ARE FORECASTS REALLY RATIONAL?
ASSESSING PROFESSIONAL MACROECONOMIC FORECASTS FROM THE G7 COUNTRIES

JONAS DOVERN AND JOHANNES WEISSER

This version: March 27, 2009
PLEASE DO NOT QUOTE

Abstract. In this paper, we use survey data to analyze the rationality of professional macroeconomic forecasts. We analyze both individual forecasts and average forecasts. We provide evidence on the properties of forecasts for all the G7 counties and four different macroeconomic variables. Furthermore, we present a modification to the structural model which is commonly used to model the forecast errors of fixed event forecasts in the literature. Our results confirm that average forecasts should be used with caution, since even if all individual forecasts are rational the hypothesis of rationality is often rejected by the aggregate forecasts. We find that there are not only large differences in the performance of forecasters across countries but also across different macroeconomic variables; in general, forecasts tend to be biased in situations where forecasters have to learn about large structural shocks or gradual changes in the trend of a variable.

Keywords: Evaluating forecasts, Macroeconomic Forecasting, Rationality, Survey Data, Fixed-Event Forecasts

JEL Classification: C25,E32,E37

1. Introduction

In this paper, we use survey data to analyze the efficiency and unbiasedness of professional macroeconomic forecasts in the G7 countries. We analyze both individual forecasts and average forecasts, so-called consensus forecasts, from a large data set that has not been used in the literature before. By using this large amount of disaggregate data on individual macroeconomic forecasts, we are able to provide much broader evidence on the properties of macroeconomic forecasts than has been available in the literature so far.

One weak point of the empirical literature on survey data is that there is only a limited number of non-US data sets, which provide information on forecasts. Consequentially,

Jonas Dovern, The Kiel Institute for the World Economy (IfW), `jonas.dovern@ifw-kiel.de`. Johannes Weisser, Max Planck Institute of Economics, `weisser@econ.mpg.de`. The views presented in this paper reflect the authors' opinion, and do not necessarily coincide with those of the IfW or the Max Planck Institute of Economics. We are grateful to Helmut Herwartz and Christian Merkl as well as to all participants of the Brown-Bag Seminar at Kiel University for valuable comments and suggestions.

existing evidence is predominantly based on US data. Notable exceptions are Harvey et al. (2001), who analyze a set of selected individual forecasts for the UK from the *Consensus* data set, Gallo et al. (2002), who analyze the evolution of macroeconomic forecasts for the US, the UK, and Japan, Bowles et al. (2007), who analyze the performance of forecasts summarized in the Survey of Professional Forecasters conducted by the European Central Bank, Isiklar et al. (2006) or Ager et al. (2009), who use data from the *Consensus* data set on forecasts for a set of industrialized countries, Timmermann (2007), who analyzes the performance of IMF forecasts from the World Economic Outlook for various countries, and Batchelor (2001), who compares the forecasts made by the IMF and the OECD to private sector forecasts. However, all the existing international studies, with the exception of Harvey et al. (2001), make exclusive use of consensus forecasts rather than analyzing individual forecasts. This paper is written to fill this gap by covering forecasts for all G7 countries.

Our results are based on an approach that is commonly used in the literature to model the structure of macroeconomic forecasts. This literature dates back to early contributions by Ball (1962), Mincer and Zarnowitz (1969), Figlewski and Wachtel (1981), or Nordhaus (1987), who introduced the basic model framework for analyzing fixed event forecasts.[1] Fixed event forecasts refer to the case that a forecaster constructs a sequence of forecasts over time for the same event (such as an annual figure for a macroeconomic variable). The data we are using below is of this type. A couple of more recent contributions have made proposals to improve the econometric approach for testing rationality of such large panels of fixed event forecasts. These include Keane and Runkle (1990) and Batchelor and Dua (1991), who introduce the analysis in a panel framework using the Generalized Methods of Moments (GMM) method, or Davies and Lahiri (1995), who develop the analytic framework for analyzing three dimensional panels of survey data which makes the use of information along all dimensions possible. To make our results comparable to existing studies, we follow the approach suggested by Davies and Lahiri (1995) and

[1]Pesaran and Weale (2006) and Stekler (2002) present nice summaries of the commonly used approaches. The latter contribution also provides an overview about the most prominent survey data sets that are used in empirical research on forecast efficiency.

recently used by Clemens et al. (2007) and Ager et al. (2009) very closely and suggest only minor modifications.

Using this model framework we test whether the forecasts provided by the panelists of the survey are unbiased and efficient. Assuming that forecast accuracy is the only objective of a forecaster and that her loss function is symmetric and increasing in the forecast error, these two properties are necessarily a feature of a rational forecast. Since the work by Pigou (1927) or Keynes (1936), it is widely accepted that expectations and forecasts play a crucial role in all kinds of economic dynamics. Muth (1961) introduced the notion of rational expectation, which has since played a central role in economic thinking. So, although in a strict sense the concept of rational expectations does always refer to model-consistent expectations in economic theory, we believe that it is important for econometricians to analyze observed forecasts to check whether they show at least the basic features of "rational" expectations.

It should be noted, however, at this point that there are also arguments against the assumption that published forecasts reflect true expectations and should, thus, be rational if made by rational agents. Some of the cases against this assumption are the following. First, forecasters might seek to maximize public attention. If this is the case, an unbiased forecast is not optimal anymore, since the utility of the forecaster depends on more than one argument (Laster et al., 1999). Second, forecasters might produce so-called "intentional" forecast in some situations (Stege, 1989). A forecaster could, for example, predict a specific event to provoke a policy action that actually prevents the realization of the event. Third, Forecasters might have asymmetric loss functions (Capistran and Timmermann, 2006). They could, for example, have different weights concerning a possible over- or underestimation of an outcome. We believe, however, that these arguments are not particularly strong a priori. We, therefore, abstract from them and start in this paper from the hypothesis of rational forecasts, which are unaffected by these issues.

[**Has to be adjusted**]Our findings confirm the result of previous studies that any analysis based on average forecasts should be treated with caution since even if all individual forecasts are rational the hypothesis of rationality is often rejected by the aggregate data.

Furthermore, we find that there are not only large differences in the performance of forecasters across countries but also across different macroeconomic variables; in general, the forecasts for inflation are most often consistent with the hypothesis of rational forecasts while forecasts tend to be biased in situations where forecasters have to learn about large structural shocks or gradual changes in the trend of a variable. In addition, we find some weak evidence that inefficient forecasts are more likely to be significantly biased than efficient forecasts.

The remainder of this paper is structured as follows. Section 2 briefly discusses the model we use to analyze the unbiasedness and efficiency of forecasts. Section 3 presents a brief overview on the data that we use. Section 4 elaborates on the empirical results for the individual forecasts (4.2) and the consensus forecasts (4.1). Section 5 concludes the paper.

