Median	10	10	9	17	23	18	21	21	24	17	22	22	19	9	24	15
RBS Global B&M	11	9	14	18	29	14	29	27	1	16	16	17	21	1	29	28
Credit Lyonnais	12	32	23	23	40	4	19	19	26					4	40	36
Moody's Economy	13	15						13	11	21	30	38	5	5	38	33
OEF	14	11	19	25	25	20	18	28	3	20	24	15	25	3	28	25
Abn Amro	15	8	5	12	10	21	22	33	25	6	34	28	33	5	34	29
Hermes	16	23	21	4	13	43	33	6	38	3	27	39	8	3	43	40
Standard Chartered	17	6			1	28	25	2	34	28	28	24	22	1	34	33
Experian Business Str.	18	18	2	27	22	19	31	20	30	30	6	25	16	2	31	29
HSBC	19	33	29	21	21	5	5	5	43	1	42	37	4	1	43	42
Merrill Lynch	20	37	43	9	5	26	44	45						5	45	40
CBI	21	19	16	24	31	17	20	9	27	32	23	20	23	9	32	23
Goldman Sachs	22	22	15	35	16	10	43	35	5	19	8	23	17	5	43	38
UBS	23	21	30	32	14	15	35	7	37	22	12	11	20	7	37	30
EIU	24	24	1	26	33	6	9	3	36	5	40	32	26	1	40	39
Global Insight	25	26	4	10	38	41	24	37	12	25	25	18	13	4	41	37
Norwich Union IM	26	41	37	13	11									11	37	26
Credit Suisse	27	12	8	14	20	13	34	17	7	36	15	5	36	5	36	31
NIESR	28	17	24	8	6	44	41	15	20	29	10	9	32	6	44	38
EC	29	20	18	5	19	45	40	34	15	35	5	6	30	5	45	40
Morgan Stanley	30	34	44	34	18	27	39	41	29	9	14	27	9	9	44	35
ITEM Club	31	25	17	33	36	12	3	25	17	39	21	7	29	3	39	36
IMF	32	14	22	41	26	8	28	23	32	13	11	4	37	4	41	37
ING Financial Markets	33	36					2	43	16	34	39	35	3	2	43	41
Dresdner Kleinwort W.	34	39	33	15	35	33	38	8	8	4	37	40	7	4	40	36
Liverpool Macro Res.	35	35	6	7	7	31	45	18	40	8	36	36	15	6	45	39
Barclays Capital	36	30	10	29	24	32	30	26	6	41	1	2	34	1	41	40
OECD	37	31	31	22	44	9	11	16	19	23	19	8	35	8	44	36
Schroders	38	44	12	38	34	36	6	12	35	15	41	34		6	41	35
Citigroup	39	38	39	20	9	42	32	42	9	42	17	12	27	9	42	33
Cambridge Econometrics	40	40	20	30	45	23	4	30	21	37	29	16	28	4	45	41
Bank of America	41	28					12	4	42	2	33	29	24	2	42	40
Deutsche Bank	42	42	35	42	28	37	13	32	22	18	35	14	10	10	42	32
Charterhouse	43	46	45	2	42	38	26	31						2	45	43
Societe Generale	44	47	25	31	43	25	42	1						1	43	42
J P Morgan	45	43	38	45	15	35	37	39	14	31	18	3	38	3	45	42
Lehman Brothers	46	45	40	36	37	40	46	40	31	10	38	33	11	10	46	36
CEBR	47	48	41	37	46	29	10	22	41	33	4	30	6	4	46	42
Barclays Bank	48	49	34	39	41	1	8							1	41	40
WestLB	49	50	32	43	30	47	27	36	23	12	32	42		12	47	35
Henley	50	51	7	40	32	34	47							7	47	40
Chase Manhattan	51	52	36	28	39	46								28	46	18
MacroEcon.com	52	4							4	40	13	13	39	4	40	36
Economic Perspectives	53	53	28	44	47	2	48	46	44	43	43	43	1	1	48	47

