Lesson 16: Forecasting Stationary Time Series

Umberto Triacca

Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica
Università dell'Aquila,
umberto.triacca@univaq.it

There are two main goals to record and to analyze the data of a time series:

1. to understand the structure of the time series
2. to predict future values of the time series

In this lesson, we consider the second goal:

to predict future values of a time series

h-step-ahead prediction of a stationary process

Let $\{x_t\}$ be a stationary process. Let $h > 0$ be an integer.

We consider the problem of predicting the value x_{T+h} in terms of the values

$$\Omega_T = \{x_T, x_{T-1}...\}$$

up time T.

h-step-ahead prediction of a stationary process

Let us denote this h-step-ahead forecast at time T by $\hat{x}_{T,h}$.

Any function of the random variables $x_T, x_{T-1}...$, can be considered like an h-step-ahead prediction.

Which is the best?

h-step-ahead prediction of a stationary process

The accuracy of forecasts is evaluated by some loss function, $L(x_{T+h}, \hat{x}_{T,h})$, that represents the penality or cost when we predict $\hat{x}_{T,h}$ but the outcome is x_{T+h}.

A typical loss function is the **Mean Square Error** (**MSE**) defined by

$$\mathrm{MSE}(\hat{x}_{T,h}) = E[(x_{t+h} - \hat{x}_{T,h})^2]$$

We suppose that the forecaster's objective is to minimize the MSE.

h-step-ahead prediction of a stationary process

It can be shown that the optimal (minimum MSE) h-step-ahead forecast of x_{T+h} at time T is the **conditional expectation**

$$E\left[x_{T+h}|\Omega_T\right]$$

h-step-ahead prediction of a stationary process

In other terms, if $\hat{x}_{T,h}$ is any h-step predictor at time T, we have that

$$\text{MSE}(\hat{x}_{T,h}) \geq \text{MSE}(E\left[x_{T+h}|\Omega_T\right])$$

Forecasting ARMA processes

Consider a causal ARMA(p, q) process

$$x_t = \nu + \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + u_t + \theta_1 u_{t-1} + ... + \theta_q u_{t-q}$$

To simplify the discussion, it is assumed that the actual coefficients

$$\{\nu, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q\}$$

and current and past realizations of x_t and u_t are known to the researcher.

The process is causal so that it has an MA representation

$$x_t = \mu + u_t + \psi_1 u_{t-1} + ... = \mu + \sum_{i=0}^{\infty} \psi_i u_{t-i}$$

with

$$\mu = \frac{\nu}{1 - \sum_{i=0}^{p} \phi_i}.$$

We also assume that our process is invertible so that it has an AR representation

$$x_t = \gamma + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \ldots + u_t = \gamma + \sum_{i=1}^{\infty} \pi_i x_{t-i} + u_t$$

with

$$\gamma = \left(1 + \sum_{i=1}^{\infty} \pi_i\right) \mu$$

The AR and MA representations show that for this process the information in

$$\{x_T, x_{T-1}...\}$$

can be equivalently be represented as

$$\{u_T, u_{T-1}...\}$$

since each u_t can be computated from the past and present x_s, $s \leq t$ and each x_t can be obtained from the past and present u_s, $s \leq t$.

Forecasting ARMA processes

Assuming in addition that u_t and u_s are independent and not only uncorrelated for $s \neq t$ we can now determine the conditional expectation $E\left[x_{T+h}|\Omega_T\right]$, that is optimal (minimum MSE) h-step-ahead forecast of x_{T+h} at time T.

Forecasting ARMA processes

For istance, for $h = 1$

$$\hat{x}_{T,1} = E\left[x_{T+1}|\Omega_T\right] = \nu + \phi_1 E\left[x_T|\Omega_T\right] + ... + \phi_p E\left[x_{T+1-p}|\Omega_T\right]$$

$$+E\left[u_{T+1}|\Omega_T\right] + \theta_1 E\left[u_T|\Omega_T\right] + ... + \theta_q E\left[u_{T+1-q}|\Omega_T\right]$$

$$= \nu + \phi_1 x_T + ... + \phi_p x_{T+1-p} + \theta_1 u_T + ... + \theta_q u_{T+1-q}$$

Forecasting ARMA processes

For an arbitrary positive h we get

$$\hat{x}_{T,h} = \nu + \phi_1 \hat{x}_{T,h-1} + ... + \phi_p \hat{x}_{T,h-p} + \theta_h u_T + ... + \theta_q u_{T+h-q} \text{ if } h \leq q$$

and

$$\hat{x}_{T,h} = \nu + \phi_1 \hat{x}_{T,h-1} + ... + \phi_p \hat{x}_{T,h-p} \text{ if } h > q$$

With these formulas, forecast can be computed recursively.

