Forecasting with ARMA models

Outline

Short overview:

(1) Principles of forecasting
  - Optimal predictor
  - ARMA models

(2) Forecasting with AR(p) models

(3) Forecasting with MA(q) models

(4) Forecasting with ARMA (1,1)

(5) Comparing predictive accuracy (Diebold-Mariano test)

(6) Combination of forecasts

Principles of forecasting

Remark

The main goal of building a model for the univariate time series is predicting the future path of economic variables. Even though they do not provide any economic insight, ARMA models perform quite well in this respect.

Definition

The theory of optimal prediction defines "prediction" as the estimate of a particular value of a random variable on the basis of a predetermined information set, I_t, at time t. In the case of univariate time series I_t is given by:

$$I_t = \{Y_t, Y_{t-1}, ..., Y_0, Y_{-1}, ...\}.$$

Principles of forecasting

Remark

The main goal of building a model for the univariate time series is predicting the future path of economic variables. Even though they do not provide any economic insight, ARMA models perform quite well in this respect.

Definition

The theory of optimal prediction defines "prediction" as the estimate of a particular value of a random variable on the basis of a predetermined information set, I_t, at time t. In the case of univariate time series I_t is given by:

$$I_t = \{Y_t, Y_{t-1}, ..., Y_0, Y_{-1}, ...\}.$$

Optimal Predictor

- Define $\hat{Y}_{t+h}$ as the h-step ahead predictor of the series:

$$\hat{Y}_{t+h} = g(I_t).$$

- The prediction error is given by:

$$e_{t+h} = Y_{t+h} - \hat{Y}_{t+h} = Y_{t+h} - g(I_t).$$

- The objective now is to choose the function $g(.)$. The general strategy is to choose $g(.)$ by minimizing some given loss function, which is usually linked to the prediction error.

Optimal Predictor

• Define $\hat{Y}_{t+h}$ as the h-step ahead predictor of the series:

$$\hat{Y}_{t+h} = g(I_t).$$

• The prediction error is given by:

$$e_{t+h} = Y_{t+h} - \hat{Y}_{t+h} = Y_{t+h} - g(I_t).$$

• The objective now is to choose the function $g(.)$. The general strategy is to choose $g(.)$ by minimizing some given loss function, which is usually linked to the prediction error.

Optimal Predictor

The most common measures of prediction accuracy are:

- **Mean square error (MSE)**:

$$MSE = \frac{1}{T} \sum_{t=1}^{T} e_{t+h}^2;$$

- **Square root MSE (RMSE)**:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} e_{t+h}^2};$$

- **Mean absolute error (MAE)**:

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |e_{t+h}|.$$

Optimal Predictor

The most common measures of prediction accuracy are:

- **Mean square error** (**MSE**):

$$MSE = \frac{1}{T}\sum_{t=1}^{T} e_{t+h}^2;$$

- **Square root MSE** (**RMSE**):

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_{t+h}^2};$$

- **Mean absolute error** (**MAE**):

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |e_{t+h}|.$$

Optimal Predictor

Remark (Optimal Predictor based on MSE)

In practice we solve the following problem:

$$g(.) = \underset{g}{\operatorname{argmin}}\ E(e_{t+h}^2) \equiv \underset{g}{\operatorname{argmin}}\ E\left[(Y_{t+h} - g(I_t))^2\right]. \tag{1}$$

We can show that the solution of the problem in (1), which corresponds to the optimal predictor, which minimize the mean square criterion, is given by the conditional expectation of Y_{t+h} conditional to the information set I_t:

$$\hat{Y}_{t+h} = E(Y_{t+h}|I_t). \tag{2}$$

ARMA models

- Consider the following ARMA(p,q) model:

$$\phi(L)Y_t = \theta(L)\epsilon_t$$

where

$$\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$$
$$\theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$$

- Then the h-step ahead predictor of the ARMA(p,q) model is given by

$$\begin{aligned} \hat{Y}_{t+h} &= E(Y_{t+h}|I_t) \simeq \phi_1 E(Y_{t+h-1}|I_t) + ... + \phi_p E(Y_{t+h-p}|I_t) \\ &\quad + E(\epsilon_{t+h}|I_t) + ... + \theta_q E(\epsilon_{t+h-q}|I_t). \end{aligned}$$

where:

$$E(Y_{t+j}|I_t) = \begin{cases} y_{t+j}, \text{ if } j \leq 0 \\ \hat{y}_{t+j}, \text{ if } j > 0 \end{cases} \quad \text{and } E(\epsilon_{t+j}|I_t) = \begin{cases} \epsilon_{t+j}, \text{ if } j \leq 0 \\ 0, \text{ if } j > 0 \end{cases}$$

