TI 2008-102/3
Tinbergen Institute Discussion Paper

Beating the Random Walk: A Performance Assessment of Long-term Interest Rate Forecasts

Frank A.G. den Butter[1]
Pieter W. Jansen[2]

[1] *Department of Economics, VU University Amsterdam, and Tinbergen Institute;*
[2] *Aegon Investment Management.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

This version 27 October 2008

Beating the random walk: a performance assessment of long-term interest rate forecasts

Frank A.G. den Butter and Pieter W. Jansen*

Abstract
This paper assesses the performance of a number of long-term interest rate forecast approaches, namely time series models, structural economic models, expert forecasts and combinations thereof. The predictive performance of these approaches is compared using out of sample forecast errors, where a random walk forecast acts as benchmark. It is found that for five major OECD countries, namely United States, Germany, United Kingdom, The Netherlands and Japan, the other forecasting approaches do not outperform the random walk, or a somewhat more sophisticated time series model, on a 3 month forecast horizon. On a 12 month forecast horizon the random walk model can be outperformed by a model that combines economic data and expert forecasts. Here several methods of combination are considered: equal weights, optimized weights and weights based on forecast error. It appears that the additional information contents of the structural models and expert knowledge is only relevant for forecasting 12 months ahead.

Keywords: interest rate forecasting, expert knowledge, combining forecasts, optimizing forecast errors

JEL-codes: C53, E27, E43, E47

* Den Butter is professor of economics at the Vrije Universiteit, Department of Economics, De Boelelaan 1105, NL-1081 HV Amsterdam, The Netherlands, tel. +31-20-5986044, fbutter@feweb.vu.nl; Jansen is Investment Strategist at AEGON Investment Management and guest researcher at the Department of Economics at the Vrije Universiteit,.

Beating the random walk: a performance assessment of long-term interest rate forecasts

Frank A.G. den Butter and Pieter W. Jansen

1. Introduction

Future developments of long-term interest rates are key to strategic decision making by economic agents. Financial markets analysts have set up a whole “industry” of interest rate forecasting. Even though the specialists in this area all try to develop a view on future interest rate movements, they do so for very different reasons. For instance, investors want to know the direction of interest rates so they can increase the performance of their investment portfolio. Government bond agencies on the other hand, predict the interest rate to estimate financing costs and determine when is a good time to fund their capital needs.

Mainstream economic literature distinguishes, broadly speaking, three methods of economic forecasting, namely time series models, structural models and forecasts that are (also) based on expert knowledge. The latter category uses tacit knowledge, based on intuition and experience. Quite often experts use model outcomes in combination with other factors they consider relevant for expected interest rate developments. It makes forecasting a mixture of science and art (Hendry and Clements (2003)). Science represents the econometric systems that embody consolidated economic knowledge but art (judgement) plays an important role as well.

This paper aims to assess the quality of the various economic forecasting methods by comparing the outside sample errors of long-term interest rate predictions with the random walk as benchmark prediction method. Such comparison is somewhat hindered by the fact that much interest rate forecasting is conducted in the private sector, where, in general, the models or methodologies are not published. Especially in the investment industry it is unlikely that successful interest rate forecasters would like to inform competitors about the quality of their model. In order to include as yet the quality of these forecasting models in our assessment we use the long-term interest forecasts for a large group of private forecasters, collected and published by Consensus Economics. These consensus forecasts can be seen as the output of the forecasting methodologies of experts.

The information contents of the various forecasting methods may not completely overlap. In that case combining forecasts makes use of the additional information content contained in the individual methods, so that a combined forecast is likely to outperform the individual forecasts (Bates and Granger, 1969). Therefore we also construct combined forecasts using different weighting schemes and compare the quality of these forecasts with the forecasts which stem from the single methodologies. Empirical research has convincingly shown that combining forecasts leads to a better forecast performance (e.g. Hendry and Clements (2004), Aiolfi and Timmermann (2004), and Timmermann (2005)).

Our paper is in line with other studies which compare interest rate forecasting methods (see e.g. Fauvel et al. (1999)). Pooter et al. (2007) also discuss different model specifications, but do so with regards to the term structure of interest rates. Chun (2008) surveys individual expert forecasts across the term structure. Our analysis extends these studies. Firstly, we consider long-term interest rates, whereas most studies focus on short-term interest rate or on the term premia. Secondly, we compare macroeconomic causal models, expert models and time series models in relation to long-term interest rate forecasting. This comparison relates to forecasting performance across countries (United States, Germany, United Kingdom, The Netherlands and Japan) and considers the difference between two forecast horizons: a 3 month forecast period and a 12 month forecast period.

This paper is structured as follows. Section 2 discusses various aims of interest rate forecasting. Section 3 describes our practical implementation of the forecasting models and the set up of the performance assessment. This section also introduces the benchmarks for our interest rate forecast models, i.e. the random walk model, and as alternative a somewhat more sophisticated times series model. Section 4 discusses the outcome of the expert based forecasts. Section 5 considers the structural forecasting models, where long-term interest rates forecasts stem from causal economic relationship using information on leading indicators as determinants for long-term interest rates. Section 6 assesses whether pooling of forecast models improves the predictive performance. Section 7 provides the final comparative quality judgment. Here we compare the forecasting performance of all 4 forecasting methodologies with the benchmark models. Finally, section 8 concludes.

2. Aims of interest rate forecasts

For an assessment of the quality of interest rate forecasts, it is necessary to find out about the aim of the forecasting exercise. The introduction already mentioned that economic agents have different aims in their use of interest rate forecasts. A comparative analysis of the quality of the forecasts should be aware of these different aims. A first difference relates to the time horizon of the forecast. Some investors base their investment decisions on very short horizons, others focus more on a longer time horizon. The first group can be categorised as tactical investors and the second group as strategic investors. In the short run view, interest rate volatility determines the change in bond prices and hence the bond investment return. It is relevant for the asset portfolio. However, the interest rate also feeds into the valuation of pension liabilities. Pension funds discount their future liability at the long-term interest rate.

Besides investors, interest rate forecasts are also relevant for borrowers. For instance government agencies, people that hold a mortgage on their property or corporations that are considering to finance an investment plan all take into account the interest rate forecast.

Different aims imply that different loss functions are to be considered when assessing the quality of the forecasts. Overall, the investor will be more worried about a forecast error that leads to a lower performance and mortgage holder will be more hurt by higher finance costs than anticipated. The risks are a-symmetric. In our study we do not focus on interest rate forecast for any specific purposes. Therefore, in the assessments of interest rate forecasting models, we take the “goodness of fit” as the criterion for comparison. In our comparative analysis models with the lowest (outside sample) forecast error are considered as the “best” forecasting models so that the mean squared forecast error acts as loss function. Moreover, in the last part of this paper, we also consider whether models are able to capture the direction of interest rate changes.