Figure 9.2: Ranking of forecasters according to the MSFE accuracy criterion

RANKING MAFE

FORECASTER	1998-2008	1998 - 2007	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	MIN	MAX	SPREAD
Williams de Broe	1	1	28	4	1	25	9	11	11	27	4	2		1	28	27
Greenwich Natwest	2	9	23	7	8	8								7	23	16
Fortis Bank	3	2				12	13	38	1	33	12	15	38	1	38	37
Bridgewell	4	21					1	44	39	11	2	23		1	44	43
Daiwa Institute of Res.	5	5	4	14	14	20	17	27	8	21	19	16	16	4	27	23
Lombard Street	6	14	42	1	2	38	14	23	16	39	6	40	2	1	42	41
Capital Economics	7	4			4	2	19	19	33	6	31	30	14	2	33	31
RBS Global B&M	8	8	6	23	27	13	26	26	3	14	18	18	21	3	27	24
OEF	9	7	14	21	18	24	12	29	4	20	23	13	26	4	29	25
Abn Amro	10	6	11	8	6	23	10	34	26	4	34	22	32	4	34	30
Credit Lyonnais	11	30	20	27	41	6	20	22	29					6	41	35
Goldman Sachs	12	13	5	36	17	11	39	35	2	16	8	21	13	2	39	37
HSBC	13	24	30	18	20	5	5	5	42	1	41	35	4	1	42	41
Primark WEFA	14	32	17	6	15	34	45							6	45	39
ITEM Club	15	12	8	30	37	9	3	25	17	34	20	5	29	3	37	34
Standard Chartered	16	15			3	19	31	2	32	30	28	26	9	2	32	30
NIESR	17	11	25	9	7	42	37	8	21	28	9	11	33	7	42	35
UBS	18	20	29	28	9	16	35	6	34	18	17	12	18	6	35	29
CBI	19	19	15	19	26	18	21	10	27	31	22	14	24	10	31	21
Experian Business Str.	20	22	2	29	22	22	30	15	30	29	7	25	17	2	30	28
EIU	21	23	3	26	35	7	8	3	37	7	40	33	27	3	40	37
Hermes	22	28	16	5	10	43	33	8	38	3	25	39	7	3	43	40
IMF	23	10	24	43	27	4	23	21	36	12	11	6	37	4	43	39
Barclays Capital	24	18	12	31	21	36	27	18	5	43	1	1	31	1	43	42
Global Insight	26	29	7	12	36	39	22	33	12	23	26	19	11	7	39	32
Benchmark Median	25	26	22	17	25	26	24	24	25	19	24	24	20	17	26	9
Credit Suisse	27	16	10	15	29	21	34	13	9	38	16	4	34	4	38	34
Norwich Union IM	28	37	33	13	19									13	33	20
EC	29	25	19	3	23	45	38	28	20	37	3	7	30	3	45	42
Morgan Stanley	30	31	43	32	16	14	40	42	24	9	14	29	10	9	43	34
OECD	31	27	26	16	45	10	6	12	18	26	21	8	36	6	45	39
Bank of America	33	17					18	4	40	2	32	27	25	2	40	38
Benchmark Average	32	34	27	22	30	29	28	32	31	17	27	28	19	17	32	15
Merrill Lynch	34	43	44	10	5	30	42	45						5	45	40
Citigroup	35	36	41	20	13	44	29	41	10	42	15	9	23	9	44	35
Cambridge Econometrics	36	35	21	33	43	17	4	30	23	40	29	20	28	4	43	39
ING Financial Markets	37	40					2	43	13	32	37	34	3	2	43	41
J P Morgan	38	39	36	45	11	36	36	39	15	25	13	3	35	3	45	42
Moody's Economy	39	33						17	14	24	30	38	6	6	38	32
Societe Generale	40	44	18	34	39	15	43	1						1	43	42
Dresdner Kleinwort W.	41	42	35	24	34	32	41	14	7	5	39	41	8	5	41	36
Liverpool Macro Res.	42	38	12	11	12	27	46	20	43	8	35	37	22	8	46	38
Barclays Bank	43	46	32	41	42	1	15							1	42	41
Schroders	44	47	9	39	38	31	7	16	35	15	42	36		7	42	35
Deutsche Bank	45	41	38	42	24	33	16	31	22	22	36	10	15	10	42	32
CEBR	46	45	40	37	46	28	11	7	41	35	5	32	5	5	46	41
Lehman Brothers	47	48	39	35	33	41	44	40	28	10	38	31	12	10	44	34
Charterhouse	48	49	45	2	44	40	32	36						2	45	43
WestLB	49	50	34	40	31	47	24	37	19	13	33	42		13	47	34
MacroEcon.com	50	3							6	36	10	17	39	6	39	33
Henley	51	51	1	38	32	35	47							1	47	46
Chase Manhattan	52	52	37	25	40	46								25	46	21
Economic Perspectives	53	53	31	44	47	3	48	46	44	41	43	43	1	1	48	47

