Comparing the BER's forecasts

NICOLAAS VAN DER WATH

Stellenbosch Economic Working Papers: 23/13

NOVEMBER 2013

KEYWORDS: FORECAST COMPARISON; FORECAST ACCURACY, FORECAST EVALUATION, CONSENSUS FORECAST, THEIL COEFFICIENTS, MEAN ABSOLUTE ERROR, ROOT MEAN SQUARED ERROR, LOSS FUNCTION
JEL: C53

NICOLAAS VAN DER WATH
BUREAU FOR ECONOMIC RESEARCH
10 VAN RIEBEECK STREET, 7600
STELLENBOSCH, SOUTH AFRICA
E-MAIL: WNWATH@SUN.AC.ZA

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

Comparing the BER's forecasts

NICOLAAS VAN DER WATH[1]

ABSTRACT

The Bureau for Economic Research publishes annual (and quarterly) forecasts for more than 140 macroeconomic indicators, with a forecasting horizon stretching up to 6 years ahead. These forecasts are generated with the aid of a structural macro-econometric model of the South African economy. The purpose of this research note is to test the accuracy of the BER's forecasts. Also to compare them with other published forecasts according to accuracy, forecast horizon and number of indicators. To determine the level of accuracy, we have calculated the mean absolute errors and the root mean squared errors of the BER's forecasts for a selection of five economic indicators. These statistics were also calculated for the forecasts of the selected other institutions or models. From these the relative accuracy of the different forecasts were compared to each other and ranked accordingly. The consensus forecast turned out to be the most accurate for the immediate year, followed with a narrow margin by the BER. The close proximity of these two forecasts is striking. Other conclusions are that structural forecasting models perform better than mechanical ones for the first two years, but lose their accuracy advantage from the third or fourth year onwards. They also fail to anticipate critical turning points in economic cycles.

Keywords: forecast comparison; forecast accuracy, forecast evaluation, consensus forecast, Theil coefficients, mean absolute error, root mean squared error, loss function

JEL codes: C53

---

[1] Bureau for Economic Research, University of Stellenbosch: I would like to thank Professors Ben Smit and Stan du Plessis for their valuable insight, comments and guidance. As usual I accept full responsibility for any errors.

Table of Contents

List of Tables

List of Figures

1. Introduction

Macroeconomic forecasts are a core service provided by the Bureau for Economic Research (BER). They are produced by means of two structural macro-econometric models: one quarter based and one annual based. To the best of my knowledge, the BER has the longest history of publishing a forecast of the South African economy on a regular basis. The annual model covers up to 140 macroeconomic indicators and spans a six year forecast horizon. Other publically available forecasts (such as that of National Treasury and the Sake24 consensus) are typically restricted to around ten core variables. Sake24 publishes an average of other forecasts, called the consensus forecast. The International Monetary Fund's (IMF) forecast of the South African economy is broader and covers 43 variables, now also spanning a six-year horizon.

Clearly a wide range of forecasts is available to modern decision makers. This raises the question of the relative quality of these forecasts. The National Bureau for Economic Research (NBER) in the United States developed a framework to systematically evaluate large scale econometric models. These include parametric evaluations before the release of a model such as model selection, parameter estimation and structural stability tests. They also provide tools to evaluate a model after its release such as error decomposition (Dhrymes, et al., 1972). In this class there are indicators of relative forecast quality. They may include ratios or differences of mean, mean-squared or mean-absolute forecast errors (West, 2006). For a corporate user of forecasts, the range of economic indicators and the forecast horizon will also be of practical importance.

The purpose of this research paper is to rank the BER's forecasts among that of other publishing forecasters. In other words a post release evaluation will be undertaken and according to three criteria: accuracy, forecast horizon and number of indicators covered. Most of the analysis focuses on the accuracy criteria, since the other two are trivial. The research paper is divided into four sections. The first is a brief discussion of some data issues, followed by an overview of methods to compare accuracy. The third is the actual accuracy analysis of the forecasts with respect to each of the variables. The conclusions are summarised in the final section.

2. *Data*

To measure how accurate a forecast is, it needs to be compared with the actual outcome. Therefore the relevant historical data are needed. A well-documented problem of historical data is that parts of it get revised from time to time, especially data from the national accounts. The Federal Reserve Bank of Philadelphia lists some reasons for such revisions in a working paper (Stark, 1999). These reasons include a change in the base year of a price deflator, or a change in definition or weights in a basket. Sometimes data will be revised many years down the line. Therefore it becomes impractical to compare a forecast with the latest version of data, since we don't know if it will be revised again. As such we rather compare with the very first value which was published for a specific variable. The quarter in which the data is published is called its vintage. For example, by March 2012 the first estimate of the 2011 GDP was published. We label this vintage as 2012-Q1. In this paper we note the historical data and its related forecasts as follows:

Y_{t+i} refers to the 1st value published for an economic indicator for period $t+i$.

