Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Evaluating Forecasts of a Vector of Variables: a German Forecasting Competition

---

CAMA Working Paper 55/2014
July 2014

Hans Christian Müller-Dröge
Handelsblatt Newspaper

Tara M. Sinclair
Department of Economics, The George Washington University and
Centre for Applied Macroeconomic Analysis (CAMA), ANU

H.O. Stekler
Department of Economics, The George Washington University

Abstract

In this paper we present an evaluation of forecasts of a vector of variables of the German economy made by different institutions. Our method permits one to evaluate the forecasts for each year and then if one is interested to combine the years. We use our method to determine an overall winner for a forecasting competition across twenty-five different institutions for a single time period using a vector of eight key economic variables. Typically forecasting competitions are judged on a variable-by-variable basis, but our methodology allows us to determine how each competitor performed overall. We find that the Bundesbank was the overall winner for 2013.

Keywords

Mahalanobis Distance, forecasting competition, GDP components, German macroeconomic data

JEL Classification

C5, E2, E3

Address for correspondence:

(E) cama.admin@anu.edu.au

EVALUATING FORECASTS OF A VECTOR OF VARIABLES: A GERMAN FORECASTING COMPETITION

Hans Christian Müller-Dröge
Handelsblatt Newspaper

Tara M. Sinclair*
Department of Economics,
The George Washington University and
Centre for Applied Macroeconomic Analysis (CAMA), ANU

H.O. Stekler
Department of Economics
The George Washington University

JEL Codes: C5, E2, E3

Keywords: Mahalanobis Distance, forecasting competition, GDP components, German macroeconomic data

This Version: July 18, 2014

Abstract

In this paper we present an evaluation of forecasts of a vector of variables of the German economy made by different institutions. Our method permits one to evaluate the forecasts for each year and then if one is interested to combine the years. We use our method to determine an overall winner for a forecasting competition across twenty-five different institutions for a single time period using a vector of eight key economic variables. Typically forecasting competitions are judged on a variable-by-variable basis, but our methodology allows us to determine how each competitor performed overall. We find that the Bundesbank was the overall winner for 2013.

---

*Corresponding author: tsinc@gwu.edu. Monroe Hall #340, 2115 G Street NW, Washington, DC 20052, USA. The authors thank Jonas Dovern, Neil Ericsson, Luca Guerrieri, Renee McKibbin, Carsten-Patrick Meier, Dan Waggoner, and participants at the Bank of England, EABCN, and CAMA conference on Judgement and Combination in Forecasting and Policy Models for helpful discussions.

This paper evaluates a set of forecasts for the Germany macroeconomy prepared by twenty-five different national and international institutions.[1] Most evaluations have used a univariate methodology that separately examined the forecasts of each variable. We, however, use a multivariate methodology that permits us to determine an overall winner in a forecasting competition. It is difficult to compare forecasters if some produced better predictions of inflation while others produced better predictions of growth. To evaluate the overall performance, we simultaneously judge the accuracy of forecasts using a multivariate framework of a set of eight variables that were predicted by all institutions in our dataset.

There have been a small but growing number of studies that have considered some multivariate characteristics of forecasts. For example, Komunjer and Owyang (2012) evaluate forecasts in a multivariate framework by using forecast errors to derive the weights of a utility function. Their approach permits them to determine whether the forecasts were rationalizable. The approach that we present below differs from that of Komunjer and Owyang because we instead focus on forecast comparison in order to determine an overall winner for a forecasting competition.[2]

In deciding how to evaluate a set of forecasts, there are a number of dimensions to examine. Consider a large database of forecasts prepared by a number of individuals/organizations. The database would likely consist of forecasts made for a number of variables over a number of horizons over a period of time. How should one evaluate these forecasts? There is no simple answer because there are a number of ways of doing this analysis. They range from the evaluation of a single variable at a single horizon to the more complex methods which aggregate across the various dimensions of the data.

[1] For other evaluations of German forecasts, see Dohrn and Schmidt (2011), Dopke and Fritsche (2006), Dopke et al (2009), Heilmann and Stekler (2013), and Muller and Kirchgassner (2006).
[2] For a discussion of statistical tests for forecast competitions, see Koning et al (2005).

The database of forecasts in general will have four dimensions: (1) the number of variables (*J*) that are predicted, (2) the number of horizons/periods (*H*) for which each variable is predicted, (3) the number of times (*T*) that the predictions are made, and (4) the number of forecasters (*N*). The traditional procedure for evaluating forecasts involves calculating a scalar descriptive statistic such as mean-squared error (MSE) which describes the *average* accuracy of the *T* forecasts of *each* variable that were made for each forecast horizon. This approach yields *NHJ* descriptive statistics, one for each forecaster, at each horizon, for each variable.