3. Methods and models for interest rate forecasts

As mentioned before this paper considers three categories of forecasting models:
1. Time series models
2. Structural models
3. Structural models cum additional information and tacit knowledge (intuition)

Of course, each of these categories can be split up into various subcategories.

With respect to the time series models, where long term interest rates are explained solely by past values of these rates, we restrict ourselves to the class of ARIMA-models, which are commonly used for (interest rate) forecasting. We select the models with the best possible fit given the usual identification and specification procedure of these models. We do not include structural time series models in our analysis as identification of the various components of the interest rates time series does not seem appropriate for our purposes.

With respect to structural econometric models where long-term interest rates are explained, along with other economic variables of interest, by causal relationships, a large variety in techniques and methodologies can be used: single equation approaches, structural systems of equations models, VAR, VECM models and non-linear neural network models. In this comparative analysis we concentrate on single equations and the VAR approach. There are also differences in the amount of data which is used to apply the structural models. On the one hand, there is a tendency in empirical research towards working with large datasets (see for instance Ludvigson and Ng (2005), Pooter et al (2007) and Stock and Watson (1998 and 2005)). Stock and Watson (2005), for instance, argue that with a large data set more information is included which protects the robustness of the model against structural instability. On the other hand, Boivin and Ng (2003) point out that smaller, pre-screened, datasets can lead to better results. As we elaborate in section 5, we follow Den Butter and Jansen (2004) in incorporating macro economic data in causal models. It means that we base our model specifications on a relatively small group of a qualified dataset.

In our empirical approach, we specify and compare models from all three categories, but use the time series models as a benchmark model in order to investigate whether the incorporating the additional information of the structural models and expert forecasts pays off to get a better forecasting performance. In fact we use two time series models as benchmark, namely the simple random walk model and an adequately specified ARIMA model.

The simple random walk model is specified as follows:

(1) $R_{l+n} = R_t + \varepsilon$

where R_{l+n} represents the interest rate at month t+n, R_t the interest rate at month t and ε is the disturbance term. ε as a mean of 0 and an expected value of 0. This implies that the expected value of R at time t+n is equal to R_t.

(2) $\mathrm{E}(R_{l+n}) = R_t$

Table 1 summarizes the mean squared forecast errors (MSFE) of random walk model for the two forecast horizons (3 months and 12 months) for the five countries of our study. The MSFE are presented in basispoints, where one basispoint corresponds to 1/100 of a percent. The table shows that, as should intuitively be suspected, the MSFE is higher for the longer period (18.3 to 36.8 basispoints) while the MSFE is quite low for the 3 month forecast period (6.2 to 15.3 basispoints).

Table 1: Mean squared error of random walk model in basispoints (1/100 of a percent) in the out of sample period (2003:5 – 2008:3)

	3 month forecast	12 month forecast
United States	15.3	36.8
Germany	8.4	37.6
United Kingdom	9.4	21.2
The Netherlands	6.2	39.3
Japan	8.1	18.3

As our other benchmark model we select the best fitting specification from the class of ARIMA models. E.g. Fauvel et (1999) find that ARIMA models are satisfactory and useful for interest rate forecasting. Application of the usual specification selection procedure leads to two alternative specifications of an ARIMA model, namely the AR(2) model:

(3) $R_t = \beta_1 * R_{t-1} + \beta_2 * R_{t-2_{\varepsilon}} + \varepsilon_t$

The estimation results of equation (3) are summarized in table X. The table shows the results for both the 3 month and 12 month forecast horizon. The second column shows which lags have been incorporated. The first lag to which referred is labelled variable C(1) and the second C(2). Both are variables, as described in equation (3), are lagged long-term interest rates.

Table 2: forecast results of ARIMA equation

	Lags (months) of variables C(1) and C(2)	C(1)	C(2)	R-squared	Akaike
3m forecast					
US	-3,-4	1.33 (8.42)	-0.35 (-2.20)	0.860	1.32
Germany	-3,-4	1.58 (10.2)	-0.58 (-3.78)	0.934	1.06
Netherl.	-3,-4	1.56 (10.0)	-0.56 (-3.64)	0.936	0.83
Japan	-3,-4	1.15 (7.49)	-0.17 (-1.10)	0.938	1.50
UK	-3,-4	1.26 (8.16)	-0.27 (-1.74)	0.957	1.26
12m forecast					
US	-12	0.94 (85.8)	-0.61 (-1.58)	0.403	2.70
Germany	-12,-13	1.57 (4.05)	-0.89 (-2.23)	0.607	2.87
Netherl.	-12,-13	1.85 (4.63)	-0.39 (-1.28)	0.599	2.69
Japan	-12,-13	1.28 (4.21)	-0.20 (-0.60)	0.750	2.88
UK	-12,-13	1.15 (3.43)		0.794	2.79

and the ARIMA(2,1,2) model:

(4) $\Delta R_t = \beta_1 * \Delta R_{t-1} + \beta_2 * \Delta R_{t-2_\varepsilon} + \varepsilon_t + \beta_3 * \varepsilon_{t-1} + \beta_4 * \varepsilon_{t-2_\varepsilon}$

where Δ is the first difference of the corresponding variable.

The results of this estimation for the countries over two forecast horizons are shown in the table below.

Table 3: forecast results of ARIMA equation

	Lags (months) of variables C(1) and C(2)	C(1)	C(2)	R-squared	Akaike
3m forecast					
US	-5	-0.10 (-1.36)		-0.000	-0.14
Germany	-3,-4	0.16 (2.04)	0.06 (0.77)	0.028	-0.36
Netherl.	-3,-4	0.12 (1.46)	0.12 (1.48)	0.026	-0.61
Japan	-3,-4	-0.12 (-1.59)	-0.09 (-1.12)	0.018	0.12
UK	-3,-4	0.02 (0.21)	0.08 (0.97)	-0.009	-0.10
12m forecast					
US	-13	-0.15 (-1.93)		0.007	-0.14
Germany	-12,-13	-0.13 (-1.66)	-0.09 (-1.18)	0.007	-0.51
Netherl.	-12,-13	-0.12 (-1.53)	-0.08 (-1.07)	0.002	-0.70
Japan	-12,-13	0.03 (0.42)	-0.07 (-0.94)	-0.020	0.10
UK	-12,-13	-0.09 (-1.13)	-0.02 (-0.26)	-0.019	-0.14

The observation period for estimating the above monthly ARIMA-models is 1989:1-2003:4. Table 2 presents the MSFE scores of the AR(2) and ARIMA (2,1,0) models for the two forecast horizons in the outside sample period 2003:5 – 2008:3. Table 2 shows that the ARIMA(2,1,0) model overall performs slightly better than the AR(2) model. It should be noted that the ARIMA(2,1,0) model is a somewhat more sophisticated version of the benchmark random walk model, which reads as ARIMA(0,1,0). Comparison of the results of Table 1 and Table 2 shows that the outside sample forecasting performance of the ARIMA (2,1,0) model is slightly better than that of the random walk model. So estimating the additional four parameters seems to pay off when forecasting long-term interest rates, albeit that differences in performance are minor.

