What Can We Learn from Macroeconomic Forecasting?

Empirical Evidence from German GDP Forecasting

Torsten Jochem

January 2009

Abstract

The paper analyzes the accuracy of over 300 GDP growth forecasts for Germany between 1998 and 2008 of the leading 8 German forecasting institutes. It shows that forecast correctness diminishes significantly as the forecast horizon increases, showing a significant jump when growth is predicted more than 12 months in the future. The study further compares the relative performance of forecasting institutes and compares their absolute performance with four simple benchmark models. Even though all benchmark models use last year's growth as the only economic indicator, all of them perform considerably better than the forecasts by the institutes. Among the benchmark models, a random walk with a constant drift performs best.

Section 1: Introduction

Every few months German media broadcasts the latest forecasts by the leading German economic institutes on the future GDP growth rate. The institutes employ macroeconomic models to not only forecast future GDP growth, but also employment, inflation, consumption, investments, imports and exports and in some cases foreign exchange and oil prices. To be sure, some of these indicators are in fact required to predict future GDP growth.

It should therefore come to no surprise that GDP forecasting models which are based on fundamentals require a number of assumptions and are usually subject to iterative revisions as the economic environment changes. Naturally, for instance, no one could have predicted in early 2001 that the terrorist attack on the World Trade Center later the same year would require the NYSE to shut down for several days, that oil prices would skyrocket and a short but strong recession would follow. Furthermore, it should be clear that if forecast models were right on spot in their modeling of oil prices and exchanges rates, then pots of money could be made in the short run by bidding accordingly on future oil and foreign exchange prices or stock markets indices.

In this paper, I study how accurate the German GDP growth forecasts actually are. To this end, I have collected over 400 German GDP growth forecasts with forecast horizons between 0 and 24 months of the leading major German forecast institutes between 1998 and 2008.[1] Table 1 lists the institutions whose forecasts were investigated as well as the time frames covered.

Institute/Organization	Abbreviation	Time Frame	# of forecasts
Deutsches Institut fuer Wirtschaftsforschung Berlin (=*German Institute for Economic Research Berlin*)	DIW	1998-2008	47
Hamburgisches Weltwirtschaft Institut (=*World economy Institute in Hamburg*)	HWWI (before 2005: HWWA)	1998-2008	43
Institut fuer Weltwirtschaft Kiel (=*Institute for the World Economy in Kiel*)	IfW	1998-2008	46

[1] Example: a forecast published in January 2007 on the expected GDP growth for 2007 has a forecast horizon of 11 months, since the calendar year 2007 will end 11 months later in December 2007. Likewise, publishing the 2007 expected GDP growth in December 2007 has a forecast horizon of 0, which makes sense as the official GDP statistics on the 4th quarter of 2007 is only released several months later (and further revisions may still occur later into the year). Nonetheless, it should be clear that a forecast with a lead of less than 12 months should be getting closer to the actual result since some part of the year has already passed and the approximate result of those months has been observed.

Rheinisch-Westfaelisches Institut fuer Wirtschaftsforschung Essen (=*Rhineland-Westphalian Institute for Economic Research in Essen)*	RWI	1999-2008	63
ifo Muenchen (=*ifo Institute for Economic Research in Munich*)	ifo	1998-2008	42
Sachverstaendigenrat[2] (=*German Council of Economic Experts*)	SVR	1998-2008	21
Gemeinschaftsdiagnose[3] (=*Joint Economic Forecast Group*)	JEFG	2000-2008	36
Institut fuer Wirtschaftsforschung Halle (=*Institute for Economic Research in Halle*)	IWH	1998-2008	75

Table 1: Leading Forecasting German Institutes

The paper is organized as follows: Section 2 compares past forecasts with actual growth rates. It also discusses how the forecast accuracy diminishes as the forecast horizon increases. Section 3 compares the accuracy of the forecasts among institutions and tries to determine which institute has gotten over the past 10 years the closest to the actual growth rate. Section 4 introduces several benchmark models, which are all based on statistical properties rather than economic fundamentals: one standard benchmark model is the "random walk model", which predicts next year's forecast as simply last year's forecast plus some random shock from a standard normal distribution. I use these benchmark models to create growth predictations and compare how well these "economically unsophisticated" models compare to those by economic institutes that take real economic indicators into account.