Recent research has proposed several different procedures that have been used to aggregate across the various dimensions and reduce the number of descriptive statistics. The appropriate procedure for aggregating the forecasts depends on the question that is being investigated. For example, Eisenbeis, Waggoner and Zha (2002) aggregated across variables for a single time period and a single horizon for each forecaster. Their procedure created a ranking of the quality of each of the Wall Street Journal forecasters across multiple variables for a single horizon and a single time period.[3]

On the other hand, Clements, Joutz, and Stekler (2007) and Davies and Lahiri (1995, 1999) do not pool across variables.[4] For each variable, they pool across horizons. This yields a measure of the performance of each forecaster over all horizons. One difference between those two studies is that Clements et al. evaluate only one forecaster (the Fed) whereas Davies and Lahiri consider the forecasts of multiple forecasters (from the Blue Chip surveys and the Survey of Professional Forecasters).

[3] Eisenbeis, Waggoner, and Zha also produced an average ranking of the forecasters over time. Bauer et al (2003) applied their methodology to the Blue Chip Forecasts. Their methodology is similar to ours, but there are key differences discussed below.

[4] Davies, Lahiri, and Sheng (2011) provide a useful summary of the framework used in these papers.

This paper has a different focus: an evaluation of the overall accuracy of the 2013 forecasts of eight variables of the German economy made by 25 institutions. Thus we are interested in the *overall* quality of one organization's forecasts of a number of different variables for a single time period. This yields a measure of that organization's accuracy. We then calculate the same measure for all the organizations. The institution that is closest to the actuals will be judged to be the winner of the overall forecasting competition.

Our approach differs from the traditional procedures that have been used to evaluate forecasts. The traditional approach has been to consider the forecasts of a single variable made over a number of years. We are evaluating the simultaneous forecasts of many variables made in a single year. To illustrate this issue, we start with the simplest case: an evaluation of an organization's one-period-ahead forecasts of one variable, say the growth rate of real GDP. These forecasts have been made T times. The traditional univariate procedure involves calculating the MSE which describes the *average* accuracy of the T forecasts of real GDP growth. Now let us assume that the organization also prepares forecasts of inflation and the unemployment rate. Traditionally, we would have calculated MSEs for each of these additional variables. If the MSE of one variable were "small", while that of one or both of the others were "large", how would we evaluate the overall quality of this forecast? What do we learn by saying that the errors made in forecasting one variable were small while those made in forecasting the other variables were large?

In order to determine whether the individual produced a "good" *overall* forecast, we would need to obtain an error measure from a multivariate evaluation that aggregated across the variables. This aggregation is accomplished by (1) creating a vector of forecasts, (2) creating a vector of outcomes, and (3) measuring the distance between the two vectors. This methodology provides a single error measure for each institution's set of forecasts. We can then rank the different forecasters based on this error measure to determine a winner for our forecast competition.

Our approach is related to the methodology that Sinclair and Stekler (2013) utilized to analyze early GDP component estimates from the Bureau of Economic Analysis. That methodology determined whether, for each quarter, the vector of the first vintage of BEA estimates of all the major GDP components was similar to a vector of a later vintage of BEA estimates of the same components. To determine whether the two sets of estimates are related, it was necessary to compare the difference between the two vectors. Sinclair and Stekler utilized the Mahalanobis measure for estimating the relationship of two vectors. This measure, which is well established in the natural sciences, is a generalization of the Euclidean distance and allows for the interdependence of the vectors.[5] In order to test whether there was a difference between the two vintages of estimates, they focused on the difference between the mean vectors relative to the common within-group variation. [6]

In this paper we will utilize a similar methodology to analyze the forecasts that different institutions made about German economic activity in 2013.[7] In this case, however, our approach will be more like that of Eisenbeis, Waggoner, and Zha (2002) and Bauer, Eisenbeis, Waggoner, and Zha (2003). That approach allows us to rank the forecasters for a single outturn rather than over a number of years as in Sinclair and Stekler (2013).

---

[5] See Abdi (2007) for a discussion of different distance measures.

[6] Sinclair, Stekler, and Carnow (2012) applied this methodology to the median forecasts of the Survey of Professional forecasts for GDP growth, unemployment, and inflation. And Sinclair, Stekler, and Carnow (2014) applied this methodology to a vector of the Federal Reserve's forecasts.

[7] The inspiration for focusing on these German forecasts came from Müller-Dröge's March 2014 article in the Handelsblatt. The ranking in that article differs from the ranking in this paper because it is based on Euclidian rather than Mahalanobis distance.

The methodology we use in this paper relies on the historical time series data to determine the weighting matrix for the distance measure rather than using a model-based approach. For each forecaster, one vector consists of the forecasts of eight variables that the organization made in the last weeks of 2012 that refer to Germany economic activity for 2013. The other vector is comprised of the actual outcomes for those variables. Therefore, each institution's forecast vector is compared to the same actual outcome vector, providing an error metric for each organization. By ranking the organizations by this metric, it is possible to determine the winner of the 2013 forecasting competition. Our results indicate that it was the Bundesbank.