Table 2: Mean squared error in basispoints (1/100 of a percent) in the out of sample period (2003:5 – 2008:3).

	3 month forecast horizon		1 year forecast horizon	
	AR(2)	ARIMA(2,1,2)	AR(2)	ARIMA(2,1,2)
United States	16.6	15.4	40.9	35.5
Germany	9.9	8.7	36.5	35.8
United Kingdom	10.2	9.6	20.9	20.9
Netherlands	6.3	6.3	21.5	17.6
Japan	8.9	8.3	36.1	17.6

4. Expert forecasts

The opinion of the experts is a crucial component in applied economic forecasting. Franses et al (2007) point out that "official forecasts of international institutions are never purely model-based. Preliminary results of models are adjusted with expert opinions." This is also in line with a quotation in Fauvel et al (1999) where it is stated that " judgement is a heavy component in forecasting economic data". As a matter of fact the combination of model-based forecasts and expert knowledge illustrates how economic forecasting is partly science and partly an art. The tension between the science of forecasting and the art of forecasting was already noted by Samuelson (1975) in the following quotation:

> "When Robert Adams wrote a MIT-thesis on the accuracy of different forecasting methods, he found that 'being Sumner Slichter' was apparently one of the best methods known at that time. This was a scientific fact, but a *sad* scientific fact. For Slichter could not and did not pass on his art to an assistant or to a new generation of economists. It died with him, if indeed it did not slightly predecease him. What he hoped to get by scientific breakthrough is a way of substituting for men of genius, men of talent and even just run-of-the-mill men. That is the sense in which science is public, reproducible knowledge."

Hence Samuelson's main concern with forecasting is that a forecasting artist may outperform a forecasting scientist, while the art of forecasting is non-reproducible. The quotation suggests that, in that time, experts were indeed able to outperform forecasters who rely on models. In this vein Fair and Shiller (1989, 1990) compare the informational content in forecasts from economic models and in forecasts combining economic models with judgement. Such investigation does not only indicate to what extent forecasters use different information, or all relevant information available, but it also gives a clue for combining the forecasts (see also section 6) in order to improve them. In the same vein McNees (1990) investigates the extent to which judgement is helpful to improve mechanically generated forecasts. He concludes that the historical records suggest that judgemental adjustment improves the forecasts, despite instances of success of mechanically generated forecasts. Moreover he looks at whether forecasters who combine their forecasts with judgement overadjust or underadjust. In other words, whether they put too much or too little trust in the mechanically generated forecast from their models. McNees finds a slight overadjustment. The message therefore is that forecasters should adjust their models using judgement, but that they should be very careful about it. It is, according to McNees, a mistake to accept adjustments that are made at face value, especially when the adjustments appear without any explanation of the reasoning behind them.

A number of more recent studies focus more specifically on the quality of survey based financial forecasts of experts. The outcomes of these studies are mixed. Brooks and Gray (2003) evaluated the bond yield forecast that published on a semi annually basis in The Wall Street Journal. These authors conclude that the performance is poor: in 67% of the forecasts the directional forecast is wrong. Chun (2008) finds that for long maturity interest rates econometric models consistently outperform the survey forecasters over a forecast horizon of over 3 to 4 quarters. On a short term (1 to 2 months ahead) the random walk outperforms. According to Chun (2008) market based and survey based forecasts do quite well on a short time horizon.

Kolb and Stekler (1996) investigate whether there is a consensus amongst financial forecasters at all. It seems incorrect to speak of a consensus when there is a large diversion in views. Kolb and Stekler (1996) conclude that probably in 50 to 65% of the cases there is a consensus about the direction long-term interest rates are expected to move. The consensus forecast is in fact an (unweighted) average of a panel of individual forecast. Therefore, the consensus forecast is already a combined model when assuming that the forecasters that participate in the survey use different information and different models. Bauer et al (2003) compared the consensus forecast from the Blue Chip Consensus forecasts survey for a number of economic variables. They conclude that the average forecast performs better than the best forecaster. We will discuss the advantages of combining forecasts further in section 6.

A specific characteristic of expert forecasts, which hampers a comparative analysis, is that there is entry and exit of forecasters (see also Capistran and Timmermann (2006)). The group of forecasters is not constant throughout the sample period. Information through personal communication with financial institutions reveals that interest rate forecasts are much connected to a forecaster. The experts who are responsible for forecasting time series use their own models and interpretation. This suggests that forecasts are probably more closely linked to individual forecasters and not necessarily to the institution. This would make an assessment of forecasts of individual institutions troublesome. For that reason we only consider average expert forecasts in our assessment in order to test the overall ability of experts in the industry to forecast long-term interest rates.

Our data are expert forecasts for long-term interest rates published by Consensus Economics. These data are available on a monthly basis, but only provide forecasts three month ahead and a year ahead. That is why, in our comparative analysis, we confined our quality judgement to the 3 month and 12 month ahead forecasts, whereas we could have calculated forecasts for other horizons using our own models.

Table5 shows how often the experts were able to forecast the direction of long-term interest rate changes. In accordance to findings of Brooks and Gray (2003) the outcome is disappointing. On average, in less than half of the cases the average forecaster was able to predict the direction of expected change of long-term interest rates. This holds for both the 3 and 12 month horizon.

Table 5: Correctly forecasted the interest rate direction of consensus forecasts (% total)

	3 month horizon	12 month horizon
United States	47.7	50.5
Germany	37.8	45.1
United Kingdom	45.6	52.2
The Netherlands	49.4	45.6
Japan	45.1	40.8
Average	*45.1*	*46.8*

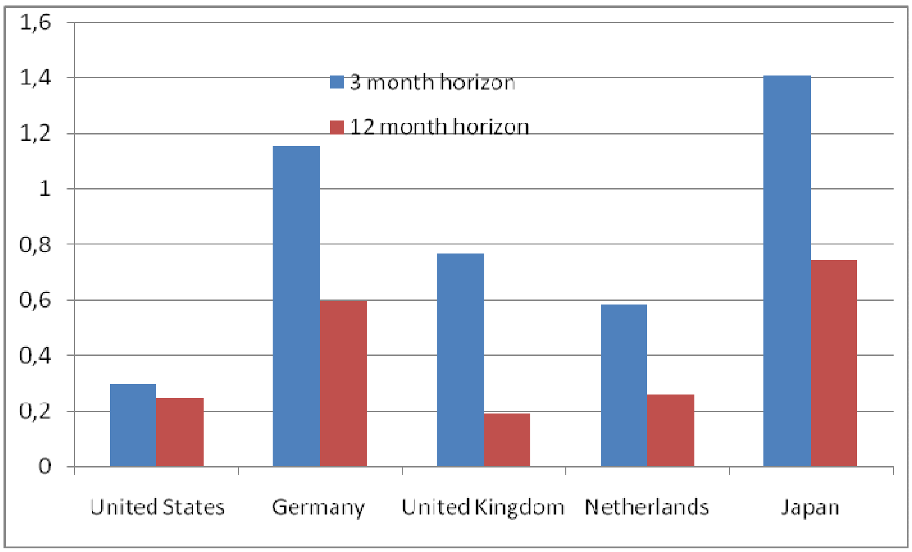
We have also tested the quality of the expert forecasts in the out of sample period. The results for the MSFE are summarised in the table 6. The table shows that on average the MSFE is equal to 22.5 basis points over a 3 month horizon and 58 basis points on a 12 month forecast horizon. The spread for the 12 month forecast horizon for the various countries is quite large.