2. Model Framework

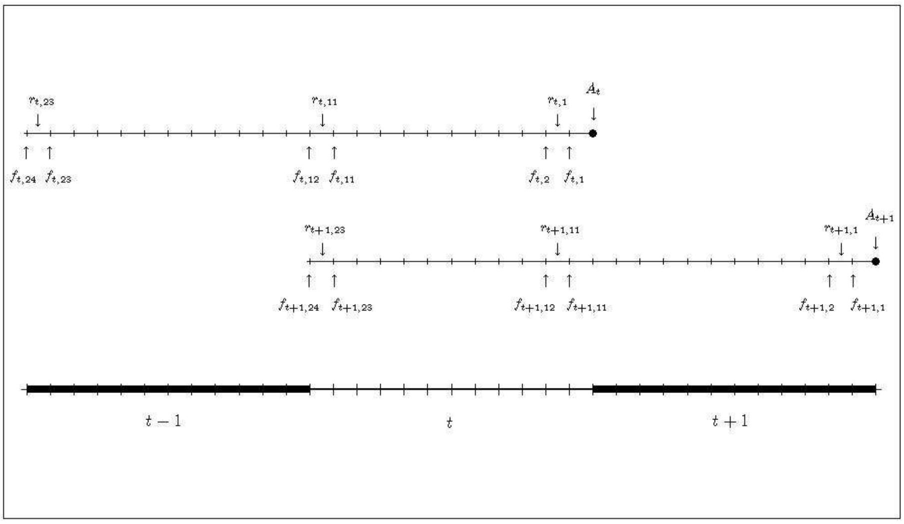
2.1. **A Structural Model for Forecast Errors.** The panel data set that we use in this paper contains so-called fixed event forecasts. Since these panels exhibit a special correlation structure, it is important to give a clear picture of the nature of the data before moving to the description of the tests that we are going to use.

The panel possesses a three dimensional structure of the kind introduced in Davies and Lahiri (1995). For each country and variable we have a NTH-vector of forecasts for T years made by N forecasters with forecast horizons ranging from one month to H months

$$F = [f_{1,1,H}, f_{1,1,H-1}, \dots, f_{1,1,1}, f_{1,2,H}, \dots, f_{1,T,1}, f_{2,1,H}, \dots, f_{N,T,1}]' \ . \tag{1}$$

In other words, for each year we collect a sequence of H forecasts from each forecaster, starting H months before the year ends and ending in the last month of the respective year. Figure 1 shows a schematic diagram of the data structure. This structure will be of importance later on when we derive the correlation between different forecast errors.

FIGURE 1. Forecast Data Structure



Following Davies and Lahiri (1995), we assume that the forecast error for each forecast can be decomposed into three different parts

$$(2) \qquad e_{i,t,h} \equiv A_t - f_{i,t,h} = \phi_i + \lambda_{t,h} + \epsilon_{i,t,h} \ ,$$

where A_t denotes the realization of a variable for year t. The first error term ϕ_i is the individual bias of the forecasts made by forecaster i. The second error term $\lambda_{t,h}$ is common to all forecasters and reflects the occurrence of macroeconomic shocks that hit an economy between the date at which the forecasts are made and the end of year t. Following the literature, we assume that these shocks are cumulated over the h months in an arithmetic way, so that the error term can be written $\lambda_{t,h} = \sum_{k=1}^{h} u_{t,k}$. We assume that $u_{t,h}$ is distributed with a zero mean and a variance of σ_u^2. Since $u_{t,h}$ and $u_{t+1,h+12}$ occur at the same point in time, they will be correlated (Davies and Lahiri, 1995).

The third error term $\epsilon_{i,t,h}$ refers to the forecaster specific error component of the forecast error. The literature proposes two alternative ways to model this error. On the one hand, it can be seen as an independently and identically distributed (*iid*) shock. This is the view taken for the estimation in Davies and Lahiri (1995). On the other hand, Davies and Lahiri mention that one could assume that over time each forecaster receives a flow of private information on the outcome. Under this assumption, one can model the individual error as $\epsilon_{i,t,h} = \sum_{k=1}^{h} \eta_{i,t,k}$, where the $\eta_{i,t,k}$ are distributed with mean 0 and variance σ_i^2.

Again, $\eta_{i,t,k}$ and $\eta_{i,t+1,k+12}$ have a non-zero correlation, since these information shocks occur at the same point in time.

It is clear that the two model variants for $\epsilon_{i,t,h}$ have very different implications. In the first case, the individual errors are assumed to be a white noise process, while in the second case they are assumed to follow a random walk for each target year t. In the first case, there would be no correlation between consecutive individual errors, while in the second case the autocorrelation would be very high and decaying only slowly for higher distances between two forecast errors for the same target year. Intuitively, the second model is much more appealing: Consider a forecaster whose forecast is above the consensus forecast in one month. Isn't it very likely that he is going to publish an above-average forecast also in the following month? That is, it is very strange to believe that individual forecasts fluctuate randomly around the consensus forecast without any persistence. Rather a forecaster is likely to be persistently more optimistic or pessimistic than the average for some time. This behavior would be better captured by the second model that implies a high autocorrelation of the individual errors.

Eventually, choosing from the two alternatives is a matter of empirical facts. In our data set, the estimates of the individual errors, say $\hat{\epsilon}_{i,t,h}$, show a fairly high degree of autocorrelation. The empirical autocorrelation functions are usually declining slowly and approach zero only after twelve month or so. So, while we have to reject the null hypothesis of a unit root in the individual errors for virtually all cases, we usually prefer the second model against the first model based on Bayesian information criteria.[2] Against this background, we think there are good reasons to modify the model used by Davies and Lahiri (1995). Hence, we will model the individual error components according to the assumption that forecasters receive private information which are accumulated over time.

2.2. **A Test of Unbiasedness.** Testing the unbiasedness of forecaster i is equivalent to testing whether $\phi_i = 0$ in (2). We can examine this hypothesis by testing the zero

[2]Detailed results are available from the authors upon request.

restriction on the elements of $\Phi = [\phi_1, \ldots, \phi_N]'$ in

$$e = A - F = \Phi \otimes i_{TH} + \underbrace{\lambda + \epsilon}_{=\nu}\,, \tag{3}$$

where e is the vector of stacked forecast errors, A is given by $i_N \otimes (A^+ \otimes i_H)$ with $A^+ = (A_1, A_2, \ldots, A_T)'$ and i_{TH}, i_N and i_H are vector of ones of dimension TH, N and H respectively.[3] λ and ϵ are vectors of length NTH in which we stack the appropriate $\lambda_{t,h}$ and $\epsilon_{i,t,h}$ respectively.

Now, while a simple OLS regression gives consistent point estimates for the bias, we cannot base our inference on the OLS standard errors, since the elements of ν are clearly not *iid* due to the special correlation structure caused by the structure of the panel data set. Davies and Lahiri (1995) show that it is neither diagonal nor homoscedastic. Recalling that due to our assumption about the individual errors our specification differs from their model, we formally have the following elements of $\Sigma = E[\nu\nu']$ for two forecasters, say i and j:

$$Cov\left(\nu_{i,t_1,h_1}, \nu_{j,t_2,h_2}\right) = Cov\left(\sum_{k=1}^{h_1} u_{t_1,k} + \sum_{k=1}^{h_1} \eta_{i,t_1,k}, \sum_{k=1}^{h_2} u_{t_2,k} + \sum_{k=1}^{h_2} \eta_{j,t_2,k}\right) \tag{4}$$

$$Cov(\nu_{i,t_1,h_1}, \nu_{j,t_2,h_2}) = \begin{cases} min\{h_1, h_2\}\,[\sigma_u^2 + \sigma_i^2] & \text{if } i = j, t_1 = t_2, h_1 = h_2 \\ min\{h_1, h_2 - 12\}\,[\sigma_u^2 + \sigma_i^2] & \text{if } i = j, t_1 = t_2 - 1, h_2 \geq 12 \\ min\{h_1, h_2\}\,\sigma_u^2 & \text{if } i \neq j, t_1 = t_2, h_1 = h_2 \\ min\{h_1, h_2 - 12\}\,\sigma_u^2 & \text{if } i \neq j, t_1 = t_2 - 1, h_2 \geq 12 \\ 0 & \text{else} \end{cases}$$

Clearly, the different non-zero cases deserve some more explanation. The forecast errors ν are correlated across several dimensions. First, they are correlated within the maximum forecast horizon H since $\lambda_{t,h}$ and $\epsilon_{i,t,h}$ are the accumulation of period-specific shocks; this refers to the first case shown in (4). Second, the forecast errors are correlated between

[3]The operator $\otimes$ denotes the Kronecker Product.

subsequent years since the forecast horizons are of overlapping nature; this refers to the second case shown in (4). Finally, the forecast errors are correlated across different forecasters, since forecast errors are produced at the same time and are all subject to the same subsequent aggregate shocks summarized by $\lambda_{t,h}$; this refers to the third and fourth case shown in (4).