[2] A group of economic experts advising and funded by the federal government.

[3] The Joint Economic Forecast Group (*Gemeinschaftsdiagnose*) is a joint effort by several of the leading German forecasting institutes, funded by the German Federal Economic Ministry. Access to the group is restricted and based on a public competitive tender. Previous to October 2006, it was comprised by ifo, IfW, RWI, IWH, DIW and HWWI. HWWI dropped out of the group in October 2006 so did DIW in October 2007.

Section 2: A Descriptive Look at the Data

Chart 1 and 2 show 12 month and 18-24 month forecasts versus the actual GDP growth rates (black line). As can be readily seen in Chart 1, forecasts seem to have an attenuation effect: during periods of high growth rates they underestimate growth while growth is overestimated during periods of low growth. This is unfortunate from a policy-maker's viewpoint who tries to enact counter-cyclical economic policies. If this pattern of attentuation held true for 2009 as well, the dim forecasts for 2009 may still overstate the current economic downturn.

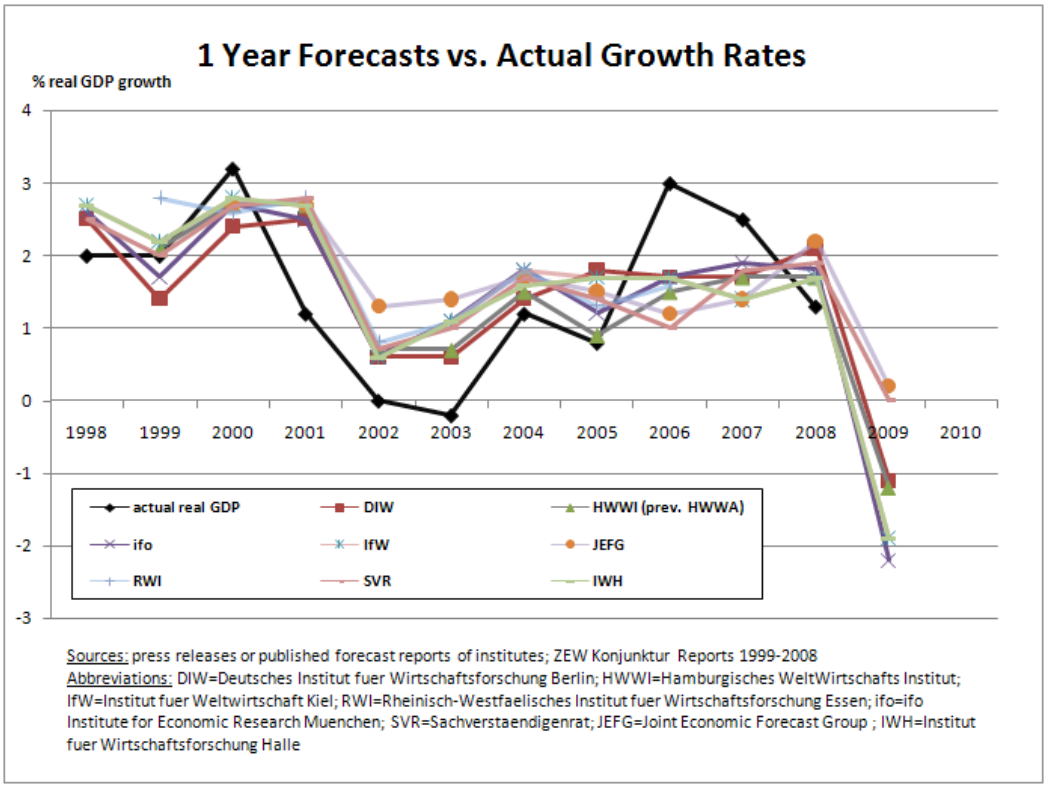

Chart 1: 1 Year Forecasts vs. Actual Growth Rates

Chart 2 shows that forecasts with a time horizon of more than 1 year are no close esimates of the actual growth rate. While this is not meant to discredit macroeconomic forecasting models, it shows that one should be wary of basing economic and policy decisions on any forecasts that exceed 12 months. The uncertainty from the long time horizon seems simply too great to draw any strong enough conclusions on future growth rates.

