CONVERGENCE, RATIONALITY AND ACCURACY IN SOUTH AFRICAN CONSENSUS FORECASTS

N. L. Samouilhan[†]

Abstract

The average values of numerous forecasts about South African economic variables are often calculated to give a summary measure of all the forecasters' views. This mean forecast is found to increasingly represent the information held by the various forecasters as the forecast horizon declines. The mean forecast is also found to systematically underpredict large actual outcomes and overpredict low actual outcomes at long forecast horizons, with the opposite behaviour being found at short horizons. The mean forecasts are found to be rational at all horizons and overall. Forecasts of growth in GDP do not become monotonically more accurate as the forecast horizon declines, though forecasts of inflation do. The mean forecast is shown to be more accurate at all horizons than two extrapolation models. No relationship is found between the degree of dispersion of the forecasts around the mean and the means's accuracy for forecasts of both variables.

Keywords: Forecasting, Rationality, Accuracy, Forecast horizon

1. Introduction

As in all countries, many South African institutions require estimates of the future value of certain economic variables in order to operate. The government, for example, requires forecasts of the coming years' GDP growth for the estimation of tax receipts, and exporting companies need forecasts of the Rand/Dollar exchange rate for planning purposes. To address this information need, economists provide forecasts of key economic variables for the coming year. Often, the forecasts of many different economists are averaged in order to arrive at a mean 'consensus' forecast.

This paper investigates some aspects of this 'consensus' forecast. Firstly, the paper looks at whether or not it is appropriate to interpret the mean forecast as a consensus forecast in South Africa. The mean could represent nothing more than a simple statistical construct if the forecasts are widely distributed in an asymmetric pattern. Alternatively the mean would represent a valuable summary of the forecasters' views if the individual forecasts are closely distributed around it and if they all move closely together. Stated differently and looked at dynamically, a convergence by all the forecasts over the horizon towards a certain value would imply that the mean increasingly represents a summary of the heterogeneous information held by the individual forecasters. The paper investigates the behaviour of the forecasts around this 'mean forecast' over the forecast horizon for each year in order to observe the summary ability of the mean forecast.

Secondly, the paper explores whether this 'consensus' forecast is an accurate and rational

[†]*Graduate student, School of Economics, UCT.* *The author is indebted to Mr C.P. Van Walbeek for many helpful comments and for the overall supervision of this paper; and to Professor E. Smit of the Stellenbosch GSB for the data used in this study. As always, any mistakes remain the authors' responsibility.*
Correspondence: nsamouilhan@commerce.uct.ac.za

predictor of the actual outcome for different forecast horizons and different years. While it is impossible for forecasters to accurately know the inherently uncertain future, they should at the very least not consistently over- or under-predict the future value. Each horizon forecast should, in theory, be an unbiased predictor of the future, being accurate on average. Evidence of systemic under- or over-prediction of the variable implies irrational behaviour by the forecaster (provided of course that the goal of the forecasting is to be correct, which may not always be the case) and provides useful information to users of such forecasts.

Forecasters make their best estimates of the future values of variables based on limited information of the future. However, as the forecast horizon declines more and more information about the future value of the variable becomes available. It would therefore be expected that forecast revisions based on more information should be more accurate than those based on less information. At the very least, rational forecasters shouldn't become more inaccurate as the forecast horizon declines. This information rationality aspect is the third characteristic tested in the paper.

Another useful inquiry is whether the forecast mean, and hence South African forecasters, provide useful information about the future value of the variable under investigation. This value-added of economist knowledge is the fourth aspect tested; it is done so by evaluating the forecasting record of the mean forecast against that of two simple pure extrapolation forecasts.

Lastly, the relationship between the dispersion of the forecasters around a mean forecast and the accuracy of that mean forecast is investigated. It would be expected that if there is a high level of agreement amongst economists on a certain future value of a variable then there should be a higher probability of it being correct. If so, users of forecasts could use the degree of forecast dispersion as a guide to the likely accuracy of the forecast value. This expected relationship is tested by investigating the relationship between the mean forecasts' accuracy and the variance of the forecasts around it.

2. Data

The data used in this study is drawn from the annual Beeld 'Economist of the Year' competition jointly run by the Stellenbosch GSB and the Beeld newspaper. In the competition various diverse forecasters from the public, private and academic spheres forecast ten key economic variables for the coming year, such as the growth in South African real GDP for the full year and the inflation rate for the full year. For the competition each forecaster is asked in January to make initial forecasts for the respective year for each of the ten variables. For each subsequent month until November each forecaster is sent the forecasts given by all the other forecasters the previous month, and asked for a revised set of forecasts. In the first few months of the following year the lowest total squared error between each economist's monthly forecasts and the actual is calculated; with the forecasts made at the beginning of the year being weighted more heavily than those made at the end of the year. The forecaster with the lowest weighted mean square error is named as the Beeld "Economist of the Year".

The data set drawn from this competition contains the forecasts of two variables for nine years (1994 to 2002), with forecasts from between 14 and 28 forecasters. The data set contains all the forecasts made at different horizons; the first in January and the last in November for each of the two variables for every respective year. For testing only the mean forecast made in each period, the data set thus contains 9 years of 11 observations for two different series, for a total of two series of 99 observations. When evaluating the full sample with the entire set of individual forecaster's values, the sample increases to 4556 observations. The two variables under examination in this study are the forecasts of the percentage growth of South African real GDP for the full year and the forecasts of the average inflation rate for the full year.[1]

This paper follows the methodology of Keane and Runkle (1990) by evaluating the forecasts against the unrevised GDP and inflation figures[2]. The actuals that the forecasts are measured against are the figures released in the South African Reserve Bank's (March) Quarterly Bulletin the year after the period being forecast. For instance, the figure used to evaluate the GDP forecasts made in 1994 is the figure released in the SARB's 1995 March Quarterly Bulletin for GDP in 1994. Subsequent revisions to the 1994 GDP are ignored as the tests of rationality and accuracy depend critically on knowing exactly what the economists tried to forecast and what information they had when they made those forecasts. Using the revised data essentially changes the rules of the game, making evaluation of the forecasts inconsistent.

