Forecasting

Claudia Foroni
Norges Bank

17 November 2014

An "optimality theory" for forecasting

- A forecast is any statement about the future.
- An "optimal" forecast requires that:
  1. the empirical model is a good representation of the economy;
  2. the structure of the economy will remain unchanged.

General properties

- An in-sample congruent encompassing model should dominate in forecasting.
- Forecast accuracy should decline as the forecast horizon increases, because more innovation errors accrue and so predictability falls.
- While adding causally relevant variables should improve forecasts, adding irrelevant variables should generally worsen forecasts from unnecessary parameter-estimation uncertainty.

Forecasting from an AR model

- Assumptions:
  - The AR model holds for $t = 1, 2, ..., T, T + 1, ..., T + H$.
  - We know the parameters of the process (we abstract from the estimation problem).
  - We know the history of the time series y_t without measurement error (we abstract from the real-time problem).

Loss function

- The forecasting **loss function is quadratic**, i.e. we are interested in minimizing the **mean of squared forecast errors** (MSFE):

$$MSFE = \frac{1}{H}\sum_{h=1}^{H}\left(y_{T+h} - \widehat{y}_{T+H|T}\right)^2 . \qquad (1)$$

- **The optimal forecasts are the conditional expectations based on the period T information set** $(\mathcal{I}_T)$:

$$\widehat{y}_{T+H|T} = E\left(y_{T+h} \,|\mathcal{I}_T\right) \qquad (2)$$

Forecasting an AR(1) process I

- 1-step-ahead:

$$\begin{aligned} y_{t+1} &= \mu + \theta y_t + \varepsilon_{t+1} && (3) \\ E\left(y_{t+1} \,|X_t\right) &= \mu + \theta y_t && (4) \end{aligned}$$

- 2-step-ahead:

$$\begin{aligned} y_{t+2} &= \mu + \theta y_{t+1} + \varepsilon_{t+2} && (5) \\ &= \mu + \theta\left(\mu + \theta y_t + \varepsilon_{t+1}\right) + \varepsilon_{t+2} \\ &= \left(\mu + \theta\mu\right) + \theta^2 y_t + \theta\varepsilon_{t+1} + \varepsilon_{t+2} \\ E\left(y_{t+2} \,|\mathcal{I}_t\right) &= \left(\mu + \theta\mu\right) + \theta^2 y_t && (6) \end{aligned}$$

Forecasting an AR(1) process II

- **Note:** this is a "recursive" approach: equivalent to **(only for linear models!)**

$$\begin{aligned} y_{t+2} &= \mu + \theta y_{t+1} + \varepsilon_{t+2} && (7) \\ E\left(y_{t+2} \,|\mathcal{I}_t\right) &= \mu + \theta \widehat{y}_{t+1} && (8) \\ &= \mu + \theta\left(\mu + \theta y_t\right) \end{aligned}$$

- h-step-ahead:

$$\begin{aligned} y_{t+h} &= \mu + \theta y_{t+h-1} + \varepsilon_{t+h} && (9) \\ &= \left(\mu + \theta\mu + \ldots + \theta^{h-1}\right) + \theta^h y_t \\ &\quad +\theta^{h-1}\varepsilon_{t+1} + \ldots + \varepsilon_{t+h} \\ E\left(y_{t+h} \,|\mathcal{I}_t\right) &= \left(\mu + \theta\mu + \ldots + \theta^{h-1}\right) + \theta^h y_t && (10) \end{aligned}$$

Forecasting an AR(1) process III

- The above formulas hold even in the case of $\theta = 1$ (non-stationarity, Random Walk).
- However, for the stationary case we have that when $h \longrightarrow \infty$

$$E\left(y_{t+h} \left| \mathcal{I}_t \right.\right) \rightarrow \frac{\mu}{1-\theta} = E\left(y_t\right) \tag{11}$$

which is the unconditional mean of the process.