The rest of the paper proceeds in this way: We first describe the data and the methodology and then evaluate each institution's forecasts. We then provide a rank of the institutions and declare a winner to our forecasting competition.

I. Data

We consider forecasts for the German economy made by 25 different institutions (see Appendix for the full list). The forecasts of the eight variables that we consider in the competition are for: GDP, Private Consumption, Gross Fixed Capital Formation, Exports, Imports, the Government Surplus (as a percentage of GDP), Consumer Price Inflation (year over year), and the Unemployment Rate. GDP and its components are measured as year over year growth rates in real terms. Graphs of the actual historical data for these variables are presented in Figure 1.

The institutions' forecasts were made around December of 2012 and are presented in Table 1. We also include a number of benchmark forecasts in the comparison including a random walk forecast (using the actual values for 2012 as the forecast for 2013) and a naïve forecast consisting of a vector of zeros.

These forecasts are compared with both the first release of actual data for 2013 from January of 2014, and the "final" (thus second) release in February of 2014. The realized values come from the German Federal Statistical Office.

II. Methodology

As mentioned above, we use a distance measure to determine the accuracy of the forecasts, i.e. the difference of the vectors. There are two common measures of distance, Euclidean and Mahalanobis, that differ in the assumptions made about the statistical independence of the vectors. Assume that we have two independent vectors, $\boldsymbol{F}$ and $\boldsymbol{A}$. representing the forecasts and outcomes consisting of *n* variables in each vector. The difference between the two vectors can be measured by the Euclidean distance between them:

$$d(\boldsymbol{F}, \boldsymbol{A}) = \sqrt{(\boldsymbol{F} - \boldsymbol{A})'(\boldsymbol{F} - \boldsymbol{A})}. \qquad (4)$$

This procedure is only applicable to vectors that are independent and that are scaled so that they have unit variances. These assumptions do not apply in this analysis. Thus, we will use a generalization of the Euclidian distance that allows for the scale to differ across the different variables and for nonzero correlation between the variables. In order to measure the distance between each set of forecasts and the actual realizations of the series, we will focus on the difference between the vectors of each set of data relative to the historical variation of the actual series. This measure is called the Mahalanobis Distance, D^2:[8]

$$D^2 = (\boldsymbol{F} - \boldsymbol{A})'\boldsymbol{W}(\boldsymbol{F} - \boldsymbol{A}), \quad (5)$$

where $\boldsymbol{W}$ is the inverse of the sample variance-covariance matrix which we construct based on 20 years of historical actual data (with a robustness check using 10 years of historical data), and $\boldsymbol{F}$ and $\boldsymbol{A}$ are the mean vectors of the forecasts and outcomes, respectively.[9]

A few key properties of this measure are worth noting. First of all, D^2 will equal zero if the forecast vector exactly matches the actual vector. Once the forecasts differ, however, the correlation between the variables as well as the historical variance of each variable matter critically for determining the rank. For example, if a forecast is in a direction where there is less correlation, then a larger distance is assigned (De Maesschalck et al, 2000). [10]

III. Results

Table 2 presents for each variable the absolute forecast errors made by each institution. In addition to the forecasts of our 25 institutions, we also included in the table: (1) a “consensus” forecast, (2) the preliminary data as a “forecast” for the final data, and (3) two naïve forecasts. The “consensus” was the average of the predictions of the 25 institutions. The first naïve forecast was the random walk (same change) forecast, with the 2012 realized values treated as

[8]Mahalanobis distance is also associated with discriminant analysis. For other economic forecast applications of this measure, see Banternghansa and McCracken (2009) and Jordá et al (2010). For a useful overview of the measure, see De Maesschalck et al (2000).

[9] We estimate the sample covariance matrix as the (bias-corrected) sample covariance matrices from 20 years of actual data (10 years for a robustness check). It is assumed that the forecasts and the actuals have a common covariance matrix in the population.

[10] One interesting result arises from Kiel Economics. This private research company had one of the worst GDP forecasts in our set of institutions, and they actually never ranked as high as third for any of the individual forecasts (they were 4th for CPI), but the others that did better than them for some variables did worse for other variables so given the weights they ended up ranking third (or second using the 10 year weights).

the forecasts for 2013. The other naïve forecast assumed that there would be no change in each variable, i.e. each forecast value was set equal to zero.

We then sorted all the forecasts of each variable to show the variation in the ranking of the different institutions depending on which variable was being evaluated. (See Tables 2a and 2b.) Some results stand out, with the exception of the government surplus variable, the preliminary figures are very closely related to the final figures.[11] A comparison of the institutions' forecasts with those of the naïve models yields mixed results. Only 14 of 25 institutions had more accurate GDP forecasts than those that could have been generated by a random walk model. The institutions' consumption, CPI, and unemployment forecasts fared better, while those made for the other variables were considerably worse. In some cases, only one or two institutions beat one of the naïve models.