The dataset used was almost entirely complete, missing only 49 observations out of 4556 (1.08% of the total). The missing data was dealt with as follows. If a data entry was missing for a particular forecaster the average value of both the previous and next forecasts from that forecaster was used. If the series was missing the initial forecasts, i.e. missing the first n forecasts, then the first known forecast was assumed to be the missing forecasts. Similar methodology was used to complete series missing the last n observations: the last known forecast of that forecaster was assumed to be the missing forecasts.

It was decided to evaluate the forecasts of growth in GDP and the average inflation rate for the year because they meet two key conditions. Firstly, they are two of the most important economic variables whose future values are understood and widely sought. Secondly, compared to other well known economic variables such as the Rand/Dollar exchange rate, they are far 'better behaved' over the sample period. This relative 'stability' is important as the difference amongst forecasts is affected by two principal components: the variability amongst the forecasters and the variability around the forecast variable. By focusing on variables with the most 'stability' the differences amongst the forecasters can be more clearly investigated.

[1] These are the variables defined as code 6006Z and 7032A in the SARB Quarterly Bulletin, respectively.

[2] The CPI figures are never revised during the relevant period under examination, and as such this only affects the GDP figures.

3. A 'Consensus' Forecast?

A problem facing users of forecasting information is that there are very many forecasters whose forecasts usually differ. To save time and effort, many users of such information rely simply on the average forecast as an aggregator of the information provided by the forecasters. Additionally, the media often only present the average forecast when discussing expected future values; the Reuters poll of economists is a good example of this. However, it needs to be investigated whether consumers of such information are correct in treating the mean forecast as a convenient summation of the views of forecasters. If it is a good information summary then the average forecast can indeed be meaningfully interpreted as a 'consensus' forecast. Alternatively, the mean could simply be a statistical construct, the forecasts being widely and asymmetrically distributed around the mean and behaving in an erratic manner not shown in the movement of the mean.

If we look at it dynamically, a convergence by all the forecasts over the horizon towards a certain value would imply that the mean increasingly becomes a useful summary of the heterogeneous information held by the individual forecasters. Theoretically, if there is a point where all forecasters fully converge on a certain value the mean would be a perfect summary of the information provided by all the forecasters. In reality, forecasters never come to a full agreement on a certain value, even towards the end of the forecast period. If, however, it can be shown that forecasters converge towards a value as the horizon declines then the mean forecast can be viewed as an increasingly appropriate summary of the individual forecasts.

This definition of consensus is the dominant view in the forecasting literature. McNees (1997) and Zarnowitz (1985), for instance, used this definition of consensus to test US forecasts, finding growing consensus amongst forecasters. However, while convergence on a certain value is the most frequent interpretation of a growing consensus, it is certainly not the only view. The literature provides another three definitions of consensus. Lahiri and Teigland (1987) and Schnader and Stekler (1991) propose a less strict definition of consensus. Here, a consensus amongst forecasters exists if the distributions of the forecasts are unimodal, symmetric around the mean and at least as peaked as the normal distribution. This is a far broader definition of consensus, requiring a general agreement around rather than exact concurrence on a certain value. Gregory and Yetman (2001) propose yet another definition of forecast consensus. In their view a consensus exists if the individual forecasters only differ from a latent, common variable by an orthogonal component with a zero mean. Finally, Gregory, Smith and Yetman (2001) hold that forecasters are in consensus if their forecasts are insignificantly different from the forecast mean.

Unfortunately, all three of those definitions of consensus require more observations than are available in this data set to be tested rigorously. For specific periods, for example, the dataset only contains between 13 and 30 observations (forecasts), making tests of normality impossible. The test for consensus in this paper will therefore use the dominant strict definition of convergence: whether or not individual forecasters are converging on a certain variable as the forecast horizon declines. While this does not allow us to say whether or not consensus exists at a certain forecast horizon, it does allows us to talk of a certain forecast period exhibiting greater consensus and hence aggregation of the forecasters' private views than another forecast horizon.

As explained by Gregory and Yetman (2004), there are three main reasons why it should be expected that such convergence over the horizon will occur. Firstly, there is less uncertainty with respect to the forecast variable as the horizon decreases. As the year progresses information regarding factors that affect the relevant variables, such as the level of the Rand/Dollar or the inflation figures for past months, become known with more certainty. This leads to more uniformity of the information held by forecasters whereupon their forecasts are based. Secondly, forecasters have the ability to see other forecasters' forecasts and take this information into account when revising their own forecasts. In the competition that the dataset is drawn from all forecasters receive the other forecasters' values from the previous revision before they make their next submission, so this could be a factor. Lastly, new information that affects the variable under consideration is available to all, and this common information is taken into account universally. Provided economists believe that this new information affects the variables in the same way the information will lead them to revise their forecasts in the same direction.

Operationally, the strict definition of a growing consensus is tested by examining whether a significant relationship exists between the horizon and the variance of the sample. Formally, we estimate:

$$\sigma_{k,t}^{2} = \alpha_{k,t} + \beta_{k,t} t + \mu_{k,t} \qquad (1)$$

where $\sigma_{k,t}^{2}$ is the variance in year k and horizon t.

Growing convergence amongst the forecasters would require the horizon coefficient, $\beta_{k,t}$, in equation (1) to be both positive and significant, with the forecast horizon running at monthly intervals from November (t=1) to January (t=11). An insignificant horizon coefficient would imply that there was no growing convergence amongst the forecasts, and a coefficient that was both significant and negative would imply a growing divergence. In addition, the magnitude of the estimated coefficient is an indication of the rate of convergence/divergence of the various forecasters on a consensus forecast. It would also be expected that the intercept be both significant and positive; it is unrealistic to expect forecasters to come to universal agreement on an exact value, even at the end of the forecast horizon.