Forecasting an AR(2) process I

- 1-step-ahead:

$$
\begin{aligned}
y_{t+1} &= \mu + \theta_1 y_t + \theta_2 y_{t-1} + \varepsilon_{t+1} && (12)\\
E\left(y_{t+1} \,|X_t\right) &= \mu + \theta_1 y_t + \theta_2 y_{t-1} && (13)
\end{aligned}
$$

- 2-step-ahead:

$$
\begin{aligned}
y_{t+2} &= \mu + \theta_1 y_{t+1} + \theta_2 y_t + \varepsilon_{t+2} && (14)\\
E\left(y_{t+2} \,|X_t\right) &= \mu + \theta_1 \widehat{y}_{t+1} + \theta_2 y_t && (15)\\
&= \mu + \theta_1 \left(\mu + \theta_1 y_t + +\theta_2 y_{t-1}\right) + \theta_2 y_t
\end{aligned}
$$

Evaluate the forecasts I

- Forecast error:

$$\varepsilon_{t+h|t} = y_{t+h} - y_{t+h|t} \tag{16}$$

- Forecasts are unbiased:

$$E\left(y_{t+h} - y_{t+h|t}\right) = 0 \text{ for all } h \tag{17}$$

- Variance of the forecast error:

$$Var\left(\varepsilon_{t+h|t}\right) = Var\left(y_{t+h} - y_{t+h|t}\right) \underset{H\rightarrow\infty}{\longrightarrow} Var\left(y_t\right) \tag{18}$$

- The variance of the forecast error is used in constructing prediction confidence intervals.
- The prediction intervals typically increase with the forecast horizon, but they stabilize as the horizon gets longer.

Evaluate the forecasts: AR(1) example I

- Forecast error:

$$y_{t+h} = \theta^h y_{t+h} + \varepsilon_{t+h} + \theta\varepsilon_{t+h-1} + \ldots + \theta^{h-1}\varepsilon_{t+1} \quad (19)$$

$$y_{t+h|t} = \theta^h y_{t+h} \quad (20)$$

$$\varepsilon_{t+h|t} = \varepsilon_{t+h} + \theta\varepsilon_{t+h-1} + \ldots + \theta^{h-1}\varepsilon_{t+1} \quad (21)$$

- Variance of the forecast error:

$$Var\left(\varepsilon_{t+h|t}\right) = \left(1 + \theta^2 + \ldots + \theta^{2(h-1)}\right)\sigma_\varepsilon^2 \quad (22)$$

Failures in forecasting

- In practice:
  1. We do not know the structure of the true model.
  2. We do not know the parameter values.
  3. Data are revised and measured with errors.
  4. The are structural breaks in the forecast period.
- 1. and 2. can be solved by methodology.
- 3. is nuisance (it matters mainly at short horizons).
- 4. is the main problem!

Forecasting bias due to structural breaks I

- We look again at a stationary AR(1) example.
- Let us first write the model in equilibrium form (remember: the mean of the process is $\frac{\mu}{1-\theta}$):

$$y_t = \mu + \theta y_{t-1} + u_t, \quad (23)$$

$$y_t = \frac{\mu}{1-\theta} + \theta \left(y_{t-1} - \frac{\mu}{1-\theta} \right) + u_t, \quad (24)$$

$$y_t = y^* + \theta \left(y_{t-1} - y^* \right) + u_t. \quad (25)$$

Forecasting bias due to structural breaks II

- We assume that there is a structural break: $E(y_t)$ changes from y^* to $y^* + d$.
- The true conditional expectation is:

$$E\left(y_{T+h|T}\right) = y^* + d + \theta^h \left(y_T - \left(y^* + d\right)\right). \tag{26}$$

- Our forecast will be

$$E\left(\widehat{y}_{T+h|T}\right) = y^* + \theta^h \left(y_T - y^*\right). \tag{27}$$

- The forecast error will then be:

$$E\left(y_{T+h|T} - \widehat{y}_{T+h|T}\right) = \left(1 - \theta^h\right) d \neq 0. \tag{28}$$

- There is a systematic **bias**.

Forecasting bias due to structural breaks III

- The forecast corrects towards the wrong equilibrium.
- After one period has passed, we can produce a new forecast that condition on y_{T+1}, where the break appears.
- However, it will still not be enough to get the correct forecast.
- Model based forecasts are not good at adapting to structural breaks.

