FEW DIFFERENCES – ACCURACY OF INTERNATIONAL FORECASTS AND FORECASTERS[1]

Ullrich Heilemann[2]

Outline

[1] Presentation at the Project LINK Meeting 2019, June 17 – 19, 2019, Glen Cove Mansion, NY

[2] Universität Leipzig, Grimmaische Straße 12-14, 04105 Leipzig. E-Mail: heilemann@wifa.uni-leipzig.de.

1 Introduction

A. Many institutions provide several times a year macroeconomic forecasts for a large number of countries

   The accuracy of forecasts by IMF, OECD, World Bank, European Commission has been widely examined (e. g. Genberg, Martinez 2014, Pain 2014, Pingfan and Tan 2014)

   Comparisons with the accuracy of national forecasts of the international economy are very rare and often confined to a small set of countries

B. Here we compare the accuracy of German forecasts of GDP and inflation (GDP deflator/deflator of Private Consumption) in the period 2001– 2015 for 19 countries by

   - the „Joint Diagnosis“ a major German, government sponsored forecaster,
   - FERI a small private, profit-oriented forecaster (FERI), with that of
   - IMF and
   - OECD

1 Introduction

D. The comparison covers forecasts in spring/autumn for the current and for the coming year

E. The study focuses on three questions:
   - How accurate were the forecasts for the 19 countries?
   - How accurate were the forecasts of FERI, JD, IMF and OECD?
   - Can the results be explained with reference
     - to countries' macroeconomic features such as size, income level, export ratio and foreign trade share and
     - – as perhaps in the case of Germany, the United States and France – by forecasters "distance"/"home advantages"/"home bias"?

F. The study follows mostly a user/policy perspective
   Shifts/break points (e.g. after the Great Recession (2007–2010) are not examined, some comparison with previous results for the G7 countries are made

2 Data and methods

A. Data

- As actual data we used the first releases of NA data available in early spring for the past year as collected in the OECD database (real-time data)
- Spring forecasts are published in April (OECD: June), autumn forecasts in October (IMF, JD) and in November (FERI, OECD), only the OECD can rely on national accounts (NA) data for the 1st quarter (spring) forecast) for the 3rd quarter (autumn)

B. Methods

To examine forecast accuracy we use standard methods such as the Mean absolute error (MAE), the Root mean square error (RMSE), both also in relative terms (MPE, RMSPE),

- the bias = $\frac{1}{T}\sum_{t=1}^{T} e_t$

- a benchmark of comparative accuracy is Theil's U coefficient:
  U = $\sqrt{\frac{\frac{1}{T}\sum_1^T (p_t - a_t)2}{\frac{1}{T}\sum_1^T a_{t-1}2}}$, comparing actual forecasts with a "naïve" forecast
  $p_t = a_{t-1}$
- The forecast performance associated with the difficulty of the task is measured by the relationship of RMSE/σ = $\frac{\mathrm{RMSE}}{\sigma}$ with $\sigma = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(a_t - \bar{a})^2}$ und $\bar{a} = \frac{1}{T}\sum_1^T a_t$
- To examine the accuracy of turning point forecasts, we first classify forecast error as over-predictions, under-predictions and turning-point errors (Lamberts and Schüssler 1967): $Q_t = \frac{p_t - a_{t-1}}{a_t - a_{t-1}}$,
  where $Q_t < 0$ indicates whether a turning point had been missed or having a turning point predicted that did not occur. This is used to calculate W = $\frac{\sum_{t=1}^{T} Q_t<0}{n}$ with n = number of (actual) turning points. A value for W=0.1, for example, means that ten percent of the turning points were not predicted or were falsely predicted. W >1 results when more turning points were predicted than had actually occurred

2 Data and methods of analysis

- To examine whether forecasts capture the dynamics of the development, we calculate the 'informational content' (IC), comparing the number of accelerations (decelerations) of change rates predicted with their realizations (see e.g. Diebold and Lopez 1996):

$$\mathrm{IC} = \frac{\mathrm{AC}}{\mathrm{AC} + \mathrm{AW}} + \frac{\mathrm{DC}}{\mathrm{DC} + \mathrm{DW}}$$

  with: AC: increase forecast and realized; AW: increase forecast and decrease realized; DC: decrease forecast and realized; and DW: decrease forecast, increase realized. For a forecast to have 'informational content', IC has to be > 1 (Merton, 1981).