Given Σ, the covariance matrix of the Generalized Methods of Moments (GMM) estimator is given by

$$Var(\hat{\Phi}) = \left[(I_N \otimes i_{TH})'(I_N \otimes i_{TH})\right]^{-1} \left[(I_N \otimes i_{TH})'\Sigma(I_N \otimes i_{TH})\right] \left[(I_N \otimes i_{TH})'(I_N \otimes i_{TH})\right]^{-1} \tag{5}$$

and can be used to derive valid t-statistics for testing $\phi_i = 0$. Naturally, Σ is not observed and has to be replaced by a consistent estimate, say $\hat{\Sigma}$, before computation of the test statistics is possible.

Though Σ has a complicated pattern, it depends only on $N + 1$ parameters, namely $\sigma_1^2, \ldots, \sigma_N^2$ and σ_u^2. Davies and Lahiri (1995) propose to obtain a consistent estimate by first estimating these $N + 1$ parameters and then replacing the parameters in Σ by the corresponding estimates. We will follow this approach. Note that an estimator of ϕ_i is simply given by the average forecast error of forecaster i and that we can estimate the two other parts of the forecasts errors by

$$\hat{\lambda}_{t,h} = \frac{1}{N}\sum_{i=1}^{N}\left(A_t - f_{i,t,h} - \hat{\phi}_i\right) \tag{6}$$

and

$$\hat{\epsilon}_{i,t,h} = A_t - f_{i,t,h} - \hat{\phi}_i - \hat{\lambda}_{t,h} \; . \tag{7}$$

We can obtain estimates for the unknown parameters as the estimated coefficients from the following regressions:

$$\hat{\lambda} \odot \hat{\lambda} = \sigma_u^2\, \kappa_H + \omega_\lambda \tag{8}$$

$$\hat{\epsilon} \odot \hat{\epsilon} = (I_N \otimes \kappa_H)\, \sigma^2 + \omega_\epsilon \quad , \tag{9}$$

where $\kappa_H = i_T \otimes [H, H-1, \ldots, 1]'$ and $\sigma^2 = [\sigma_1^2, \ldots, \sigma_N^2]'$. Clemens et al. (2007) and Ager et al. (2009) show how the problem simplifies when $N = 1$. Essentially, all terms that refer to the individual component of the forecasting errors vanish if one assumes that there simply is no "private" shocks in this setup with one agent only (Ager et al., 2009) or if one models these "private" shocks like we do – so that they become observationally indistinguishable from the aggregate shocks (Clemens et al., 2007).[4]

2.3. **Test of (Weak) Efficiency.** For testing the efficiency of the forecasts, we use the concept of weak-form efficiency that has been originally proposed by Nordhaus (1987). The concept starts from the notion of strong efficiency of forecasts which requires that all information, which has been revealed at the time a forecast is made, is taken into account during the forecasting process. In other words: If a series of forecasts is strongly efficient, it would have not been possible to improve the forecast performance by using any information available also to the forecaster. Since the amount of potentially relevant information is immense and any selection for an empirical analysis would be ad-hoc,[5] Nordhaus (1987) proposes to restrict the relevant information set to lagged values of the forecasts themselves. He shows that under weak form efficiency the revisions of forecasts should be uncorrelated under certain assumptions. It should be intuitively clear that for efficient forecasts the current forecast should not reveal any information on future revisions – or as Nordhaus states (p. 673):

> If I could look at your most recent forecasts and accurately say, "Your next forecast will be 2% lower than today's", then you can surely improve your forecasts.

[4]Ager et al. (2009) analyze consensus forecasts with this $N = 1$ setup by noting that the individual errors should cancel out in the aggregate. This is true, however, only when the consensus forecast is based on a large sample of panelists. For small panels, one gets components of the covariance matrix of the aggregate error that reflect the aggregated variance of the individual errors. Hence, when the consensus forecast is based on a small number of panelists, treating the former as an $N = 1$ case does bias the estimated variances and covariances downward leading to too many rejections of unbiasedness (or efficiency as we will see below).
[5]Not to mention the problem of constructing large data sets with real-time vintages.

Against this background, weak-form efficiency of a sequence of forecasts can be formally tested using an equation of the form

$$r_{i,t,h} = \beta_i \, r_{i,t,h+k} + \xi_{i,t,h} \ , \tag{10}$$

where $r_{i,t,h}$ is defined as $f_{i,t,h} - f_{i,t,h+1}$, $k \geq 1$, and $\xi_{i,t,h}$ is the error term. The hypothesis of weak-form efficiency implies $\beta_i = 0$; a consistent estimate of β_i can be obtained by the OLS estimator treating $\xi_{i,t,h}$ as white noise. But again – due to the special structure of the fixed event forecasts – the covariance matrix of $\xi = \left[\xi_{1,1,H-(k+1)}, \ldots, \xi_{N,T,1}\right]'$, say $\Xi = E[\xi\xi']$, is non-diagonal and heteroscedastic.

To derive the exact form of Ξ, we first note that, using (2), we can re-write the forecast revisions as

$$r_{i,t,h} = f_{i,t,h} - f_{i,t,h+1} = \lambda_{t,h+1} - \lambda_{t,h} + \epsilon_{i,t,h+1} - \epsilon_{i,t,h} = u_{t,h+1} + \eta_{i,t,h+1} \ . \tag{11}$$

Now, it is evident that under the Null hypothesis $\beta_i = 0$ we obtain the following expressions for the elements of Ξ:[6]

$$Cov\left(\xi_{i,t_1,h_1}, \xi_{j,t_2,h_2}\right) = Cov\left(u_{t_1,h_1+1} + \eta_{i,t_1,h_1+1}, u_{t_2,h_2+1} + \eta_{j,t_2,h_2+1}\right) \tag{12}$$

$$Cov\left(\xi_{i,t_1,h_1}, \xi_{j,t_2,h_2}\right) = \begin{cases} \sigma_u^2 + \sigma_i^2 & \text{if } i = j, t_1 = t_2, h_1 = h_2 \\ \sigma_u^2 + \sigma_i^2 & \text{if } i = j, t1 = t_2 - 1, h_1 = h_2 - 12 \\ \sigma_u^2 & \text{if } i \neq j, t_1 = t_2, h_1 = h_2 \\ \sigma_u^2 & \text{if } i \neq j, t_1 = t_2 - 1, h_1 = h_2 - 12 \\ 0 & \text{else} \end{cases} \tag{13}$$

[6]Note that at this point the assumption of private information for $\epsilon_{i,t,h}$ is crucial for the result that under weak-form efficiency $\beta_i = Cov(r_{i,t,h}, r_{i,t,h+1}) = Cov(u_{t,h+1} + \eta_{i,t,h+1}, u_{t,h+2} + \eta_{i,t,h+2}) = 0$. Under the assumption that the $\epsilon_{i,t,h}$ represent ordinary iid shocks we would get $\beta_i = Cov(u_{t,h+1} + \epsilon_{i,t,h+1} - \epsilon_{i,t,h}, u_{t,h+2} + \epsilon_{i,t,h+2} - \epsilon_{i,t,h+1}) = -\sigma_i^2 \neq 0$.