To illustrate these formulas we consider the MA(3) process

$$x_t = \nu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \theta_3 u_{t-3}$$

For this process

$$\hat{x}_{T,1} = \nu + \theta_1 u_T + \theta_2 u_{T-1} + \theta_3 u_{T-2}$$

$$\hat{x}_{T,2} = \nu + \theta_2 u_T + \theta_3 u_{T-1}$$

$$\hat{x}_{T,3} = \nu + \theta_3 u_T$$

$$\hat{x}_{T,h} = \nu \ \ \forall h \geq 4$$

As the MA(3) process has a memory of only 3 periods, all forecasts 4 or more steps ahead collapse to the intercept (the mean).

Now, we consider a causal AR(2) process

$$x_t = \nu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + u_t$$

For this process

$$\hat{x}_{T,1} = \nu + \phi_1 x_T + \phi_2 x_{T-1}$$

$$\hat{x}_{T,2} = \nu + \phi_1 \hat{x}_{T,1} + \phi_2 x_T$$

$$\hat{x}_{T,3} = \nu + \phi_1 \hat{x}_{T,2} + \phi_2 \hat{x}_{T,1}$$

$$\hat{x}_{T,h} = \nu + \phi_1 \hat{x}_{T,h-1} + \phi_2 \hat{x}_{T,h-2} \quad \forall h \geq 3$$

Forecasting ARMA processes

Alternatively, the optimal predictor can be determined using the AR or MA representation of x_t

$$\hat{x}_{T,h} = \gamma + \sum_{i=1}^{\infty} \pi_i \hat{x}_{T,h-i}$$

$$\hat{x}_{T,h} = \mu + \sum_{i=h}^{\infty} \psi_i u_{T+h-i}$$

The forecast error

From the formula

$$\hat{x}_{T,h} = \mu + \sum_{i=h}^{\infty} \psi_i u_{T+h-i}$$

the forecast error is easy to obtain.

The forecast error

We have

$$e_{x,T+h} = x_{T+h} - \hat{x}_{T,h} \quad = \sum_{i=0}^{h-1} \psi_i u_{T+h-i}$$

The forecast error

The forecast is unbiased since the expected error is zero

$$E[x_{T+h} - \hat{x}_{T,h}] = \sum_{i=0}^{h-1} \psi_i E[u_{T+h-i}] = 0$$

The forecast error

The variance of the forecast error is

$$\sigma_h^2 = E\left[\left(\sum_{i=0}^{h-1}\psi_i u_{T+h-i}\right)^2\right] = \sigma_u^2\sum_{i=0}^{h-1}\psi_i^2$$

Interval forecast

We note that if $u_t \sim i.i.N(0, \sigma^2)$, then

$$\frac{x_{T+h} - \hat{x}_{T,h}}{\sigma_h} \sim N(0, 1)$$

where σ_h is the square root of σ_h^2.

Interval forecast

Denoting by z_α the upper $\alpha 100$ percentage point of the standard normal distribution we get

$$1-\alpha = P\left(-z_{\alpha/2} \leq \frac{x_{T+h} - \hat{x}_{T,h}}{\sigma_h} \leq z_{\alpha/2}\right).$$

Hence, a $(1-\alpha)100\%$ interval forecast h periods ahead for x_{T+h} is

$$\left[\hat{x}_{T,h} - z_{\alpha/2}\sigma_h, \hat{x}_{T,h} + z_{\alpha/2}\sigma_h\right].$$

Interval forecast

The meaning of this interval is the following. If the forecast interval is computed repeatedly from a large number of realizations of the considered stochastic process, then $(1-\alpha)100\%$ of the intervals will contain the actual value (the realization) of the random variable x_{T+h}.