Table 6: Mean Squared Forecast Error of consensus experts forecasts (out of sample period)

	3 month horizon	12 month horizon
United States	24.6	70.7
Germany	35.8	99.8
United Kingdom	17.0	27.2
The Netherlands	22.6	22.4
Japan	12.7	70.3
Average	*22.5*	*58.1*

One key element of the expert forecast data is that the expert forecast consistently underestimate interest rate changes. This may be the result of two factors. First, these survey data are averages of expert forecasts where, in case of diverging views, expected interest rate changes are smoothed out. A second reason could be that, given uncertainties, the interest rate forecast also reflects probabilities risk scenarios which are other than the base forecast. Chart 1shows ratios of the means square of the expert forecast (expected interest rate change) relative to the mean square of the actual interest rate changes. In all but two cases the ratio has a value of significantly lower than 1. It means that the expected interest rate change of the experts is smaller than the mean squared of the actual change. Hence the experts underestimate the size of interest rate changes. By the way, it should be noted that, when it is indeed hard to beat the random walk, this seems a wise strategy.

Chart 1: ratio of mean squared expected interest rate change versus mean square of actual interest rate changes



Yet it appears to be of interest to correct for the bias of the consensus forecast of experts to underestimate interest rate changes. For that reason we estimated for each country the following specification which relates the true interest change to the expected interest change:

(5) $R_{t+n} - R_t = \alpha + \beta_1 * (E(R_{t+n}) - R_t) + \varepsilon$

In this equation the coefficient α symbolizes the structural bias in the expert forecast. This is quite similar to the approach of Capistran and Timmermann (2006). They suggest that the best way to deal with the problem of entry and exit of forecasters is equal weighting and adding a constant variable to adjust for noise in the aggregate forecast. The expert forecast dataset we use is equally weighted. Table 7 shows the MSFE for the out of sample period of this corrected expert consensus forecast for both time horizons. For nearly all countries the MSFE is lower in the out of sample period for the corrected expert forecast than for the expert forecast of table 4.

How do the corrected expert perform in relation to the benchmarks? Table 7 shows that the average MSFE of the models of expert forecasts is higher both on a 3 and 12 month forecast horizon than the random walk and the ARIMA (2,1,0) model in the out of sample period. Even though the corrected expert forecasts perform better than the original expert forecasts, the corrected expert forecasts are still not good enough to beat the benchmarks in the out of sample period.

Table 7: Mean Squared Forecast Error of corrected expert forecast (out of sample period)

	3 month horizon			12 month horizon		
	Expert corrected	Random walk	ARIMA (2,1,2)	Expert corrected	Random walk	ARIMA (2,1,2)
United States	18.1	15.3	15.4	3.2	36.8	35.5
Germany	9.5	8.4	8.7	9.9	37.6	35.8
United Kingdom	10.7	9.4	9.6	17.0	21.2	20.9
Netherlands	15.9	6.2	6.3	98.7	39.3	17.6
Japan	7.4	8.1	8.3	33.1	18.3	17.6
Average	12.3	9.5	9.7	32.4	30.6	25.5

It is remarkable that correcting the expert forecasts for underestimation bias also increased the score of these forecasts with respect to the direction of the interest change. This is illustrated in table 8. Whereas the expert forecasts without bias correction indicated the right direction below 50% of the cases in the out of sample period, the corrected forecasts obtain a score is above 50%. They especially improve for the 12 month horizon.

Table 8: Number of periods in which direction of forecast was correct as a share of total forecast periods

	Original expert forecast		Corrected expert forecast	
	In sample	Out of sample	In sample	Out of sample
3 month horizon				
United States	42.9	58.3	57.9	55.0
Germany	33.8	46.7	57.1	51.7
United Kingdom	45.1	48.3	64.7	51.7
The Netherlands	50.0	46.7	69.2	51.7
Japan	42.9	50.0	69.2	51.7
Average	*42,9*	*50,0*	*63,6*	*52,4*
12 month horizon				
United States	52.4	46.7	70.2	70.0
Germany	53.2	50.0	68.5	63.3
United Kingdom	43.8	48.3	64.5	65.0
The Netherlands	44.4	46.7	68.6	51.7
Japan	36.3	50.0	70.2	63.3
Average	*46,0*	*48,3*	*68,4*	*62,7*

5. Structural model forecasts

Most structural econometric models used for macroeconomic policy analysis contain interest rate equations which can be used for forecasting interest rates. There are also a number of studies which focus more specifically on interest rate forecasting, using causal economic relationships. Ludvigson and Ng (2005) detect several economic variables that have forecasting ability for future excess returns on US government bonds. Also Ang and Piazzesi (2003) find evidence of macro economic factors for long-term interest rates (even though forecasting power is weaker for long maturity bonds than shorter maturity bonds). Pooter et al (2007) show that macro economic factors improve the forecast performance as far as term structure models are concerned. Dewachter and Lyrio (2006) find that long-run inflation expectations are important in the modelling of long-term bond yields, although this relationship is better in explaining than on forecasting interest rates, Finally, Bikbov and Chernov (2006) also find that macro factors and the term structure are useful for forecasting long-term bond yields.

Here we follow the methodology of Den Butter and Jansen (2004) who estimated structural long-term interest rate equations through encompassing five partial interest rate theories, namely the theories of the interest rate parity, term structure theory, classical theory of capital, Fisher's interest rate theory and portfolio theory. The explanatory variables of their specification can be used for forecasting as well. We have selected variables that are available on a monthly basis that link with these theories. The following variables have been taken as possible explanatory variables in our interest rate equation: short-term interest rate, inflation, oil price, leading indicator for economic activity, earnings return on equities and finally a lagged long-term interest rate variable.

Two alternative structural models are specified and estimated using the above explanatory variables, namely a VAR specification and a single interest equation where changes in interest rates act as dependent variable.