These tables also demonstrate the great variability in the rankings. For example, the Bundesbank ranks at the top for GDP growth and CPI inflation, but near the bottom for the government surplus. In constructing our measure for the overall accuracy of the forecasts, such differences must be taken into account. This variation suggests that the results will be sensitive to the importance or weights that are assigned to each of the variables that were forecast. Theoretically, the weights should be those of the user of the forecasts, but we do not know the future use of the forecasts or the loss function of the forecasters. Without this knowledge, we adopt an agnostic way to judge overall forecasting ability.

The Mahalanobis distance presents such a measure, with weights based on the historical patterns of the actual data. Forecasters that are consistent with these patterns are ranked more highly than those that perform well on some variables but poorly on others that are historically connected with those same variables. Weighting by the inverse of the historical variance-

[11] See Sinclair and Stekler (2013) for a similar result for the US GDP data.

covariance matrix also accounts for the historical relative predictability of the variables in terms of the variables' variability.

Table 3 presents the sorted Mahalanobis distances of each of the forecasts from the actuals. As shown there, we considered two different ways of constructing the weights for the Mahalanobis. Both are based on the historical data but one relies on only the last 10 years where the other relies on 20 years. For both weighting matrices, the preliminary data are the best forecast of the final data, and the naïve all zeros forecast is the worst. Seven institutions rank better than the random walk using the 20-year matrix (8 for the 10-year matrix). Although both weighting matrices result in the same top institution (the Bundesbank), the rest of the ranking is sensitive to the choice of the matrix. The Spearman rank correlation is 0.62. The weights are different as a result of the volatility that occurred during the global financial crisis that dominates the 10-year weights.[12]

IV. Conclusions

In this paper we showed how a new mutlivariate approach for evaluating economic forecasts permitted us to evaluate the predictions of several variables jointly. We then applied this approach to forecasts for the German economy made by 25 institutions and determined that the Bundesbank made the most accurate overall forecast for 2013.

[12] As a comparison, the Spearman rank correlation between the 20-year weights ranking and the Euclidean distance ranking (which was reported in Müller-Dröge, 2014), is 0.38.

Table 1: Forecasts for Germany for 2013 by the Different Organizations

Institutions	GDP	Private Consumption	Gross fixed capital formation	Exports	Imports	Government Surplus	Consumer Prices	Unemployed quota
Bundesbank	0.4	1	-0.1	1.9	3	-0.75	1.5	7.2
Commerzbank	0.5	1.3	0.1	2.8	4.1	-0.5	1.9	7.1
Deka	0.7	1.1	-0.3	3.3	3.3	-0.3	1.9	6.9
Deutsche Bank	0.3	0.6	1.1	3.2	4.2	-0.5	1.7	7
DIW	0.9	1.1	0.9	4.2	4.6	0	1.8	7
DZ Bank	0.4	0.9	0.1	3	3.8	-0.7	2.1	7.1
Feri	1.2	1.2	1.9	4.1	4.1	-0.3	2	6.6
Gemeinschaftsdiagnose	1	1.1	1.9	3.8	4.6	-0.2	2.1	6.8
Helaba	1.1	1.2	2.6	5.5	5	0	2	7
HSBC	0.6	1	0.5	2.9	4.1	-0.4	2	7
HWWI	0.5	1	0.7	3.5	4.4	-0.1	1.9	6.6
Ifo	0.7	0.7	0.7	3	3.3	-0.1	1.6	6.9
IfW	0.3	0.6	0.6	2.9	3.9	-0.5	2	7
IKB	0.8	0.9	-0.6	4.3	4	-0.1	1.9	6.9
IMK	0.8	0.7	0	3.5	3.6	0	1.7	7
ING	0.8	0.4	1.2	0.6	0.4	-0.2	2	6.9
IW	0.75	0.5	1.5	4	4	0	1.75	6.5
IWH	0.7	0.4	0.3	4	3.7	-0.3	2	6.7
Kiel Economics	1	1.4	1	3.6	4.1	-0.1	1.6	6.8
Landesbank Berlin	0.5	1	0.7	2.2	2.9	-0.8	1.6	6.8
MM Warburg	0.6	0.7	-0.8	4.2	1.6	0	1.5	7
Postbank	0.6	1	0.4	3.7	4.3	-0.3	2	7.1
RWI	0.3	0.2	-1.1	3	3.2	-0.5	1.7	7
UBS	0.8	0.9	1.7	2.6	3.9	0.2	2.1	7.2
Wirtschaftsweise	0.8	0.8	1.4	3.8	4.2	-0.5	2	6.9
Other Benchmark Forecasts								
Random Walk (2012 actual values)	0.7	0.8	-2.1	3.2	1.4	0.1	2	6.8
Naïve Zeros	0.0	0	0	0	0	0	0	0
Preliminary	0.4	0.9	-0.8	0.6	1.3	-0.1	1.5	6.9
Average over 25 Institutions	0.7	0.9	0.7	3.3	3.7	-0.3	1.9	6.9

