How far can we forecast?

Statistical tests of the predictive content

Jörg Breitung and Malte Knüppel

University of Cologne and
Deutsche Bundesbank

9. September 2017

Workshop on Forecasting
Deutsche Bundesbank, Frankfurt

The null hypothesis

- Assume that Y_t is stationary and ergodic
- Let $\widehat{Y}_{t+h|t}$ denote the forecast based on the information set $\mathcal{I}_t$
- The forecast is uninformative if

$$H_0: \quad var(\underbrace{Y_{t+h} - \widehat{Y}_{t+h|t}}_{e_{t+h|t}}) = var(\underbrace{Y_{t+h} - \mu}_{u_{t+h}})$$

- Since

$$\mathbb{E}(e^2_{t+h|t}) = \mathbb{E}[(Y_{t+h} - \mu) - (\widehat{Y}_{t+h|t} - \mu)]^2$$

$\Rightarrow$ sufficient (but not necessary) condition for an uninformative forecast is $\widehat{Y}_{t+h|t} = \mu$

- For rational forecasts with $\mathbb{E}(e_{t+h}|\widehat{Y}_{t+h|t}) = 0$ it follows that

$$\begin{aligned}\mathbb{E}(Y_{t+h} - \mu)(\widehat{Y}_{t+h|t} - \mu) &= \mathbb{E}(e_{t+h|t} + \widehat{Y}_{t+h|t} - \mu)(\widehat{Y}_{t+h|t} - \mu)\\ &= \mathbb{E}(\widehat{Y}_{t+h|t} - \mu)^2\end{aligned}$$

$\Rightarrow$ H_0 is equivalent to $cov(Y_{t+h}, \widehat{Y}_{t+h|t}) = 0$.

- **Maximum forecast horizon**

  There exists some h^* such that

$$H_0: \quad var(e_{t+h|t}) \geq var(u_{t+h}) \text{ for } h > h^*$$

  h^* is called the maximum forecast horizon

- Sequential test of H_0 for $h = 1, 2, \ldots, h_{\max}$. Stop when H_0 is not rejected for first time. Previous horizon is $\hat{h}^*$.
- Non-stationary variables:

$$Y_{t+h|t} = Y_t + \sum_{s=1}^{h} \Delta \widehat{Y}_{t+s|t}$$

$$e_{t+h|t} = \sum_{s=1}^{h} \Delta e_{t+s|t} = \sum_{s=1}^{h} (\Delta Y_{t+s} - \Delta \widehat{Y}_{t+s|t})$$

  $\Rightarrow$ Non-predictability of Y_{t+h} equivalent to non-predictability of ΔY_{t+s} for $s = 1, \ldots, h$

Earlier work

a) Theil's (1958) inequality coefficient:

$$U2(h) = \frac{\sqrt{\sum_{t=1}^{n}(Y_{t+h} - \widehat{Y}_{t+h|t})^2}}{\sqrt{\sum_{t=1}^{n}(Y_{t+h} - Y^0_{t+h})^2}}$$

where Y^0_{t+h} denotes some "naive forecast" (typically "no-change forecast") $\Rightarrow$ forecast uninformative if $U2(h) = 1$

b) Nelson (1976) or Granger-Newbold (1986) measure:

$$R^2(h) = 1 - \frac{var(e_{t+h|t})}{var(Y_{t+h})}$$

c) Diebold-Kilian (2001) forecastability measure:

$$Q(\mathcal{L}, h, k) = 1 - \frac{\mathbb{E}[\mathcal{L}(e_{t+h|t})]}{\mathbb{E}[\mathcal{L}(e_{t+k|t})]} \quad \text{where } k > h$$

- Note that for (i) stationary variables and (ii) MSE as the loss function:

$$\lim_{k\to\infty} Q(MSE, h, k) \ = \ R^2(h)$$

- Our approach is based on $R^2(h)$ (resp. MSE DIFF)
- We propose tests for the limiting horizon h^* beyond which forecasts become uninformative
- Empirical work suggests that economic forecasts of macroeconomic key variables (output growth, inflation) are informative 2-6 quarters ahead (or even less)
- Our empirical application based on survey forecasts from Consensus Economics indicates a maximum forecast horizon of typically less than one year

