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

Hans Christian Müller-Dröge
Handelsblatt Newspaper

Tara M. Sinclair
The George Washington University

H. O. Stekler
The George Washington University

Motivation and Overview

- Typically forecasts are evaluated variable-by-variable.
- For a forecasting competition, however, **we may want to determine a <u>single winner</u> for a <u>set of forecasts</u>.**
- We propose a specific formulation of the Mahalanobis Distance measure which is useful in this context.
- We apply this method to a forecasting competition comparing forecasts from 25 institutions of 8 annual macroeconomic outcomes for Germany in 2013.

Mahalanobis Distance

- We want to measure the distance between each set of forecasts (F) and the actual realizations (A) of the series and then rank the forecasters by their respective distances from the actuals.
- Mahalanobis distance (D^2) is a generalization of Euclidean distance for interrelated series, such as several macroeconomic variables for the same country:

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

- D^2 will equal zero if the forecast vector exactly matches the actual vector.
- Forecasts that have a strictly smaller squared error for all variables included in the evaluation, will have a smaller D^2.
- We must determine the appropriate weighting matrix, W. We use the inverse of the historical (20 years) variance-covariance matrix of the realized data.
  - Others have used variance-covariance matrix of forecast errors from a naïve model instead of from the data.
  - Ours is more transparent and consistent with the statistical literature
  - We can compare with naïve models by including them as additional forecasts to evaluate.

The Data

- One vector (F) consists of the forecasts of eight variables that each 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 (February 2014 vintage) for those variables.

2013 Actuals and Forecasts

	GDP growth y/y	Private Consumption growth y/y	Gross fixed capital formation growth y/y	Exports growth y/y
Actual	0.4%	0.9%	-0.7%	0.8%
Forecast Min	0.3%	0.2%	-1.1%	0.6%
Forecast Max	1.2%	1.4%	2.6%	5.5%

	Imports growth y/y	Government Surplus in % of GDP	Consumer Prices growth y/y	Unemployed quota
Actual	0.9%	0.0%	1.5%	6.9%
Forecast Min	0.4%	-0.8%	1.5%	6.5%
Forecast Max	5.0%	0.2%	2.1%	7.2%

Univariate Results

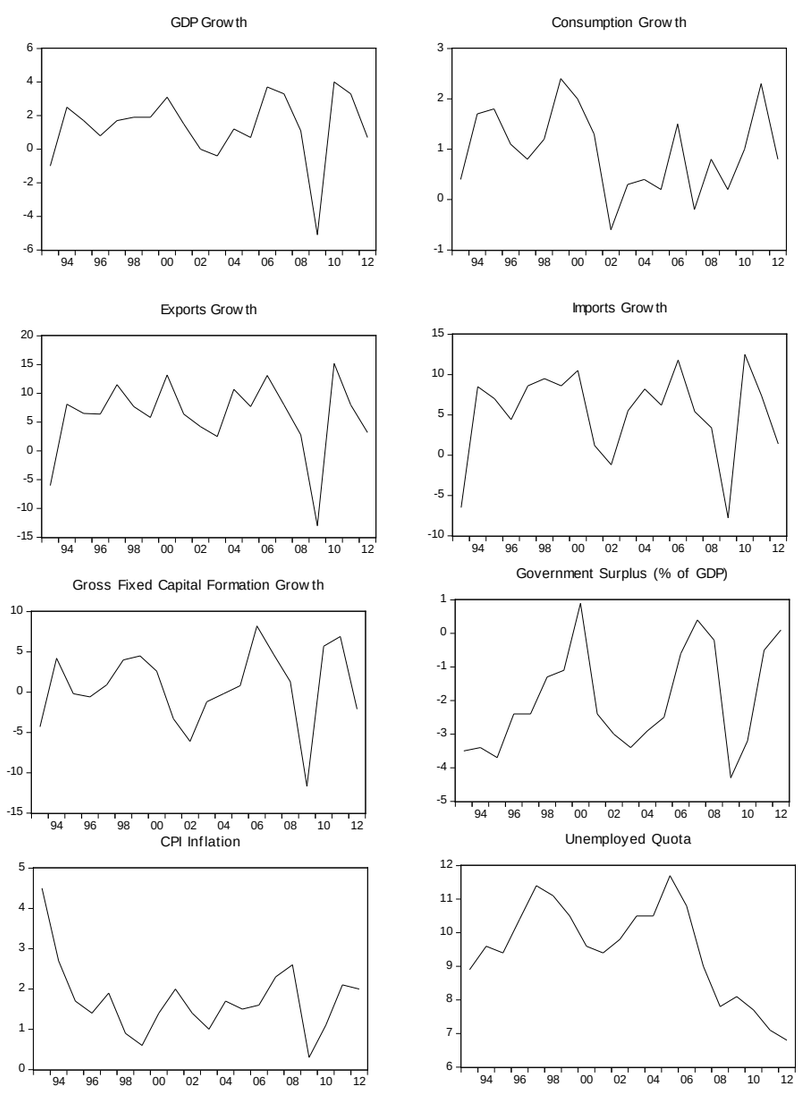
Table 2a: Sorted Top 10 by Absolute Forecast Errors: GDP, Consumption, Investment, and Exports

