ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Forecasting ***(a brief introduction)***

Andrea Guizzardi
Department of Statistical Sciences
University of Bologna

Sources:

Diebold (200X), Elements of forecasting, South Western ed.

- Cap.2 Six consideration basic to succesful forecasting
- Cap 4. Modelling and forecasting trend
- Cap 12 Units roots, Stocastic trends, ARIMA Forecasting Models and Smoothing

Guizzardi A. (2002) « La previsione Economica » Guaraldi Ed. Rimini
Capp.2,6,12

Blaskowitz, O; Herwartz, H (2008) Testing directional forecast value in the presence of serial correlation, SFB 649 Discussion Paper 2008-073

What does forecasting means?

(The forecast statements)

Pr(Y∈A)=? (unconditional with A a fixed interval)

Pr(Y∈A|I)=? (conditional to the information set I)

Measure of this probability via P.d.F. **(**Some examples….)

This is also known as **probability forecast**; frequently used in the scenario analysis.

Practitioners are generally more interested in a point, not in a probability; (they have to take decisions).

Pr(Y∈?)=(1-α) (unconditional, with α the probability of forecasting failure)

Pr(Y∈?|I)=(1-α) (conditional to the information set I)

What does forecasting means?
(The forecast statements)

Predictor: a function of observation and parameters

Point forecast: the value assumed by predictor given specific sample and parameters. It is a random variable when the **HORIZON** is greater than 0.

Horizon is definend as the number of periods between today and the date of the forecast we make. If we are in 2014 (having annual data) the 2016 forecast has a forecast horizon of 2 steps. In general we speak of an h-step ahead (**static**) forecast.

h-step ahead forecast can also be **dynamic** (extrapolation forecasts) when previous forecasted values are used to forecast step by step along a fixed horizon.

The way to measure point forecast uncertain is to calculate confidence intervals

Interval forecast: The set of value "around" a point forecast that - with prob. equal to 100(1-α)% - will contain the future value of Y.

What does forecasting means?
(The forecast statements)

Problem: which P.d.F. to probabilize intervals?

If we could assume normality: → we need to forecast only conditional mean and conditional variance to identify a **density (forecast) function**

In more complex empirical setting:

- Quantile estimation via residual analysis
- Calibration of intervals based on dispersion of point (subjective forecasts) (pay attention to the dynamics in the variance reduction of the experts' subsequent opinions).

The forecasting error: $$e_{T+H} = y_{T+H} - \hat{y}_{T+H}$$

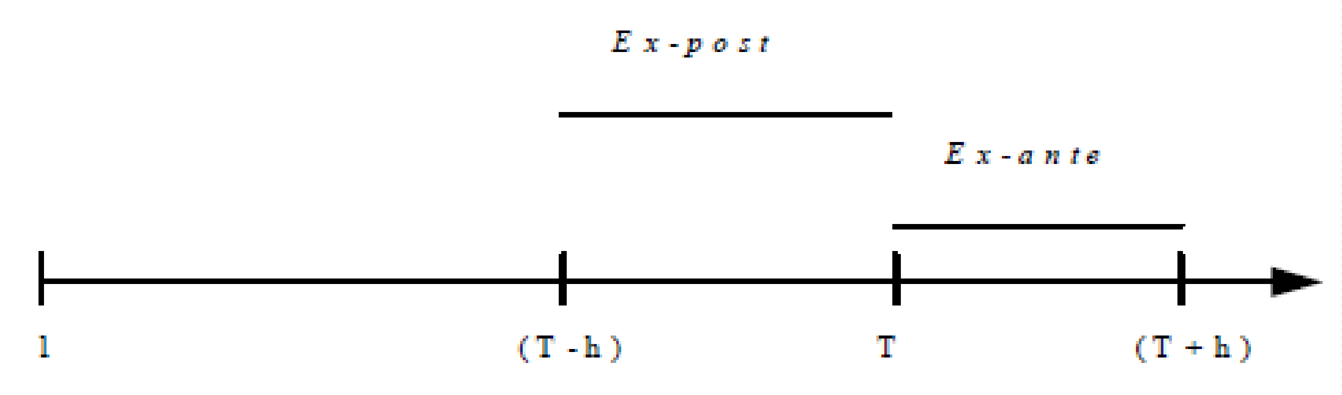
- is not the best way to evaluate a forecast. There are always different decision environment, i.e. loss - or a cost – function.