TABLE 1: Average Variance Regressed on Forecast Horizon, GDP Forecasts 1994 to 2002

Year	Prob(F-Stat)	R^2	$\alpha_{k,t}$	$\beta_{k,t}$	Behaviour
1994	0.045	.374	0.01 (0.25)	0.01 (2.32)	Convergence
1995	0.008	.554	0.02 (2.00)	0.01 (3.34)	Convergence
1996	0.000	.774	0.01 (1.18)	0.01 (5.55)	Convergence
1997	0.005	.541	0.02 (1.64)	0.01 (3.57)	Convergence
1998	0.941	.001	0.19 (3.48)	0.00 (0.07)	No Relationship
1999	0.000	.960	-0.01 (0.5)	0.04 (14.8)	Convergence
2000	0.002	.666	0.04 (.033)	0.01 (4.23)	Convergence
2001	0.009	.543	0.16 (7.53)	-0.01 (-3.27)	Divergence
2002	0.000	.741	0.06 (7.26)	0.01 (5.08)	Convergence
Avg	*0.000*	*.855*	*0.06 (5.88)*	*0.01 (7.28)*	Convergence

Note:
t-statistics in brackets after coefficients

TABLE 2: Average Variance Regressed on Forecast Horizon, Inflation Forecasts 1994 to 2002

Year	Prob(F-Stat)	R^2	$\alpha_{k,t}$	$\beta_{k,t}$	Behaviour
1994	0.355	.095	0.19 (2.70)	-0.01 (-0.9)	No relationship
1995	0.008	.555	0.26 (6.12)	0.02 (3.35)	Convergence
1996	0.109	.259	0.81 (4.61)	-0.04 (-1.7)	Divergence
1997	0.000	.834	-0.06 (-2.2)	0.02 (6.73)	Convergence
1998	0.000	.810	-0.00 (-0.23)	0.02 (6.20)	Convergence
1999	0.000	.729	0.02 (0.62)	0.03 (4.92)	Convergence
2000	0.000	.903	-0.07 (-2.05)	0.04 (9.16)	Convergence
2001	0.001	.683	0.05 (2.23)	0.01 (4.40)	Convergence
2002	0.000	.735	1.35 (18.2)	0.01 (-4.9)	Convergence
Avg	*0.018*	*.475*	*0.02 (18.45)*	*0.01 (2.85)*	Convergence

Note:
t-statistics in brackets after coefficients

There is strong evidence that the forecasts converge for nine of the eleven years for the GDP forecasts, although at very different rates (see Table 1). In 1995 for example, forecasters converge on a 'consensus' at a far slower rate (.01) than in 1999 (.04). In addition, in one year, 1998, there is no evidence either way for convergence or divergence of the GDP forecasts. In another, 2001, there is evidence of a divergence of the GDP forecasts, implying a growing disagreement over the variable's value by the forecasters. In general though, the years show a positive relationship between the forecast horizon and the level of variance.

The forecasts for inflation show a similar pattern of general convergence both overall and for nine of the eleven years, though, as found with the GDP forecasts, at very different rates. In addition, again as found with the GDP forecasts, there is evidence of divergence in one year (1996) and no evidence either way for another (1994). It should be noted however that the years where non-convergence is found among the inflation forecasts are not the same years where non-convergence is found among the GDP forecasts.

This overall pattern of growing convergence is widely found internationally. Amongst others, Gregory and Yetman (2004), Batchelor (1990), Spiro (1989) and Zarnowitz (1984) found convergence in US forecasts.

In summary, even though we have not tested whether or not the mean forecast is a consensus forecast, we can say the following regarding it: In general, as the forecast horizon declines the mean forecast increasingly represents a useful summary of the views held by the forecasters. As such, users of such information can increasingly rely on the mean as an appropriate aggregator of the heterogeneous information held by the different forecasters.

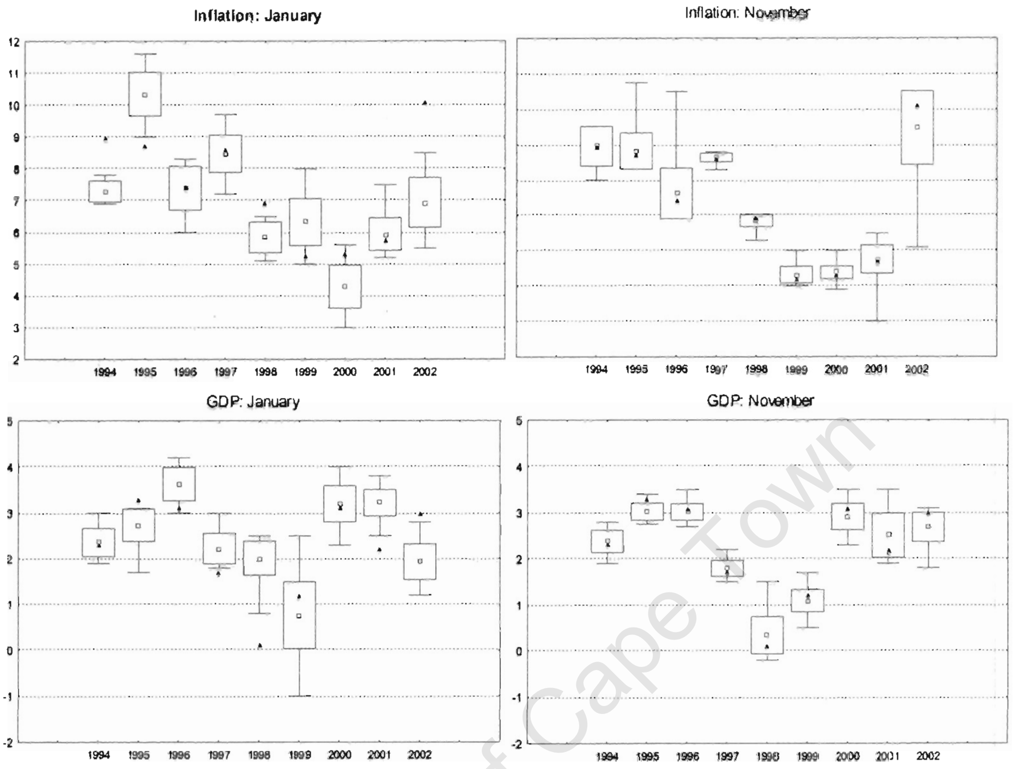
4. Aspects of the Mean Forecast

4.1 Horizon Dynamics

While the mean forecast can be increasingly interpreted as a useful summary of the forecasters' views by users of such information, it obviously needs to be tested how accurate this mean forecast really is. Using the mean as a measure that captures the available forecast information so as to minimise costs (in its broadest term) is of no use if the mean forecast is generally inaccurate. As an initial evaluation of the accuracy of the mean forecasts the comparison of the Box and Whisker plots of January forecasts to the November forecasts (Figure 1, below) provides some key initial insights about the accuracy of the forecasts over the forecast horizon.