Maximum forecast horizons in quarters

	US	EA	JP	DE	UK	IT	CA	FR	median
GDP q-o-q									
DM-type test	2	2	1	1	3	1	1	2	1.5
encompassing test	2	2	1	2	3	5	1	4	2
CPI y-o-y									
DM-type test	3	5	3	2	2	3	4	3	3
encompassing test	3	3	4	3	3	3	4	3	3
PrivCons q-o-q									
DM-type test	3	1	0	0	−1	1	0	2	0.5
encompassing test	3	3	0	3	3	3	1	5	3
d(3m rate)									
DM-type test	1		0	3	2	2	1	2	2
encompassing test	2		0	2	6	3	1	2	2

Note: Regions considered are USA, Euro area, Japan, Germany, UK, Italy, Canada, and France. Variables are growth rates of real GDP, CPI, and real private consumption, and 1st differences of interest rates. $h = 0$ refers to the nowcast.

Notation and assumptions

- Forecast results from replacing θ by some estimator $\widehat{\theta}$ such that

$$\widehat{Y}_{t+h|t} = Y^{\widehat{\theta}}_{t+h|t}$$

- The forecast evaluation may be based on three different schemes:

recursive:	$\{-T+1, -T+2, \ldots, t\}$
rolling:	$\{t-T+1, t-T+2, \ldots, t\}$
fixed:	$\{t-T+1, \ldots, 0\}$

- We assume that we only observe actual values and forecasts but do not know (i) the forecasting model and (ii) the data used for estimating the model
- Forecast errors:

$$e_{t+h|t} = Y_{t+h} - Y^{\theta}_{t+h|t}$$

$$\widehat{e}_{t+h|t} = Y_{t+h} - \widehat{Y}_{t+h|t}$$

Assumption 1: (Time series process for Y_t)

Let $Y_t = \mu + u_t$ with $u_t = \phi(L)\varepsilon_t$, $\phi(L) = 1 + \phi_1 L + \phi_2 L^2 + \cdots$ is a lag polynomial with all roots outside the unit circle, $\sum_{i=1}^{\infty} |\phi_i| < \infty$ and ε_t is an i.i.d. white noise process with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$. Furthermore $\mathbb{E}|\varepsilon_t|^{2+\delta} < \infty$ for some $\delta > 0$.

Assumption 2: (Properties of the forecast)

(i) *Under H_0: $u_{t+h} = Y_{t+h} - \mu$ is independent of the past estimation error $\widehat{\theta}_t - \theta$, $\widehat{\theta}_{t-1} - \theta, \ldots$.*

(ii) *The parameters are estimated consistently with*

$$a) \quad \widehat{\theta}_0 - \theta \; = \; O_p(T^{-1/2}), \qquad b) \quad \widehat{\theta}_t - \widehat{\theta}_0 \; = \; O_p\left(\frac{\sqrt{t}}{T}\right)$$

(iii) *Let $D_{t+h}(\theta) = \partial Y_{t+h|t}^{\theta} / \partial\theta$ and $\overline{D}_h(\theta) = n^{-1} \sum_{t=1}^{n} D_{t+h}(\theta)$*

$$\frac{1}{n} \sum_{t=1}^{n} (D_{t+h}(\theta) - \overline{D}_h)^2 \; \stackrel{p}{\to} \; \overline{D}^2 \quad \text{with } 0 < \overline{D}^2 < \infty$$

$$\mathbb{E}|D_{t+h}(\theta) u_{t+h}|^{2+\delta} < \infty \quad \text{for some } \delta > 0 \text{ and all } t.$$

Diebold-Mariano-type test

Comparing model forecast with unconditional mean $\overline{Y}_h$:

$$\text{loss diff.} \quad \delta_t^h = \widehat{e}_{t+h|t}^2 - \left(Y_{t+h} - \overline{Y}_h\right)^2$$

DM (1995) statistic:

$$d_h = \frac{1}{\widehat{\omega}_\delta \sqrt{n}} \sum_{t=1}^{n} \delta_t^h \ ,$$

where $\widehat{\omega}_\delta^2$ denotes the estimated long-run variance of δ_t^h.

Theorem 1: *Asymptotic distribution of the DM statistic:*

If $T \to \infty$, $n \to \infty$, $n/T \to 0$ we have

$$d_h = \sqrt{n}\frac{|\overline{u}|}{2\widehat{\omega}_u} + O_p\left(\frac{n}{T}\right) \ \stackrel{d}{\to} \ \frac{|z|}{2} \ ,$$

where $z \sim \mathcal{N}(0,1)$.