- Efficiency: Forecasts are (weak) efficient if two conditions hold: first, the forecast error cannot be explained by systematic forecast errors, so that in $e_t = \alpha_0 + \beta_1 p_t + \varepsilon_t$ the condition $\beta_0 = 0$ holds, second, the forecast error cannot be explained by auto correlated errors, so that in $e_t = \alpha_1 + \gamma_1 e_{t-1} + \varepsilon_t$ the condition $\gamma_1 = 0$ holds (Holden and Peel 1990). (Like Theil's inequality coefficient, this test is based on the assumption that the previous year's actual data are known, which is rarely the case, in particular not in the spring and autumn forecasts for the coming year.

- Finally, we examine how quickly forecasts are revised or forecast errors decrease. We compare the ratio of the MAE of the respective forecast with the MAE of the spring forecast for the coming year:

  $Rev = MAE_{\text{forecast t, t+k}} / MAE_{\text{spring forecast t, t+1}}$. with k = 0 and 1.

  Correspondingly, for example, the autumn forecast for the respective year (t) results:

  $Rev = MAE_{\text{forecast autumn t, t}} / MAE_{\text{spring forecast t, t+1}}$.

  If the revisions are unidirectional, which has been the case with only a few exceptions, the smaller the Rev, the faster the final forecast value is reached.

3 Results

We concentrate here on autumn forecasts for the coming year – first we look at countries than at institutions

A. Growth forecasts (Table 2. Fig. 4, below)

- The MAE („error“) for all countries is rather high (1.2) and varies little (σ: 0,5) – notable exemptions are shown in Figure 4 (below)
- Classifying the errors into groups with a distance of 0.5 (Figure 2) shows an approximately normal distribution of the errors within the range of +/-0.5.
- The bias is clearly positive and varies considerably (ø: 0.4, σ: 0.4). Efficiency tests, however, reject the hypothesis of unbiasedness only for Italy.
- Accelerations/decelerations are mostly correctly predicted – as the IC of 1,5 signals)
- The share of turning point errors is on average 60%
- U is on average 0.5 – 50% better than (very) naïve forecasts

- The picture improves when errors are standardized (RMSE improves from ø: 1.4, σ: 0.9 to ø: 0.9, σ: 0.2 when standardized
- The revision of the forecasts is considerably but results do not differ much. The largest corrections take place with the autumn forecasts for the coming year, all in all following an inverted L-curve (Lahiri, Sheng 2008).
- As to the “sticky processing of information/noisy information”-debate, the results seem to support the latter
- The Great Recession (here: 2008–2010) was, of course (1/5th of the sample period), of considerable influence. When it is excludes, errors shrink by 30%, in half the cases by up to 50%
-

Institutions

- Differences rarely exceed 0.5. The slightly higher accuracy of FERI/OECD forecasts than that of IMF/JD may result from more actual data
- Regarding forecasting accelerations/decelerations, turning points, recessions and forecast revisions the differences to the growth forecasts are very small
- The null-hypothesis of equal forecast accuracy can be rejected only for few countries as the Diebold-Mariano test reveals (Table 4).

3 Results

3.2 Inflation forecasts (Table 5)

- On average, the accuracy is surprisingly low (ø: 1.1, σ: 0.5) and hardly differs from that of growth forecasts
- The bias is negative and small (-0.1) but varies very much (σ: 0.4) but tests of unbiasedness reject the null hypothesis only for few countries
- The classification of errors is similar to that of growth forecasts but the shape is sometimes flatter
- The proportion of turning-point errors (60%) is also rather high while IC is as high as with growth forecasts
- Relative accuracy (U) is somewhat lower but varies less (ø: 0.8; σ: 0.2) – two thirds of the countries seem not take properly account of the variability of the development: UC is much lower than with growth forecasts
- Standardization does not much improve accuracy but reduces very much the σ (ø: 1.0, σ: 0.1 vs. ø: 1,1, σ: 0.8)

3 Results

- Most revisions are also unidirectional and similar to that of growth forecasts and major revisions take place with the autumn forecasts for the coming year
- The accuracy of inflation forecasts also suffered from the Great Recession – excluding it reduces the error by 40%
- Correlation between the errors of growth and inflation forecasts is about 0.457 and increases to 0.559 when the Great Recession is excluded

Institutions

- Accuracy of Inflation forecasts of the institutions are even more homogenous than growth forecasts
- The small differences between institution may be due to different actuality of the data
- Comparisons with other studies are difficult because of different country sets and sample periods – for the G7 countries results are much in line with that of Heilemann/Stekler (2013) for 2000–2010

3.3 Similarities and explanations

3.3.1 Similarities

- Errors of both forecasts in autumn for the coming year are about 1 percentage point
- Accelerations and decelerations are mostly seen correctly
- Recessions/crisis are seen only when countries are already in
- Forecasters have a low prior about the probability of recessions – hardly any recession had been predicted that did not happen
- The Great Recession had a major influence on the accuracy of both forecasts
- The Diebold-Mariano tests rejects the hypothesis of equal accuracy of growth and inflation forecasts only for few countries and forecasters