Broadly speaking, the spread of the forecasts tends to narrow as the forecast horizon declines; an aspect that concurs with earlier evidence of a growing consensus. In addition, the mean of the forecasts also tends to move towards the actuals as the horizon declines. Forecasters appear to be interpreting new information about the GDP correctly as the year progresses and become increasingly accurate as the forecast horizon declines. (This relationship between the forecast horizon and forecast accuracy is considered in more detail later on)

FIGURE 1: Initial (January) and Final (November) GDP and Inflation Forecasts, 1994 to 2002



Note: The box represents 1 standard deviation above and below the forecast mean, 95% of the forecasts thus lie within the box limits. The whiskers represent the highest and lowest forecasts, the small square the mean consensus forecast and the solid triangle the actual value of the forecast variable.

For the GDP forecasts the initial January forecasts are, in general, not very accurate at all. For five of the nine years the entire spread of the January forecasts fails to include the actual. In 1994 and 2000, however, the mean January forecasts are remarkable close to the actual; and in 1999 the actual is within one standard deviation of the mean, though from a very wide spread. Lastly, in no single year is the actual completely unexpected; in each year the actual falls inside the full spread of the forecasts made in November. In fact, with the exception of 1995, the actual always falls within one standard deviation of the end of year mean forecast.

The January forecasts of inflation are also not very accurate on the whole. The entire spread of forecasts fails to include the actual for four of the nine years. The actual forecast was, however, very close to the mean for 1996, 1997 and 2001. As with the GDP forecasts, the actual falls within the complete spread of the November forecasts for every single year, in this case always within one standard deviation of the mean forecast.

4.2 Forecast Rationality

One of the major questions this paper seeks to address is whether the forecasts are rational. In testing for this the paper follows the interpretation of the rational expectations literature in defining rationality to mean that rational expectations are exactly the mathematical conditional expectations inferred by the model (Bonham and Cohen 2000), a widely accepted definition of rationality associated with John Muth (1961).

Before evaluating the forecasts though, two qualifications about the results of the rationality test have to be made. The first is that, as Bonham and Cohen (2000) point out, we cannot infer characteristics about the rationality of individual forecasters by investigating the rationality of consensus forecasts because of aggregation bias. This aggregation bias is the difference between the parameter estimates of the average forecast and the average of all the parameter estimates obtained from the analysis of individual forecasters. The second caveat is that the rationality test assumes that the objective of forecasters is to be as accurate as possible. This may not always be the case; Lomant (2002), for instance, observes that forecasters may use their forecasts to differentiate themselves and influence beliefs about their abilities.

The first and most important question of the mean forecast when evaluating rationality is whether or not it is consistently too optimistic or pessimistic. While the visual plots suggest that there is no apparent systematic under- or over-prediction of GDP or inflation it can be tested formally whether the mean forecast is indeed an unbiased predictor of the actual. Formally, the relation between an unbiased forecast and the actual is:

$$E[A_T - F_{T-t,T} | I_{T-t}] = 0 \tag{2}$$

where A_T is the actual value of the variable known at time T, $F_{T-t,T}$ is the consensus forecast of the variable for period ending T made at period T-t, t is the forecast horizon, and I_{T-t} the information set available to the forecasters at time T-t.

Operationally, we test the above formulation of unbiasedness of the forecasts by estimating the following model and testing the joint unbiased hypothesis that $\alpha_t = 0$ and $\beta_t = 1$:

$$A_T = \alpha_t + \beta_t F_{T-t,T} + \mu_{T-t,T} \tag{3}$$

Using this formulation, a value for α_t of 0 would imply that there was no overall under or over prediction of the relevant variable, irrespective of the actual. A value for β_t of 1 would imply that there was no actual-dependant inaccuracy on the part of the forecasters. Both are required for the forecasts to be unbiased.

4.2.1 An Estimation Problem

A problem, as stated previously by Swidler and Ketcher (1990), Hansen and Hodrick (1980) and Brown and Maital (1981), is that ordinary-least-squares (OLS) estimation of equation (3) may be inappropriate due to serial correlation of the error terms, violating a key assumption of OLS estimation. Serial correlation of the error terms would be expected in this sample as the forecasters are unaware until after period T what their forecast errors are for all their previous forecasts made at periods T-t. For instance, the forecast error made in February 1994 is likely to be strongly correlated with the forecast error made in January 1994 as forecasters will not know the magnitude of both errors until the actual 1994 GDP figure is released at the beginning of the following year. Adding to this predisposition for serial correlation is that the forecasts are often not revised from one period to the next. Estimating equation (3) for the full sample using OLS will still generate unbiased estimators for α_t and β_t but have inefficient standard errors, making accurate inference impossible. As such, the Newey-West Heteroscedasticity-and-Autocorrelation–Consistent (HAC) standard errors were calculated (shown below the normal OLS estimates).

However, serial correlation and heteroscedasticity are not expected to be present when testing the rationality of forecasts made at certain horizons, such as for the January forecasts for example. There is no reason to expect a correlation between, say, the January 1994 error and the January 1995 error; and no reason to expect the accuracy in January 1994 to be systematically different from the accuracy in January 1995[3].

Tables 3 and 4, below, show the results of the estimation of equation (3) for both the inflation and the GDP forecasts. These coefficients, along with the other statistics in the table, are explained in the sections that follow.

[3] Statistical testing confirmed this: neither of the two series showed evidence of serial correlation or heteroscedasticity at any horizon at the 5 % significance level.

