Motivation and Overview

- Typically forecasts are evaluated variable-by-variable.
- For a forecasting competition, however, we may want to determine a single winner for a set of forecasts.
- 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:

  \[ D^2 = (F - A)^TW(F - A) \]

- $D^2$ 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

### Table 2a: Sorted Top 10 by Absolute Forecast Errors:

<table>
<thead>
<tr>
<th>Institutions</th>
<th>20 year weighting matrix</th>
<th>Rank</th>
<th>10 year weighting matrix</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bundesbank</td>
<td>0.07</td>
<td>1</td>
<td>Bundesbank</td>
<td>2.49</td>
</tr>
<tr>
<td>MM Warburg</td>
<td>1.20</td>
<td>2</td>
<td>Bundesbank</td>
<td>3.07</td>
</tr>
<tr>
<td>Landesbank</td>
<td>2.16</td>
<td>3</td>
<td>Kieler Economies</td>
<td>3.81</td>
</tr>
<tr>
<td>Deutsche Bank</td>
<td>2.29</td>
<td>4</td>
<td>Deka</td>
<td>4.07</td>
</tr>
<tr>
<td>Deka</td>
<td>2.44</td>
<td>5</td>
<td>IKB</td>
<td>6.24</td>
</tr>
<tr>
<td>IKB</td>
<td>2.76</td>
<td>6</td>
<td>Commerzbank</td>
<td>8.06</td>
</tr>
<tr>
<td>Commerzbank</td>
<td>2.80</td>
<td>7</td>
<td>Landesbank</td>
<td>12.18</td>
</tr>
<tr>
<td>Landesbank</td>
<td>2.95</td>
<td>8</td>
<td>DIK</td>
<td>17.66</td>
</tr>
<tr>
<td>DIK</td>
<td>3.03</td>
<td>9</td>
<td>Bundesbank</td>
<td>21.37</td>
</tr>
<tr>
<td>Bundesbank</td>
<td>3.03</td>
<td>10</td>
<td>Bundesbank</td>
<td>22.10</td>
</tr>
</tbody>
</table>

### Table 2b: Sorted Top 10 by Absolute Forecast Errors:

- We compare with naïve models by projecting the historical data.
- We base our weights on the inverse of the covariance matrix used.
- Ranking will be consistent with what we determine by 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.

### Results

- 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 reflects 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.

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, the Bundesbank was declared the winner.
  - 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).