The system of VAR equations incorporates all possible explanatory variables from interest rate theories mentioned above. We specified the VAR models with a maximum lag length of two month, so that:

(6) $Y_{1t} = \alpha + \beta_1 * Y_{1t-1} + \beta_2 * Y_{1t-2} + \beta_3 * X_{1t-1} + \beta_4 * X_{1t-2} + \beta_5 * X_{2t-1} + \beta_6 * X_{2t-2} + \varepsilon$,

where :

(7) $X_{1t} = \alpha + \beta_1 * Y_{1t-1} + \beta_2 * Y_{1t-2} + \beta_3 * X_{2t-1} + \beta_4 * X_{2t-2} + \beta_5 * X_{2t-1} + \beta_6 * X_{2t-2} + \varepsilon$

When estimating these systems of VAR equations we did not look at the economic plausibility of the estimates, e.g. by a further structuring of the VAR models, but just accepted the estimation results whenever they are statistically feasible. Table 7 summarises the results the MSFE of the VAR equations. The table shows that on 3 month forecast horizon, the out of sample results are quite good as compared to the random walk model, especially for Japan and Germany. On a 12 month horizon the performance of the VAR model forecasts is substantially poorer, and shows a large variation between countries. This may be caused by the unstructured character of the VAR-models, so that on the longer run, the forecasts diverge strongly from the range of plausible realisations.

Table 9: Outside sample MSFE of VAR equations

	3 month horizon	12 month horizon
United States	33.4	174.0
Germany	6.2	52.2
United Kingdom	14.8	37.1
Netherlands	21.0	125.3
Japan	4.5	12.3
Average	*16.0*	*80.2*

For that reason we have estimated a second type of specifications to forecast long-term interest rates using the other macroeconomic variables suggested by interest rate theories. These structured interest equations use the information in the explanatory variables as follows. First it is investigated how a change in a level of a variable over a period preceding the time of forecast is able to explain interest rate movements over the two forecast horizons. This specification is as follows for a one variable model:

(8) $R_{t+n} - R_t = \alpha + \beta_1 * (X_{t-1} - X_{t-p}) + \varepsilon$

In this equation we optimize lag length p, considering changes in the level of the explanatory variables form levels at the month of forecast up to 6 months preceding the month of forecast. This means that when we forecast the interest rate at month t for t+3 we analyse changes in our explanatory dataset from t-1 to t=t up to t-6 to t=t. The result is that through this method both our set of explanatory variables and actual and forecast interest rate changes are stationary (which we tested through an Augmented Dickey-Fuller test).

Similar to Pooter et al (2007) we acknowledge that macro economic data is published with a delay and we follow Pooter et al in the specifying all real sector explanatory variables with a one month lag (see also the annex). However, such publication delay is not, in order to mimic the actual forecasting process, is not required for financial market variables (10yrs yield, 3month yield and the oil price), as they are published without delay.

Table 10 shows which variables have been incorporated in the country models. For each variable it shows how many months change (t – t-p) led to the best within sample performance. It also shows the t-value for each variable and finally the R-squared of the model and the Akaike information criterion. Leading indicators and oil prices are represented in most equations. On the other hand it appears that equity market developments do not improve the explanation by the models of long-term interest rates.

Table 10: Structured interest equations for 3 month ahead forecasts

		US	Germany	UK	Netherl	Japan
Leading indicator	no months change T-value	1 1.20	2 2.80		2 2.47	5 2.93
Inflation	no months change T-value				3 3.06	3 -2.86
Oil	no months change T-value		6 -2.00	4 -2.42	6 -3.78	3 -3.05
3month rate	no months change T-value	3 -2.19				6 3.44
10yr rate	no months change T-value	1 2.23	2 1.79	1 1.87	3 2.84	
C	T-value	-2.31	-0.07	-2.45	-2.68	-1.61
R-Squared		0.070	0.140	0.056	0.228	0.167
Akaike info		1.30	0.91	1.24	0.68	1.37

Table11 gives the results for the models which can be used to predict the change of the long term interest rate in the next 12 months. A striking difference with the 3 month horizon models is that the oil price appears not to provide any additional information to explain 12 months interest rate changes.

Table 11: Structured interest equations for 12 month ahead forecasts

		US	Germany	UK	Netherl.	Japan
Leading indicator	no months change T-value		3 1.99	2 1.47		
Inflation	no months change T-value		6 -2.83			
Price/Earnings ratio	no months change T-value			5 -1.33		
Oil	no months change T-value					
3month rate	no months change T-value			4 2.64		4 3.29
10yr rate	no months change T-value	6 -3.75			1 1.96	
C	T-value	-0.38	-0.34	-0.36	-3.94	-4.02
R-Squared		0.084	0.090	0.060	0.047	0.066
Akaike info		2.70	2.64	2.65	2.59	2.76

Table 12 presents the outside sample MSFE results for the structural equations of interest changes of tables10 and 11. For reasons of comparison also the MSFE results of the VAR models of table 9 are reproduced in table 12. The table shows that the structural interest equations yield, on average, better forecasts than the VAR models. The exception is the 3 month forecast for Japan, where the VAR model also outperforms the random walk. The table shows the poor performance of the VAR model as compared to the structured interest equation over the 12 month forecasting horizon.

Table 12: Outside sample MSFE of the VAR models and the structured interest equations

	3 month horizon		12 month horizon	
	Structured interest equation	VAR model	Structured interest equation	VAR model
United States	18.1	33.4	17.7	174.0
Germany	10.3	6.2	9.8	52.2
United Kingdom	10.1	14.8	11.8	37.1
Netherlands	12.4	21.0	10.8	125.3
Japan	10.9	4.5	9.1	12.3
Average	*12.4*	*16.0*	*11.8*	*80.2*

6. Combining interest rate forecast

In case the information content of forecasting methods differs, a combined forecast can be superior to both individual forecasts. Empirical evidence confirms that combined forecasts often lead to better forecasting performance. Hendry and Clements (2003) point out that structural breaks are major source of forecast failure. Timmermann (2005) adds that models differ due to a different information set and different modeling approaches, so that they may generate a diverging view on the occurrence of structural breaks. In that case combining forecasts can be seen as a form of diversification, which reduces the chance that the model is wrong about a structural breakdown.

Then, the question is how to combine separate forecasts. Methods that are used, apart from equal weighting, are historical performance weighted, optimal weighted, trimming, shrinkage, and time varying weights (see for instance Stock and Watson (2005), Chan, Stock and Watson (1999), Aiolfi and Timmermann (2004), Timmermann (2005), Poorter et al (2007) and Chun (2008)). Studies that use a large group of forecasts have found positive outcomes for trimming and shrinkage (Aiolfi and Timmermann (2004)). Other studies find positive results for equal weighting (Chan, Stock and Watson (1999) and Timmermann (2005)).