TABLE 3: GDP Forecasts Rationality Tests
By Horizon

Horizon, T-t	α_t	β_t	R^2	MSE	Unbiased*	n
January, 11	0.33 (0.34)	0.76 (-0.63)	.372	0.738	0.552	9
February, 10	0.43 (0.95)	0.72 (-0.73)	.357	0.759	0.575	9
March, 9	0.40 (0.42)	0.71 (-0.79)	.364	0.801	0.844	9
April, 8	0.35 (0.35)	0.74 (-0.69)	.366	0.778	0.733	9
May, 7	-0.02 (-0.03)	0.93 (-0.20)	.506	0.554	0.296	9
June, 6	-0.22 (-0.32)	1.07 (0.14)	.668	0.355	0.185	9
July, 5	-0.26 (-0.46)	1.03 (0.33)	.749	0.266	0.163	9
August, 4	-0.24 (-0.62)	1.08 (0.50)	.865	0.145	0.226	9
September, 3	-0.27 (-0.78)	1.13 (0.91)	.891	0.123	0.430	9
October, 2	-0.21 (-0.59)	1.12 (0.80)	.884	0.131	0.418	9
November, 1	-0.19 (-1.24)	1.08 (1.38)	.976	0.030	0.960	9
Full Sample *HAC Standard Errors Adj. Figure*	0.02 (0.14) *(0.08)*	0.93 (-0.90) *(-1.16)*	.603		2.67* *(1.33)*	99

TABLE 4: Inflation Forecasts Rationality Tests
By Horizon

Horizon, T-t	α_t	β_t	R^2	MSE	Unbiased$^{\Psi}$	n
January, 11	3.52 (1.62)	0.55 (-1.40)	.305	2.77	1.49	9
February, 10	3.33 (1.53)	0.58 (-1.35)	.336	2.56	1.32	9
March, 9	2.21 (1.11)	0.72 (-1.03)	.516	1.63	0.87	9
April, 8	1.95 (1.04)	0.74 (-1.03)	.571	1.42	0.57	9
May, 7	1.72 (0.92)	0.76 (-0.98)	.600	1.26	0.48	9
June, 6	1.23 (0.78)	0.82 (-0.79)	.663	1.02	0.32	9
July, 5	0.52 (0.37)	0.92 (-0.39)	.769	0.65	0.09	9
August, 4	0.05 (0.53)	0.98 (-0.13)	.868	0.37	0.08	9
September, 3	-0.39 (-0.56)	1.00 (0.52)	.947	0.15	0.15	9
October, 2	-0.37 (-1.57)	1.04 (1.35)	.993	0.02	1.66	9
November, 1	-0.12 (0.87)	1.00 (0.42)	.997	0.01	2.34	9
Full Sample *HAC Standard Errors Adj. Figure*	1.44 (3.16)*** *(3.37)****	0.81 (-0.92) *(-1.48)*	.645		5.03*** *(1.96)*	99

Notes for Tables 3 and 4:
n: number of observations in sample
t-statistics in brackets after coefficients. Values calculated for $\alpha_t = 0$ and $\beta_t = 1$, respectively
Unbiased$^{\Psi}$: F-stat of Joint Wald test for $\alpha_t = 0$, $\beta_t = 1$ in eq. $A_T = \alpha_t + \beta_t F_{T-t,T} + \mu_{T-t,T}$
*** significant at the 1% level
** significant at the 5% level
* significant at the 10% level

4.2.2 Systemic Biasedness

The estimated coefficients of the horizon forecasts display an interesting pattern. Generally speaking, we can separate both the GDP and inflation forecasts into two broad periods: a period where forecasters are overly cautious and anchor their forecasts to a historical mean and a period where they overreact and take extreme positions.

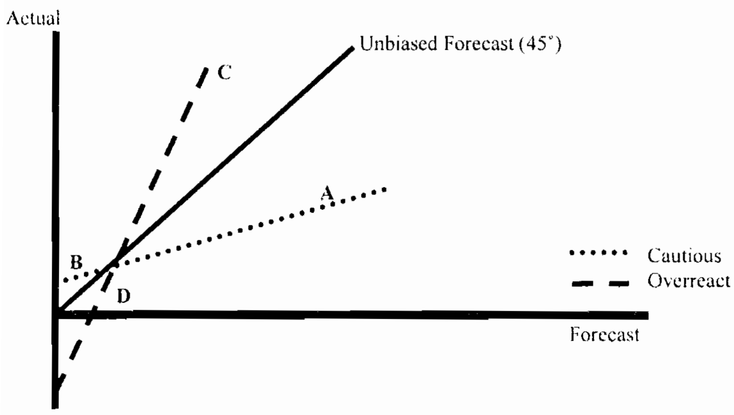
Figure 2 (below) aids in explaining the behaviour of the forecasts over the forecast horizon. From January to May for the GDP forecasts and January to August for the Inflation forecasts it is found that $\beta_t < 1$ and $\alpha_t > 0$. Forecasters tend to systematically underpredict high levels of inflation (GDP growth) and overpredict low levels of inflation (GDP growth) in the beginning of the year. Stated differently, when the actual inflation rate is low, forecasters will initially forecast values that are too high (Consider Pt. D in Fig 2). If the actual inflation is high, then forecasters will behave as if they are at Pt C in Fig 2; they will forecast values that are systematically too low.

This 'cautious' period could be present because forecasters lack sufficient information at the beginning of the year and will forecast Inflation (GDP growth) values that are not too far from the average value of Inflation (GDP growth) in the past, just to be prudent and keep credibility. Essentially, when having to forecast in an information-poor period forecasters revert to mean values, using them as an anchor that their forecasts are contingent upon.

Conversely, it is found that from June to November for GDP and September to October for Inflation $\beta_t > 1$ and $\alpha_t < 0$. The evidence suggests that in the latter part of the year forecasters tend to systematically overpredict high levels of inflation (GDP growth) and underpredict low levels of inflation (GDP growth). In reference to Fig. 2, when the actual inflation rate is high, forecasters will in the latter part of the year forecast values that are too high (Pt. A). If the actual inflation is low, then forecasters will behave as if they are at Pt. B; they will forecast values that are systematically too low in the final months of the year. A possible explanation for this is that forecasters are overreacting to past information regarding Inflation (GDP growth) as it becomes known later on in the year. A further reason might be a function of the competition from where the data is drawn; forecasters could be taking extreme positions in order to better their chances of winning the competition[4].