Parameters of interest when forecasting

- When we estimate a model for policy purposes or for testing a hypothesis, the parameters of interest are the derivative coefficients.
- When the purpose is forecasting, the parameters of interest are the conditional means of the endogenous variables.
- There can be breaks in the conditional means and not in the derivative coefficients, e.g. a break in the constant.
- This means that a model can be good for policy analysis, even though it forecasts badly.

Forecasting a time series model in practice I

Example: *AR(1) process. Sample: 1990m1 - 2010m12.*
1-step-ahead recursive forecast

- Divide the sample in estimation sample (1990m1-1999m12) and evaluation sample (2000m1-2010m12)
- Recurive forecasting exercise:
  1. Estimate the AR(1) on the sample 1990m1-1999m12 $\Longrightarrow$ obtain $\widehat{\theta}^{(1)}, \widehat{\sigma}^{(1)}$
  2. With $\widehat{\theta}^{(1)}, \widehat{\sigma}^{(1)}$ and $y_{1999m12}$ forecast $\widehat{y}_{2000m1}$
  3. Estimate the AR(1) on the sample 1990m1-2000m1 $\Longrightarrow$ obtain $\widehat{\theta}^{(2)}, \widehat{\sigma}^{(2)}$
  4. With $\widehat{\theta}^{(2)}, \widehat{\sigma}^{(2)}$ and y_{2000m1} forecast $\widehat{y}_{2000m2}$
  5. ...

Forecasting a time series model in practice II

Example: *AR(1) process. Sample: 1990m1 - 2010m12.*
2-step-ahead recursive forecast

- Divide the sample in estimation sample (1990m1-1999m12) and evaluation sample (2000m1-2010m12)
- Recurive forecasting exercise:
  1. Estimate the AR(1) on the sample 1990m1-1999m12 $\Longrightarrow$ obtain $\widehat{\theta}^{(1)}, \widehat{\sigma}^{(1)}$
  2. With $\widehat{\theta}^{(1)}, \widehat{\sigma}^{(1)}$ and $y_{1999m12}$ forecast $\widehat{y}_{2000m1}$
  3. With $\widehat{\theta}^{(1)}, \widehat{\sigma}^{(1)}$ and $\widehat{y}_{2000m1}$forecast $\widehat{y}_{2000m2}$
  4. Estimate the AR(1) on the sample 1990m1-2000m1 $\Longrightarrow$ obtain $\widehat{\theta}^{(2)}, \widehat{\sigma}^{(2)}$
  5. ...

How to compute the MSE in practice

- Collect all the h-step ahead forecasts you obtain in a vector: $\{\widehat{y}_{2000m1}, \ldots, \widehat{y}_{2010m12}\}$
- Compute the MSE as

$$\frac{\sum_{t=2000m1}^{2010m12} \left(\widehat{y}_t - y_t\right)^2}{N} \tag{29}$$

where N is the number of recursive samples (e.g. "months") in your forecasting sample.

Forecasts from a Keynesian type macro model

Example: Medium term macro model:

$$C_t = a + b\left(GDP_t - TAX_t\right) + cC_{t-1} + \varepsilon_{Ct} \tag{30}$$

$$TAX_t = d + eGDP_t + \varepsilon_{TAXt} \tag{31}$$

$$GDP_t = C_t + I_t \tag{32}$$

$$I_t = \mu_I + \varepsilon_{It} \tag{33}$$

- C_t, GDP_t and TAX_t are endogenous. $C_{t-1} = C_{t-1}$ is predetermined.
- I_t is strictly exogenous with $E\left(I_t\right) = \mu_I$ and $Var\left(I_t\right) = \sigma_I^2$.
- The error terms are independent.