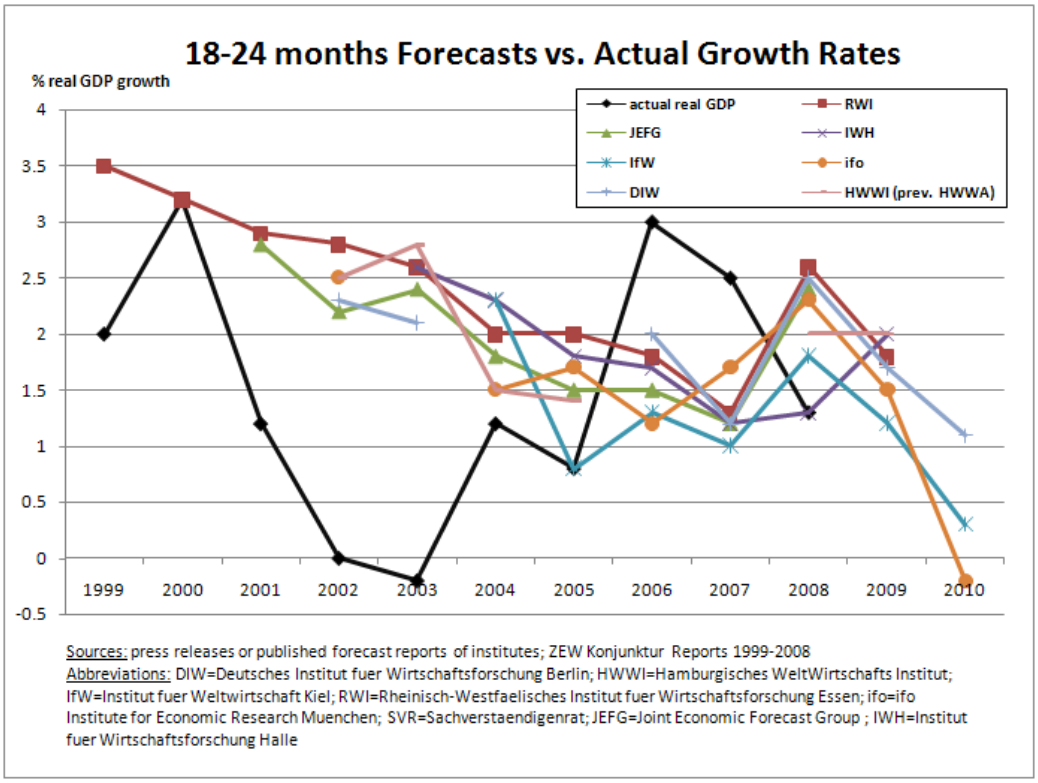

Chart 2: 18-24 months Forecasts vs. Actual Growth Rates

Interestingly, both graphs show that there seems to be an overarching consensus among economic forecasters on future growth rates. The visual impression from chart 1 and 2 on the consensus is verified in table 2, which shows the correlation of forecasts between institutions.

12 months forecast pairwise correlation

	rwi	jefg	iwh	ifw	ifo	diw	hwwi	svg
rwi	1.0000							
jefg	0.8909	1.0000						
iwh	0.9248	0.8955	1.0000					
ifw	0.9279	0.8930	0.9990	1.0000				
ifo	0.8590	0.8648	0.9762	0.9782	1.0000			
diw	0.6968	0.8763	0.9510	0.9483	0.9506	1.0000		
hwwi	0.9471	0.9208	0.9600	0.9587	0.9601	0.9282	1.0000	
svg	0.8978	0.9513	0.8970	0.8966	0.8804	0.8945	0.9397	1.0000

18-24 months forecast pairwise correlation

	rwi	jefg	iwh	ifw	ifo	diw	hwwi
rwi	1.0000						
jefg	0.9463	1.0000					
iwh	0.4091	0.4135	1.0000				
ifw	0.4959	0.6424	0.3978	1.0000			
ifo	0.7879	0.7564	-0.5421	0.6212	1.0000		
diw	0.9006	0.8574	0.2087	0.9115	0.7488	1.0000	
hwwi	0.7236	0.8321	0.4095	0.0749	0.7445	0.0865	1.0000

Table 2: Pairwise Correlation between forecasts by time horizon

Table 2 shows that among the 12 month forecasts (upper part) the correlation between forecasts is 85% or higher. The consensus however diminishes when examining the pairwise correlation for forecasts with horizons between 18 and 24 months.