[4] If a forecaster participating in the competition wishes to 'win' the forecasting of a certain variable, in the sense of being the closest to it, a certain strategy is to take an extreme position on the variable, with extreme being a value significantly above or below what all the other participating forecasters are expected to forecast for the variable this time round, with the hope that everyone else will severely under/over forecast it. Forecasting a similar value to all the other forecasters might be the safest, in that it minimises the risk of being wrong, but it removes the chance of being the closest forecaster of that variable.

FIGURE 2: Systemic Under and Over Prediction of the Actual



4.3 Information Rationality: Revisions and Accuracy

A rational consensus forecasts implies another feature beyond unbiasedness. If the mean forecast is a rational processor of information and uses all available information available then the mean forecast made at time T-t must take into consideration an information set that contains all previous information sets. It would consequently be expected that the forecasts will become more accurate as the forecast horizon declines as the forecasters gain more and more information; allowing them to make more and more accurate forecasts. Formally, this aspect of rationality requires:

$$E[A_T - F_{T-t,T} | I_{T-t}] - E[A_T - F_{T-t-j,T} | I_{T-t-j}] \leq 0 \quad (4)$$

with $j \geq 0$

This requirement of increasing accuracy of the forecasts can easily be tested using the dataset; the Beeld competition allows forecasters to revise their forecasts each month from January until November.

The Mean Squared Error (MSE) was used to gauge the relative accuracy of the forecast revisions; it is the measure of accuracy that is most frequently used to evaluate forecast accuracy in the literature. It is defined as:

$$MSE = \frac{[\sum (A_T - F_{T-t})^2]}{n} \quad (5)$$

where A_t is the actual value of the forecast variable for period T, F_{T-t} is the forecast value for respective variable and n is the number of years over which the horizons tests are estimated.

As can be seen in tables 3 and 4 the accuracy of the forecasts displays a peculiar pattern for the GDP forecasts, at odds with the constant increase in forecast accuracy required for rationality. The accuracy of the forecasts initially decreases from January to March, where they are the most inaccurate, then become increasingly more accurate from April until September, with large increases in accuracy in August (45%), June (36%) and May (29%). Then, after a very slight decrease in accuracy in October, they reach their most accurate period in November.

The inflation forecasts on the other hand show full rationality with the use of information. The forecasts become continuously better as the forecast horizon declines, and although it is a monotonic increase it is far from being smooth. There are large increases in accuracy in October (83%), November (60%), September (59%) and August (42%). Nevertheless, the inflation forecasts' accuracy is monotonically increasing as the forecast horizon declines, suggesting that the mean inflation forecasts make full and rational use of all available information. A possible contributing factor is that, in contrast to the information available on GDP, information releases on inflation are frequent: figures for previous months' inflation rates are available the very following month for use by forecasters.

International evidence of the accuracy of revisions is mixed. Swidler and Ketcher (1990) test forecasts of growth in real US GDP from 1976 to 1988 for horizons of one to eleven months and find that the forecasts improve monotonically as the horizon declines. In contrast, Kolb and Steckler (1990), testing a single major forecaster of US GNP growth from 1972 to 1983, find an overall increase in accuracy but with periods of decreasing accuracy. Cho (2002), in contrast, finds that revisions to forecasts of US GDP growth do not improve in accuracy significantly at all.

In summary, although both GDP and inflation forecasts become more accurate as the forecast horizon declines only the inflation forecasts exhibit rational use of available information. The lack of full rational use of all available information in the GDP forecasts implies that there is possibly some room for improvement in the accuracy of them.

4.4 Relative Accuracy of the Consensus Forecast to Simple Forecasts: Naïve and Trend Forecast Alternatives

While the MSE measure gives an indication of the how the accuracy of the mean forecasts behaves over the different horizons, it does not allow us to evaluate the accuracy of the mean forecast by itself. In order to do so one can compare the mean forecast with that of other forecasts. Here, the consensus forecast is compared to two momentum models, where the only information used to generate them are their past, publicly known values.

The first extrapolation forecast that the consensus forecast is compared against is a naïve forecast, where the expected future value of the relevant variable is equal to its value the previous year. The second comparative extrapolation series is a trend forecast that has as its forecast a moving three-year average of the past values of the series being forecasted.

A key problem when evaluating both of these constructed series against the consensus forecast is that the consensus forecast has the advantage of increasing information as the forecast horizon declines, information that it incorporates with each of its revisions. The two extrapolation series are thus heavily penalised when evaluating their forecasting ability over the full year horizon compared to that of the consensus forecast as they are not revised to include any new relevant information. Useful inferences about comparative forecasting ability can therefore only really be made about the initial forecasts at the beginning of the forecast period.

To address this problem and provide for a better evaluation of the consensus forecasts record over the full horizon, each of the two series were augmented with a simple learning revision to incorporate new information, but done so in a way to keep them both as simple models.
The learning heuristic is based on the fact that, as the forecast horizon declines, part of the forecast variable becomes known with certainty and as such the amount of the variable that needs to be forecast declines. For example, when forecasting the inflation figures (which are an average for the full year) in January the entire figure is unknown and needs to be forecasted, but by February only 11/12^ths^ of its value is unknown and needs to be forecasted, January's inflation rate (1/12^th^ of the full years average value) having been publicly released before the revised forecasts are submitted. By November, only one twelfth of the years' inflation rate is unknown and needs to be forecasted, the inflation rate for each of the previous eleven months is already publicly known with complete certainty.

Based on this principle, both the naïve and trend forecasts are revised each month by including the information known about their forecast variable that is available at that time, and leaving the rest as the naïve/trend forecast. The naïve forecast of inflation in March 2004, for example, is constructed by adding the previous full year's inflation figure (weighted by 10/12) to the released inflation figures for February 2004 (weighted by 1/12) and January 2004 (weighted by 1/12). Importantly, the monthly releases of inflation figures allow the naïve and trend inflation forecasts to be updated each month.