$p_{t,i}$ refers to the forecast of an economic indicator Y, done at the period t, for the i-th period into the future. (In the case of a six-year forecast horizon, i will run from 0 to 5.)

To be able to compare the forecasts of different forecasters with each other, the same economic indicators should be used. Whilst the BER's forecast covers more than 140 economic variables, other publicly available forecasts have a much smaller scope. They are forecasts by the National Treasury, the IMF and the Sake24 consensus[2]. The indicators which were common to most of these models are GDP growth, inflation, the prime interest rate, the exchange rate and the gold price. These are the five indicators which will be analysed in this paper.

The BER's annual-model forecasts are for six years ahead and are generated bi-annually at the end of quarter one and quarter three. Records of these forecasts are available from 1982. From them the historical vintages and forecasts of the five selected variables were extracted, ordered in a pivot table and processed for comparison. Data from the other forecasters were processed in the same way such that the BER's performance could be compared to them.

We obtained the forecasts of National Treasury from the annual Budget Review published every year in February, and the Medium Term Expenditure Framework (MTEF) published in October. They forecast three years ahead in February and years ahead in October. The IMF update their own

[2] An average of around 20 forecasters who participate in Sake24's 'Economist of the Year' forecasting competition.

forecasts bi-annually in April and October, and their first vintage is for 2003-Q1. Initially they forecasted only two years ahead, but since 2008-Q1 they started to make a six-year forecast. The Sake24 consensus forecast is an average of around 20 external forecasts – making it the collective wisdom of a group of experts in their field. Research has shown (Clemen, 1989) that such consensus forecasts normally turn out to be the most accurate. The Sake24 consensus forecasts are for one year ahead on most of its indicators, and two years for GDP and inflation. They publish a new forecast every month which makes their frequency much higher than the other models above, but only their April and October forecasts were considered in this analysis.

3. Measuring Forecast Accuracy

The first step in the evaluation process is to identify a method that can measure forecast accuracy. The method must also provide for the comparison of forecasts with each other. The Austrian Institute of Economic Research (WIFO) groups the measures of forecast accuracy into three categories. They are statistical accuracy tests, Theil coefficients and tests for the correct sign (Krainz, 2011).

3.1. Statistical Accuracy

The mean absolute error (MAE) is an example of a statistical test. In the same group and commonly used is the Root Mean Squared Error (RMSE). It is very similar to the MAE, but puts more emphasis on large forecast errors and cannot be interpreted directly and as intuitively. Another statistical measurement is the simple error average. In this case positive and negative errors could cancel each other, making it less useful as a measure of comparison between different forecasts. However, it can serve as an indication if a specific model has a bias to over or under estimate.

Intuitively the accuracy of a forecast will increase as the forecast moves closer to the actual outcome of its indicator. This error can be expressed by different types of formulas, called loss functions. The absolute error, an example of a simple loss function, is the absolute distance between the forecast and the realised value. This measure accords equal weight to an under- and an overestimate. A perfect forecast will have an error of 0. When considering a whole series of forecasts for one variable, the average of these absolute errors can be calculated as the MAE. In comparing the forecasts of different models, the one with the lowest MAE will be considered the most accurate. The loss function of the MAE can be expressed as follows:

$$L_{t,i}(Y_{t+i}, p_{t,i}) = |p_{t,i} - Y_{t+i}| \qquad \text{if Y is a percentage-type indicator (e.g. inflation)}$$

$$L_{t,i}(Y_{t+i}, p_{t,i}) = \left|\frac{p_{t,i}}{Y_{t+i}} - 1\right| \qquad \text{if Y is a level-type indicator (e.g. gold price)}$$

Y: actual outcome of an economic indicator

p: forecast of the economic indicator

t: vintage of the data

i: future period on the forecast horizon for which the forecast is done.

For each of the *m* different periods on the forecast horizon a MAE can be calculated (each with its own *n* vintages):

$$MAE_i = \frac{1}{n_i}\sum_{t=1}^{n_i} L_{t,i} \qquad i=1,\dots,m$$

Similarly for the RMSE:

$$L_{t,i}(Y_{t+i}, p_{t,i}) = (p_{t,i} - Y_{t+i})^2$$ if *Y* is a percentage-type indicator (e.g. inflation)

$$L_{t,i}(Y_{t+i}, p_{t,i}) = \left(\frac{p_{t,i}}{Y_{t+i}} - 1\right)^2$$ if *Y* is a level-type indicator (e.g. gold price)

$$RMSE_i = \sqrt{\frac{1}{n_i}\sum_{t=1}^{n_i} L_{t,i}} \qquad i=1,\dots,m$$

The statistical tests above are all examples of symmetric loss functions. Asymmetric loss functions have also been developed (Legerstee, Paap, & Franses, 2011), such as the lin-lin and quad-quad loss functions (counterparts to the MAE and RMSE respectively). These loss functions provide for an asymmetric parameter as well as a systematic bias which can be estimated using least squares techniques. This type of in-depth analysis is beyond the scope of this paper.