Figure 9.3: Ranking of forecasters according to the MAFE accuracy criterion

Performance of different forecast combination methods measured in MSFE

Method		Year	2002	2003	2004	2005	2006	2007	2008	2001-2008
Y	Data Average Benchmark		0.233	0.148	0.282	0.330	0.236	0.214	1.427	0.394
Y	Data Median Benchmark		0.273	0.173	0.235	0.426	0.184	0.153	1.590	0.410
Trimming based on Last Year's Best Yearly Forecasters (= rolling window length of 1 year)										
Y	Data Average Best"-2"		0.273	0.135	0.239	0.381	0.193	0.175	1.580	0.410
Y	Data Median Best"-2"		0.285	0.167	0.235	0.437	0.173	0.149	1.624	0.423
Y	Data Average Best30		0.311	0.109	0.257	0.382	0.196	0.180	1.651	0.425
Y	Data Median Best30		0.323	0.158	0.239	0.472	0.182	0.168	1.674	0.444
Y	Data Average Best20		0.346	0.106	0.320	0.337	0.195	0.236	1.779	0.457
Y	Data Median Best20		0.359	0.166	0.280	0.438	0.197	0.212	1.833	0.480
Y	Data Average Best10		0.309	0.094	0.385	0.293	0.218	0.321	1.601	0.444
Y	Data Median Best10		0.248	0.145	0.359	0.260	0.221	0.329	1.743	0.453
Y	Data Average Best05		0.316	0.155	0.389	0.226	0.252	0.459	1.627	0.472
Y	Data Median Best05		0.219	0.112	0.266	0.180	0.264	0.471	1.763	0.447
Trimming based on Last 3 Years' Best Yearly Forecasters (= rolling window length of 3 years)										
Y	Data Average Best"-2"		0.270	0.171	0.239	0.380	0.204	0.176	1.586	0.417
Y	Data Median Best"-2"		0.284	0.172	0.231	0.445	0.183	0.150	1.628	0.427
Y	Data Average Best30		0.278	0.171	0.231	0.420	0.196	0.176	1.649	0.430
Y	Data Median Best30		0.291	0.156	0.236	0.470	0.187	0.148	1.700	0.440
Y	Data Average Best20		0.265	0.170	0.266	0.422	0.172	0.176	1.667	0.432
Y	Data Median Best20		0.270	0.160	0.272	0.514	0.175	0.141	1.693	0.445
Y	Data Average Best10		0.264	0.181	0.353	0.378	0.165	0.223	1.666	0.444
Y	Data Median Best10		0.236	0.163	0.310	0.461	0.176	0.228	1.612	0.440
Y	Data Average Best05		0.181	0.183	0.430	0.253	0.210	0.277	1.540	0.422
Y	Data Median Best05		0.204	0.150	0.410	0.230	0.226	0.259	1.556	0.417
Trimming based on All Years' Best Yearly Forecasters (= computed recursively, i.e. expanding window)										
Y	Data Average Best"-2"		0.276	0.170	0.224	0.380	0.205	0.173	1.564	0.412
Y	Data Median Best"-2"		0.289	0.166	0.226	0.447	0.174	0.142	1.603	0.420
Y	Data Average BEST 30		0.277	0.168	0.214	0.395	0.180	0.174	1.619	0.417
Y	Data Median Best 30		0.291	0.169	0.211	0.471	0.152	0.144	1.685	0.430
Y	Data Average BEST 20		0.286	0.146	0.268	0.332	0.168	0.202	1.621	0.415
Y	Data Median Best 20		0.290	0.155	0.284	0.443	0.156	0.181	1.660	0.437
Y	Data Average BEST 10		0.311	0.148	0.287	0.296	0.204	0.211	1.476	0.405
Y	Data Median Best 10		0.281	0.144	0.337	0.400	0.185	0.196	1.556	0.428
Y	Data Average Best05		0.310	0.136	0.247	0.335	0.136	0.201	1.019	0.332
Y	Data Median Best05		0.299	0.137	0.326	0.430	0.160	0.248	0.878	0.348

Figure 9.4: Trimming: comparison of the window length (yearly based)