Given Ξ, the covariance matrix for the GMM estimator of β can be written as

$$(14) \qquad Var(\hat{\beta}) = \left(r'_{+k}r_{+k}\right)^{-1} r'_{+k}\, \Xi\, r_{+k} \left(r'_{+k}r_{+k}\right)^{-1} \ ,$$

where $r_{+k} = \left[r_{1,1,H-1}, \ldots, r_{1,1,(k+1)}, r_{1,2,H-1}, \ldots, r_{N,T,(k+1)}\right]'$ and $\beta = [\beta_1, \ldots, \beta_N]'$. $Var(\hat{\beta})$ can be used to derive valid t-statistics for testing $\beta_i = 0$. Naturally, Ξ is not observed and has to be replaced by a consistent estimate, say $\hat{\Xi}$, before computation of the test statistics is possible.

To obtain $\hat{\Xi}$ we can use the same method that we used to derive $\hat{\Sigma}$. First, we derive estimates for the single elements of Ξ and replace these elements in a second step by their estimates to consistently estimate Ξ. Note that the structure of Ξ is much more simple than that of Σ so that its elements are simply given by

$$(15) \qquad \hat{\sigma}_u^2 = \frac{1}{T(H-(k+1))} \sum_{t=1}^{T} \sum_{h=1}^{H-(k+1)} (\hat{u}_{t,h+1}^2)$$

and

$$(16) \qquad \hat{\gamma}_i^2 = \frac{1}{T(H-(k+1))} \sum_{t=1}^{T} \sum_{h=1}^{H-(k+1)} \left(\hat{\eta}_{i,t,h+1}^2\right) \ ,$$

where $\hat{u}_{t,h}$ and $\hat{\eta}_{i,t,h}$ are consistently estimated by

$$(17) \qquad \hat{u}_{t,h} = \frac{1}{N} \sum_{i=1}^{N} r_{i,t,h-1}$$

and

$$(18) \qquad \hat{\eta}_{i,t,h} = r_{i,t,h-1} - \hat{u}_{t,h} \ .$$

Given this formal framework, we will now move to the empirical analysis of the macroeconomic forecasts in the G7-countries.

3. Data

In this study, we rely on data from the surveys conducted by *Consensus Economics*, a London-based firm.[7] Each month, starting in October 1989, *Consensus Economics* polls institutions like investment banks or research institutes for economics about their forecasts for the most common macroeconomic variables. The largest samples are available for the G7 countries, on which we concentrate in this paper.[8] A big advantage of the data set is that estimates are comparable across countries as well as across panelists.

We concentrate on forecasts for three variables: the annual growth rate of gross domestic product (GDP), the annual inflation rate, and the annual growth rate of private consumption expenditure. It is important to note that there occurred some changes in the definition of the target variables in some of the countries. More specifically, while the inflation forecasts refer to the consumer price inflation in general, the relevant figure which had to be forecasted in the United Kingdom (UK) was based on the Retail Price Index at the beginning of our sample. Forecasts for CPI inflation in the UK were introduced in 2004.[9] Furthermore, forecaster were asked to target the annual growth rate of the gross national product (GNP) rather than that of GDP in Germany and Japan until 1992 and 1993 respectively. For German forecast, there is another break in the data due to the switch from West-German data to data for the reunified Germany. In our data set forecasts for GDP growth and inflation refer to West-Germany until 1996; for forecasts on private consumption expenditures the change was made in 1995.

A final issue regarding data concerns the realizations that we use to evaluate the forecast errors. It has become standard in the literature to use data from the initial releases rather than revised ex-post data for the evaluation of macroeconomic forecasts (see e.g.

[7]Information from the *Consensus* data set have been used in a sequence of papers during recent years to analyze the properties of macroeconomic forecasts. Most contributions, however, consider only data on the average forecasts and do not analyze individual forecasts. Notable exceptions are Lahiri and Sheng (forthcoming), who propose a model for disagreement among forecasters and estimate it based on individual forecasts on GDP growth from the Consensus data set, Batchelor (2007), who uses a similar disaggregated data set to analyze the bias in forecasts for GDP growth, and to some extend Harvey et al. (2001), who analyze the properties of forecasts of a selected group of panelists form the Consensus data for the UK.

[8]The average numbers of panelists range from 15 for Canada to 30 for United Kingdom.

[9]An additional break occurred in May 1997 when the underlying Retail Price Index changed to a version that excludes interest payments on mortgages.

Croushore, 2006). We follow this approach and compute forecast errors based on the historical data as they are listed in the publications of *Consensus Economics* in May of the subsequent year since those figures should reflect the initial releases in all cases. To give an example: We use the relevant annual figures for 1996 as they are reported by *Consensus Economics* together with the forecasts made in May 1997 to evaluate all forecasts that have been made for 1996 during the years 1995 and 1996.

The structure of the survey data set fits exactly the framework discussed in section 2. In each month, the participating institutions are asked to state their forecasts for the current and the subsequent calender year. That leaves us with a sequence of $H = 24$ forecasts from each panelist for each annual figure starting with the first forecast made in January of the preceding year and ending with the last forecast made in December of the year for which the forecast is made. We include in our sample forecasts for the years 1991–2005, i.e. $T = 15$ in our analysis. The number of panelists covered by the data set varies considerably across countries but also over time; but usually $N \geq 10$ even for the smaller countries.

4. Results

4.1. **Individual Forecasts.** In this section we present all results concerning the properties of the individual forecasts. For both – the tests of unbiasedness (Section 4.1.1) as well as the tests of weak efficiency (Section 4.1.2) – we include those panelists in the sample who made a forecast at more than 50% of the possible dates. Thereby, we avoid the influence of small sample problems which could arise from those panelists that reported only a few forecasts.[10]

An additional feature of the data that we have to deal with is given by the fact that the record of most of the forecasters includes a bunch of missing values, i.e. the panel is heavily unbalanced. There are two reasons for that. First, the set of panelists who take part in the *Consensus* survey changes continuously. Hence, there are some forecasters that enter the panel at a later stage, while other forecasters leave the panel after the first

[10]The threshold of 50% is of course arbitrary. Results for the included panelists are, however, robust to the inclusion of more forecasters in the used sample.

part of the time period covered by our data set. Second, some forecasters do not submit there forecast on a regular basis, i.e. some of them do not provide their current forecasts for some of the months. To minimize the reduction of our data base due to the second issue, we interpolate a missing value in all those cases, in which a forecast is missing only for one month in a row *and* the two adjacent forecasts are equal to each other. Formally, if $f_{i,t,h}$ is missing *and* $f_{i,t,h+1} = f_{i,t,h-1}$, we set the missing forecast equal to $f_{i,t,h+1}$.