Ex ante and ex post forecast
(The forecast statements)

In latin “ex-post” means "after the event‟ while “ex-ante”, means "before the event".

Ex post forecast is a forecast that uses information beyond the time at which the actual forecast is prepared

Ex ante forecast is a forecast that uses information available only by time T (the time of the actual forecast. Is the only genuine forecast of the future



Ex ante and ex post forecast

(The forecast statements)

Ex post forecast are used to evaluate and compare rival forecasting model performances given the assumptions that:

- The future will repeat the past (DGP will not change in the future)
- The more a model is accurate in predicting the more it is correctly specified (includes and utilizes effectively the information necessary to explain the dynamics of the phenomenon under observation)

Ex post forecasts - for which more information is available, should be more accurate than ex ante forecasts.

A comparative evaluation of the two types of forecasts can be used to find out what actually causes higher forecast errors:

- bad forecasting model
- bad forecast of the variables used in the model.

The optimal predictor
(The loss function)

- 3 characteristics of a loss function (mandatory):
  1. no intercept: *if e=0: C(0) = 0;*
  2. positive function: *C(e) = 0;*
  3. monotone function in the two quadrants.
- Optional characteristics.
  - Simmetry *C(e) = C(-e);*
  - Homogeneity for some function Q(a) *C(ae)=Q(a)C(e)*
  - Differentiability (continuous function, that is nearly identical forecast errors should produce nearly identical losess)

- "Universal" function: Square function $C(e) = e^2 \rightarrow FMSE = \frac{1}{H}\sum_{h=1}^{H}\left(y_{t+h} - \hat{y}_{t+h}\right)^2$

  Our focus will be to minimize the FMSE (forecasting mean square error) → expected value of the forecasting squared error.

- In case of square cost function the conditional expected value is always the optimal predictors.

Examples
(loss function *(examples)*)

- Others loss function examples….

Absolute function $C(e)=|e|$ $MAE = \frac{1}{H}\sum_{h=1}^{H}|y_{t+h}-\hat{y}_{t+h}|$

Absolute relative function $MAPE = \frac{1}{H}\sum_{h=1}^{H}\left|\frac{(y_{t+h}-\hat{y}_{t+h})}{y_{t+h}}\right|$

Subjective function $C(e)=\ln(e+1) \quad e \geq 0$ (stored products)

$C(e)=2\cdot|e| \quad e<0$

- Scaled (relative) loss functions

Mean Absolute Scaled Error
(the mean of ex-post errors, scaled by the *in-sample* MAE from the naïve (random walk) forecast method.)

$$\frac{1}{H}\left[\sum_{h=1}^{H}\frac{e_h}{\frac{1}{T-H-1}\sum_{t=2}^{T-H-1}|y_t-y_{t-1}|}\right]$$

Examples
(loss function in classification *(particulary useful financial markets)*)

- Exact classification rate $$ECR = \frac{1}{H}\left(\sum_{h=1}^{H} D_h\right)$$

- Trend accuracy (direction of change)

$$TA = \frac{1}{H}\sum_{h=1}^{H} sign\left[\left(y_t - y_{t-1}\right)\cdot\left(\hat{y}_t - \hat{y}_{t-1}\right)\right]$$

- Confusion Matrix

	Positive outcome	**Negative outcome**
Positive forecast	True Positive	False Positive
Negative forecast	False Negative	True Negative

Useful for asymetrics loss function
It generalize trend accuracy measure as: *TA=(TP+TN)/H*

is the number of **incorrect** predictions that an instance is **positive**

The optimal predictor

Optimal predictor: $$\int C\left[\left(y_{t+1} - y_t \mid I_{t+1}\right)\right] \cdot D\left(y_{t+1} \mid I_t\right) dy_{t+1} = \min \quad \rightarrow$$

$$\int C'\left[\left(y_{t+1} - y_t \mid I_{t+1}\right)\right] \cdot D\left(y_{t+1} \mid I_t\right) dy_{t+1} = 0$$

- In case of square cost function the conditional expected value is always the optimal predictors.