Quarterly GDP figures are released just before the revised forecasts in May (1st quarter), August (2nd quarter) and November (3rd Quarter)[5], and as such the simple GDP forecasts can only be revised three times. The naïve forecasts of GDP for May 2004, for example, are the summation of the previous years GDP figure (weighted 3/4) and the GDP for the first quarter (weighted 1/4). GDP forecasts made before the May release are completely naïve, and forecasts made between GDP quarterly releases stay as the previous revised figure. June and July, for instance, have the same value as their forecasts as May, as they are made before the August release of the second quarter GDP figures.

[5] Importantly, both series are augmented using the unrevised figures, continuing the rule of using only the information known at that specific time.

Theil's (1996) U-Statistic was used to evaluate the forecasting record of the mean forecast against that of the two augmented simple extrapolation forecasts; it is by far the most frequently used measure of comparison between two forecasts. Theil's U-Statistic is defined as:

$$U_T = \frac{\sqrt{\sum (F_{T-t,T} - A_T)^2}}{\sqrt{\sum (E^*_{T-t,T} - A_T)^2}} \tag{6}$$

Here, Theil's U-Statistic evaluates the mean forecast at each period against the relevant revised extrapolation forecast, E^*. A U-statistic of less than 1 would imply that the mean forecast is a more accurate predictor of the actual than the simple extrapolation; a value of greater than 1 would imply that the opposite was true. Stated differently, a U-statistic of greater than one would imply that there was no value added by the information provided by the forecasters. Theil's U was estimated for both the inflation and GDP forecasts, comparing the revised trend and naïve forecasts to the consensus forecast over both the forecast horizon and for individual years. The results are given in tables 5 and 6, below.

TABLE 5: Relative Accuracy of the Consensus forecast to Extrapolation forecasts, By Horizon

	GDP		Inflation	
	Naive	Trend	Naive	Trend
January	0.707	0.462	0.927	0.640
February	0.718	0.470	0.895	0.638
March	0.738	0.483	0.728	0.533
April	0.727	0.476	0.691	0.521
May	0.622	0.421	0.676	0.526
June	0.502	0.340	0.633	0.508
July	0.434	0.294	0.523	0.436
August	0.361	0.278	0.412	0.395
September	0.334	0.257	0.311	0.271
October	0.344	0.265	0.151	0.134
November	0.212	0.193	0.098	0.101

Over the full forecast horizon (Table 5, above) the mean consensus forecast is consistently better than both the naïve and the trend forecasts for both GDP and inflation. Between the two extrapolation series, the naïve forecasts are, in general, more accurate than the trend forecasts. Importantly, both are relatively more accurate at long horizons than short. The naïve forecast in January for inflation, particularly, is almost as accurate as the consensus forecast, while it becomes a very poor forecast by November. This general decline in relative accuracy as the horizon declines implies that forecasters are utilising more information when making their forecasts than only the past known values of the relevant variable concerned.

Looking at the relative forecast accuracy for particular years (Table 6, below), the pattern is very different. The naïve forecast of GDP is more accurate than the consensus forecasts in 1996, and both the trend and the naïve forecasts are relatively more accurate in 2001 and 2002. For the inflation forecasts, the naïve forecasts of inflation are better in 1994 and substantially better in 1995, and the trend forecasts better in 2001. In 1998 both extrapolation series are relatively more accurate than the consensus forecasts. Conversely, in 1997, 1999 and 2000 the consensus forecast is far more accurate than both the naïve and the trend for both variables.

Comparatively, although no single forecast is always the most accurate, in general the consensus forecasts can be said to be the most accurate, the naïve forecasts the second most accurate and the trend forecasts the least accurate of the three. Users of mean forecasts are thus justified in using them as they provide relevant additional information on the future value of the respective variable.

TABLE 6: Relative Accuracy of the Consensus forecast to Extrapolation forecasts, By Year

	GDP		Inflation	
	Naive	Trend	Naïve	Trend
1994	0.213	0.135	1.849	0.388
1995	0.432	0.187	2.461	0.850
1996	1.247	0.489	0.413	0.321
1997	0.446	0.525	0.313	0.723
1998	0.983	0.636	1.120	1.383
1999	0.446	0.531	0.453	0.363
2000	0.154	0.141	0.735	0.649
2001	1.188	1.389	1.456	0.936
2002	1.137	1.112	0.525	0.502

4.5 Degree of Forecast 'Consensus' and Relative Accuracy

A useful additional question regarding the mean forecasts is whether meaningful information can be gained about the accuracy of the mean forecast from the level of dispersion of the forecasts around it. A greater 'consensus' amongst the forecasters (i.e. less dispersed forecasts) should imply less uncertainty around the variable being forecasted, and hence possibly a greater probability of it being more accurate, whereas less 'consensus' around the forecast value should be associated with more uncertainty and hence less accuracy. Stated differently, does a tight grouping of forecasts provide evidence of a more accurate mean forecast, and a large dispersion of the forecasts evidence of a poor mean forecast? If such a relationship exists, users of forecasts could employ the level of agreement of the forecasts as a guide to how much confidence to put into the mean forecast.

This relationship was tested by looking at the relationship between the absolute error of the forecasts and the variance of the economists' forecasts for each period, specifically:

$$E[A_t - F_{T-t,t} \mid \sigma^2_{T-t,T}] \tag{7}$$

A problem is that both the forecast accuracy and the dispersion are affected by the forecast horizon; both of them declining as the year progresses. As such, the association can only be tested within each horizon separately.

The variance-accuracy relationship of the forecasts was examined in two ways. Firstly, simple correlations between the absolute error of the forecasts and the variance amongst the forecasters at that period were calculated for each horizon. Secondly, to address the problem that strong isolated correlations in one year might distort the overall results, Spearman's Rank correlations were also calculated for each horizon. In both cases the correlations would be expected to be positive, as greater agreement would be expected to imply greater accuracy. Both figures are shown for all horizons in table 7 below.