Reduced form

- Our purpose is to forecast $C_{T+1}, C_{T+2}, ..., C_{T+H}$ based on data up to period T.
- We generate our forecasts from the reduced form equation for C_t :

$$C_t = \underbrace{\frac{a + bd}{1 - b(1-e)}}_{\beta_0} + \underbrace{\frac{b(1-e)}{1 - b(1-e)}}_{\beta_1} I_t + \underbrace{\frac{c}{1 - b(1-e)}}_{\beta_2} C_t + \underbrace{\frac{\varepsilon_{Ct} - be\varepsilon_{TAXt}}{1 - b(1-e)}}_{\varepsilon_t}. \quad (34)$$

- The model is an ARDL, with the reduced form coefficients β_0, β_1 and $\beta_2 = \beta_2$ estimated by OLS from a sample $t = 1, 2, ..., T$.

1-step ahead forecast

- The 1-step ahead forecast for C_{T+1} is:

$$E\left(C_{T+1} \mid C_T, I_T\right) = \beta_0 + \beta_1 E\left(I_{T+1} \mid C_T, I_T\right) + \beta_2 C_T. \quad (35)$$

- We need to compute the forecast for I_{T+1} (we need the forecast also for exogenous variables!):

$$E\left(I_{T+1} \mid C_T, I_T\right) = \mu_I. \quad (36)$$

- Therefore the forecast for C_{T+1} is:

$$E\left(C_{T+1} \mid C_T, I_T\right) = \beta_0 + \beta_1 \mu_I + \beta_2 C_T. \quad (37)$$

- The parameters in the forecast are unknown: typically we replace them with the parameter estimates:

$$E\left(\widehat{C}_{T+1} \mid C_T, I_T\right) = \widehat{\beta}_0 + \widehat{\beta}_1 \widehat{\mu}_I + \widehat{\beta}_2 C_T. \quad (38)$$

h-step ahead forecast

- The h-step ahead forecast for C_{T+1} is:

$$E\left(\widehat{C}_{T+h}\,|\,C_T, I_T\right) = \sum_{j=0}^{h-1}\widehat{\beta}_2^j\left(\widehat{\beta}_0 + \widehat{\beta}_1\widehat{\mu}_I\right) + \widehat{\beta}_2^h C_T, \quad (39)$$

where

$$\widehat{\gamma} = \left(\widehat{\beta}_0 + \widehat{\beta}_1\widehat{\mu}_I\right) \quad (40)$$

- For $h \rightarrow \infty$

$$E\left(\widehat{C}_{T+h}\,|\,C_T, I_T\right) = \frac{\widehat{\beta}_0 + \widehat{\beta}_1\widehat{\mu}_I}{1-\widehat{\beta}_2} \quad (41)$$

Forecast errors: Bias

- Forecast error:

$$C_{T+h} - \widehat{C}_{T+h} = (\gamma - \widehat{\gamma}) \sum_{j=0}^{h-1} \widehat{\beta}_2^j + \left(\beta_2^h - \widehat{\beta}_2^h \right) C_T + \sum_{j=0}^{h-1} \beta_2^j \varepsilon_{t+j}. \tag{42}$$

- Bias:

$$E\left(C_{T+h} - \widehat{C}_{T+h} \right) = \sum_{j=0}^{h-1} E\left[(\gamma - \widehat{\gamma}) \widehat{\beta}_2^j \right] + C_T E\left(\beta_2^h - \widehat{\beta}_2^h \right). \tag{43}$$

- The bias should be small if the estimation sample is long enough.

Forecast errors: Variance

- Variance:

$$Var\left(C_{T+h}-\widehat{C}_{T+h}\right)=\underbrace{Var\left(\sum_{j=0}^{h-1}\left[(\gamma-\widehat{\gamma})\,\widehat{\beta}_2^{j}\right]\right)}_{\text{estimation uncertainty}}+\sigma^2\sum_{j=0}^{h-1}\beta_2^{2j}. \tag{44}$$

- The estimation uncertainty should be small if the estimation sample is long enough.
- For $h\rightarrow\infty$

$$Var\left(C_{T+h}-\widehat{C}_{T+h}\right)=Var\left(C_t\right)=\frac{\sigma^2}{1-\beta_2^2}, \tag{45}$$

that is we expect that the variance of the forecast error converge to the theoretical variance of the forecasted variable.