3.3.2 Explanations

- Numerous objective and subjective determinants of forecast accuracy – even within the widely used framework of the GDP-model
- Not much is known about the production of forecasts in the four institutions: All use the GDP-model framework, a rather similar set and apparently not all too different assumptions and the (final) forecast is more or less the outcome of a negotiation process
- Here we examine the role of a number of structural characteristics and have a brief look on institution related factors

National structural characteristics (Table 1)

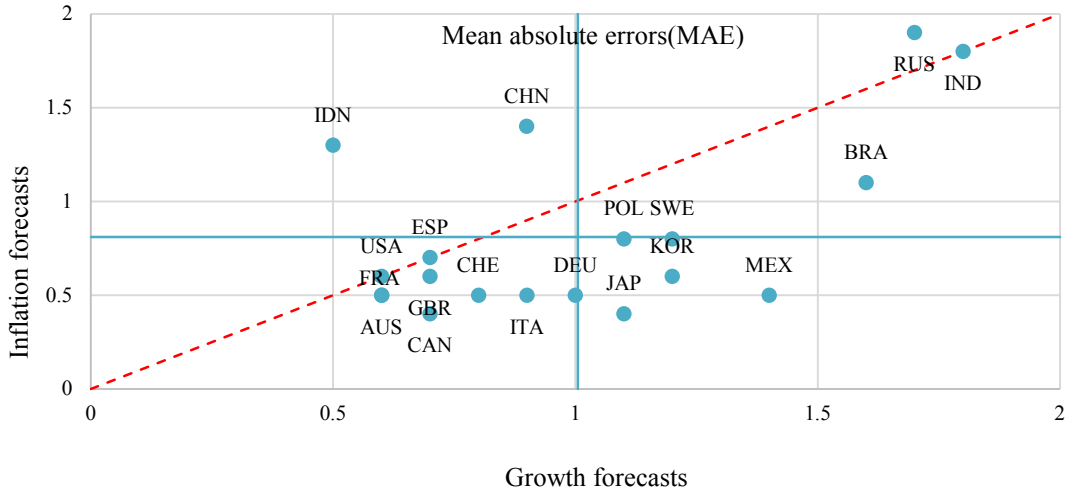
- Per capita income is negatively correlated with the accuracy of growth and inflation forecasts (the idea behind this relationship is that expenditure for statistics increases with per capita income) (actual comparisons between expenditure for official statistics and its quality seem to be missing)
- There seems to be no relationship between export-share and accuracy even for countries with an export share (GDP) > 25% (IMF results differ from this)

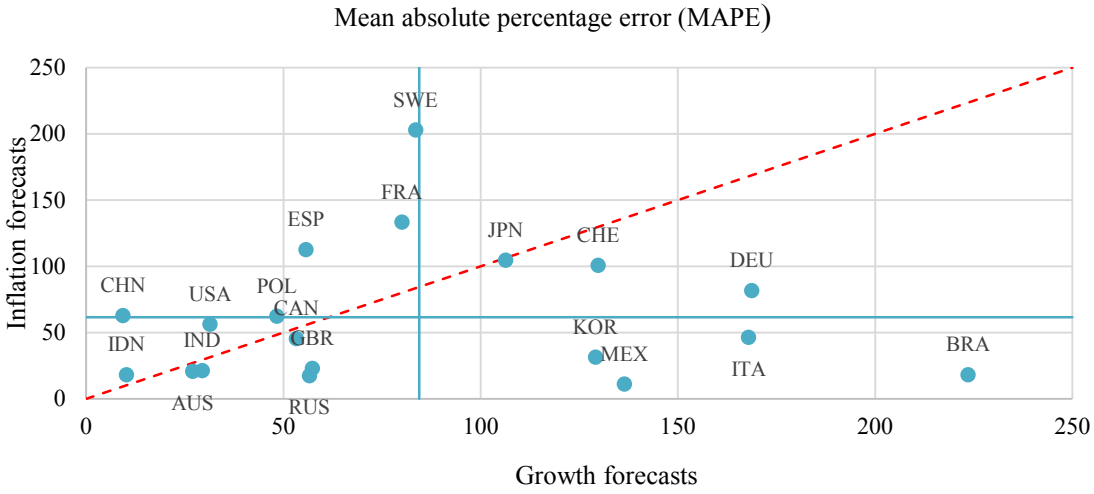
Table 1: Macroeconomic indicators, selected countries, 2001–2015