Table 2a: Sorted Absolute Forecast Errors: GDP, Consumption, Investment, and Exports

GDP growth y/y %		Private Consumption growth y/y %		Gross fixed capital formation growth y/y %		Exports growth y/y %	
Bundesbank	0.0	DZ Bank	0.0	IKB	0.1	ING	0.2
DZ Bank	0.0	IKB	0.0	MM Warburg	0.1	**Preliminary**	0.2
Preliminary	0.0	UBS	0.0	**Preliminary**	0.1	**Naïve Zero**	0.8
Commerzbank	0.1	**Preliminary**	0.0	Deka	0.4	Bundesbank	1.1
HWWI	0.1	**Average over 25 Institutions**	0.0	RWI	0.4	Landesbank Berlin	1.4
Landesbank Berlin	0.1	Bundesbank	0.1	Bundesbank	0.6	UBS	1.8
Deutsche Bank	0.1	HSBC	0.1	IMK	0.7	Commerzbank	2.0
IfW	0.1	HWWI	0.1	**Naïve Zero**	0.7	HSBC	2.1
RWI	0.1	Landesbank Berlin	0.1	Commerzbank	0.8	IfW	2.1
HSBC	0.2	Postbank	0.1	DZ Bank	0.8	DZ Bank	2.2
MM Warburg	0.2	Wirtschaftsweise	0.1	IWH	1.0	Ifo	2.2
Postbank	0.2	**Random Walk (2012 actual values)**	0.1	Postbank	1.1	RWI	2.2
Average over 25 Institutions	0.3	Deka	0.2	HSBC	1.2	Deutsche Bank	2.4
Deka	0.3	DIW	0.2	IfW	1.3	**Random Walk (2012 actual values)**	2.4
Ifo	0.3	Gemeinschaftsdiagnose	0.2	**Average over 25 Institutions**	1.4	Deka	2.5
IWH	0.3	Ifo	0.2	HWWI	1.4	**Average over 25 Institutions**	2.5
Random Walk (2012 actual values)	0.3	IMK	0.2	Ifo	1.4	HWWI	2.7
IW	0.4	MM Warburg	0.2	Landesbank Berlin	1.4	IMK	2.7
IKB	0.4	Feri	0.3	**Random Walk (2012 actual values)**	1.4	Kiel Economics	2.8
IMK	0.4	Helaba	0.3	DIW	1.6	Postbank	2.9
ING	0.4	Deutsche Bank	0.3	Kiel Economics	1.7	Gemeinschaftsdiagnose	3.0
UBS	0.4	IfW	0.3	Deutsche Bank	1.8	Wirtschaftsweise	3.0
Wirtschaftsweise	0.4	Commerzbank	0.4	ING	1.9	IW	3.2
Naïve Zero	0.4	IW	0.4	Wirtschaftsweise	2.1	IWH	3.2
DIW	0.5	Kiel Economics	0.5	IW	2.2	Feri	3.3
Gemeinschaftsdiagnose	0.6	ING	0.5	UBS	2.4	DIW	3.4
Kiel Economics	0.6	IWH	0.5	Feri	2.6	MM Warburg	3.4
Helaba	0.7	RWI	0.7	Gemeinschaftsdiagnose	2.6	IKB	3.5
Feri	0.8	**Naïve Zero**	0.9	Helaba	3.3	Helaba	4.7