Performance of different forecast combination methods measured in MSFE

Method		Year	2002	2003	2004	2005	2006	2007	2008	2001-2008
Y	Data Average Benchmark		0.233	0.148	0.282	0.330	0.236	0.214	1.427	0.394
Y	Data Median Benchmark		0.273	0.173	0.235	0.426	0.184	0.153	1.590	0.410
Trimming based on Last Year's Best Yearly Forecasters (= rolling window length of 1 year)										
Y	Data Average Best"-2"		0.273	0.135	0.239	0.381	0.193	0.175	1.580	0.410
Y	Data Median Best"-2"		0.285	0.167	0.235	0.437	0.173	0.149	1.624	0.423
Y	Data Average Best30		0.311	0.109	0.257	0.382	0.196	0.180	1.651	0.425
Y	Data Median Best30		0.323	0.158	0.239	0.472	0.182	0.168	1.674	0.444
Y	Data Average Best20		0.346	0.106	0.320	0.337	0.195	0.236	1.779	0.457
Y	Data Median Best20		0.359	0.166	0.280	0.438	0.197	0.212	1.833	0.480
Y	Data Average Best10		0.309	0.094	0.385	0.293	0.218	0.321	1.601	0.444
Y	Data Median Best10		0.248	0.145	0.359	0.260	0.221	0.329	1.743	0.453
Y	Data Average Best05		0.316	0.155	0.389	0.226	0.252	0.459	1.627	0.472
Y	Data Median Best05		0.219	0.112	0.266	0.180	0.264	0.471	1.763	0.447
Trimming based on Last 3 Years' Best Yearly Forecasters (= rolling window length of 3 years)										
Y	Data Average Best"-2"		0.270	0.171	0.239	0.380	0.204	0.176	1.586	0.417
Y	Data Median Best"-2"		0.284	0.172	0.231	0.445	0.183	0.150	1.628	0.427
Y	Data Average Best30		0.278	0.171	0.231	0.420	0.196	0.176	1.649	0.430
Y	Data Median Best30		0.291	0.156	0.236	0.470	0.187	0.148	1.700	0.440
Y	Data Average Best20		0.265	0.170	0.266	0.422	0.172	0.176	1.667	0.432
Y	Data Median Best20		0.270	0.160	0.272	0.514	0.175	0.141	1.693	0.445
Y	Data Average Best10		0.264	0.181	0.353	0.378	0.165	0.223	1.666	0.444
Y	Data Median Best10		0.236	0.163	0.310	0.461	0.176	0.228	1.612	0.440
Y	Data Average Best05		0.181	0.183	0.430	0.253	0.210	0.277	1.540	0.422
Y	Data Median Best05		0.204	0.150	0.410	0.230	0.226	0.259	1.556	0.417
Trimming based on All Years' Best Yearly Forecasters (= computed recursively, i.e. expanding window)										
Y	Data Average Best"-2"		0.276	0.170	0.224	0.380	0.205	0.173	1.564	0.412
Y	Data Median Best"-2"		0.289	0.166	0.226	0.447	0.174	0.142	1.603	0.420
Y	Data Average BEST 30		0.277	0.168	0.214	0.395	0.180	0.174	1.619	0.417
Y	Data Median Best 30		0.291	0.169	0.211	0.471	0.152	0.144	1.685	0.430
Y	Data Average BEST 20		0.286	0.146	0.268	0.332	0.168	0.202	1.621	0.415
Y	Data Median Best 20		0.290	0.155	0.284	0.443	0.156	0.181	1.660	0.437
Y	Data Average BEST 10		0.311	0.148	0.287	0.296	0.204	0.211	1.476	0.405
Y	Data Median Best 10		0.281	0.144	0.337	0.400	0.185	0.196	1.556	0.428
Y	Data Average Best05		0.310	0.136	0.247	0.335	0.136	0.201	1.019	0.332
Y	Data Median Best05		0.299	0.137	0.326	0.430	0.160	0.248	0.878	0.348

Figure 9.5: Trimming: comparison of extent of trimming (yearly based)