GDP	Private Consumption	Gross Fixed Capital Formation	Exports
Bundesbank	DZ Bank	IKB	ING
DZ Bank	IKB	MM Warburg	**Preliminary**
Preliminary	UBS	**Preliminary**	**Naïve Zero**
Commerzbank	**Preliminary**	Deka	Bundesbank
HWWI	**Average over 25 Institutions**	RWI	Landesbank Berlin
Landesbank Berlin	Bundesbank	Bundesbank	UBS
Deutsche Bank	HSBC	IMK	Commerzbank
IfW	HWWI	**Naïve Zero**	HSBC
RWI	Landesbank Berlin	Commerzbank	IfW
HSBC	Postbank	DZ Bank	DZ Bank

Table 2b: Sorted Top 10 by by Absolute Forecast Errors: Imports, Government, CPI, and Unemployed

Imports	Government Surplus	Consumer Prices	Unemployed Quota
Preliminary	DIW	Bundesbank	Deka
Random Walk (2012 actual values)	Helaba	MM Warburg	Ifo
ING	IMK	**Preliminary**	IKB
MM Warburg	IW	Ifo	ING
Naïve Zero	MM Warburg	Kiel Economics	Wirtschaftsweise
Landesbank Berlin	**Naïve Zero**	Landesbank Berlin	**Preliminary**
Bundesbank	HWWI	Deutsche Bank	**Average over 25 Institutions**
RWI	Ifo	IMK	Deutsche Bank
Deka	IKB	RWI	DIW
Ifo	Kiel Economics	IW	Helaba

Historical Data



- We base our weights on the inverse of the variance-covariance matrix from 20 years of historical data
  - Robustness check using just 10 years of historical data.
- This neglects the relative importance of the series to the end user.

The Competition

- A comparison of the D^2 for each institution gives us which institution's (overall) forecast is closest to the actuals.
  - We use this to rank the institutions by accuracy and to determine the winner of our forecasting competition.

Mahalanobis Distance Results

Table 3: Sorted Mahalanobis Distance (D^2)

20 year weighting matrix		Rank	10 year weighting matrix	
Institution	**D^2**		**Institution**	**D^2**
Preliminary Data	0.47	1	**Preliminary Data**	2.49
Bundesbank	1.20	2	Bundesbank	3.07
Ifo	1.58	3	Kiel Economics	3.81
Kiel Economics	1.64	4	Deka	4.07
Landesbank Berlin	2.16	5	IKB	6.24
ING	2.29	6	Commerzbank	8.06
IMK	2.44	7	IMK	12.19
Deka	2.76	8	HSBC	17.65
Random Walk (2012 #s)	2.80	9	DIW	21.57
Average over 25 Institutions	2.95	10	Feri	22.10
Feri	3.03	11	**Random Walk (2012 #s)**	25.92
DIW	3.41	12	Postbank	25.98
UBS	3.60	13	DZ Bank	26.12
MM Warburg	4.03	14	ING	26.25
Gemeinschafts-diagnose	4.45	15	Landesbank Berlin	29.75
Wirtschaftsweise	4.48	16	**Average over 25 Institutions**	35.01
HSBC	4.63	17	MM Warburg	36.33
Commerzbank	4.65	18	UBS	38.49
IKB	4.84	19	Gemeinschafts-diagnose	41.53
RWI	5.12	20	RWI	46.56
IWH	5.23	21	Ifo	49.17
Postbank	5.25	22	Wirtschaftsweise	67.26
DZ Bank	5.59	23	IWH	82.95
IW	5.68	24	HWWI	93.29
Deutsche Bank	6.48	25	IfW	114.23
IfW	6.53	26	Helaba	121.27
HWWI	6.73	27	Deutsche Bank	195.14
Helaba	8.43	28	IW	217.08
Naïve Zero	58.79	29	**Naïve Zero**	2021.66

Sensitivity Issues

- The rankings depend substantially on both the variables chosen and the specific covariance matrix used.
- Ranking will be consistent with what we determine from looking at absolute forecast errors only with strict inequality.
  - It is possible for a forecaster to be worse on all forecasts except one where they have the same forecast as another contestant and be ranked above the weakly "better" forecaster.

Conclusions

- Based on our analysis, **<u>the Bundesbank was declared the winner.</u>**
  - Based on the univariate forecast errors, the Bundesbank ranks at the top for GDP growth and CPI inflation, but near the bottom for government surplus.
- One interesting result arises from Kiel Economics:
  - one of the worst GDP forecasts in our set of institutions
  - never ranked as high as third for any of the individual forecasts (they were 4th for CPI),
  - but ended up ranking 3rd (or 2nd using the 10 year weights).