$\Rightarrow$ non-standard as under H_0 the forecasts are *nested*

Modified DM statistic

Theorem 1 suggests the 2 adjusted DM statistics

$$2d_h \xrightarrow{d} |\mathcal{N}(0,1)|$$

$$\widetilde{d}_h = \frac{1}{\widehat{\omega}_u^2} \sum_{t=1}^{n} \delta_t^h \xrightarrow{d} \chi_1^2$$

where $\widehat{\omega}_u^2$ is a consistent estimator for the long-run variance of $u_t = y_t - \overline{y}$.

- 5% critical values are 0.0627 and 0.0039, resp. $\Rightarrow$ large size distortions
- Under $H_1 : 2d_h = O_p(\sqrt{n})$ and $\widetilde{d}_h = O_p(n)$. Nevertheless the **local power** is identical.
- If the model-based forecast is biased, the tests become conservative

Actual sizes for various n/T combinations ($\alpha = 0.05$)

	$n=25$		$n=50$		$n=100$		$n=200$	
T	$2d_1$	$\widetilde{d}_1$	$2d_1$	$\widetilde{d}_1$	$2d_1$	$\widetilde{d}_1$	$2d_1$	$\widetilde{d}_1$
50	0.089	0.094	0.066	0.070	0.044	0.047	0.027	0.029
100	0.105	0.110	0.089	0.093	0.065	0.069	0.043	0.046
200	0.114	0.121	0.105	0.111	0.088	0.093	0.065	0.069
500	0.116	0.123	0.115	0.122	0.106	0.112	0.094	0.099
1000	0.110	0.117	0.116	0.123	0.114	0.121	0.108	0.114
∞	0.049	0.049	0.049	0.049	0.050	0.050	0.051	0.051

Note: For $T = \infty$ the test statistics are computed using the true parameter values. Results are based on 100,000 replications.

Encompassing test

- Modifications based on the following decomposition:

$$\sum_{t=1}^{n} \delta_t^h = \sum_{t=1}^{n} \left[Y_{t+h} - \overline{Y}_h - (\widehat{Y}_{t+h|t} - \overline{Y}_h) \right]^2 - \left(Y_{t+h} - \overline{Y}_h \right)^2$$

$$= \underbrace{\sum_{t=1}^{n} (\widehat{Y}_{t+h|t} - \overline{Y}_h)^2}_{\text{always positive}} - 2 \sum_{t=1}^{n} (Y_{t+h} - \overline{Y}_h)(\widehat{Y}_{t+h|t} - \overline{Y}_h)$$

- First term does not contribute to power
- Reject H_0 if correlation between Y_{t+h} and $\widehat{Y}_{t+h|t}$ is large
- LM-type test statistic:

$$\varrho_h = \frac{1}{\sqrt{n}\,\widehat{\omega}_\xi} \sum_{t=1}^{n} \xi_t^h$$

where $\xi_t^h = (Y_{t+h} - \overline{Y}_h)(\widehat{Y}_{t+h|t} - \overline{\widehat{Y}}_h)$

and $\widehat{\omega}_\xi^2$ denotes the corresponding long-run variance

$$\widehat{\omega}_\xi^2 = \widehat{\gamma}_0^\xi + 2\sum_{j=1}^{k} w_j^k \widehat{\gamma}_j^\xi$$

$$\widehat{\gamma}_j^\xi = \frac{1}{n}\sum_{t=j+1}^{n} \xi_t \xi_{t-j} \ .$$

- Note that this test is asymptotically equivalent to the Mincer-Zarnowitz regression:

$$Y_{t+h} \ = \ \beta_{0,h} + \beta_{1,h}\widehat{Y}_{t+h|t} + u_{t+h}$$

but with $H_0 : \beta_{1,h} = 0$ instead of $\beta_{1,h} = 1$

- Encompassing test:

$$Y_{t+h} = \lambda \widehat{Y}_{t+h|t} + (1-\lambda)\overline{Y}_h + v_{t+h}$$
$$Y_{t+h} - \overline{Y}_h = \lambda(\widehat{Y}_{t+h|t} - \overline{Y}_h) + v_{t+h}$$

with $H_0 : \ \lambda = 0$

Theorem 2:

Under Assumption 1–2, a recursive forecasting scheme with $h > h^*$, $T \to \infty$, $n \to \infty$ and $n/T \to 0$ we have

$$\varrho_h \xrightarrow{d} \mathcal{N}(0,1)$$

- This might look trivial but is not. In the proof we show that

$$\widehat{Y}_{t+h|t} - \overline{Y}_h \approx (\widehat{\theta}_0 - \theta) D_{t+h}(\theta)$$

  and thus the regressor tends to zero as $\widehat{\theta} \xrightarrow{p} \theta$
- Asymptotically, the test is equivalent to the regression

$$Y_{t+h} = \beta^*_{0,h} + \beta^*_{1,h} D_{t+h}(\widehat{\theta}) + \eta_{t+h}$$

Local power

Assume that the target value is generated as

$$Y_{t+1} = \mu + \left(\frac{c}{\sqrt{n}}\right) X_t + u_{t+1}$$

such that

$$\begin{aligned} \text{regression forecast error:} &\quad u_{t+1} + O_p(n/\sqrt{T}) \\ \text{unconditional forecast error} &\quad u_{t+1} - \overline{u}_1 + (c/\sqrt{n})(X_t - \overline{X}) \end{aligned}$$

Theorem 4:
Under the sequence of alternatives $\beta = c/\sqrt{n}$, $X_t \sim iid(0, \sigma_x^2)$, *Assumptions 1 – 2 and* $n/\sqrt{T} \to 0$ *it follows that*

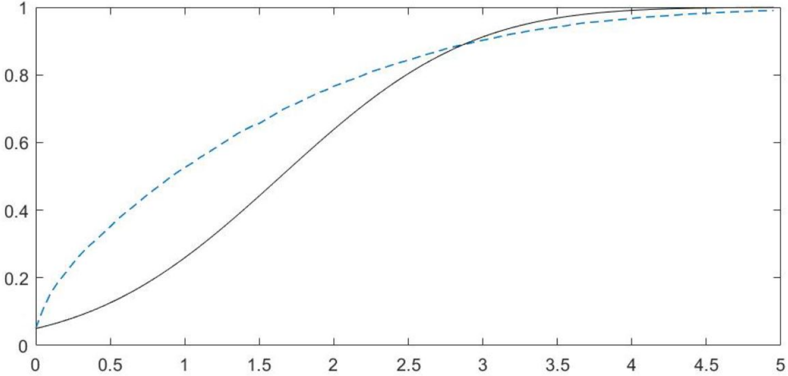
$$\widetilde{d}_1 \xrightarrow{d} z_1^2 - 2\lambda z_2 - \lambda^2$$

$$\widehat{\varrho}_1 \xrightarrow{d} sign(c) z_2 + \lambda$$

where $\lambda^2 = c^2\sigma_x^2/\sigma_u^2$ *is the signal-to-noise ratio and* z_1 *and* z_2 *are independent* $\mathcal{N}(0,1)$

$\Rightarrow$ tests are NOT asymptotically equivalent

Figure 1: Local power curves



Note: Broken line: DM-type test. Solid line: encompassing test

Cases considered for Monte Carlo simulations

case	DGP	forecast model
$MA(1)$-$AR(1)$	$y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$	$\widehat{y}_{t+h} = \widehat{\theta}_1^h + \widehat{\theta}_2^h y_t$
$MA(2)$-$AR(1)$	$y_t = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.3\varepsilon_{t-2}$	$\widehat{y}_{t+h} = \widehat{\theta}_1^h + \widehat{\theta}_2^h y_t$
$AR(1)$-$AR(1)$	$y_t = 0.8y_{t-1} + \varepsilon_t$	$\widehat{y}_{t+h} = \widehat{\theta}_1^h + \widehat{\theta}_2^h y_t$
multivar. 1	$y_t = 0.5x_{t-1} + \varepsilon_t$	$\widehat{y}_{t+h} = \widehat{\theta}_1^h + \widehat{\theta}_2^h x_t$
multivar. 2	$y_t = 0.5x_{t-1} + 0.3x_{t-2} + \varepsilon_t$	$\widehat{y}_{t+h} = \widehat{\theta}_1^h + \widehat{\theta}_2^h x_t + \widehat{\theta}_3^h x_{t-1}$

Results for case 'MA(1)-AR(1)'