Performance of different forecast combination methods measured in MSFE

Method		Year	2002	2003	2004	2005	2006	2007	2008	2001-2008
M	Data Average Benchmark		0.233	0.148	0.282	0.330	0.236	0.214	1.427	0.394
M	Data Median Benchmark		0.273	0.173	0.235	0.426	0.184	0.153	1.590	0.410
Trimming based on Last Year's Best Montly Forecasters (= rolling window length of 1 year)										
M	Data Average Best"-2"		0.227	0.165	0.228	0.374	0.217	0.160	1.590	0.401
M	Data Median Best"-2"		0.279	0.175	0.230	0.442	0.186	0.141	1.660	0.415
M	Data Average Best30		0.208	0.155	0.248	0.420	0.242	0.144	1.583	0.410
M	Data Median Best30		0.252	0.175	0.234	0.444	0.199	0.129	1.670	0.414
M	Data Average Best20		0.180	0.131	0.288	0.490	0.268	0.106	1.772	0.439
M	Data Median Best20		0.212	0.150	0.276	0.466	0.257	0.099	1.777	0.429
M	Data Average Best10		0.126	0.099	0.433	0.503	0.387	0.119	1.886	0.481
M	Data Median Best10		0.175	0.105	0.383	0.464	0.415	0.099	1.856	0.464
M	Data Average Best05		0.073	0.116	0.426	0.513	0.463	0.100	1.885	0.488
M	Data Median Best05		0.108	0.133	0.433	0.458	0.513	0.073	1.864	0.483
Trimming based on Last 3 Years' Best Monthly Forecasters (= rolling window length of 3 years)										
M	Data Average Best"-2"		0.267	0.170	0.226	0.369	0.211	0.169	1.566	0.408
M	Data Median Best"-2"		0.286	0.173	0.227	0.436	0.183	0.143	1.614	0.416
M	Data Average Best30		0.278	0.163	0.232	0.417	0.221	0.158	1.564	0.418
M	Data Median Best30		0.291	0.167	0.229	0.455	0.189	0.130	1.617	0.419
M	Data Average Best20		0.278	0.151	0.260	0.444	0.220	0.138	1.617	0.430
M	Data Median Best20		0.307	0.162	0.252	0.486	0.200	0.117	1.651	0.431
M	Data Average Best10		0.279	0.142	0.344	0.503	0.266	0.153	1.668	0.471
M	Data Median Best10		0.287	0.139	0.314	0.540	0.253	0.107	1.666	0.455
M	Data Average Best05		0.241	0.146	0.489	0.634	0.294	0.177	1.703	0.503
M	Data Median Best05		0.242	0.122	0.531	0.635	0.342	0.121	1.673	0.504
Trimming based on All Years' Best Montly Forecasters (= computed recursively, i.e. expanding window)										
M	Data Average Best"-2"		0.258	0.167	0.227	0.371	0.206	0.168	1.575	0.407
M	Data Median Best"-2"		0.282	0.175	0.227	0.434	0.182	0.140	1.642	0.418
M	Data Average Best30		0.278	0.164	0.241	0.394	0.204	0.158	1.572	0.416
M	Data Median Best30		0.295	0.175	0.239	0.451	0.175	0.131	1.629	0.421
M	Data Average Best20		0.288	0.155	0.260	0.398	0.205	0.152	1.627	0.430
M	Data Median Best20		0.320	0.172	0.255	0.447	0.174	0.127	1.683	0.432
M	Data Average Best10		0.271	0.138	0.344	0.362	0.223	0.167	1.593	0.440
M	Data Median Best10		0.268	0.134	0.348	0.417	0.232	0.133	1.656	0.439
M	Data Average Best05		0.238	0.144	0.372	0.353	0.224	0.144	1.586	0.431
M	Data Median Best05		0.217	0.133	0.323	0.396	0.275	0.113	1.740	0.449

Figure 9.6: Trimming: comparison of window length (monthly based)

Performance of different forecast combination methods measured in MSFE

Method		Year	2002	2003	2004	2005	2006	2007	2008	2001-2008
M	Data Average Benchmark		0.233	0.148	0.282	0.330	0.236	0.214	1.427	0.394
M	Data Median Benchmark		0.273	0.173	0.235	0.426	0.184	0.153	1.590	0.410
Trimming based on Last Year's Best Montly Forecasters (= rolling window length of 1 year)										
M	Data Average Best"-2"		0.227	0.165	0.228	0.374	0.217	0.160	1.590	0.401
M	Data Median Best"-2"		0.279	0.175	0.230	0.442	0.186	0.141	1.660	0.415
M	Data Average Best30		0.208	0.155	0.248	0.420	0.242	0.144	1.583	0.410
M	Data Median Best30		0.252	0.175	0.234	0.444	0.199	0.129	1.670	0.414
M	Data Average Best20		0.180	0.131	0.288	0.490	0.268	0.106	1.772	0.439
M	Data Median Best20		0.212	0.150	0.276	0.466	0.257	0.099	1.777	0.429
M	Data Average Best10		0.126	0.099	0.433	0.503	0.387	0.119	1.886	0.481
M	Data Median Best10		0.175	0.105	0.383	0.464	0.415	0.099	1.856	0.464
M	Data Average Best05		0.073	0.116	0.426	0.513	0.463	0.100	1.885	0.488
M	Data Median Best05		0.108	0.133	0.433	0.458	0.513	0.073	1.864	0.483
Trimming based on Last 3 Years' Best Monthly Forecasters (= rolling window length of 3 years)										
M	Data Average Best"-2"		0.267	0.170	0.226	0.369	0.211	0.169	1.566	0.408
M	Data Median Best"-2"		0.286	0.173	0.227	0.436	0.183	0.143	1.614	0.416
M	Data Average Best30		0.278	0.163	0.232	0.417	0.221	0.158	1.564	0.418
M	Data Median Best30		0.291	0.167	0.229	0.455	0.189	0.130	1.617	0.419
M	Data Average Best20		0.278	0.151	0.260	0.444	0.220	0.138	1.617	0.430
M	Data Median Best20		0.307	0.162	0.252	0.486	0.200	0.117	1.651	0.431
M	Data Average Best10		0.279	0.142	0.344	0.503	0.266	0.153	1.668	0.471
M	Data Median Best10		0.287	0.139	0.314	0.540	0.253	0.107	1.666	0.455
M	Data Average Best05		0.241	0.146	0.489	0.634	0.294	0.177	1.703	0.503
M	Data Median Best05		0.242	0.122	0.531	0.635	0.342	0.121	1.673	0.504
Trimming based on All Years' Best Montly Forecasters (= computed recursively, i.e. expanding window)										
M	Data Average Best"-2"		0.258	0.167	0.227	0.371	0.206	0.168	1.575	0.407
M	Data Median Best"-2"		0.282	0.175	0.227	0.434	0.182	0.140	1.642	0.418
M	Data Average Best30		0.278	0.164	0.241	0.394	0.204	0.158	1.572	0.416
M	Data Median Best30		0.295	0.175	0.239	0.451	0.175	0.131	1.629	0.421
M	Data Average Best20		0.288	0.155	0.260	0.398	0.205	0.152	1.627	0.430
M	Data Median Best20		0.320	0.172	0.255	0.447	0.174	0.127	1.683	0.432
M	Data Average Best10		0.271	0.138	0.344	0.362	0.223	0.167	1.593	0.440
M	Data Median Best10		0.268	0.134	0.348	0.417	0.232	0.133	1.656	0.439
M	Data Average Best05		0.238	0.144	0.372	0.353	0.224	0.144	1.586	0.431
M	Data Median Best05		0.217	0.133	0.323	0.396	0.275	0.113	1.740	0.449