TABLE 7: Spearman's Rank Correlation and Normal Correlation between the Variance and Accuracy of Per Horizon Forecasts

	GDP		Inflation	
Horizon, T-t	Spearman's Rank R	Correlation r	Spearman's Rank R	Correlation r
January, 11	0.00	-0.13	0.26	0.43
February, 10	-0.16	-0.15	0.33	0.67*
March, 9	-0.03	0.08	0.40	0.76*
April, 8	-0.06	-0.05	0.55	0.69*
May, 7	0.29	0.11	0.26	0.53
June, 6	0.34	0.55	0.37	0.57
July, 5	0.38	0.25	0.36	0.35
August, 4	0.27	0.24	0.41	0.46
September, 3	0.29	0.29	0.23	0.62
October, 2	0.49	0.48	0.56	-0.09
November, 1	0.70***	0.58	0.51	-0.04

Note:

*** significant at the 1% level
** significant at the 5% level
* significant at the 10% level

It appears that, in general, no meaningful information regarding a forecasts' accuracy can be gained from examining the dispersion of the forecasts for GDP. Firstly, no pattern of the direction of the relationship between the two emerges. Secondly, the correlations between them are not significant for all horizons except the very last forecast period (November). Its positive values of 0.70 and 0.58 imply that a large agreement amongst the forecasters in November is an indication of a relatively more accurate November forecast than if there was a large dispersion of the forecasts that month. Unfortunately, not only is the relationship relatively weak, but it is right at the end of the forecast period and as such this finding is of very limited service to the users of forecasts.

The inflation forecasts also do not appear to be generally more accurate if there is more agreement about its future value. Only in February, March and April is there a significant relationship, but this is only implied in the simple correlations and not the Spearman's R. As such, the correlations cannot be said to provide tangible evidence of a relationship even for these few months for the inflation forecasts.

This pattern of no general link between dispersion and accuracy was also found by Swidler and Ketcher (1990) in US GDP forecasts using similar methodology. They found that only forecasts made at a horizon of six months had a positive correlation with the degree of agreement and level of accuracy, with none of the other months showing any pattern or significance.

5. Concluding Comments

This paper used the Beeld/Stellenbosch GSB Economist of the Year competition forecasts to test various aspects of the average value of the forecasts of various forecasters. When forecasters make repeated revised forecasts of future GDP and inflation figures, their revised forecasts tend to converge on each other as the forecast horizon declines. It could therefore be argued that the mean forecast thus represents an increasingly appropriate summary of the information held by the individual forecasts.

While both the consensus forecasts of GDP and Inflation forecasts are found to be rational overall, there is evidence that the mean forecast under-predicts historically large values and over-predicts historically small values of both variables at long forecast horizons, and does the opposite at short horizons. If forecasters expect a relatively high inflation or GDP growth figure early on in the year by historical standards, users of such forecasts should therefore view them as being probably too low, and if the initial forecast values of inflation and GDP growth are historically low forecast users should view them as being probably too high. In the latter part of the year, users of mean forecasts should view historically high forecasts are probably being too high, and historically low figures as being too low.

Overall the mean forecast is observed to become increasingly more accurate as the forecast horizon declines, so forecasts at large horizons cannot be relied upon as much forecasts at short horizons. Lastly, there appears to be no relationship between the level of consensus around a certain forecast value and its accuracy. Users of forecasts cannot interpret strong agreement around a certain value as being indicative of a more accurate forecast than if there was less agreement around it. The mean forecast generated from a sample of widely dispersed forecasts appears to be no less accurate as a mean forecast generated from a highly concentrated field of forecast values.

References

Batchelor, R (1990): "All forecasters are equal", Journal of Business & Economic Statistics, 8, 143-144

Bonham, C and Cohen, R (2000): "Rational consensus forecasts?", The Journal of Economic Perspectives, 14(4), 228-230

Brown, B and Maital, S (1981): "What do economists know? An empirical study of experts expectations", Econometrica, 49, 491–504

Gregory, A and Yetman, J (2001): "Testing for consensus", Journal of Business and Economic Statistics, 19, 34-42

Gregory, A and Yetman, J (2004): "The evolution of consensus in macroeconomic forecasting", International Journal of Forecasting, 20, 461–473

Gregory, A, Smith G and Yetman, J (2001): "Testing for forecast consensus", Journal of Business and Economic Statistics, 19(1), 34-43

Hansen, L and Hodick, R (1980): "Forward exchange rates as optimal predictors of future spot rates: An economic analysis", Journal of Political Economy, 88, 829-853

Keane, M and Runkle E (1990): "Testing the rationality of price forecasts: New evidence from panel data", The American Economic Review, 80, 714-735

Lahiri, K and Teigland, C (1987): "On the normality of probability distributions of inflation and GNP forecasts", International Journal of Forecasting, 3, 269-279

Lamont, O (2002): "Microeconomic forecasters and macroeconomic forecasts", Journal of Economic Behaviour & Organization, 48, 265-280

McNees, S (1987): "Consensus Forecasts: Tyranny of the majority?", New England Economic Review, Nov/Dec, 15-21

Muth, J (1961): "Rational Expectations and the theory of price movements", Econometrica, 29, 1961, 315-335

Schnader, M and Stekler, H (1991): "Do consensus forecasts exist?", International Journal of Forecasting, 7, 165-170

Spiro, P (1989): "Improving a group forecast by removing the conservative bias in its components", International Journal of Forecasting, 5(1), 127-131

Swidler, S and Ketcher, D (1990): "Economic forecasts, rationality and the processing of new information over time", Journal of Money, Credit and Banking, 22(1), 65-76

Theil, H (1996): "Applied economic forecasting", Amsterdam, North Holland Publishing Co.

Zarnowitz, V (1984): "The accuracy of individual and group forecasts from business outlook surveys", Journal of Forecasting, 3, 11-23

Zarnowitz, V (1985): "Rational expectations and macroeconomic forecasts", Journal of Business & Economic Statistics, 3, 293-311