This consensus of economic experts in particular for the 12 month forecasts may generate greater credibility among the media and the society, but it should be taken with a grain of salt: little diversity in forecasts can be troublesome since policy-makers may miss the point of time when counter-cyclical economic policies are most fruitful. One could conjecture three reasons for this consensus: (1) The institutes use very similar forecasting models. (2) The institutes do not work in vacuums but instead observe the same official statistics and are likely to read similar expert viewpoints on key economic indicators (future oil prices, exchange rates, etc.). Furthermore, institutes have talks with one and another as for instance through their joint work in the JEFG. (3) No institute wants to be an outlier in its forecasts since it fears for its credibility. If however all institutes are equally wrong, this will not hurt the credibility and standing of any single institute.

Section 3: Average Forecast Error and Institutes' Performances

Chart 3 illustrates the size of the average forecast error for different forecast horizons. It is computed by running for each forecast horizon a separate OLS regression with the absolute forecast error on an intercept term only. We thereby get the equally-weighted mean forecast error by forecast horizon with a 95% confidence interval.

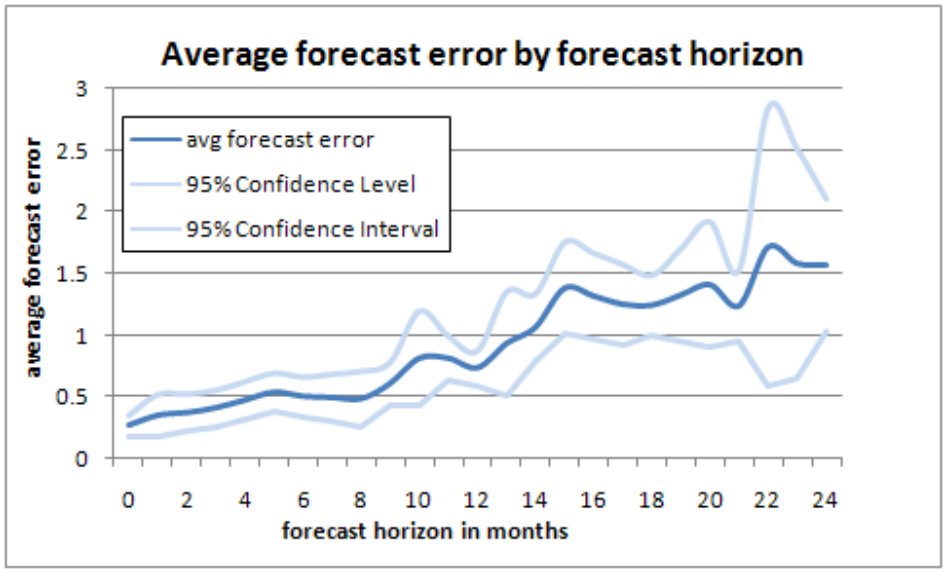

Chart 3: Average forecast error by forecast horizon

Two features stand out in this chart:

First, there is an upward jump in the forecast error between a forecast horizon of 12 and 15 months from 0.7 to 1.4 percent. Therafter the average forecast error remains on a plateau between the horizons 17 and 22 months.

Second, the longer the forecast horizon the closer we get to the 1.6% forecast error line. This is interesting from a practical viewpoint: Growth rates in Germany between 1998 and 2008 ranged between -0.2% and 3.2%. A naïve approach to estimate GDP could therefore be to simply estimate 1.6% for each year since this is the middle value between -0.2% and 3.2%. Obviously such an estimate would be off in average by less than 1.6%. It is disturbing that forecasts horizons greater than 14 months are so close to this value.