To shed some light on the biasedness of a particular forecast, the mean error (ME) and its related variance can be calculated:

$$L_{t,i}(Y_{t+i}, p_{t,i}) = (p_{t,i} - Y_{t+i})$$ if *Y* is a percentage-type indicator (e.g. inflation)

$$L_{t,i}(Y_{t+i}, p_{t,i}) = \left(\frac{p_{t,i}}{Y_{t+i}} - 1\right)$$ if *Y* is a level-type indicator (e.g. gold price)

$$ME_i = \frac{1}{n_i}\sum_{t=1}^{n_i} L_{t,i} \qquad i=1,\dots,m$$

$$Var_i(L_{t,i}) = \frac{1}{n_i}\sum_{t=1}^{n_i}(L_{t,i} - ME_i)^2 \qquad i=1,\dots,m$$

A *ME* > 0 indicates a positive bias (a tendency to overestimate). Similarly will a *ME* < 0 indicate a negative bias. The variance of the simple errors is not only a measure of the forecast accuracy itself, but also of the volatility. A high variance indicates that the forecast errors will differ a lot from vintage to vintage. Such a forecast carries more risk with it.

Note that when a forecast is not biased (*ME* = 0), the variance will be the same as the mean squared error ($RMSE^2$). These two indicators thus carry similar information and it would not be of significant benefit to calculate both. The concepts of accuracy, variance and volatility are converging in the case of MAE, RMSE and Variance.

3.2. Theil Statistics

The second category of forecast performance measurement is Theil statistics. They are standardised around one, which makes them easy to interpret (Krainz, 2011). They compare a specific model's forecast to that of a benchmark forecast (such as a naïve forecast, but other forecast types can also be used). A naïve forecast assumes that the last recorded value will continue into the future:

$$p_{t,i} = Y_{t-1} \text{ for } i=0,\ldots m$$

For example, if the economic growth rate in 2011 were 3.1%, then a naïve forecast will predict a 3.1% growth rate for 2012 and beyond. A Theil statistic will then express the MAE of different forecasts as a ratio to the naïve's MAE:

$$T_r = \frac{MAE_r}{MAE_{naïve}}, \text{ where } r \text{ is the forecaster being compared}$$

If the MAE of the forecast is the same as that of the naïve, then $T = 1$, indicating these two models have the same level of accuracy. In other words, we can calculate a standardised score for every forecast. The MAE ratio presented in the tables of this paper is an example of such a Theil statistic.

3.3. Tests for the Correct Sign

Other methods of comparison have also been developed, such as tests for the correct sign. Some researchers argue that the forecast error is not of such a great importance. They prefer the ability to anticipate turning points in economic cycles, for example the ability to predict a recession. This is done through a directional analysis – counting the percentage of times that a forecast correctly predicts an increase or decrease in an economic indicator. Directional accuracy is important for variables that do not trend over time, such as the interest-, inflation- and economic growth rates. Level-type indicators like the national aggregates are nearly always trending upwards, thus showing positive changes. The difficulty is to predict the few times when they will have negative changes (Van Walbeeck & Sessions, 2007, p. 120).

This paper will not do a directional analysis on the BER's forecasts since Van Walbeeck and Sessions already published one in 2007. They did this by testing for dependence between the forecast and actual values. They found the BER to outperform a naïve forecast in the case of interest rates, but not for economic growth.

4. The Comparison

Now that we have identified the data and methods, the actual analysis can be done. The historical data was ordered into pivot tables and arranged in a usable manner in Excel. The naïve and trend forecasts were also done in Excel. These data files are available from the BER on request. Each variable's comparison follows a similar pattern.

The BER's forecast is compared with each of the other forecasts, in turn, in the tables below. For each of the six years forecasted ahead (t+0 to t+5) the mean absolute error is presented, followed by the overall MAE and root mean squared error (RMSE). In order to allow for a normalised measure between different forecast comparisons, a Theil-type statistic is calculated, namely the MAE ratio. In each case the MAE of the BER is expressed as a ratio of the forecast to which it is compared at that moment (benchmark). First the naïve, followed by the IMF, Treasury and lastly Sake24. A ratio smaller than one is indicative of a forecast more accurate than the benchmark, and larger than one of the opposite. For example: a MAE ratio of 0.95 would imply that the BER's forecasts was on average 5% closer to the target than the benchmark forecast.

Note that the different forecasters all started their forecasting in different years and have different horizons. For example: the BER's economic growth forecast is rated from 1999-Q1 to 2012-Q1, and there are six years in the forecast horizon, bringing the total number of forecasts rated to 126. The IMF started in 2003 over a two-year horizon, totalling 40 forecasts to be rated. To compare on a similar level, only the equivalent matching 40 observations from the BER will be used for comparison with the IMF in this particular case. We applied this principle also when comparing to Treasury and the Sake24 consensus. It explains why the MAE of the BER is different between the models, even though it is for similar leads (*t+i's*) on the forecasting horizon.