Following this literature, our performance assessment considers three ways of combining: (i) equal weighting, (ii) historical performance based weighting and (iii) optimized weighting. Method (ii) calculates the weights based on the relative strength of the forecasting error. The method with the lower forecasting error is given a higher weight. Method (iii) calculates the optimal weight within the in sample period regressing through an regression that minimizes the forecasting error. Using these three weighting schemes, we consider two types of combined forecasts, namely:
Combination 1: the corrected expert forecast combined with the structured interest equation;
Combination 2: the original expert forecast combined with the structured interest equation.

Table 13 gives the weights in the combined models according to the relative performance of the models in the sample period. It shows that, for a 3 month forecast period, the weight of the models based on the corrected expert forecast is on average larger than that of the structured interest equations. For the United Kingdom and The Netherlands the weights of the corrected expert forecasts is unity versus zero for the structured interest equation. It indicates that the structured interest rate equations, in the sample period contain no additional information vis a vis the corrected expert forecast. This is different for the other countries, where there seems to be a good balance between the models which indicates that both methods contain about the same amount of additional information. Apparently the correction for underestimation bias is largely responsible for the good contribution of the expert forecasts, as in Combination 2, with the uncorrected expert forecast, these forecast obtain relatively low weights as compared to the structured interest equation.

The results for the 12 month horizon are somewhat different. On average, in both combined forecasts, the structured interest rate equations obtain the largest weights. However, for the United States and Germany the corrected experts forecasts obtains a weight of one, so that here the structured interest rate equation provides no additional information (within the observation period) to the corrected expert forecasts. On the other hand, it is the other way around for the Netherlands. Also in case of the United Kingdom and Japan the structured interest rate equation contributes most to the combined forecast. With respect to Combination 2 it appears that all weights are about the same for the 12 month ahead forecast: the structured interest equation contributes almost 80% to the combined forecast and the uncorrected expert forecast somewhat over 20%.

Table 13: Model weightings based on within sample forecast errors

	3 month horizon				12 month horizon			
	Combination 1		Combination 2		Combination 1		Combination 2	
	Interest equation	Corr. expert forecast	Interest equation	Expert forecast	Interest equation	Corr. expert model	Interest equation	Expert forecast
United States	0.421	0.579	1.000	0.000	0.000	1.000	0.794	0.206
Germany	0.442	0.558	0.974	0.026	0.000	1.000	0.775	0.225
United Kingdom	0.000	1.000	0.964	0.036	0.819	0.181	0.736	0.264
Netherlands	0.000	1.000	0.991	0.009	1.000	0.000	0.797	0.203
Japan	0.357	0.643	0.919	0.081	0.813	0.187	0.775	0.225
Average	0.234	0.766	0.958	0.042	0.685	0.315	0.775	0.225

Below we present the MSFE of the combined forecast in three ways for both horizons. We present the results for combination 1: economic change models and models based on expert forecasts. The first column with MSFE findings (second and fifth column in the table) show the MSFE for the optimized weights. These weights are based on the weights presented in the previous table. The table also shows the MSFE in case we apply equal weights (50% economic change model and 50% model based on expert forecasts). Finally, we also present the MSFE in the out of sample period for a third measure of combining: performance based weights. The weight through this measure for the economic change model is the inverse of the ratio MSFE economic change model/(MSFE economic change model + MSFE model based on expert forecasts).

Table 14 compares the forecasting performance of the Combination 1 models with different weighting schemes in the outside sample period. It shows that the performance based weights yield the lowest MSFE for a 3 month forecast horizon (in the out of sample period) and that the optimized weighting system has the lowest MSFE for a 12 month forecast horizon.

Table 14: MSFE for Combination 1 models in out of sample period

	3 month forecast horizon			12 month forecast horizon		
	Optimized weights	Equal weights	Performance based weights	Optimized weights	Equal weights	Performance based weights
US	17.6	17.6	17.6	3.2	6.6	3.5
Germany	9.0	9.0	9.0	9.8	8.8	8.8
UK	10.7	9.9	9.9	12.6	14.1	13.6
Netherlands	16.0	12.1	12.0	10.8	19.1	15.1
Japan	7.5	8.2	7.9	12.4	42.7	12.9
Average	12.2	11.4	11.3	9.8	18.3	13.3

Table 15 gives the same indicators of forecasting performance for the Combination 2 models with the original instead of the corrected expert forecasts. Although these models give a lower weighting to the expert forecast and a higher weighting to structured interest equation forecasts, the conclusion almost the same as for the Combination 1 forecasts: the performance based weightings lead to better results in the out of sample period for the 3 month forecast and the optimized weights perform better with the 12 month forecast horizon. It is remarkable that the average level of MSFE is lower for Combination 2 than Combination 1, whereas the original expert forecasts perform worse than the corrected expert forecasts.

Table 15: MSFE for Combination 2 models in out of sample period

	3 month forecast horizon			12 month forecast horizon		
	Optimized weights	Equal weights	Performance based weights	Optimized weights	Equal weights	Performance based weights
US	18.3	15.0	15.0	5.9	9.0	5.8
Germany	9.5	10.8	8.8	3.4	15.9	5.6
UK	9.7	7.8	7.7	6.0	6.8	5.8
Netherlands	12.5	9.5	9.5	3.8	10.8	5.1
Japan	9.9	6.6	6.5	4.8	5.0	4.2
Average	11.9	9.9	9.5	4.8	9.5	5.3

7 Summary comparison and lessons for future forecasting exercises

Table 16 summarizes the results of all relevant forecasting models by comparing the MSFE of the 3 month forecasting horizon in the out of sample period 2004:5 – 2008:3. The best benchmark model is the random walk, but the ARIMA (2,1,0) model has a very similar average MSFE. We find similar low values for all countries which are under 10 basis points, except for the United States. It appears that almost all other models are unable to beat the random walk. A notable exception is Japan, where the corrected expert forecast, and 3 out of 6 combination models yield lower MSFE's than the random walk. Quite surprisingly the lowest value here is found for the VAR model, a model with a poor performance in most other countries except Germany.

All combined models outperform the best expert forecasts and the best structural models. Yet there is only one combination model, which even on average can match the random walk, namely the combination of the uncorrected expert forecasts and the structured interest rate equation, where the weights are calculated using within sample forecasting errors. This model beats the random walk in the United States, in the United Kingdom and in Japan. However, it is impossible to calculate whether such differences are statistically significant.