Methods and complexity
(The parsimony principle)

- One can guess that to model real (complex) world phenomena are necessary complex specification…
- … decades of professional experience suggest just the opposite. Hence the parsimony principle: “other things the same, simple models are preferable to complex models”.

- There is a trade-off between model fitting and forecasting accuracy

Y=signal+error

- With #parameters/#data → 0

  very low fit. Simplest models are even unusefull in forecasting because they give no signal description)
- With #data = #parameters

  perfect fit: $$R^2 = 1 - \left(\sum_{t=1}^{T} e^2 \Big/ \sum_{t=1}^{T} (y_t - \bar{y}) \right) = 1$$ of both signal and error

  (but less usefull in forecasting because error will not be the same in the future).

Methods and complexity
(The parsimony principle)

- FMSE can be seen as the expected value of squared error (future error is a random variable)

$$C(e) \equiv FMSE = E\left(e^2\right)$$

- Following the usual sum of square decomposition:

$$E\left(e^2\right) = \left[E(e)\right]^2 + E\left[\left(e - E(e)\right)^2\right] = bias(e) + Var(e)$$

is demonstrated that FMSE is equal to the sum of forecasts bias (squared), and the forecasting error variance.

- **Trade-off:** simple model are expected to show low forecasting variance of errors (at least equal to 0) but a high bias. Complex model are expected to show e low bias but a high error variance
- The compromise between model complexity and #data, depend from:
  - variance of errors (variance of Y)
  - available information set
- Practitioner suggestion: never over 1/3
- The **K**eep **i**t **S**ophisticatedly **S**imple principle (shrinkage principle) →

imposing restriction of forecasting models often improve forecasting performance. If restriction are on the number of parameters we get the parsimony principle

Methods and complexity
(The parsimony principle)

- During specification is better to select (forecasting) models not only looking at fitting indexes.
- To avoid overfitting problem it could be useful penalize a fitting criterion with “model complexity (a function of the number of parameters).

- **Akaike Information Criterion (AIC):** $AIC = 2k - 2\ln(L)$

  with L the maximum vale of the likelihood function; *k= #parameters*
- **Schwarz (Bayesian) Information Criterion (BIC):**

  $$BIC = \ln(n)k - 2\ln(L)$$

  With n #data points

The min value is better (for each criterion)

Penalyzing the (in sample) MSE allows more accurate estimation of the 1-step ahead FMSE.

Methods and complexity
(The parsimony principle)

- **Consistency** helps to evaluate a model selection criterion (the penalty factor). **Suppose that the generating model is of a finite dimension**. A consistent criterion will asymptotically select the fitted candidate model having the correct structure with probability one.
  - *"When the true model (the DGP) is among the alternative considered models, the probability of selecting the true DGP approaches one as the sample size gets large."*
  - *When the true model (the DGP) is not among the alternative considered models, the probability of selecting the best approximation to the true DGP approaches one as the sample size gets large."*

- MSE is inconsistent because it doesn't penalize for degree of freedom.
- AIC penalize for D. of F. but not enough to avoid large model selection when the sample size gets large.
- **SIC is consistent.**

Methods and complexity

(The parsimony principle)

- **If the true DGP** (or the best approximation to it) are much more complicated than any fitted model (**is of an infinite dimension**), THEN we need to consider **asymptotic efficiency** instead of consistency.

- Suppose that **the generating model is of an infinite dimension**, and therefore lies outside of the candidate collection under consideration. An asymptotically efficient criterion will asymptotically select the fitted candidate model whose 1-step ahead forecast error variances approach the one that would be obtained using the true model with known parameters, at a rate at least as fast as that of any other model selection criterion.

- AIC is asymptotical efficent
- SIC is not growth asymptotical efficent