As expected, chart 1 and 2 show that the accuracy of forecasts declines with an increasing time horizon. Theoretically, one can separate the forecast error (defined as the absolute difference of forecasted and actual growth rate) as the result of two error sources:

1. The forecasting model is incomplete in modeling the world, which causes an error in the forecast. As an example, the model may assume an incorrect *relationship* between the movement of key economic indicators, such as exchange rates, oil prices, etc. This introduces a forecast error which is idiosyncratic to the institute's model.

2. Unforeseeable year-specific macroeconomic shocks change the economic environment unexpectedly so that the model uses an incorrect assumption on the *level* of key economic indicators.

It is worthwhile to note that the first error source can be attributed to the institute's model, while the second error source can only be partly attributed to the institute as an error of its forecasting model since some macroeconomic shocks are less prone to be forecasted.

As an example: Naturally, the 9/11 attacks could hardly be included in any of the forecasts before the attacks had actually occurred. So it is understandable that the economic forecasts for the fourth quarter of 2001 and for early 2002 may have been off-track. In contrast, other shocks are more prone to be included: economists have argued well before the crisis hit that a housing bubble was existent in the United States.[4] Hence, this shock could be considered more prone to be included in any forecast. To be sure, in reality we likely deal with a continuum of shocks between the two extremes foreseeable and unforseeable with no clear boundary.

While an institute may want to argue that *any* deviation from the assumed level of oil and exchange prices should be regarded as an unforseeable shock, such a viewpoint would probably let forecasters get off the hook too easily. Forecasts are "sold" to the public as some insight on the economy's future and so their quality should also be judged at a reasonable level against how close they then match up with reality.

[4] For instance, Dean Baker, *The Run-Up in Home Prices: Is it Real or Is It Another Bubble?*, CEPR, 2002.

This section applies two tests to measure the performance of institutes' forecasts:

Test 1 measures an institute's performance by running regression (1) for each institution's forecasts:

$$\text{Forecast error} = \alpha + \beta \, \text{forecast\_horizon} + \varepsilon \qquad (1)$$

This regression estimates in β by how much the accuracy of one institute's forecasts diminishes when the forecast horizon (measured in months) increases by one month. The "forecast error" is measured as the absolute difference between the forecast and the actual growth rate. I therefore define the "quality" of an institution's forecasts through the estimate β: the lower β, the lower is the forecast error for a given time horizon and the more accurate was its forecast.

Note that this regression does *not* account for unforeseeable year-specific macroeconomic shocks, but instead compares only the forecast "sold" to the public to the actual outcome. While this may affect the absolute value of β, it will not change the relative ranking among the institutes since all the institutes face the same uncertainty about future shocks.

Test 2 measures the *relative* performance of the institutes to one another by running regression (2) and on the forecast error.

$$\text{Forecast error} = \alpha + \beta \, \text{forecast\_horizon} + \Sigma_{i=1}^{7} \, \text{institution\_dummy\_i} + \varepsilon \qquad (2)$$

The regression includes 7 institution dummies capturing the *relative* performance of the institution to the omitted 8th institution, while controlling for the forecast horizons of the institutes' forecasts.

Results

Table 3 shows the result of test 1 where we measure the performance as explained above through regression (1).

Table 3: Quality of forecasts when comparing forecast with actual GDP growth

	RWI	JEFG	IWH	IfW	Ifo	DIW	HWWI	SVG	All institutes
forecast horizon (t-statistic)	0.066 (5.33)	0.062 (5.02)	0.052 (6.55)	0.044 (5.54)	0.056 (4.63)	0.051 (5.35)	0.066 (3.86)	0.040 (2.58)	0.056 (16.74)
F-statistic (Prob>F)	28.45 (0.000)	25.23 (0.000)	42.88 (0.000)	30.65 (0.000)	21.46 (0.000)	28.59 (0.000)	14.92 (0.000)	6.65 (0.018)	280.14 (0.000)
Ranking	7	6	4	2	5	3	7	1	
R-squared	0.40	0.47	0.36	0.35	0.40	0.44	0.37	0.24	0.38
Observation	63	36	74	52	42	47	40	21	375

All columns with White heteroskedastic-robust standard errors. Column on "all institutes" also with clustering by forecasting institution

To interpret the result, consider for example the final column which features the results for all institutes. The forecast-horizon estimate is 0.056, which means that a 18-month forecast (e.g. forecasting in June 2008 the GDP growth rate for 2009) will be in average off by $18*0.056 \approx 1\%$.[5] As Germany has growth rates range between -0.2% and 3.2% in the sample period, an average error of 1% is quite considerable, since a very naïve approach of estimating 1.7% for each year would be off in average by less than 1.7%. A 12-month forecast will be off in average by roughly 0.7%. We can see that the *Sachverstaendigenrat* (SVG; German Council of Economic Experts) and the Institute fuer Weltwirtschaft Kiel (=IfWl; Institute for the World Economy in Keil) are doing comparably the best job.