4.1. The GDP Growth Rate

We start with economic growth. Note that in this case the forecast error is defined as the difference between the forecasted and actual growth rates in percentage points (%pts). For example in Table 1, in forecasting the economic growth rate of the nearest year (*t*+0), the BER was on average 0.48%pts off target compared to 1.59%pts for a naïve forecast. The IMF, Treasury and Sake24 all came close to 0.60%pts, which is only marginally worse than the BER forecast. It seems that at *t*+2 onwards the BER loses its advantage as Treasury, the IMF and the trend performed better. On average the BER's MAE ratio is 0.69 times the value of the naïve forecast, indicating that it came 31% closer to the target than the naïve forecast. Similarly it came 13% closer than the IMF and 4.3% closer than Sake24. Treasury improves on the BER's forecast for economic growth by a small margin (2%).

Note the big difference in accuracy between the BER forecast of the nearest year (*t*+0) and the next year (*t*+1), jumping from 0.48%pts to 1.27%pts. From this one can conclude that most of the BER's forecast value lies in the nearest year, some of it in the next, but from the third onwards it follows mostly mechanical patterns.

Table 1: Evaluation of the margin of error of GDP growth forecasts (percentage points)

MAE	*t*+0	*t*+1	*t*+2	*t*+3	*t*+4	*t*+5	MAE	RMSE	MAE Ratio	n
Naive	1.59	1.96	2.17	2.34	2.38	2.28	2.08	2.64		126
Trend	1.80	1.75	1.79	1.83	1.89	2.11	1.85	2.38	0.89	126
BER	0.48	1.27	1.87	2.34	2.34	2.00	1.43	2.09	0.69	126
IMF	0.65	1.58	1.17	1.65			1.12	1.56		40
BER	0.49	1.37	1.34	1.40			0.98	1.44	0.87	40
Treasury	0.61	1.20	1.58	1.71			1.18	1.87		82
BER	0.50	1.19	1.69	1.98			1.20	1.85	1.02	82
Sake24	0.56	1.26					0.89	1.41		42
BER	0.48	1.27					0.85	1.33	0.96	42

Figures 1 and 2 below show the distance (errors) by which different forecasts missed their targets since 1999. The sharp spike in the BER's forecast error seen in Figure 1 is due to the financial crises. Even up to the end of 2008 growth rates were not anticipated to decrease by so much during 2009. The forecast for *t*+1 represents a huge overestimation, indicating the financial crisis was not foreseen by more than one year. Looking at the nearest year forecast (*t*+0), the pattern is different: the BER overestimated 2009 growth by only 1%pt (and Treasury by 3%pts).

Figure 1: BER forecast error in %pts

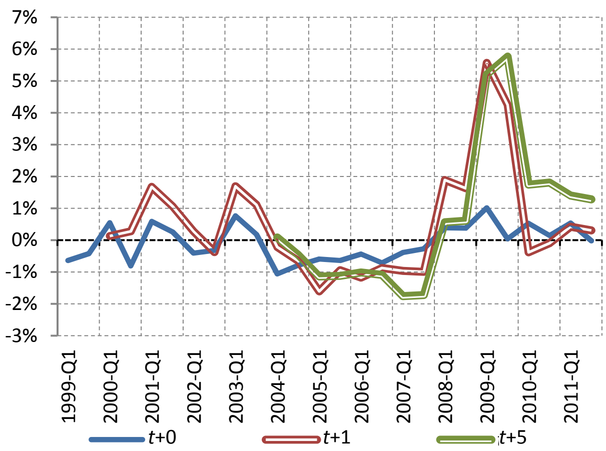
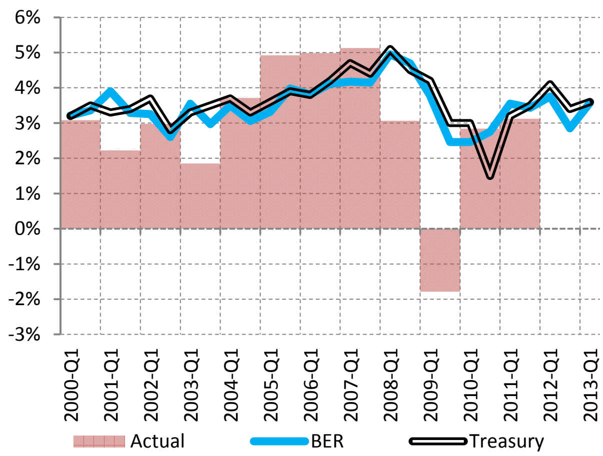


Figure 2: Growth forecast (*t*+1 estimates)



Interestingly, the average error for the BER (over its entire forecast horizon) is 0.5%pts, indicating a bias to overestimate economic growth by 0.5%pts. However, if the 2009 recession is excluded the average error drops to 0%pts, indicating no bias. To summarise, when it comes to economic growth, we can conclude that all the forecasters' models come remarkably close to each other in terms of accuracy, and that in the first two years all of them perform significantly better than a purely trended or naïve forecast. For periods longer than that they lose their competitive advantage (in terms of pure accuracy).