For the estimation, we follow Davies and Lahiri (1995) and deal with missing values by simply deleting the appropriate elements in the vectors of forecast errors or revisions and the corresponding rows and columns in the covariance matrices respectively. Those compressed matrices can be directly used in the GMM estimation procedure (Blundell et al., 1992).

4.1.1. *Unbiasedness.* The analysis of the biases present in the individual forecasts reveal some interesting differences across countries as well as variables. The results are summarized in Table 1.[11] In general, most of the individual forecasts are unbiased. The overall performance is best for the inflation forecasts; only for France, the UK and the US there are one or two forecasters respectively who produce biased inflation forecasts. Especially surprising is the good performance of inflation forecasts for Italy that underwent a significant transition from a high inflation regime towards a low inflation regime during the early sample period. One could imagine that forecasters adjusted only slowly to this new environment causing forecasts to be biased upwards.

This behavior is actually what can be observed for the inflation forecasts in the UK, where inflation was also very high at the beginning of our sample period and declined considerably to low levels in the mid of the 1990s. All but three panelist, which entered the sample rather late, have overestimated inflation on average. After all, only 2 out of 30 did so significantly on a 95% confidence level.

A similar argument applies to the bias that is found for most of the forecasts for GDP growth in the European countries. Here, the wide majority of forecasters overestimate growth on average. This phenomenon is most pronounced in Germany and Italy but

[11]Detailed results on basis of individual panelists are available on request.

applies to a lesser extend also to France. The same is also true for the forecast for growth of private consumption in Germany. Batchelor (2007) shows that this kind of bias can be inevitable in an environment of declining trend growth rates since forecasters have to gradually learn about the new trend.

A very special picture is given by the combination of forecasts for GDP growth and for growth of industrial production in the UK. While forecasts for the former are generally unbiased, the results show strong evidence for rejecting the hypothesis of unbiased forecasts for the latter forecasts; most panelist on average overestimate growth of industrial production by about 1 to 1.5 percentage points. This might reflect the fact that although the trend growth of overall output remained relatively constant over the sample, there has been a shift of the structural composition of the economy in the UK from production oriented sectors towards services – especially towards the financial sector – which had to be learned by the forecasters. A similar phenomenon can be observed when comparing forecasts on GDP growth for the US, which are generally unbiased, to forecasts for growth of private consumption in the US, which tend to underestimate consumption growth. Again it seems that it has been hard for a large number of panelists to anticipate the gradual decline in the saving rate of private households as well as to proper estimate additional consumption effects of huge increases in household wealth that was caused by the stock market boom of the late 1990s and the real estate booms during the time from 2002 until the end of our sample.

In general, we can conclude that biased forecasts seem to be produced in times of structural changes or gradual developments that have to be learned by the forecasters; this source for bias in macroeconomic forecasts is also supported by the results in Andolfatto et al. (2008) who analyze the properties of artificial forecast generated within a standard dynamic stochastic general equilibrium model. On the contrary, forecasts seem to be unbiased in general for stable economies without large structural shocks. One example is Canada where the structure of the economy and the medium term growth trend have not fundamentally changed since the introduction of inflation targeting in 1991. As a consequence, there is not one single case among all forecasts for the Canadian economy in which the panelist produced biased forecasts.

Table 1. Bias of Individual Forecasts

	# obs	# bias	# bias pos	# bias neg	mean(bias)	var(bias)	cons biased?
Gross Domestic Product							
Germany	29	7	0	7	-0.49	0.021	no
Canada	14	0	0	0	-0.27	0.020	no
France	17	4	0	4	-0.40	0.016	no
Italy	11	10	0	10	-0.61	0.002	yes (-)
Japan	13	0	0	0	-0.22	0.042	no
UK	30	0	0	0	-0.25	0.025	no
USA	22	0	0	0	0.23	0.023	no
Inflation							
	# obs	# bias	# bias pos	# bias neg	mean(bias)	var(bias)	cons biased?
Germany	30	0	0	0	-0.06	0.011	no
Canada	14	0	0	0	0.25	0.028	no
France	17	2	0	2	-0.14	0.021	no
Italy	11	0	0	0	0.20	0.010	no
Japan	13	0	0	0	-0.13	0.011	no
UK	30	2	0	2	-0.20	0.025	no
USA	22	1	0	1	-0.14	0.027	no
Industrial Production							
	# obs	# bias	# bias pos	# bias neg	mean(bias)	var(bias)	cons biased?
Germany	28	0	0	0	-1.00	0.078	no
Canada	4	0	0	0	-0.93	0.120	no
France	8	2	0	2	-0.98	0.093	yes (-)
Italy	8	4	0	4	-1.53	0.048	no
Japan	13	0	0	0	-1.26	0.089	no
UK	28	22	0	22	-1.40	0.056	yes (-)
USA	21	0	0	0	-0.46	0.077	no
Private Consumption							
	# obs	# bias	# bias pos	# bias neg	mean(bias)	var(bias)	cons biased?
Germany	29	18	0	18	-0.52	0.039	yes (-)
Canada	14	0	0	0	0.03	0.051	no
France	17	0	0	0	-0.18	0.008	no
Italy	11	0	0	0	-0.40	0.005	no
Japan	13	0	0	0	-0.31	0.012	no
UK	29	0	0	0	0.15	0.033	no
USA	22	10	10	0	0.51	0.019	no

Notes: *#obs* indicates the number of individual panelists, *#bias* the number of them which provides significantly biased forecasts and *#bias pos/neg* the direction of the biases. $Mean(bias)$ and $Var(bias)$ indicate the mean and the variance of the individual biases and *cons biased?* provides comparison to the corresponding consensus forecast.

4.1.2. *Weak Efficiency.* For testing weak efficiency of individual forecasts we followed the literature (Clemens, 1995, Harvey et al., 2001, Isiklar et al., 2006) by setting k in (10) equal to 1. This makes indeed sense, since by the time a new revision is made each forecaster knows about his most recent previous forecast revision. The results are summarized in Table 2.[12]

[12]Again, detailed results for all individual panelists are available on request.

The analysis of the individual forecasts' properties in terms of weak efficiency reveals an interesting contrast between the forecasts made for GDP growth and those for the other variables under investigation in this paper. For the majority of forecasts for growth of industrial production and private consumption as well as for the inflation rate we cannot reject the hypothesis of weakly efficient forecasts; only few series of forecast show a significant correlation between proceeding forecast revisions. In those cases, the estimated coefficient is mostly negative which means that the corresponding forecasters tend to overreact to incoming news, i.e. they initially revise their forecasts by an amount that is too large and undo part of this revision during the next forecasting round.

In contrast, we find more evidence for deviations from weak efficiency for forecasts of GDP growth in all counties but Italy and Japan.[13] The main difference, however, is that the estimated coefficients are positive in all but one of the significant cases. Thus, those forecasts for GDP growth, which deviate from weak efficiency, show a strong tendency towards forecast smoothing in general. This means that forecasters tend to process new information only slowly which results in positively autocorrelated revisions.[14] Also Gallo et al. (2002) find that forecasters tend to stick to their past forecast even when the authors control in their study for the most recently observed average forecast and the dispersion of forecasts. Batchelor and Dua (1992) rationalize such a forecasting behavior by noting that in reality forecasters might not have a single objective which is minimizing the expected squared errors. They are likely to take into account as well that their clients might "mistrust forecasters who make frequent [erratic] revisions to forecasts" (p. 179). The fact that usually the forecast for GDP growth is the part of a comprehensive macroeconomic forecast published by a forecaster, which is most widely anticipated and discussed by clients or the media, might bring about that it is exactly this forecast, for which incentive and reputation considerations make forecasters deviate most from their

[13]The fact that we find weakly efficient forecasts for GDP growth for Japan is in contrast to the results of Ashiya (2003), who analyzes the reaction to news of forecasters for GDP growth in a slightly different modeling framework and based on a different set of private sector forecasts; he concludes that the individual forecasters tend to significantly overreact to new information.