Predictability of a stochastic process

Following Granger and Newbold (1976), we define as a measure of predictability of a stochastic process $x_t; t \in \mathbb{Z}$, with finite variance, the index

$$R_x^2 = 1 - \frac{\mathrm{var}(e_{x,T+1})}{\mathrm{Var}(x_t)}$$

Predictability of a stochastic process

We note that $R_x^2 \in [0, 1]$.

- If $R_x^2 = 0$ the process x_t $t \in \mathbb{Z}$ is unpredictable;
- If $R_x^2 = 1$ the process x_t $t \in \mathbb{Z}$ is perfectly predictable.

Now, we suppose that x_t $t \in \mathbb{Z}$ is an ARMA(p, q) causal and invertible stochastic process,

$$x_t = \nu + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + u_t + \theta_1 u_{t-1} + \ldots + \theta_q u_{t-q}, \quad u_t \sim WN(0, \sigma^2)$$

The innovation u_{T+1} is the corresponding one-step forecast error.Thus we have

$$R_x^2 = 1 - \frac{\sigma^2}{\mathrm{Var}(x_t)}$$

Predictability of a stochastic process

The index R_x^2 can be related to the coefficients $\{\psi_1, \psi_2, ...\}$ of the MA representation of x_t.

Predictability of a stochastic process

Since

$$\mathrm{Var}(x_t) = \left(1 + \sum_{j=1}^{\infty} \psi_j^2\right) \sigma^2,$$

we obtain

$$R_x^2 = 1 - \frac{1}{1 + \sum_{j=1}^{\infty} \psi_j^2} = \frac{\sum_{j=1}^{\infty} \psi_j^2}{1 + \sum_{j=1}^{\infty} \psi_j^2}.$$

We call the sequence $\{\psi_j\}$ ψ-weights; they represent the dependence structure of the series.

Predictability of a stochastic process

A series with small ψ-weights (with a few structure) will be less predictable than one with large ψ-weights (a series with more structure). Thus, this predictability measure provides a synthetic evaluation of the dependence structure of a stationary time series.

Predictability of a stochastic process

In general, we can define a measure of predictability relative to a h-steps forecast by

$$R_x^2(h) = 1 - \frac{\mathrm{Var}(e_{x,T+h})}{\mathrm{Var}(x_t)} = 1 - \frac{1 + \sum_{j=1}^{h-1} \psi_j^2}{1 + \sum_{j=1}^{\infty} \psi_j^2} = \frac{\sum_{j=h}^{\infty} \psi_j^2}{1 + \sum_{j=1}^{\infty} \psi_j^2}.$$

This predictability index is utilised by Hong and Billings (1999) and for $h = 1$ coincides with the index proposed in Granger and Newbold (1976).

Evaluating the performance of forecasting methods

Suppose that

$$\hat{x}_{T+h} = g(x_T, x_{T-1}, ...; \hat{\theta}),$$

is a forecasting methods used to forecast x_{T+h} for $h = 1, 2, ..., N$.How good is the forecast?

Evaluating the performance of forecasting methods

After we have observed the real values x_{T+h}, $h = 1, 2, ..., N$, we can calculate the forecast errors

$$e_{x,T+h} = x_{T+h} - \hat{x}_{T+h} \quad h = 1, 2, ..., N$$

Various accuracy measures, based on these errors, have been used to evaluate the performance of forecasting methods. We will present three performance measures

- mean squared error (MSE)
- root mean squared error (RMSE)
- mean absolute percent error (MAPE)

The most common measure is the mean squared error (MSE). It is defined as

$$MSE = \frac{1}{N}\sum_{h=1}^{N} e_{x,T+h}^2$$

Evaluating the performance of forecasting methods

A widely used measure of overall accuracy of a forecasting method is the root mean squared error (RMSE). The RMSE statistic is defined as follows

$$RMSE = \sqrt{\frac{\sum_{h=1}^{N} e_{x,T+h}^2}{N}}$$

Evaluating the performance of forecasting methods

Another common criterion used in comparing performance of forecast models is the mean absolute percent error (MAPE).The MAPE statistic is defined as follows

$$MAPE = \frac{100}{N} \sum_{h=1}^{N} \left| \frac{e_{x,T+h}}{x_{T+h}} \right|$$

The MSE, RMSE, and MAPE measure the magnitude of the forecast errors. Better models will show smaller values for these statistics.