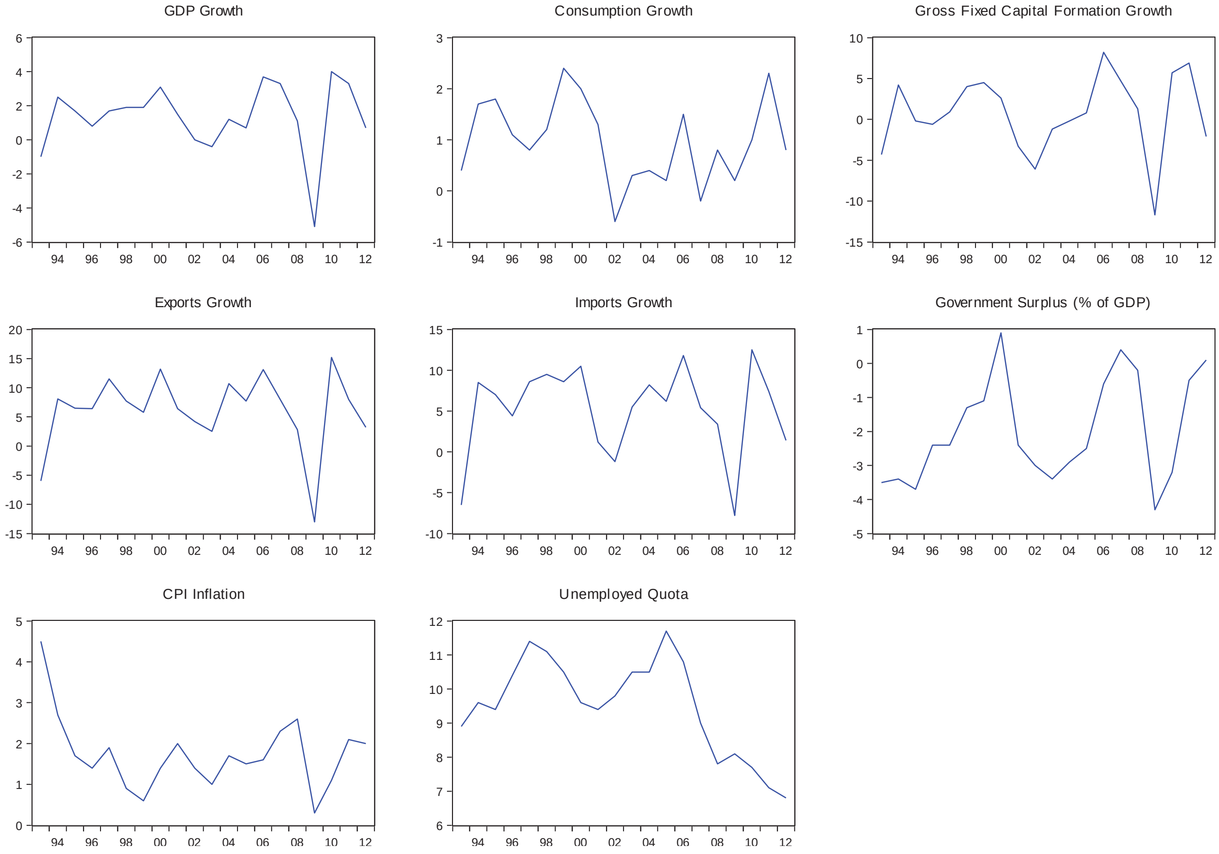
Table 2b: Sorted Absolute Forecast Errors: Imports, Government, CPI Inflation, and Unemployment

Imports growth y/y %		Government Surplus in % of GDP		Consumer Prices growth y/y %		Unemployed quota %	
Preliminary	0.4	DIW	0.0	Bundesbank	0.0	Deka	0.0
Random Walk (2012 actual values)	0.5	Helaba	0.0	MM Warburg	0.0	Ifo	0.0
ING	0.5	IMK	0.0	**Preliminary**	0.0	IKB	0.0
MM Warburg	0.7	IW	0.0	Ifo	0.1	ING	0.0
Naïve Zero	0.9	MM Warburg	0.0	Kiel Economics	0.1	Wirtschaftsweise	0.0
Landesbank Berlin	2.0	**Naïve Zero**	0.0	Landesbank Berlin	0.1	**Preliminary**	0.0
Bundesbank	2.1	HWWI	0.1	Deutsche Bank	0.2	**Average over 25 Institutions**	0.0
RWI	2.3	Ifo	0.1	IMK	0.2	Deutsche Bank	0.1
Deka	2.4	IKB	0.1	RWI	0.2	DIW	0.1
Ifo	2.4	Kiel Economics	0.1	IW	0.3	Helaba	0.1
IMK	2.7	**Random Walk (2012 actual values)**	0.1	DIW	0.3	HSBC	0.1
Average over 25 Institutions	2.8	**Preliminary**	0.1	**Average over 25 Institutions**	0.4	IfW	0.1
IWH	2.8	Gemeinschaftsdiagnose	0.2	Commerzbank	0.4	IMK	0.1
DZ Bank	2.9	ING	0.2	Deka	0.4	MM Warburg	0.1
IfW	3.0	UBS	0.2	HWWI	0.4	RWI	0.1
UBS	3.0	**Average over 25 Institutions**	0.3	IKB	0.4	Gemeinschaftsdiagnose	0.1
IKB	3.1	Deka	0.3	Feri	0.5	Kiel Economics	0.1
IW	3.1	Feri	0.3	Helaba	0.5	Landesbank Berlin	0.1
Commerzbank	3.2	IWH	0.3	HSBC	0.5	**Random Walk (2012 actual values)**	0.1
Feri	3.2	Postbank	0.3	IfW	0.5	Commerzbank	0.2
HSBC	3.2	HSBC	0.4	ING	0.5	DZ Bank	0.2
Kiel Economics	3.2	Commerzbank	0.5	IWH	0.5	Postbank	0.2
Deutsche Bank	3.3	Deutsche Bank	0.5	Postbank	0.5	IWH	0.2
Wirtschaftsweise	3.3	IfW	0.5	Wirtschaftsweise	0.5	Bundesbank	0.3
Postbank	3.4	RWI	0.5	**Random Walk (2012 actual values)**	0.5	UBS	0.3
HWWI	3.5	Wirtschaftsweise	0.5	DZ Bank	0.6	Feri	0.3
DIW	3.7	DZ Bank	0.7	Gemeinschaftsdiagnose	0.6	HWWI	0.3
Gemeinschaftsdiagnose	3.7	Bundesbank	0.8	UBS	0.6	IW	0.4
Helaba	4.1	Landesbank Berlin	0.8	**Naïve Zero**	1.5	**Naïve Zero**	6.9