Figure 9.7: Trimming: comparison of extent of trimming (monthly based)

Month-based vs. year-based forecast methods

Method		Year	2002	2003	2004	2005	2006	2007	2008	2001-2008
M-Y	Data Average Benchmark		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M-Y	Data Median Benchmark		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Trimming based on Last Year's Best Montly Forecasters (= rolling window length of 1 year)										
M-Y	Data Average Best"-2"		-0.05	0.03	-0.01	-0.01	0.02	-0.02	0.01	-0.01
M-Y	Data Median Best"-2"		-0.01	0.01	0.00	0.00	0.01	-0.01	0.04	-0.01
M-Y	Data Average Best30		-0.10	0.05	-0.01	0.04	0.05	-0.04	-0.07	-0.01
M-Y	Data Median Best30		-0.07	0.02	0.00	-0.03	0.02	-0.04	0.00	-0.03
M-Y	Data Average Best20		-0.17	0.03	-0.03	0.15	0.07	-0.13	-0.01	-0.02
M-Y	Data Median Best20		-0.15	-0.02	0.00	0.03	0.06	-0.11	-0.06	-0.05
M-Y	Data Average Best10		-0.18	0.01	0.05	0.21	0.17	-0.20	0.28	0.04
M-Y	Data Median Best10		-0.07	-0.04	0.02	0.20	0.19	-0.23	0.11	0.01
M-Y	Data Average Best05		-0.24	-0.04	0.04	0.29	0.21	-0.36	0.26	0.02
M-Y	Data Median Best05		-0.11	0.02	0.17	0.28	0.25	-0.40	0.10	0.04
Trimming based on Last 3 Years' Best Monthly Forecasters (= rolling window length of 3 years)										
M-Y	Data Average Best"-2"		0.00	0.00	-0.01	-0.01	0.01	-0.01	-0.02	-0.01
M-Y	Data Median Best"-2"		0.00	0.00	0.00	-0.01	0.00	-0.01	-0.01	-0.01
M-Y	Data Average Best30		0.00	-0.01	0.00	0.00	0.03	-0.02	-0.09	-0.01
M-Y	Data Median Best30		0.00	0.01	-0.01	-0.02	0.00	-0.02	-0.08	-0.02
M-Y	Data Average Best20		0.01	-0.02	-0.01	0.02	0.05	-0.04	-0.05	0.00
M-Y	Data Median Best20		0.04	0.00	-0.02	-0.03	0.02	-0.02	-0.04	-0.01
M-Y	Data Average Best10		0.01	-0.04	-0.01	0.13	0.10	-0.07	0.00	0.03
M-Y	Data Median Best10		0.05	-0.02	0.00	0.08	0.08	-0.12	0.05	0.02
M-Y	Data Average Best05		0.06	-0.04	0.06	0.38	0.08	-0.10	0.16	0.08
M-Y	Data Median Best05		0.04	-0.03	0.12	0.40	0.12	-0.14	0.12	0.09
Trimming based on All Years' Best Montly Forecasters (= computed recursively, i.e. expanding window)										
M-Y	Data Average Best"-2"		-0.02	0.00	0.00	-0.01	0.00	-0.01	0.01	-0.01
M-Y	Data Median Best"-2"		-0.01	0.01	0.00	-0.01	0.01	0.00	0.04	0.00
M-Y	Data Average Best30		0.00	0.00	0.03	0.00	0.02	-0.02	-0.05	0.00
M-Y	Data Median Best30		0.00	0.01	0.03	-0.02	0.02	-0.01	-0.06	-0.01
M-Y	Data Average Best20		0.00	0.01	-0.01	0.07	0.04	-0.05	0.01	0.01
M-Y	Data Median Best20		0.03	0.02	-0.03	0.00	0.02	-0.05	0.02	-0.01
M-Y	Data Average Best10		-0.04	-0.01	0.06	0.07	0.02	-0.04	0.12	0.04
M-Y	Data Median Best10		-0.01	-0.01	0.01	0.02	0.05	-0.06	0.10	0.01
M-Y	Data Average Best05		-0.07	0.01	0.13	0.02	0.09	-0.06	0.57	0.10
M-Y	Data Median Best05		-0.08	0.00	0.00	-0.03	0.12	-0.14	0.86	0.10