4.2. The Consumer Inflation Rate

In the case of inflation, the comparison of the BER with the trend, naïve and IMF forecasts only considers the headline inflation. For Treasury and Sake24 a switch to CPIX had to be incorporated from 2003 to 2008. For those years in which CPIX were forecasted, equivalents were compared even though only one result is presented. For inflation the forecast error is defined as the difference between the forecasted and actual inflation rates in percentage points (%pts). From Table 2 the BER's MAE ratio is 0.81 compared to the IMF (a 19% improvement) and 0.64 compared to the naïve forecast (a 36% improvement). For both Treasury and Sake24 the MAE was very close to that of the BER. Strictly speaking the Sake24 consensus was the best forecast (by a non-significant 0.004 margin in the MAE ratio).

Table 2: Evaluation of the margin of error of inflation forecasts (percentage points)

MAE	***t*+0**	***t*+1**	***t*+2**	***t*+3**	***t*+4**	***t*+5**	**MAE**	**RMSE**	**MAE Ratio**	***n***
Naive	2.57	3.44	3.97	3.21	2.50	2.19	3.02	3.80		126
Trend	3.16	2.97	2.70	2.60	2.36	2.57	2.76	3.54	0.91	126
BER	0.67	1.80	2.36	2.75	2.32	2.18	1.94	2.76	0.64	126
IMF	0.97	2.09	0.56	0.47			1.35	1.91		38
BER	0.56	1.91	0.59	0.51			1.09	1.77	0.81	38
Treasury	0.57	1.82	2.13	2.64			1.55	2.48		69
BER	0.50	1.66	1.97	2.64			1.43	2.30	0.93	69
Sake24	0.74	1.88					1.25	1.99		39
BER	0.66	2.03					1.26	2.00	1.00	39

Figure 3 below shows the errors of the BER for its current, next and sixth year forecasts. Note the spikes that occur regularly, indicating that the forecasts fail to anticipate directional changes. In Figure 4 the forecasts of the BER and Treasury for *t*+1 seem to follow the same general pattern, both missed the spikes of inflation in 2002 and 2008.

Figure 3: BER forecast error in %pts

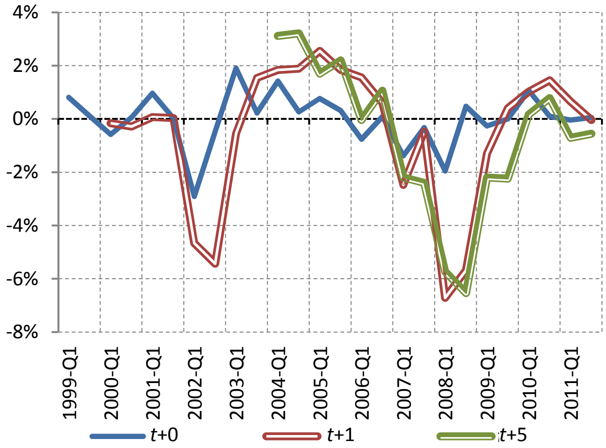
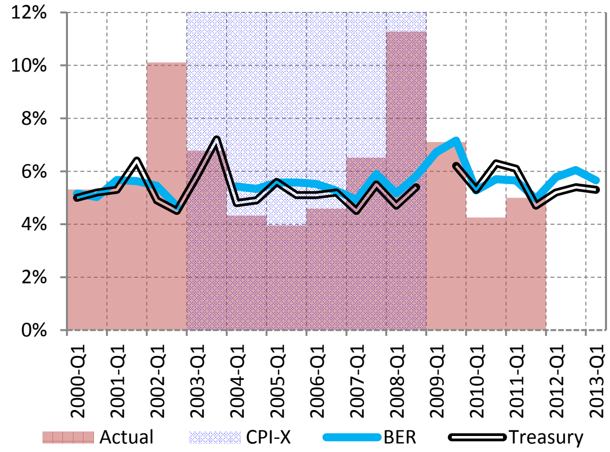


Figure 4: Inflation forecast (*t*+1 estimates)



4.3. The Prime Interest Rate

It is only the BER and the Sake24 consensus that publish forecasts for the prime interest rate. In this case the forecast error is defined as the difference between the forecasted and actual interest rates in percentage points (%pts). The BER's forecast seems to perform very well for t+0, scoring a low 0.23%pt mean absolute error. (However, it should be kept in mind that this is a forecast for the annual average of the prime interest rate. More of the interest rate becomes known as the year progresses, especially by the time of the October forecast.) The BER retains this advantage up to *t*+3, after which the naïve forecast becomes a better estimate of future interest rates (perhaps indicating a four-year cycle in interest rates). The naïve forecast also performs better than the trend. Sake24 only has a forecast for the last quarter of the nearest year (*t*+0).