[14]This phenomenon is also known as *conservatism* in psychology (Phillips and Edwards, 1966, Edwards, 1968).

true expectations. This would explain why we find the strong tendency for forecast smoothing only for forecasts on GDP growth.

Table 2. Efficiency of Individual Forecasts

	# obs	# ineff	# ineff pos	# ineff neg	mean(beta)	var(beta)	cons eff?
Gross Domestic Product							
Germany	24	9	9	0	0.10	0.007	no (+)
Canada	12	5	5	0	0.13	0.011	no (+)
France	15	9	9	0	0.15	0.014	no (+)
Italy	10	2	2	0	0.05	0.010	no (+)
Japan	10	0	0	0	0.02	0.003	no (+)
UK	26	7	6	1	0.07	0.012	no (+)
USA	21	8	8	0	0.10	0.009	no (+)
Inflation							
	# obs	# ineff	# ineff pos	# ineff neg	mean(beta)	var(beta)	cons eff?
Germany	26	4	0	4	-0.03	0.009	no (+)
Canada	13	1	0	1	-0.03	0.008	no (+)
France	15	2	1	1	-0.01	0.008	no (+)
Italy	10	1	0	1	-0.01	0.004	yes
Japan	11	1	0	1	-0.07	0.004	no (+)
UK	25	7	0	7	-0.08	0.023	yes
USA	20	4	1	3	-0.05	0.013	no (+)
Industrial Production							
	# obs	# ineff	# ineff pos	# ineff neg	mean(beta)	var(beta)	cons eff?
Germany	23	6	3	3	-0.03	0.019	no (+)
Canada	2	0	0	0	0.02	0.014	yes
France	7	2	1	1	0.01	0.014	no (+)
Italy	7	3	1	2	-0.01	0.020	no (+)
Japan	10	0	0	0	0.02	0.003	no (+)
UK	24	4	2	2	0.02	0.011	no (+)
USA	20	4	1	3	-0.02	0.013	no (+)
Private Consumption							
	# obs	# ineff	# ineff pos	# ineff neg	mean(beta)	var(beta)	cons eff?
Germany	27	9	0	9	-0.10	0.009	no (+)
Canada	12	1	0	1	-0.02	0.006	no (+)
France	15	3	2	1	0.02	0.007	no (+)
Italy	10	1	0	1	-0.03	0.006	no (+)
Japan	10	0	0	0	-0.03	0.002	no (+)
UK	24	3	1	2	-0.01	0.010	no (+)
USA	20	2	1	1	0.02	0.007	no (+)

Note: *#obs* indicates the number of individual panelists, *# ineff* the number of them which provides significantly weakly inefficient forecasts and *# ineff pos/neg* the sign of the estimated coefficient. $Mean(beta)$ and $Var(beta)$ indicate the mean and the variance of the estimated individual coefficients and *cons eff?* provides comparison to the corresponding consensus forecast.

4.2. **Consensus Forecast.** In this section, we present the results concerning the properties of the average forecast, the so-called *consensus forecast.*[15] Average forecasts have been frequently used in empirical research although results based on them should be treated

[15]Note that all panelists are included in the computation of the average forecast. Hence, unlike in the analysis of individual forecasts we do not exclude those panelists who reported less than 50% of all possible forecasts over the entire sample.

with caution because of inconsistency problems due to the aggregation bias (Bonham and Cohen, 2001) caused for instance by not-accounting for private information (Figlewski and Wachtel, 1981) or the fact that the aggregation might cancel out deviations from unbiasedness of individual forecasters (Keane and Runkle, 1990). We, therefore, present the results based on average forecasts to compare their properties to those of the individual forecasts.

4.2.1. *Unbiasedness.* We can simply use the framework presented in Section 2.2 with $N = 1$ to analyze the bias of the consensus forecasts. In order to be able to compare the results for consensus and private forecasts, we start by discussing the results under the assumption of a homogenous bias across horizons. They are given in Table 3. The estimation outcomes show that for all countries we cannot reject the hypothesis that the consensus forecasts for inflation are unbiased. For all other variables the picture is mixed. First, the average forecasts for growth of private consumption are biased upwards in Germany while they are significantly, but only weakly, too pessimistic in the US. For the other five countries the corresponding forecasts are unbiased. Second, consensus forecasts for GDP growth are unbiased in all countries but Germany, France and Italy where they tend to be too optimistic on average. And finally, the average forecasts for growth of industrial production are biased upwards in France, Italy, and the UK while being unbiased for the remaining four countries. These figures broadly reflect the picture given by the individual forecasts, i.e. in cases where a sizable fraction of individual forecasts was found to be biased also the consensus forecast is bases.

Table 3. Bias of Consensus Forecasts

		Germany	Canada	France	Italy	Japan	UK	USA
GDP	phi	-0.52*	-0.28	-0.44*	-0.63***	-0.28	-0.24	0.18
	t-stat	-1.79	-0.90	-1.66	-2.64	-0.67	-0.83	0.62
Inflation	phi	-0.03	0.23	-0.15	0.18	-0.13	-0.20	-0.13
	t-stat	-0.19	0.54	-1.09	0.61	-0.99	-1.31	-0.88
Ind. Prod.	phi	-1.00	-0.91	-1.06**	-1.56**	-1.35	-1.43**	-0.48
	t-stat	-1.56	-1.52	-1.96	-2.25	-1.28	-2.85	-0.83
Priv. Cons.	phi	-0.56**	0.04	-0.19	-0.44	-0.36	0.18	0.48*
	t-stat	-2.24	0.14	-1.07	-1.50	-1.27	0.60	1.87

Notes: *,**,*** denotes rejection of the null at the 10%, 5% and 1% significance level, respectively.

So far we have assumed that bias does not vary with respect to the forecast horizon. Since it is a reasonable hypothesis that this might be wrong, we relax the restriction of a constant bias for the average forecasts.[16] To do this robustness analysis, we write the aggregated forecast errors as

$$e_{t,h} = A_t - f_{t,h} = \phi_h + \sum_{k=1}^{h} u^*_{t,k} \ ,$$

where now ϕ_h is a horizon-specific bias and $u^*_{t,k} = u_{t,k} + \eta_{t,k}$. A consistent estimator for ϕ_h is given by $\phi_h = 1/T \sum_{t=1}^{T} e_{t,h}$. Note that under the joint hypothesis that the forecasts are unbiased for all forecast horizons ($H_0 : \phi_H = \cdots = \phi_1 = 0$), the structure for the residual covariance matrix of this regression is equal to the one derived above for the case, in which the bias is restricted to be equal across horizons.

As it is not the central focus of this paper, we refrain from tabulating the results for horizon-specific biases in detail.[17] Instead we give a broad picture by referring to the sequential test of forecast unbiasedness as introduced by Ager et al. (2009). The null-hypothesis of this test is that all horizon-specific biases up to some horizon h are jointly equal to 0 ($H_0 : \phi_1 = \ldots = \phi_h = 0$). The Wald statistic is given by $W = (\phi_h)'(Var(\phi_h))^{-1}(\phi_h)$ and χ^2-distributed with h degrees of freedom. The corresponding F-statistic, which possesses better small sample properties is obtained by dividing the Wald statistic by h.