Interestingly, the JEFG (the Joint Economic Forecast Group consisting of 4-6 of the economic forecasting institute) is doing a rather poor job; too many cooks may spoil the broth.

[5] This assumes that there is a uniform impact of the forecast horizon for all the forecasts independently from the length of the forecast. We will relax this assumption later.

Test 2 measures the *relative* performance of the institutes to one other by running regression (2) on the forecast error. The Institute "ifo Munich" is the omitted institute and coefficients need to be interpreted as the relative performance of a given institute relative to the omitted institutes "ifo Munich".

Table 4: Comparative Performance of Institutes

	Forecast error	Rank
forecast horizon	0.056 (17.79)	-
DIW	0.030 (7.41)	6
HWWI	-0.005 (-2.91)	3
IWH	0.037 (17.19)	7
IfW	-0.127 (-21.25)	1
JEFG	0.027 (8.17)	5
RWI	0.040 (4.72)	8
SVG	-0.042 (-5.94)	2
Omitted Institute: ifo	-	4
R-squared	0.39	
Observation	375	

From the two test we can compile the following ranking:

Table 5: Ranking

	Test 1	Test 2
IfW	2	1
JEFG	6	5
RWI	7	8
SVG	1	2
ifo	5	4
DIW	3	6
HWWI	7	3
IWH	4	7

The ranking of institutes by the two tests gives a sense of relative performance. Tests 1 and 2 provide overall similar results. (Major changes only occurs between the DIW, HWWI and IWH.) Importingly, the ranking should probably not given too much weight since – as shown above in chart 1 and table 1 – there is a strong correlation between institutes' forecasts (>85% for 1 year forecasts).

Section 4: Benchmark Models

This section compares the forecasts by the economic institutions to five very simple benchmark models:

(1) a "no change" model: we predict next year's growth to be the same as this year's growth

(2) a "random walk" model: we predict next year's growth to be the same as this year's growth plus some random shock drawn from a standard normal distribution

(3) a "constant drift" model: we predict next year's growth to be the same as this year's growth plus some constant factor that is either +0.5 or -0.5 (each factor is equally likely to occur)

(4) a "random walk with a constant drift" model: we predict next year's growth to be the same as this year's growth plus some random shock drawn from a standard normal distribution, plus some constant factor that is either +0.5 or -0.5 (each factor is equally likely to occur)

It is noteworthy that all benchmark models are purely statistical models which do not derive any prediction from economic fundamentals/indicators other than this year's growth rate. Furthermore, none of the models are statistically sophisticated in a way that they would include statistical information about the length or intensity of past business cycles or the long term growth rate.

Consequently, we would assume that due to its informational disadvantage, these benchmark models should not perform very well and be outperformed by the institutes' forecasts.

Moreover, comparing the benchmark model predictions to the forecasts by institutions with a forecast horizon horizon of less than 12 months, the benchmark models suffer a *real* informational disadvantage because they do not include any information of the months that have already past. [6] Instead, it duly continues to use only last year's growth rate as the only fundamental.

Table 6, column (1) shows the result for the no change benchmark model. The estimator on the time horizon is 0.035 (with a 99% significance level) compared to time horizons between 0.043 and 0.066 for the forecast instituts. Recall that this means that for each additional month in the forecast horizon the expected error increases by 0.034. Hence, while the no change benchmark model has for a 1 year forecast horizon an expected error of 12*0.034=0.408%, the institutes have expected errors between 12*0.043=0.516% and 12*0.066=0.792% (see also chart 3). Hence the simple no change benchmark model seems to be in fact more accurate than the forecasting models of the institutes.