Table 3: Evaluation of the prime interest rate forecasts (percentage points)

MAE	*t*+0	*t*+1	*t*+2	*t*+3	*t*+4	*t*+5	MAE	RMSE	MAE Ratio	*n*
Naive	1.75	3.10	3.63	2.82	1.25	1.67	2.46	2.93		78
Trend	3.02	3.82	4.71	3.71	3.99	4.68	3.89	4.58	1.58	78
BER	0.23	1.52	2.48	2.33	2.21	1.72	1.63	2.13	0.66	78
Sake24*	0.47						0.47	0.67		16
BER*	0.51						0.51	0.78	1.09	16

** Forecast for the 4th quarter average only*

Note how the accuracy of the *t*+1 forecast is but a shadow of that of the *t*+0 forecast in Figure 5. This is typical of a model with a strong trended component and a tendency to miss changes in the cycle. In general the BER would miss the interest rate outcome for the nearest (*t*+0) fourth quarter by 0.51%pts, as opposed to 0.47%pts in the case of Sake24. Thus the BER has been 9% less accurate in relative terms (MAE ratio of 1.09). In Figure 6 the closeness of the two forecasts is striking.

Figure 5: Interest rate forecast (BER)

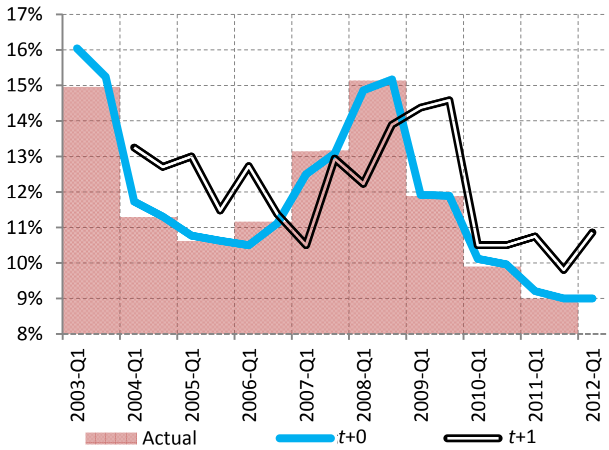
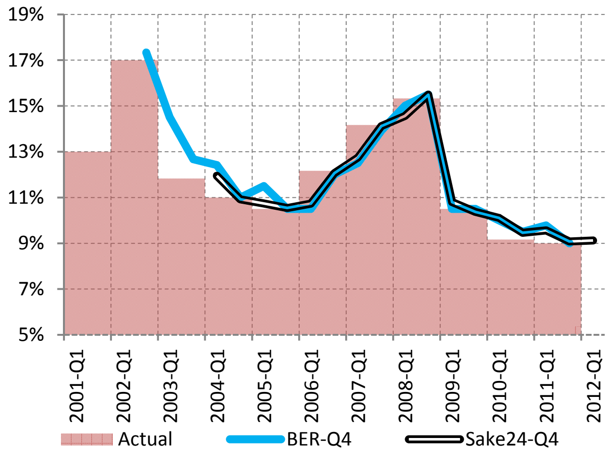


Figure 6: Interest rate forecast (*t*+0 estimates)



4.4. The Exchange Rate

The exchange rate is perhaps the most difficult variable to forecast. In this case (and for the gold price in Section 4.5) the error is defined as the percentage that the forecast deviates from the actual exchange rate level. Once again the BER outperforms the naïve and trended forecast for the first two years, something which is not expected given the research results of Mees and Rogoff (Meese & Rogoff, 1983) on exchange rate models (although they looked at monthly models). However, the BER loses its advantage at *t*+2 as the naïve's accuracy bypasses it. Over the total forecast horizon the BER had a MAE that was 60% larger compared to the naïve, indicating less accuracy.

Table 4: Evaluation of the R/$ exchange rate forecasts (percentage)

MAE	*t*+0	*t*+1	*t*+2	*t*+3	*t*+4	*t*+5	MAE	RMSE	MAE Ratio	n
Naive	12.2	21.3	23.9	22.8	18.7	13.3	18.7	23.8		126
Trend	20.2	33.8	43.8	52.6	64.0	70.8	44.7	56.4	2.39	126
BER	4.0	16.7	27.6	30.0	31.1	36.5	30.0	38.0	1.60	126
IMF	5.8	13.9	27.6	16.0			11.8	16.7		38
BER	3.9	12.6	26.7	33.0			11.3	16.1	0.95	38
Sake24*	11.4						11.4	14.9		19
BER*	12.0						12.0	15.2	1.06	19

** Forecast for the 4th quarter average only*

The Sake24 consensus forecast is only for the last quarter of each year, necessitating it to be compared with the BER's quarterly model. The forecast accuracy for Q4 is much lower than for the full year since quarterly data is more volatile than annual data. In general the consensus missed the R/$ exchange rate by 11.4%, slightly less than the BER's 12%. In Figure 7 the *t*+5 forecast error lies above the 0-line all the way, indicating a BER-bias to forecast a depreciating rand in the long run.