Figure 9.8: Trimming competition: monthly-based vs. yearly-based

Organisation

(Source: HMT, March 2010)

Bank of America - Merrill Lynch	Global Insight
Barclays Capital	Goldman Sachs
Beacon Economic Forecasting	HSBC
BNP Paribas	ING Financial Markets
British Chambers of Commerce	ITEM club
Cambridge Econometrics	J P Morgan
Capital Economics	Liverpool Macro Research
Citigroup	Lombard Street Research
CBI	Morgan Stanley
CEBR	NIESR
Commerzbank	Oxford Economics
Credit Suisse	Royal Bank of Canada
Daiwa Capital Markets	Royal Bank of Scotland
Deutsche Bank	Schroders Investment Management
Experian Business Strategies	Societe Generale
EC	Standard Chartered Bank
EIU	UBS
Economic Perspectives	

Figure 9.9: List of agencies

Bibliography

[1] World economic outlook, 1998-2009.

[2] M. Aiolfi and C. A. Favero. Model uncertainty, thick modeling and the predictability of stock returns. *Journal of Forecasting*, 24:233–254, 2005.

[3] J. S. Armstrong. Combining forecasts: The end of the beginning or the beginning of the end? *International Journal of Forecasting*, 5(4):585–588, 2001.

[4] M. Ashiya. Forecast accuracy and product differentiation of japanese institutional forecasters. *International Journal of Forecasting*, 22(2):395 – 401, 2006.

[5] R. Batchelor. How useful are the forecasts of intergovernmental agencies? the imf and oecd versus the consensus. *Applied Economics*, 33:225–235, 2001.

[6] J. M. Bates and C. W. J. Granger. The combination of forecasts. *Operational Research Society*, 20:451–468, 1969.

[7] R. T. Clemen. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583, 1989.

[8] M. P. Clements and D. I. Harvey. Forecast combination and encompassing. *Palgrave Handbook of Econometrics: Volume 2 Applied Econometrics,*, 2, 2008.

[9] M. P. Clements and D. F. Hendry. *Forecasting economic time series.* Cambridge University Press, 2000.

[10] M. Costantini and R. Kunst. Combining forecasts based on multiple encompassing tests in a macroeconomic core system. *Economics Series, Institute for Advanced Studies, Vienna*, 2009.

[11] C.W.J. Granger and Y. Jeon. Thick modeling. *Economic Modelling*, 21:323–343, 2004.

[12] S. Gupta and P.C. Wilton. Combination of forecasts: An extension. *Management Science*, 33:356–372, 1987.

[13] D.F. Hendry and M.P. Clements. Pooling of forecasts. *Econometrics Journal*, 5:1–26, 2002.

[14] S. Makridakis and et al. The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, 1:111–153, 1982.

[15] S. Makridakis and R.L. Winkler. Averages of forecasts: Some empirical results. *Management Science*, 29:987–996, 1983.

[16] P. Newbold and C.W.J. Granger. Experience with forecasting univariate time series and the combination of forecasts. *Journal of Royal Statistical Society A*, 137:131–146, 1974.

[17] F. C. Palm and A. Zellner. To combine or not to combine? issues of combining forecasts. *Journal of Forecasting*, 11::687–701, 1992.

[18] J.H. Stock and M. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. *NBER Working Paper Series*, Working Paper 6607, 1998.

[19] J.H. Stock and M. Watson. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23:405–430, 2004.

[20] A. G. Timmermann. Forecast combinations. *SSRN eLibrary*, November 2005.

[21] R.L. Winkler and S. Makridakis. The combination of forecasts. *Journal of the Royal Statistical Society Series A*, 146:150–57, 1983.