Table 3: Sorted Mahalanobis Distance (D^2)

20 year weighting matrix			10 year weighting matrix	
Institution	**D^2**		**Institution**	**D^2**
Preliminary Data	0.4682		**Preliminary Data**	2.4945
Bundesbank	1.1973		Bundesbank	3.0692
Ifo	1.5763		Kiel Economics	3.8082
Kiel Economics	1.6410		Deka	4.0672
Landesbank Berlin	2.1605		IKB	6.2440
ING	2.2874		Commerzbank	8.0585
IMK	2.4383		IMK	12.1918
Deka	2.7624		HSBC	17.6526
Random Walk (2012 #s)	2.7980		DIW	21.5701
Average over 25 Institutions	2.9478		Feri	22.0972
Feri	3.0313		**Random Walk (2012 #s)**	25.9210
DIW	3.4148		Postbank	25.9825
UBS	3.6007		DZ Bank	26.1239
MM Warburg	4.0346		ING	26.2496
Gemeinschaftsdiagnose	4.4498		Landesbank Berlin	29.7532
Wirtschaftsweise	4.4844		**Average over 25 Institutions**	35.0089
HSBC	4.6321		MM Warburg	36.3310
Commerzbank	4.6454		UBS	38.4889
IKB	4.8429		Gemeinschaftsdiagnose	41.5261
RWI	5.1225		RWI	46.5641
IWH	5.2292		Ifo	49.1720
Postbank	5.2491		Wirtschaftsweise	67.2605
DZ Bank	5.5921		IWH	82.9496
IW	5.6753		HWWI	93.2919
Deutsche Bank	6.4792		IfW	114.2302
IfW	6.5306		Helaba	121.2678
HWWI	6.7250		Deutsche Bank	195.1387
Helaba	8.4295		IW	217.0846
Naïve Zero	58.7932		**Naïve Zero**	2021.6592

Figure 1: Actual Historical Data for Germany



Appendix: List of Forecasting Institutions

Short Name	Full Name	Type of Institution
Bundesbank	Deutsche Bundesbank	national bank
Commerzbank	Commerzbank Aktiengesellschaft	international bank
Deka	DekaBank Deutsche Girozentrale	umbrella organization of the german savings bank
Deutsche Bank	Deutsche Bank AG	international bank
DIW	Deutsches Institut für Wirtschaftsforschung	economic research institute
DZ Bank	Deutsche Zentral-Genossenschaftsbank	umbrella organization of the german cooperative banks
Feri	Feri Finance AG	wealth management
Gemeinschaftsdiagnose	Gemeinschaftsdiagnose der führenden Wirtschaftsforschungsinstitute	advisory board for the government (business cycle topics)
Helaba	Landesbank Hessen-Thüringen Girozentrale	regional state bank
HSBC	Hongkong & Shanghai Banking Corporation Holdings PLC	international bank
HWWI	Hamburgische WeltWirtschaftsInstitut	economic research institute
Ifo	Institut für Wirtschaftsforschung an der Universität München	economic research institute
IfW	Institut für Weltwirtschaft	economic research institute
IKB	Deutsche Industriebank	bank for sme-financing
IMK	Institut für Makroökonomie und Konjunkturforschung	economic research institute
ING	ING Groep N.V.	international bank
IW	Institut der Deutschen Wirtschaft	economic research institute
IWH	Institut für Wirtschaftsforschung Halle	economic research institute
Kiel Economics	Kiel Economics	private research company
Landesbank Berlin	Landesbank Berlin AG (LBB)	regional state bank
MM Warburg	M.M.Warburg & CO KGaA	wealth management
Postbank	Deutsche Postbank AG	international bank
RWI	Rheinisch-Westfälisches Institut für Wirtschaftsforschung	economic research institute
UBS	Union Bank of Switzerland	international bank
Wirtschaftsweise	Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung	advisory board for the government (general economic topics)

References

Abdi, Herve (2007). Distance. In N.J. Salkind (Ed.) Encyclopedia of Measurement and Statistics. Thousand Oaks CA: Sage, pp. 280-284.