Each of these measurements has different advantages and limitations. Often the square root of MSE (RMSE) is used so as to preserve the units. RMSE, however, does not provide information about the relative magnitude of the forecast error. Hence, using more than one performance measure is always recommended.

The reliability of ARMA time series forecasts

Some example of forecasts:

- "I think there is a world market for about five computers" - Founder of IBM in 1947
- "There is no reason for any individual to have a computer in their home" - President Digital Equipment in 1977
- "Stock prices have reached what looks like a permanently high plateau" - Yale Professor of Economics in September 1929

It's not easy to make good forecasts!

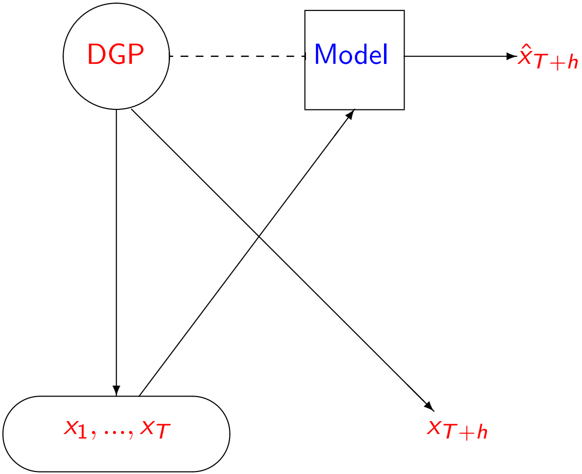
However, these are judgmental forecasts.In this lecture we are interested to the ARMA time series forecasts.

Are the ARMA time series forecasts reliable?

The reliability of ARMA time series forecasts

A reliable ARMA time series forecast requires that the future is not too different from the past.

In other terms, it requires that x_{T+h} will be drawn from the same DGP that generated the previous observations.

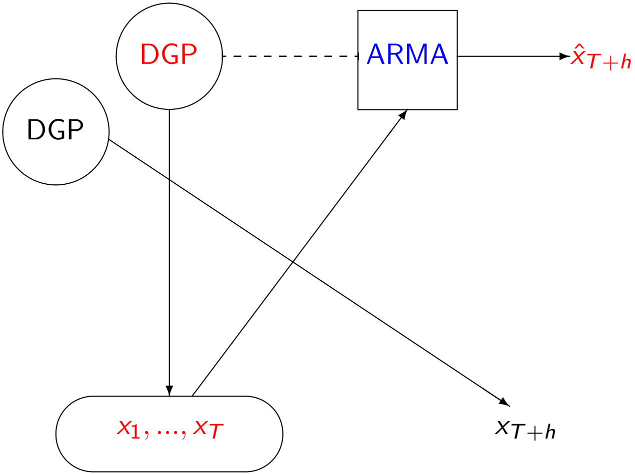

DGP
Model
$\hat{x}_{T+h}$
$x_1, \ldots, x_T$
x_{T+h}

The reliability of ARMA time series forecasts

If the future is too different from the past the ARMA model will produce biased forecasts.

In particular, if a structural break will occur between T and $T + h$, then x_{T+h} will be drawn from a different DGP from that has generated the previous observations.

The reliability of ARMA time series forecasts



The reliability of ARMA time series forecasts

The ARMA model is mis-specified for new DGP and then any forecast, based on this model, is wrong.

Advantages of ARMA models

Advantages:

1. constitutes a flexible class of models;
2. provide unconditional forecasts;
3. models are parsimonious with respect to coefficients;

Disadvantages of ARMA models

Disadvantages:

1. requires large number of observations for model identification (at least 50 and preferably 100 observations should be available to build a proper model);
2. need a long series of data without structural change;
3. the amount of subjective input at the identification stage make them somewhat more of an art than a science