Figure 7: BER forecast error in percentage

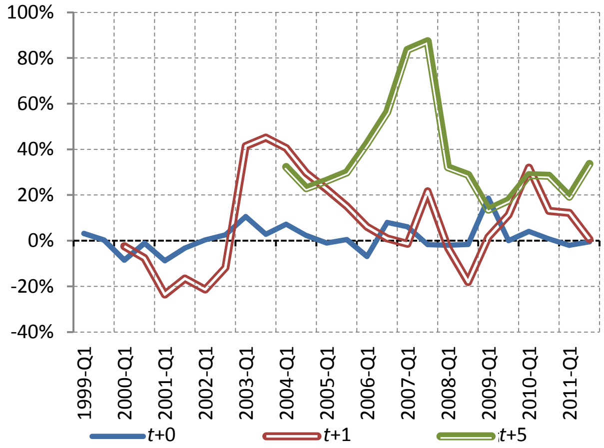
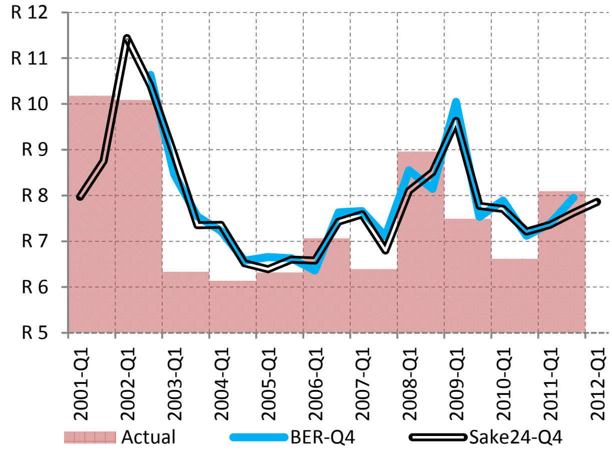


Figure 8: R/$ forecast (*t*+0 estimates)



4.5. The Gold Price

The gold price exhibits the same level of volatility as the exchange rate, since they are both asset prices. Their relative high MAEs reflect the challenge to forecast these variables. The MAE of the BER (to forecast the gold price) was 23% higher than that of the Sake24 consensus, an indication of better accuracy by the consensus. However, in this case the BER's long term forecast remains better than that of the trend and naïve, right up to six years ahead (*t*+5). The downside is that at *t*+5 even the BER missed the actual gold price by more than 50%.

Table 5: Evaluation of the gold price forecasts (percentage)

MAE	*t*+0	*t*+1	*t*+2	*t*+3	*t*+4	*t*+5	MAE	RMSE	MAE Ratio	*n*
Naive	12.9	23.4	32.8	40.9	48.9	55.7	33.4	38.6		126
Trend	11.4	19.6	29.7	41.7	53.4	63.5	33.6	39.1	1.01	126
BER	4.1	17.7	28.6	36.3	44.3	51.2	27.8	34.2	0.83	126
Sake24*	8.7						8.7	10.8		19
BER*	10.7						10.7	13.9	1.23	19

** Forecast for the 4th quarter average only*

Figure 9 shows that the BER has a bias to underestimate the gold price slightly for the nearest year (3.1%), but much more so for the further years (at times more than 60% for *t*+5). From Figure 10 the close proximity between the BER and consensus forecasts is clear and it has been the case for the interest and exchange rates. With this in mind, it may be that the participants keep the consensus in mind when making their own forecasts. Both the BER and consensus forecasts lie mostly on the lower side of the actual: exhibiting the same bias.

Figure 9: BER forecast error in percentage

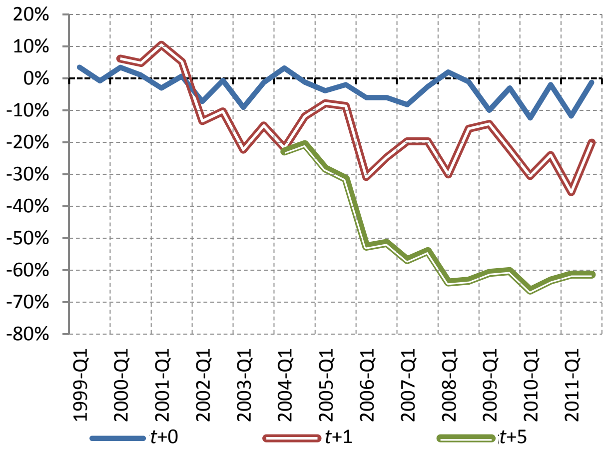
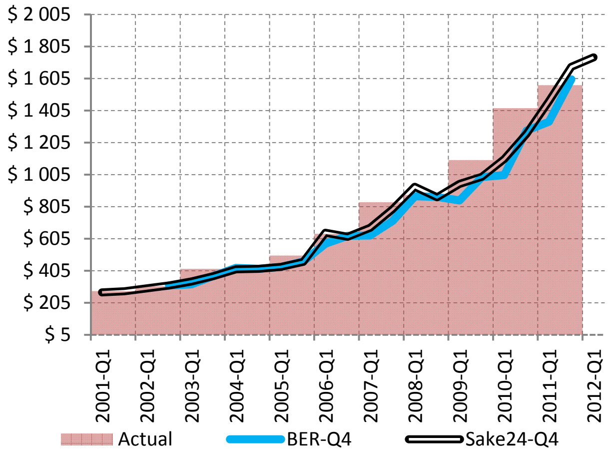