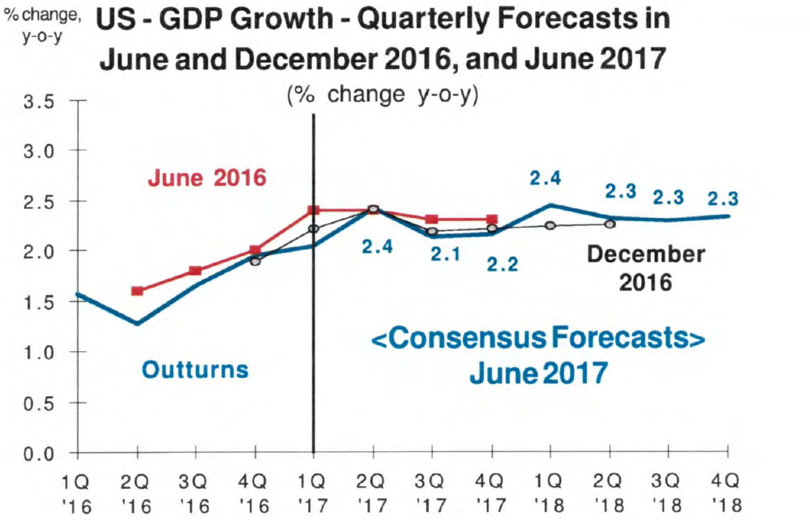
forecast horizon h	0	1	2	3	4	0	1	2	3	4
		$T = 100$, $n = 50$					$T = 100$, $n = 100$			
MSPE / Variance		0.89	1.05	1.05	1.05		0.87	1.03	1.03	1.03
rejections										
DM-type tests										
$\widetilde{d}_h$		0.87	0.10	0.09	0.10		0.98	0.08	0.08	0.08
$2d_h$		0.88	0.10	0.10	0.10		0.98	0.08	0.08	0.08
encomp. tests										
$\beta_{\mathbf{1},h}$		0.92	0.06	0.05	0.05		1.00	0.04	0.04	0.03
ϱ_h		0.81	0.03	0.02	0.02		0.99	0.03	0.02	0.02
classic DM test		0.26	0.00	0.00	0.00		0.57	0.00	0.00	0.00
$\hat{h}^*$										
DM-type tests										
$\widetilde{d}_h$	0.13	**0.79**	0.05	0.02	0.01	0.02	**0.90**	0.05	0.02	0.00
$2d_h$	0.12	**0.79**	0.06	0.02	0.01	0.02	**0.90**	0.06	0.02	0.00
encomp. tests										
$\beta_{\mathbf{1},h}$	0.08	**0.87**	0.04	0.01	0.00	0.00	**0.96**	0.03	0.01	0.00
ϱ_h	0.19	**0.80**	0.01	0.00	0.00	0.01	**0.97**	0.02	0.00	0.00
classic DM test	0.74	**0.26**	0.00	0.00	0.00	0.43	**0.57**	0.00	0.00	0.00

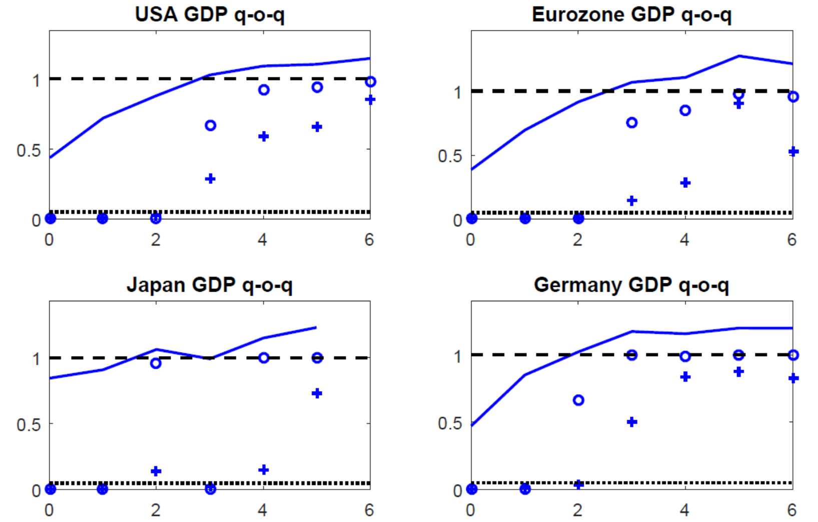
Note: Values displayed in category 'rejections' denote percentage of rejections for each horizon h, values displayed in category '$\hat{h}^*$' denote percentage of cases in which h is identified as maximum forecast horizon. Bold entries refer to the true h^*. If test rejects for all horizons, $\hat{h}^*$ is set equal to $h = 4$.