Table 6, column (2) shows the result for the random walk benchmark model, which generates a GDP growth forecast through a simple random walk process. We estimate regression (1) by a simulation with 10,000 repetitions.[7] In each repetition we create over 300 forecast values – one for each of the institutes' forecasts between 1998 and 2008. The random walk benchmark model has a time horizon of 0.021 only and hence performs better than the no change benchmark model as well as the institutes' forecasting models.

Table 6, column (3) shows the result of the constant drift benchmark model, which adds or subtracts with equal likelihood a constant drift value of 0.5% to last year's GDP growth value. The simulation was again run 10,000 times. The model performs better than the no change benchmark model but worse than the random walk model. With a time horizon estimate of 0.028 it fares better than the forecasting models of the institutes.

The final benchmark model combines model (2) and (3): it adds or subtracts with equal likelihood a simple drift value of 0.5% to last year's GDP growth value as well as adds a random shock from a standard normal distribution. As can be seen in table 6, column (4) this benchmark model performs slightly better than the random walk benchmark model with an average expected error per month of 0.020 and also represents an improvement over the institutes' forecasts.

[6] Example: if a GDP forecast for 2006 is published in June 2006, its time horizon is 6 months.

[7] Since the shock is randomly drawn, a simulation based on a single run would not provide the correct estimation. Repeating instead the simulation 10,000 times we obtain (by the law of large numbers) a close approximation of the correct estimate.

Table 6: Benchmark results

	(1) No-Change Benchmark model	(2) Random Walk Benchmark model	(3) Constant Drift (+/- 0.5) Benchmark Model	(4) Random walk with Drift (+/- 0.5) Benchmark Model
Forecast horizon (t-statistic)	0.035 (6.67)	0.021 (2.95)	0.028 (4.76)	0.020 (2.63)
Observation	374	374 (each repetition)	374 (each repetition)	374 (each repetition)

Model (1) with heteroskedastic-robust standard errors. Model (2), (3) and (4) with 10,000 repetitions; each repetition runs regression (1) with White heteroskedastic-robust standard errors.

Section 5: Conclusion - What can we learn from Macroeconomic Forecasts?

The following lessons cans be learnt from the analysis:

1. Not surprisingly, the accuarcy of economic forecasts declines with increasing forecast horizons.
2. Chart 3 illustrates that the expected forecast error increases considerably among forecasts with horizons greater than one year.
3. There is a great deal of correlation between institutes' forecasts, in particular for the important 1-year forecasts. Three conjectures have been made why this could be the case: same models, same information and/or fear of losing credibility by being an outlier.
4. There seems to be an attenuation factor in the forecasts: in a boom, institutes underestimate economic growth, in a crisis institutes overestimate economic growth.
5. Comparably, the IfW and SVG have performed best in forecasting; this relative performance difference should however not be over-interpreted since the correlation among all forecasts is very high.
6. Comparing forecast models with several benchmark models, the forecasts perform rather poorly. Using a simple random walk process (i.e. estimating next year's GDP growth rate to be this year's GDP growth rate plus some random shock from a standard normal distribution)

performs much better than the institutes. In fact, even a "no-change" model (next year's GDP growth = this year's GDP growth) is a closer estimation than the institutes' forecasts.

7. In regard to the current year 2009, the latest forecasts range between 0% (SVG) and -2.7% (IfW) with a mean of -1.5%. Weighting these forecasts (with horizons of 11, 12 and 13 months) by their respective average time horizon error, we get that the real GDP growth rate for 2009 should be anywhere between -0.7% and -2.3%.[8]

The benchmark results need to be seen as sobbering news for the German economic forecasting profession: all simple benchmark models perform considerably better than the forecasts by economic institutions despite not using any real economic indicators other than the previous year's growth rate.

While this evidence is not meant to debase economic forecasting models in general (after all, institutes cannot be made responsible for clearly unforseeable macroeconomic shocks), it should remind the public to be wary of forecasts with greater time horizons and to take this into account when basing economic or political decision-making upon economic forecasts.

[8] The 95% confidence interval derived from past forecast accuracy is large, ranging from -0.1% to -3.2%. This reflects the little accuracy forecasts have had in the past.