Figure 10: Gold price forecast (*t*+0 estimates)



5. *Conclusion*

The BER's forecasting performance has been compared with those of other models/institutions: mechanical, structural, and consensus. Five indicators were evaluated: economic growth, inflation, the prime interest rate, exchange rate and the gold price. The table below combines the results according to all three evaluating criteria: accuracy, forecast horizon and number of variables:

Table 6: Summary of best forecasters for each year of the forecast horizon

MAE	*t*+0	*t*+1	*t*+2	*t*+3	*t*+4	*t*+5
GDP Growth	BER	Sake24	IMF	Treasury	Trend	BER
Inflation	BER	Sake24	IMF	Trend	BER	BER
Prime interest rate	Sake24	BER	BER	BER	Naïve	Naïve
Exchange Rate	Sake24	BER	Naïve	IMF	Naïve	Naïve
Gold Price	Sake24	BER	BER	BER	BER	BER

For the first two years on the forecast horizon the Sake24 consensus and BER turned out to be the best forecasters, with Sake24 slightly ahead. This is not unexpected as researchers (Clemen, 1989) have shown that consensus estimates are more accurate. However, one drawback is that the Sake24 consensus forecast is mostly for a few selected indicators and at most for two years ahead, whereas the BER and others forecast for three to six years ahead. Out of the 25 windows that we evaluated, the BER was the best in 14 of them. Second was Sake24 with five, followed by the naïve with five, the IMF with three, the trend with two and Treasury with one. However, in many of the cases the margins of winning were very close. The BER gains it advantage from all three dimensions: it is comparatively accurate, has the longest forecast horizon and covers the most variables.

The BER was the best economic growth forecaster for the current year (*t*+0), but over the total four-year horizon National Treasury came the closest to the target. For inflation, the BER also performed the best in the current year, yet the consensus did better in the next (*t*+1). In the case of the prime interest rate, exchange rate and gold price the BER lost its first position in the nearest year to Sake24. Neither the IMF nor Treasury make forecasts for the interest rate or gold price. Another interesting result is the great similarity between the BER and Sake24 consensus forecasts.

In combination, all the forecasts evaluated above suffer from the same inability to anticipate directional changes. None could foresee the Great Recession by more than one year in advance, and none anticipated any of the spikes in inflation or interest rates. The surge in the gold price was not foreseen by the BER or any of the other forecasts. Further analysis shows that for most of the five indicators presented, by the fifth and sixth year a naïve forecast is more accurate compared to a trend: confirming the notion that no trend continues for too long.

6. Bibliography

Clemen, R. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 559-583.

Dhrymes, P., Howrey, E., Hymans, S., Kmenta, J., Leamer, E., Quanot, R., et al. (1972). Criteria For Evaluation Of Econometric Models. In *Annals of Economic and Social Measurement, Volume 1, number 3* (pp. 291-325). NBER.

Krainz, D. M. (2011). An Evaluation of the Forecasting Performance of Three Econometric Models for the Eurozone and the USA. *WIFO Working Papers*.

Legerstee, R., Paap, R., & Franses, P. H. (2011). *Estimating Loss Functions of Experts.* Amsterdam: Tinbergen Institute.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economics*, 14, 3-24.

Stark, D. C. (1999). *A Real time dataset for Macroeconomists.* Philadelphia: Federal Reserve Bank of Philadelphia.

Van Walbeeck, M., & Sessions, C. (2007). A directional analysis of the Bureau for Economic Research's quarterly forecasts. *Journal for Studies in Economics and Econometrics*, 119-137.

West, K. D. (2006). Forecast Evaluation. In *Handbook of Economic Forecasting* (pp. 100-134). University of Wisconsin: Elsevier.

Data

Data were obtained from the following sources:

Bureau for Economic Research's Economic Outlook, various years

South African Reserve Bank Quarterly Bulletin, various years.

Provincial Treasuries Budget Statement documents, various years.

International monetary Fund's World Economic Outlook, various years

Sake24 consensus forecast from the BER's website