[22] V Zarnowitz. The accuracy of individual and group forecasts from business outlook surveys. *of Forecasting,*, 3:11–26, 1984.

[23] V. Zarnowitz and P. Braun. Twenty-two years of the nber-asa quarterly outlook surveys: aspects and comparisons of forecast performance. *NBER Working Paper*, 3965, 1992.

Abstract

This work is divided in two parts. A first part concentrates on the evaluation of the forecast accuracy of around 50 agencies respective the forecasts of UK GDP growth rates in the last eleven years. For this goal, around 130 publications of "Forecasts for the UK economy" from the HM treasury were assembled in order to construct a usable data set. The individual forecasts are then used in a second part to generate combined forecasts according to the forecast combination literature. Different combination schemes are applied and compared with each other. The findings mainly confirm the evidence of comparable empirical studies.

Zusammenfassung

Diese Arbeit ist zweigeteilt. Im ersten Teil werden Institute, die sich an der Prognose zentraler, ökonomischer Kennzahlen beteiligen, auf ihre Prognosegenauigkeit hin untersucht. Zu diesem Zweck wurde ein Datensatz erstellt, der BIP-Prognosen für Großbritannien von ca. 50 verschiedenen Instituten über die letzten 11 Jahre zusammenfasst. Grundlage dieses Datensatzes sind 130 Veröffentlichungen des britischen Finanzministeriums (HMT). Im zweiten Teil wird durch Kombinationen dieser Prognosen versucht, den tatsächlichen Wert noch genauer vorherzusagen. Um dieses Ziel zu erreichen, werden verschiedene Methoden angewandt und miteinander verglichen. Die Ergebnisse bestätigen größtenteils die empirischen Befunde anderer Arbeiten.

Curriculum Vitae

Name:	Sebastian Paul Koch
Date of Birth:	03.09.1980
Nationality:	German
Email:	sebastianpaulkoch@gmail.com

EDUCATION

10/07 - present **Master in Economics at the University of Vienna, Austria**
› Title of the master thesis: "Forecast Evaluation and Forecast Combination", Supervisor Prof. Robert Kunst

10/01 - 09/06 **Bachelor of Science in Economics and Management at the Free University of Bolzano / Bozen, Italy**
› Trilingual studies: English, Italian and German

02/04 - 08/04 **Studies at the Universitat Pompeu Fabra in Barcelona, Spain**
› ERASMUS-Scholarship

WORK EXPERIENCE

08/08 – present **IHS – Institut für Höhere Studien, Vienna, Austria**
› Student assistant in the "Economics and Finance" department
› Assistant on different projects and studies, mainly in the area of regional economics and macroeconomics

08/07 – 10/07 **RWI - Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Essen, Germany**
› Student assistant in the department of „Empirical Industrial Economics"
› Assistant on the „Innovation report 2007" and the „Report on the German technological performance"

SCIENTIFIC REPORTS AND PUBLICATIONS (at IHS)

Wolfgang Polasek, Wolfgang Schwarzbauer, Richard Sellner, Sebastian Koch, Teresa Körner. (2010 forthcoming): *Does Globalisation affect regional growth?* IHS-Study with support of the „Jubiläumsfonds" of the Austrian National Bank.

Ulrich Schuh, Iain Paterson, Karin Schönpflug, Sebastian Koch. (2010 forthcoming): *Effizienz-potenziale in der Verwendung öffentlicher Mittel*, IHS-Study for the Austrian chamber of commerce.

Iain Paterson, Nikolaus Graf, Richard Sellner, Wolfgang Schwarzbauer, Sebastian Koch (2008): *Wettbewerb und Wohlstand.* IHS-Study for the Management Club Vienna.

Nikolaus Graf, Sebastian Koch, Iain Paterson, Wolfgang Schwarzbauer, Richard Sellner. (2008): IHS-Study about the employment opportunities in the services sector.

SCIENTIFIC ASSISTANCE FOR REPORTS AND PUBLICATIONS (at IHS)

Wolfgang Polasek, Wolfgang Schwarzbauer, Richard Sellner, Teresa Körner. (2010): *Volks-wirtschaftliche Bewertung der Projekte des Rahmenplans 2009-2014 in der Betriebsphase.* IHS-Study for the ÖEBB.

Alexander Schnabl, Wolfgang Schwarzbauer, Sarah Dippenaar, Teresa Körner, Sandra Müll-bacher, Wolfgang Polasek, Richard Sellner, Isabella Skrivanek, Irene Weberberger. (2009): *Bewertung der volkswirtschaftlichen Effekte neuer Straßeninfrastruktur.* IHS-Study for the ASFINAG.