Table 16: MSFE in out of sample period for 3 month forecast horizon

	United States	Germany	United Kingdom	Nether-lands	Japan	Average
Benchmark						
1A Random walk	15.3	8.4	9.4	6.2	8.1	9.5
1B AR	16.6	9.9	10.2	6.3	8.9	10.4
1C ARIMA	15.4	8.7	9.6	6.3	8.3	9.7
Expert forecasts						
2A Corrected expert forecast	18.1	9.5	10.7	15.9	7.4	12.3
2B Original expert forecast	24.6	35.8	17.0	22.6	12.7	22.5
Macro driven models						
3A VAR	33.4	6.2	14.8	21.0	4.5	16.0
3B Structured interest equation	18.1	10.3	10.1	12.4	10.9	12.4
Combined models						
Optimal weights: 2A and 3B	17.6	9.0	10.7	16.0	7.5	12.2
Optimal weights: 2B and 3B	18.3	9.5	9.7	12.5	9.9	11.9
Equal weights: 2A and 3B	17.6	9.0	9.9	12.1	8.2	11.4
Equal weights: 2B and 3B	15.0	10.8	7.8	9.5	6.6	9.9
MSFE weights: 2A and 3B	17.6	9.0	9.9	12.0	7.9	11.3
MSFE weights: 2B and 3B	14.9	8.8	7.7	9.5	6.5	9.5

Table 17 gives the same summary indicators for comparing the forecasting performance on a 12 month forecast horizon. Here we already noted that the ARIMA (2,1,0) model performs best amongst the benchmark models in the out of sample period. This holds both for the average and for each individual country. Obviously the MSFE is substantial higher for a 12 month horizon than for a 3 month horizon because a longer lead time is more difficult to predict.

Here the benchmark time series models are more often outperformed by other models than in case of the 3 month horizon forecasts. With the corrected expert forecasts it is the case for the United States, Germany and the United Kingdom. The structured interest rate equation forecasts outperform the benchmark models for all countries considered. The same is true for the combined models, where in case of the 12 month forecasting horizon, the structured interest rate equations carry a large weight. It is again the model that combines the original expert forecasts with the structured interest equation that performs best.

Table 17: MSFE in out of sample period for 12 month forecast horizon

	United States	Germany	United Kingdom	Nether-lands	Japan	Average
Benchmark						
1A Random walk	36.8	37.6	21.2	39.2	18.3	30.6
1B AR	40.9	36.5	20.9	21.5	36.1	31.2
1C ARIMA	35.5	35.8	20.9	17.6	17.6	25.5
Expert forecasts						
2A Corrected expert forecast	3.2	9.9	17.0	98.7	33.1	32.4
2B Original expert forecast	70.7	99.8	27.2	22.4	70.3	58.1
Macro driven models						
3A VAR	174.0	52.2	37.1	125.3	12.3	80.2
3B structured interest equation	17.7	9.8	11.8	10.8	9.1	11.8
Combined models						
Optimal weights: 2A and 3B	3.2	9.8	12.6	10.8	12.4	9.8
Optimal weights: 2B and 3B	5.9	3.4	6.0	3.8	4.8	4.8
Equal weights: 2A and 3B	6.6	8.8	14.1	19.1	42.7	18.3
Equal weights: 2B and 3B	9.0	15.9	6.8	10.8	5.0	9.5
MSFE weights: 2A and 3B	3.5	8.8	13.6	15.1	12.9	13.3
MSFE weights: 2B and 3B	5.8	5.6	5.8	5.1	4.2	5.3

The MSFE as indicators to compare forecasting performance assume a quadratic and symmetric loss function with respect to forecast errors. As an alternative table 18 gives in percentages how often the models correctly predict the direction of change of the long-term interest rate. A percentage higher than 50% outperforms the throwing of a coin (here we include zero's for the random walk because it does not predict a change). We have presented these over 50% hits in bold characters. The table shows that for all countries and for both forecasting horizons the corrected expert forecasts beat the throwing of the coin. The same is true, except for the 3 month horizon in the United States, for the structured interest equation. The best overall performance is again obtained by the combination models, where the number of correctly forecasted interest rate changes is even higher for the 12 month horizon than for the 3 month horizon. Here on average two thirds of the forecasts of the direction of the interest rate change are correct.

Table 18: Successful directional forecast as a percentage of total forecasts in outside sample period

	3m horizon						12m horizon					
	US	GE	UK	NL	JP	avg	US	GE	UK	NL	JP	avg
Benchmark												
1A Random walk	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1B AR	50.0	50.0	**55.0**	**60.0**	50.0	**53.0**	**53.3**	50.0	**53.3**	**53.3**	**55.0**	**53.0**
1C ARIMA	**61.7**	48.3	43.3	50.0	**58.3**	**52,3**	**65.0**	45.0	**58.3**	50.0	46.7	**53.0**
Expert forecasts												
2A Corrected expert forecast	**55.0**	**51.7**	**51.7**	**51.7**	**51.7**	**52,4**	**70.0**	**63.3**	**65.0**	51.7	**63.3**	**62.7**
2B Original expert forecast	**58.3**	46.7	46.7	48.3	50.0	50.0	46.7	46.7	50.0	48.3	50.0	48.3
Macro driven models												
3A VAR	**61.7**	**58.3**	43.3	**53.3**	**65.0**	**56.3**	**51.7**	48.3	50.0	**51.7**	50.0	**50.3**
3B Structured interest rate equation	43.3	**60.0**	**55.0**	**56.7**	**53.3**	**53.7**	**73.3**	**68.3**	**68.3**	**68.3**	**60.0**	**67.6**
Combined models												
Optimal weights: 2A and 3B	**53.3**	**55.0**	**53.3**	**53.3**	**51.7**	**53.3**	**70.0**	**63.3**	**63.3**	**68.3**	**61.7**	**65.3**
Optimal weights: 2B and 3B	45.0	**58.3**	**56.7**	**58.3**	50.0	**53.7**	**68.3**	**65.0**	**61.7**	**68.3**	**65.0**	**65.7**
Equal weights: 2A and 3B	**53.3**	**53.3**	**56.7**	50.0	**53.3**	**53.3**	**68.3**	**66.7**	**65.0**	**56.7**	**61.7**	**63.7**
Equal weights: 2B and 3B	48.3	48.3	**53.3**	41.7	**61.7**	**50.7**	**71.7**	**58.3**	**70.0**	**65.0**	**63.3**	**65.7**
MSFE weights: 2A and 3B	**53.3**	**53.3**	**56.7**	50.0	**53.3**	**53,3**	**71.7**	**66.7**	**65.0**	**61.7**	**61.7**	**65.4**
MSFE weights: 2B and 3B	46.7	**58.3**	**56.7**	41.7	**60.0**	52.7	**68.3**	**66.7**	**63.3**	**68.3**	**61.7**	**65.7**

8. Conclusions

The comparative analysis of this paper shows that it is hard to beat the random walk when forecasting long-term interest rates. Especially when the forecasting horizon is relatively short – 3 month – the random walk model and another, somewhat more sophisticated ARIMA-model almost consistently outperform other forecasting methodologies investigated in this paper. A combination of expert consensus forecasts and forecasts calculated by structured interest equations, where the explanatory variables are suggested by interest rate theories, comes second best in our assessment based on the criterion of lowest squared forecast errors. These combined models appear to beat the random walk more often when the longer forecasting horizon of 12 month is looked upon. In that case the additional information of other relevant leading macroeconomic variables and of (tacit) expert knowledge seems to pay off. We acknowledge that our assessment relates to a specific reference period (1989:1 – 2003:4) and to a specific outside sample forecasting period (2003:5- 2008: 3), and that we only consider 5 major OECD countries with well developed capital markets (United States, Germany, United Kingdom, The Netherlands and Japan), so that in a strict sense, our results are only applicable to those periods and countries. However, the scope of our assessment seems sufficiently broad that our results may contribute to knowledge on the adequacy of forecasting methods. In that sense it corroborates with results from other studies that combination of forecasts may be useful to enhance the forecasting performance (see e.g. Den Butter and Van de Gevel (1989), Den Butter and Van Dijken 1997)).