Results for (most) remaining cases

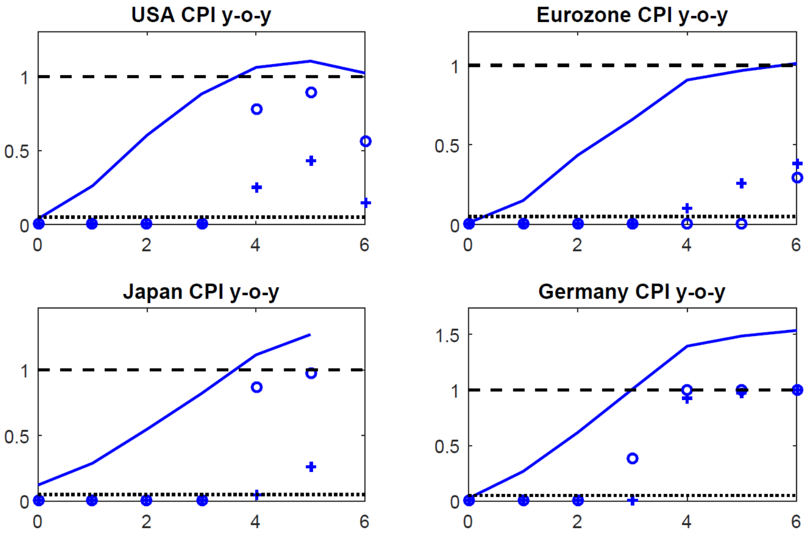
forecast horizon h	0	1	2	3	4	0	1	2	3	4
			$T=100$, $n=50$					$T=100$, $n=100$		
						MA(2)-*AR*(1)				
MSPE / Variance		0.82	1.01	1.07	1.07		0.79	0.99	1.04	1.04
rejections										
DM-type test		0.91	0.39	0.10	0.10		0.99	0.58	0.09	0.09
encomp. test		0.95	0.34	0.07	0.06		1.00	0.55	0.05	0.05
$\hat{h}^*$										
DM-type test	0.09	0.53	**0.35**	0.02	0.02	0.01	0.41	**0.54**	0.02	0.02
encomp. test	0.05	0.62	**0.31**	0.01	0.01	0.00	0.45	**0.53**	0.01	0.01
						AR(1)-*AR*(1)				
MSPE / Variance		0.44	0.72	0.90	1.01		0.40	0.65	0.82	0.92
rejections										
DM-type test		0.99	0.85	0.62	0.42		1.00	0.98	0.86	0.68
encomp. test		1.00	0.95	0.76	0.50		1.00	1.00	0.95	0.78
$\hat{h}^*$										
DM-type test	0.01	0.14	0.23	0.20	**0.42**	0.00	0.02	0.12	0.19	**0.68**
encomp. test	0.00	0.05	0.20	0.26	**0.50**	0.00	0.00	0.05	0.18	**0.77**
						multivar. 1				
MSPE / Variance		0.83	1.04	1.04	1.04		0.82	1.02	1.02	1.02
rejections										
DM-type test		0.93	0.09	0.09	0.09		0.99	0.07	0.07	0.06
encomp. test		0.95	0.03	0.03	0.03		1.00	0.02	0.02	0.02
$\hat{h}^*$										
DM-type test	0.07	**0.84**	0.07	0.01	0.00	0.01	**0.93**	0.06	0.01	0.00
encomp. test	0.05	**0.93**	0.03	0.00	0.00	0.00	**0.98**	0.02	0.00	0.00

Note: The DM-type test uses $\widetilde{d}_h$, the encompassing test employs $\beta_{\mathbf{1},h}$.


% change, y-o-y
US - GDP Growth - Quarterly Forecasts in June and December 2016, and June 2017
(% change y-o-y)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
June 2016
2.4
2.3
2.3
2.3
2.4
2.1
2.2
December 2016
Outturns
<Consensus Forecasts> June 2017
1Q '16
2Q '16
3Q '16
4Q '16
1Q '17
2Q '17
3Q '17
4Q '17
1Q '18
2Q '18
3Q '18
4Q '18



— MSPE variance ratio , + encompassing test , ○ DM-type test


USA CPI y-o-y
Eurozone CPI y-o-y
Japan CPI y-o-y
Germany CPI y-o-y
0
0.5
1
1.5
0
2
4
6

Conclusions

- Test for predictability at horizon h
- Determine h^*: maximum forecast horizon
- Problem: comparison of nested forecasts
- DM-type test and encompassing test
- Encompassing test outperforms the DM-type test
- We found h^* between 1 and 5 quarters for macroeconomic key variables
- Extension to any comparison of nested forecast comparisons?