The results are summarized in Table 4.[18] They show that for most variables and countries the null hypothesis of joint insignificance of the horizon specific biases can be rejected for horizons greater than 14 to 16. Noteworthy exceptions are the forecasts on GDP for Canada, on inflation for Germany and Italy, as well as the ones on industrial production for France and Japan, where joint insignificance can not be rejected for any horizon. Furthermore the test rejects the null hypothesis also for low horizons in some

[16] Note that we cannot do the same in the analysis of individual forecasts since for the wide majority of panelists the data sets includes so many missing values that the estimate for each horizon-specific bias would be based on 10 or even less observations.
[17] The detailed results are available on request.
[18] More detailed results are available on request.

cases. These are the forecasts on inflation in Canada and France and those for private consumption for Japan.

Table 4. Sequential Test of Forecast Unbiasedness for Consensus Forecasts

	Germany	Canada	France	Italy	Japan	UK	US
GDP	15-24		14-24	14-24	15-24	15-24	16-24
Inflation		1-4; 24	1; 14; 15; 19-24		15-24	14-24	14-24
Ind. Prod.	16-24	15-24		24		14-24	16-24
Priv. Cons.	15-24	14-24	15-24	15-24	1-3; 14-24	24	15-24

Notes: Each cell shows the horizons h for which the null-hypothesis of joint insignificance of biases can be rejected at the 5% level.

4.2.2. *Weak Efficiency.* The results from the tests of weak efficiency for the consensus forecasts demonstrate very well that caution is required when working with average forecast data. In contrast to the setup for the analysis of individual forecasts we set $k = 2$ for the analysis of the consensus forecasts (Isiklar, 2005). The reason is that it is not clear whether each forecaster knows already about the most recent consensus forecast when a new forecast is produced, since in the extreme case the forecasts have to be reported two weeks before a new consensus forecast is published and additionally the production process for each forecast might last more than a week depending on the institutional framework of a specific forecaster. In any case, each forecaster should know about the average forecast published two months ago.

Table 5 shows the results for implementations of the test based on (10). It is obvious that the results taken at face value would lead to completely different conclusions than those seen in Section 4.1.2. Clearly, all average forecasts except in two cases[19] show evidence for forecast smoothing, i.e. incoming information gets reflected in the average forecast in a very sluggish way. The effect is indeed most strongly visible for forecasts for GDP growth also here, but even for variables for which the individual forecasters tend to overreact to news we find the opposite deviation from weak efficiency in the consensus forecast (e.g. growth of private consumption in Germany).

Like mentioned already in the previous section it is possible to estimate horizon specific biases for the consensus forecasts. If we relax the assumption of one single bias for all

[19]Those are the inflation forecasts in Italy and the forecasts for growth of industrial production in Canada.

forecast horizons, this has implications for the construction of the test on weak efficiency. Namely, the unconditional expectation for a revision is no longer equal to zero under the null hypothesis in that case, since $r_{t,h} = f_{t,h} - f_{t,h+1} = \phi_{h+1} - \phi_h + u^*_{t,h+1}$. We, therefore, expand (10) by including constant terms for each forecast horizon. Dropping the index i, the new equation on which we base the robustness check for our results is

$$r_{t,h} = \beta r_{t,h+2} + \phi_{h+1} - \phi_h + \xi_{t,h} \ . \tag{19}$$

The coefficient β can be consistently estimated by regressing the consensus forecasts' revisions on its second lags and a set horizon-specific dummy-variables using OLS. An estimate for the variance of $u^*_{t,h+1}$ can be obtained by regressing the revisions on the set of horizon-specific dummy-variables and taking the mean of the squared residuals from that regression.

The results that are given in Table 6 confirm the evidence presented in the previous paragraph.[20] The point estimates for the correlations between two subsequent revisions do not change qualitatively. The maximum difference between two corresponding point estimates is 0.12 in absolute values.

TABLE 5. Efficiency of Consensus Forecasts (Without Horizon Effects)

		Germany	Canada	France	Italy	Japan	UK	USA
GDP	beta	0.49***	0.27***	0.34***	0.34***	0.25***	0.37***	0.23***
	t-stat	7.79	4.31	4.92	5.23	3.94	6.49	3.67
Inflation	beta	0.19***	0.17***	0.14**	0.06	0.20***	0.09*	0.23***
	t-stat	2.99	3.09	2.23	1.02	3.27	1.66	3.65
Ind. Prod.	beta	0.46***	0.00	0.33***	0.38***	0.33***	0.49***	0.28***
	t-stat	6.94	0.06	5.02	5.46	4.91	8.10	4.67
Priv. Cons.	beta	0.32***	0.17***	0.24***	0.19***	0.18***	0.29***	0.16**
	t-stat	4.94	2.71	3.79	3.13	2.82	4.92	2.53

Notes: *,**,*** denotes rejection of the null at the 10%, 5% and 1% significance level, respectively.

4.2.3. *Comparison of individual and consensus forecasts.* To facilitate comparison, Tables 1 and 2 also provide information about the results for consensus forecasts in their last columns. As stated in section 4.1.2, almost all consensus forecasts exhibit characteristics of forecast smoothing and are, thus, not weakly efficient. It is evident that this holds even

[20]To save space the estimates of the horizon-effects are not presented in this paper. They are, however, available from the authors upon request.

Table 6. Efficiency of Consensus Forecasts (With Horizon Effects)

		Germany	Canada	France	Italy	Japan	UK	USA
GDP	beta	0.43***	0.26***	0.29***	0.24***	0.26***	0.36***	0.22***
	t-stat	6.94	4.11	4.27	3.78	4.01	6.30	3.55
Iflation	beta	0.20***	0.16***	0.14**	0.10*	0.19***	0.06	0.27***
	t-stat	3.07	2.98	2.25	1.68	3.22	1.13	4.27
Ind. Prod.	beta	0.44***	-0.02	0.30***	0.28***	0.35***	0.37***	0.27***
	t-stat	6.67	-0.37	4.44	4.06	5.28	6.21	4.44
Priv. Cons.	beta	0.23***	0.17***	0.21***	0.12**	0.16***	0.30***	0.12**
	t-stat	3.69	2.71	3.37	1.99	2.55	4.97	1.96

Notes: *,**,*** denotes rejection of the null at the 10%, 5% and 1% significance level, respectively.

if the wide majority of individual forecasters is weakly efficient. This phenomenon can be explained by the fact that new information is processed by some forecasters slower than by others. It results in positive autocorrelation of the revisions of the consensus forecasts. Similar results hold for the analysis of biases. Naturally, the consensus is unbiased if there are only a few individual panelists who produce biased forecasts. As soon as there is a significant fraction of forecasters who report biased forecasts, it depends on the correlation of their biases whether the consensus will be biased or not. If forecasters deviate into both directions from unbiasedness, the biases might cancel out in the aggregate. This, however, is not the case in our sample. As already mentioned in Section 4.1.1, forecasters tend to be biased into the same direction for a specific target variable. Therefore, the consensus is biased if there is a sizable fraction of biased individual forecasts.