Banternghansa, Chanont and Michael W. McCracken (2009). Forecast Disagreement Among FOMC Members. Federal Reserve Bank of St. Louis Working Paper No. 2009-059.

Bauer, Andy, Robert A. Eisenbeis, Daniel F. Waggoner, and Tao Zha, 2003. "Forecast evaluation with cross-sectional data: The Blue Chip Surveys," Economic Review, Federal Reserve Bank of Atlanta, issue Q2, pages 17-31.

Clements, Michael P., Fred Joutz and Herman O. Stekler, (2007). An Evaluation of the Forecasts of the Federal Reserve: A Pooled Approach. Journal of Applied Econometrics, 22(1), 121-136.

Davies, Antony and Kajal Lahiri, (1995). “A new framework for testing rationality and measuring aggregate shocks using panel data.” Journal of Econometrics 68: 205-227.

Davies, Antony and Kajal Lahiri, (1999). “Re-examining the rational expectations hypothesis using panel data on multi-period forecasts.” In Hsaio, C., K. Lahiri, L.F. Lee, and M.H. Pesaran (eds.), Analysis of Panels and Limited Dependent Variable Models, 226-254. Cambridge University Press.

Davies, Antony, Kajal Lahiri, and Xuguang Sheng (2011). “Analyzing Three-Dimensional Panel Data of Forecasts.” Oxford Handbook on Economic Forecasting, 473-495: 1-13

De Maesschalck, R., D. Jouan-Rimbaud, and D.L. Massart (2000). “The Mahalanobis distance.” Chemometrics and Intelligent Laboratory Systems, Volume 50, Issue 1, Pages 1-18

Dohrn, R. and C. M. Schmidt (2011), ‘Information or Institution – on the Determinants of Forecast Accuracy’, Jahrbucher fur Nationalokonomie und Statistik, 231, 9–27.

Dopke, J. and U. Fritsche (2006), ‘Growth and Inflation Forecasts in Germany: A Panel based Assessment of Accuracy and Dispersion’, Empirical Economics, 31, 777–798.

Dopke, J., U. Fritsche and B. Siliverstovs (2009), ‘Evaluating German Business Cycle Forecasts under an asymmetric loss function’, KOF Working Paper. KOF, ETH Zurich, Zurich.

Eisenbeis, Robert A., Daniel F. Waggoner, and Tao Zha (2002). “Evaluating Wall Street Journal survey forecasters: a multivariate approach”. Federal Reserve Bank of Atlanta Working Paper 2002-8.

Groen, Jan J. J., George Kapetanios and Simon Price (2009).Real Time Evaluation of Bank of England Forecasts for Inflation and Growth. International Journal of Forecasting, 25, 74-80.

Heilmann, U. and H. Stekler (2013), Has the Accuracy of Macroeconomic Forecasts for Germany Improved?, German Economic Review, 14, 235-253.

Holden, K. and D.A. Peel (1990).On testing for unbiasedness and efficiency of forecasts. Manchester School, 48, 120–127.

Jordá, Òscar, Malte Knüppel, and Massimiliano Marcellino (2010).Empirical Simultaneous Confidence Regions for Path-Forecasts. European University Institute Economics Working Papers No.ECO2010/18.

Komunjer, Ivana and Michael T. Owyang (2012). Multivariate Forecast Evaluation and Rationality Testing. Review of Economics and Statistics. Vol. 94, No. 4: 1066-1080.

Koning, Alex J., Philip Hans Franses, Michele Hibon, H.O. Stekler. The M3 competition: Statistical tests of the results. International Journal of Forecasting. 21: 397-409.

McLachlan, G. J. (1999) "Mahalanobis Distance" Resonance, pp. 20 -26.

Mincer, J. and V. Zarnowitz (1969).The Evaluation of Economic Forecasts. In J. Mincer (Eds.) Economic Forecasts and Expectations. National Bureau of Economic Research, New York.

Müller-Dröge, Hans Christian (2014) "Weidmann ist der beste Hellseher" Handelsblatt. March 17, 2014.

Muller, U.K. and G. Kirchgassner (2006), Are forecasters reluctant to revise their predictions? Some German evidence, Journal of Forecasting, 25, 401-413.

Sinclair, T. M., and Stekler, H.O. (2013). Examining the Quality of Early GDP Component Estimates. International Journal of Forecasting. Vol 29, No. 4, pp. 736-750.

Sinclair, Tara M., H. O. Stekler and Warren Carnow (2012). "A New Approach for Evaluating Economic Forecasts" Economics Bulletin, Vol. 32 No. 3 pp. 2332-2342.

Sinclair, Tara M., H. O. Stekler and Warren Carnow (2014). "Evaluating a Vector of the Fed's Forecasts" International Journal of Forecasting. Forthcoming.