	Growth[1]		Inflation[2]		Real GDP[3] 2010	Real GDP p. capita.[4] 2010	Export-share[5] 2010	World trade share[6] 2010
	Ø	Σ	Ø	Σ				
Australia	2.7	0.7	2.6	0.8	0.94	42.2	21.1	1.4
Brazil	2.5	2.3	6.1	1.6	2.8	14.3	10.7	1.3
China	8.9	1.3	3.2	2.2	12.36	9.2	28.8	9.7
Germany	1.1	1.9	1.6	0.8	3.24	40.4	42.3	7.6
France	1.0	1.3	1.6	0.9	2.33	35.9	26.0	3.7
India[7]	7.2	2.0	7.5	2.7	5.42[6]	4.4	22	1.8
Indonesia[7]	5.3	0.7	7.0	2.4	2.05[6]	8.5	24.3	0.9
Italy	0.0	1.9	2.0	1.0	2.06	34.4	25.2	3.0
Japan	1.0	2.1	0.2	0.9	4.32	33.8	15.2	4.8
Canada	1.9	1.6	1.8	0.7	1.36	40.1	29.1	2.6
Mexico	2.2	3.0	4.2	0.6	1.73	15.1	29.9	2.0
Poland	3.5	1.6	2.3	1.4	0.79	20.6	40.1	1.1
Russia	3.5	4.2	10.0	3.1	3.12	21.9	27.3	2.1
Sweden	2.0	2.3	1.1	1.1	0.39	41.8	46.2	1.0
Switzerland	1.4	1.3	0.4	0.9	0.4	51.2	64.2	1.2
Spain	1.4	2.1	2.2	1.5	1.51	32.4	25.5	1.8
South Korea	3.7	2.0	2.7	1.0	1.51	30.5	49.4	2.9
United Kingdom	1.5	1.9	2.3	1.1	2.28	36.3	28.3	3.2
United States	2.1	1.6	2.1	1.2	14.96	48.3	12.4	10.6
Ø (σ)	2.8 (2.2)	1.9 (0.8)	3.2 (2.5)	1.4 (0.7)	3.3 (1.9)	29.5 (3.7)	29.9 (3.6)	3.3 (1.7)

Authors' computations, according to OECD, WTO and World Bank. – [1] 2000–2015. [2] 2003–2015. – [3] Billion US$, purchasing power parities, constant prices (2010). – [4] In 1000 US$, purchasing power parities, constant prices (2010). – [5] GDP share of exports, in %. – [6] share in world trade in %. – [7] According to World Bank.

Figure 4: Errors of growth and inflation errors, 2001–2015





Authors' computations according to Table 7 and 8. – Averages of all forecasts of all institutions

- Continental patterns are not detected (for the between war period see e.g. Wagemann 1928) though there are more or less clear classes of forecast accuracy (Figure 4)
- There is strong relation between trade links and the accuracy of growth and inflation forecasts (Table 6) – forecasts for the United States and Germany (and probably China) are of pivotal influence

Institutions

- Size/resources seem to be of no importance for the accuracy of growth and inflation forecasts. Of course this may be different when looking at aggregates or at the quality of forecasting reports
- "Distance does not matter": "home advantages" do not exist – the forecasts for Germany by German institutes are in general not better than that of OECD or IMF and the same holds for the German forecasts e.g. for the United States (again, things might be different when looking at aggregates such as Government deficit)

4 Summary and conclusions

- The accuracy of both growth and inflation forecasts is modest, even for OECD-countries – this is much more the case when looking at the spring forecasts for the coming year and even autumn forecasts for the current year are far from flawless (not addressed here – see Table 7)
- The error of forecasts in autumn for the coming year is about 1 percentage point. It is much higher for the spring forecasts for the coming year but still 0.3 for the autumn forecasts for the current year. The variance is considerable
- Accelerations and decelerations are mostly predicted but turning-points/recessions are not predicted and diagnosed only when countries are already in
- Though each decade had its crisis, the influence of the Great Recession on forecast accuracy was large
- Differences of countries‘ forecast accuracy are to a large degree the result of larger variances of growth/inflation

4 Summary and conclusions

- Though the results are difficult to compare with that of other studies but where possible, the differences seem to be small
- The differences between forecasters are small
- They do hardly reflect the institutions size or resources but the availability of a more actual data base
- „Home advantages“ or a home bias of forecasters were not detected

Improvements

- While there seems to be room for improvement for a small number of countries, however, looking at the relative accuracy (Figure 4) of their forecasts the potential for this appears to be limited
- Macro econometric model simulations (for Germany) suggest that this may also hold for many of the OECD-countries
- Some improvement will result from better forecasts for key economies like the United States and Germany (and China) (ignoring aggregation gains)
- Most will be gained by better forecasts of crisis/recessions, 60 years of macroeconomic forecasting keeps such expectations also low
- „The end of history“?