5. Conclusion

In this paper, we tested the unbiasedness and efficiency of macroeconomic forecasts in G7-countries based on survey data from the *Consensus* data set. We analyzed both individual forecasts and average forecasts. The evidence on the properties of forecasts for all G7-counties and four different macroeconomic variables lead us to several conclusions.

First, our results confirm that data on average forecasts should be used with caution since even in a situation where all individual forecasts are rational the hypothesis of rationality is often rejected based on the aggregate data. Second, we find large difference in the performance of forecasters across countries and different macroeconomic variables. Third, among the four kinds of forecasts that we analyze, inflation forecasts perform best in terms of unbiasedness. Fourth, forecasts tend to be biased in situations where forecasters have to learn about large structural shocks or gradual changes in the trend of a variable. Finally, the prevalence of weak inefficiency in aggregate forecasts seems rather to be an artefact of the aggregation itself since it is not necessarily reflected by the individual forecasts' properties.

There are several dimensions along which the study could be expanded in the future. For simplicity, we have assumed that the variance of the macroeconomic shocks ($\lambda_{t,h}$) as well as the variance of the idiosyncratic forecast error ($\epsilon_{i,t,h}$) decay linear if h goes to 1. In the future, one should develop more general functional forms to match the data better.

Second, as soon as enough longer time series become available for individual forecasters one can implement the estimation of horizon specific bias also in the analysis of individual forecasts which would be more appealing from a theoretical point of view. Currently, however, the time dimension of the data set is so small that for most of the panelists the estimates would be based on less than ten observations.

Third, taking into account correlations across countries – like Isiklar et al. (2006) do in their analysis of consensus forecasts – would clearly be desirable given the probably high impact that international shocks have on the size of forecast errors. This, however, requires very large computational power for the estimation of the covariance matrices.

Finally, when modeling the individual error component, we have assumed that it is determined by the private information that forecasters receive as time advances. So far, we did not account for a possible overlap of the forecasters' information sets but rather treated those private information shocks as mutually exclusive. Although such an investigation should prove useful in modeling correlations between different forecast errors more realistically, this was not the main focus of this paper. We believe, however, that it provides a potentially fruitful avenue for future research.

References

Ager, P., Kappler, M., Osterloh, S., 2009. The accuracy and efficiency of the consensus forecasts: A further application and extension of the pooled approach. International Journal of Forecasting 25 (1), 167–181.

Andolfatto, D., Hendry, S., Moran, K., 2008. Are inflation expectations rational? Journal of Monetary Economics 55 (2), 406–422.

Ashiya, M., 2003. Testing the rationality of Japanese GDP forecasts: the sign of forecast revision matters. Journal of Economic Behavior & Organization 50, 263–269.

Ball, R. J., 1962. The prediction of wage-rate changes in the United Kingdom economy 1957-60. Economic Journal 72, 27–44.

Batchelor, R., 2001. How useful are the forecasts of intergovernmental agencies? the IMF and OECD versus the consensus. Applied Economics 33, 225–235.

Batchelor, R., 2007. Bias in macroeconomic forecasts. International Journal of Forecasting 23, 189–203.

Batchelor, R., Dua, P., 1991. Blue chip rationality tests. Journal of Money, Credit, and Banking 23, 692–705.

Batchelor, R., Dua, P., 1992. Conservatism and consensus-seeking among economic forecasters. Journal of Forecasting 11, 169–181.

Blundell, R., Bond, S., Devereux, M., Schiantarelli, F., 1992. Investment and Tobin's Q: Evidence from company panel data. Journal of Econometrics 51, 233–257.

Bonham, C. S., Cohen, R. H., 2001. To aggregate, pool, or neither: Testing the rational-expectations hypothesis using survey data. Journal of Business & Economic Statistics 19 (3), 278–291.

Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A., Rautanen, T., 2007. The ECB survey of professional forecasters (SPF). review after eight years' experience. Occasional paper series 59, European Central Bank.

Capistran, C., Timmermann, A., October 2006. Disagreement and biases in inflation expectations, mimeo.

Clemens, M. P., 1995. Evaluating the rationality of fixed-event forecasts. Journal of Forecasting 16, 225–239.

Clemens, M. P., Joutz, F., Stekler, H. O., 2007. An evaluation of the forecasts of the Federal Reserve: A pooled approach. Journal of Applied Econometrics 22 (1), 121–136.

Croushore, D., 2006. Forecasting with real-time macroeconomic data. In: Elliott, G., Granger, C. W., Timmermann, A. (Eds.), Handbook of Economic Forecasting Volume 1. North-Holland, Amsterdam.

Davies, A., Lahiri, K., 1995. A new framework for analysing survey forecasts using three-dimensional panel data. Journal of Econometrics 68, 205–227.

Edwards, W., 1968. Conservatism in human information processing. In: Kleinmuntz, B. (Ed.), Formal Representations of Human Judgements. Wiley, New York.

Figlewski, S., Wachtel, P., 1981. The formation of inflationary expectations. Review of Economics and Statistics 63, 1–10.

Gallo, G. M., Granger, C. W., Jeon, Y., 2002. Copycats and common swings: The impact of the use of forecasts in information sets. IMF Staff Papers 49, 4–21.

Harvey, D. I., Leybourne, S. J., Newbold, P., 2001. Analysis of a panel of UK macroeconomic forecasts. Econometrics Journal 4, 37–55.

Isiklar, G., 2005. On aggregation bias in fixed-event forecast effciency tests. Economic Letters 89 (3), 312–316.

Isiklar, G., Lahiri, K., Loungani, P., 2006. How quickly do forecasters incorporate news? evidence from cross-country surveys. Journal of Applied Econometrics 21, 703–725.

Keane, M. P., Runkle, D. E., 1990. Testing the rationality of price forecasts: new ecidence from panel data. American Economic Review 80, 714–735.

Keynes, J. M., 1936. The General Theory of Employment, Interest and Money. Macmillan, London.

Lahiri, K., Sheng, X., forthcoming. Evolution of forecast disagreement in a bayesian learning model. Journal of Econometrics.

Laster, D., Bennett, P., Geoum, I. S., 1999. Rational bias in macroeconomic forecasts. The Quaterley Journal of Economics, 293–318.

Mincer, J., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J. (Ed.), Economic Forecasts and Expectations. National Bureau of Economic Research, New York.

Muth, J. F., 1961. Rational expectations and the theory of price movements. Econometrica 29, 315–335.

Nordhaus, W. D., 1987. Forecasting efficiency concepts and applications. Reviev of Economics and Statistics 69, 667–674.

Pesaran, M. H., Weale, M., 2006. Survey expectations. In: Elliott, G., Granger, C. W., Timmermann, A. (Eds.), Handbook of Economic Forecasting Volume 1. North-Holland, Amsterdam.

Phillips, L. D., Edwards, W., 1966. Conservatism in a simple probaility inference task. Journal of Experimental Psychology 72, 346–357.

Pigou, A. C., 1927. Industrial Fluctuations. Macmillan, London.

Stege, C., 1989. Zur Beurteilung der Prognosetätigkeit der "führenden Wirtschaftswissenschaftlichen Forschungsinstitute" in der Bundesrepublik Deutschland. Verlag Dr. Kovac.

Stekler, H. O., 2002. The rationality and efficiency of individuals' forecasts. In: Clements, M. P., Hendry, D. F. (Eds.), A Companion to Economic Forecasting. Blackwell Publishers Inc., Oxford.

Timmermann, A., 2007. An evaluation of the World Economic Outlook forecasts. IMF Staff Papers 54 (1), 1–33.