The results of our assessment raises the question whether the whole industry of interest rate forecasting yields value for money now that it appears that simple time series models, which are to be constructed and maintained at low costs, show such good predictive performance on a relatively short horizon. Yet the fact that traders at the financial markets are willing to pay for these forecasts, and do not share them with their competitors, is a revealed preference. It may suggest that interest rate forecasts serve another aim than just be accurate with lowest possible forecast errors. Our alternative indicator of forecast performance, namely the relative amount of correct forecasts of the direction of change of the interest rate, shows that more sophisticated models using additional macroeconomic information and combining that information with expert knowledge, can be useful. A scope for future research is to see what alternative loss functions are relevant for interest rate forecasting at the financial markets. That would allow a cost benefit analysis of the forecasting industry.

Literature

Aiolfi, M and A.Timmermann, 2004, Persistence in forecasting performance and conditional combination strategies, *Journal of Econometrics*, 135, pp. 31-53.

Ang, A and M.Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics*, 50, pp 745-787.

Bates, J.M. and C.W.J. Granger, 1969, The combination of forecasts, *Operational Research Quarterly*, 20, pp. 451-468.

Bauer, A., R.A.Eisenbeis, D.F.Waggoner and T.Zha, 2003, Forecasting evaluation with cross-sectional data: The Blue Chip Survey, *Economic review*, Federal Reserve Bank of Atlanta, second quarter 2003.

Bikbov.R and M.Chernov, 2006, No-arbitrage Macroeconomic determinants, AFA 2006 Boston Meetings Paper and EFA 2005 Moscow Meetings Paper.

Boivin, J. and S.Ng, 2003, Are more data always better for factor analysis? NBER Working Paper no 9829.

Brooks, R. and B.Gray, 2003, History of the forecasters: an assessment of the semi-annual U.S.Treasury bond yield forecast survey as reported in The Wall Street Journal. Working Paper no 03-06-01, The University of Alabama.

Butter, F.A.G. den, and F.J.J.S. van de Gevel, 1989, Prediction of the Netherlands' money stock, *De Economist*, 137, pp. 173-201.

Butter, F.A.G. den, and S. van Dijken, 1997, The information contents of aggregated money demand in the EMU, *Open Economies Review* 8:3, pp. 233-244.

Butter, F.A.G. den, and P.W.Jansen, 2004, An empirical analysis of the German long-term interest rate, *Applied Financial Economics*, 14, pp.731-741.

Capistrán, C. and A.Timmermann, 2006, Forecast combination with entry and exit of experts. Banco de México working paper no 2006-08.

Chan, Y.L., J.H.Stock and M.W.Watson, 1999, A dynamic factor model framework for forecast combination, *Spanish Economic Review*, 1, 91-121.

Chun, A.L., 2008, Forecasting interest rates and inflation: Blue chip clairvoyants, econometrics or qrinkage? http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1102878.

Dewachter, H. and M.Lyrio, 2006, Macro factors and the term structure of interest rates, *Journal of Money, Credit and Banking*, 38-1, 120-140.

Fauvel, Y., A.Paquet and C.Zimmermann, 1999, A survey on interest rate forecasting. Working Paper no 87, Université du Québec à Montreál.

Fair, R.C. and R.J. Shiller, 1989, The informational content of ex ante forecasts, *Review of Economics and Statistics*, 71, pp. 325-331.

Fair, R.C. and R.J. Shiller, 1990, Comparing information in forecasts from econometric models, *American Economic Review*, 80, pp. 375-389.

Franses, P.H., H.Kranendonk and D.Lanser, 2007, On the optimality of expert-adjusted forecasts. CPB Discussion Paper no 92.

Hendry, D.F. and M.P.Clements, 2003, Economic forecasting: some lessons from recent research, *Economic Modelling*, 20, pp. 301-329.

Hendry, D.F. and M.P.Clements, 2004, Pooling of forecasts, *Econometrics Journal*, 7, pp. 1-31.

Kolb, R.A. and H.O.Stekler, 1996, Is there a consensus among financial forecasters? *International Journal of Forecasting*, 12, 455-464.

Ludvigson, S.C. and S.Ng, 2005, Macro factors in bond risk premia. NBER Working Paper no 11703.

McNees, S.K., 1990, Man vs. model? The role of judgement in forecasting, *New England Economic Review,* July/August 1990, pp. 41-52.

Pooter, M. de, F.Ravazzolo and D. van Dijk, 2007, Forecasting the term structure of interest rates. Tinbergen Institute Discussion Paper No. 2007-028/4.

Samuelson, P.A., 1975, The art and science of macromodels over 50 years, in G. Fromm and L.R. Klein (eds.), *The Brookings Model, Perspective and Recent Developments* (North-Holland, Amsterdam), pp. 3-10

Stock, J.H. and M.W.Watson, 1998, A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. NBER Working Paper no 6607.

Stock, J.H. and M.W.Watson, 2005, An empirical comparison of methods for forecasting using many predictors, http://econ.ucsd.edu/seminars/0506seminars/Stock_SP06.pdf.

Stock, J.H. and M.W.Watson, 2005, Forecasting with many predictors. In: Elliot, G., C.W.J. Granger and A.Timmermann, *Handbook of Economic Forecasting*, chapter 8.

Timmermann, A.G., 2005, Forecast combinations. CEPR Discussion Paper Series no 5361.

ANNEX

The economic data we have incorporated in our data set have monthly frequency. For data that are available at a daily frequency (such as the interest rate time series and the oil price) we have calculated monthly averages. The consensus economic forecast is published around the third week of the month and are collected throughout the month, which approximates a monthly average as well. We have delayed the economic data with one month, since most data is published with one month delay. This is required to make sure that the data is actually available at the time of the forecast.

The data series have been collected from Bloomberg and Thomson Financial Datastream. The consensus forecasts have been collected from Consensus Economics.