ARMA models

Remark

• *In practice we replace the unknown parameters* ϕ_j, θ_j, *and* σ^2 *by their estimate from the realization* $\{y_t\}_{t=1}^{T}$. *For T sufficiently large, using the estimates* $\hat{\phi}_j$, $\hat{\theta}_j$, *and* $\hat{\sigma}^2$, *instead of the true parameters, produces a good approximation.*

• *Remember that the prediction error can be found as*

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h}$$

which is a random variable. For unbiased predictors ($E(e_{t+h}) = 0$), *it is useful to look at the variance* $Var(e_{t+h})$, *which is a measure of prediction accuracy.*

Forecasting with AR models

- Consider the following simple AR(1) model:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t.$$

- We can show that the optimal prediction is given by the following function:

$$\hat{y}_{t+h} = \phi_0(1 + \phi_1 + \phi_1^2 + ... + \phi_1^{h-1}) + \phi_1^h y_t \quad (3)$$

- The prediction error which corresponds to the prediction in (3) is given by:

$$e_{t+h} = \epsilon_{t+h} + \phi_1 \epsilon_{t+h-1} + \phi_1^2 \epsilon_{t+h-2} + ... + \phi_1^{h-1} \epsilon_{t+1}$$

- Its variance is:

$$Var(e_{t+h}) = \sigma^2(1 + \phi_1^2 + \phi_1^4 + ... + \phi_1^{2(h-1)}).$$

Forecasting with AR models

Consequently, when $h \to \infty$, we have:

$$\hat{Y}_{t+h} \underset{h\to\infty}{\to} \frac{\phi_0}{1-\phi_1} = E(Y_t)$$

Which means that when the horizon h is equal to infinity the optimal prediction is given by the unconditional mean. The corresponding variance of the optimal prediction is:

$$Var(e_{t+h}) \underset{h\to\infty}{\to} \frac{\sigma^2}{1-\phi_1^2} = Var(Y_t).$$

Forecasting with AR models

Example

Consider the following AR(1) model for annual GDP growth

$$GDP_t = 0.53 + 0.76 GDP_{t-1} + \epsilon_t, \ t = 1900, .., \ 2007$$

Assume that the level of GDP growth in 2007 is $GDP_{2007} = 1.341\%$.
We can use the AR(1) model of annual GDP to predict GDP_{2008}, $GDP_{2009}, ...$

$$\widehat{GDP}_{2008} = E(GDP_{2008} \mid I_{2007}) = 0.53 + 0.76 GDP_{2007} = 1.549\%$$

Similarly, the level of GDP in 2009 will be:

$$\widehat{GDP}_{2009} = E(y_{2009} \mid I_{2007}) = 0.53 + 0.76 \ \widehat{GDP}_{2008} = 1.707\%$$

The variances of the errors forecast for these two predictions are:

$\hat{Var}(e_{2008}) = 1.06$ and $\hat{Var}(e_{2009}) = \hat{Var}(\epsilon_{2009} + 0.76\epsilon_{2008}) = 1.06(1 + 0.76^2)$

Forecasting with MA models

- Consider the simple MA(1) model:

$$Y_t = \theta_0 + \epsilon_t + \theta_1 \epsilon_{t-1}.$$

- We can show that the optimal one period ahead forecast is given by:

$$\hat{y}_{t+h} = \begin{cases} \theta_0 + \theta_1 \epsilon_t \text{ for } h = 1; \\ \theta_0 \text{ for } h \geq 1. \end{cases}$$

- The variance of the forecast error is:

$$Var(e_{t+h}) = \begin{cases} \sigma^2, \text{ for } h = 1; \\ \sigma^2(1 + \theta_1^2), \text{ for } h \geq 2 \end{cases} .$$

Forecasting with MA models

Remark

Generally, MA(q) processes give non-trivial predictions only if $h \leq q$ and for $h > q$

$$\hat{y}_{t+h} = \theta_0 = E(Y_t)$$

and

$$Var(e_{t+h}) = \sigma^2(1 + \theta_1^2 + ... + \theta_q^2) = Var(Y_t).$$

Forecasting with MA models

Example

Consider the following MA(1) model for given asset return:

$$r_t = \epsilon_t - 0.68\epsilon_{t-1},\ t = 1960, ..., 2007$$

Assume that by estimating the model of asset returns we found that

$$\epsilon_{2007} = -0.223.$$

We can use the MA(1) model to predict $r_{2008}, r_{2009}, ...$

$$\hat{r}_{2008} = E(r_{2008} \mid I_{2007}) = -0.68\epsilon_{2007} = 0.152$$

Similarly, the return rate in 2009 will be:

$$\hat{r}_{2009} = E(y_{2009} \mid I_{2007}) = E(-0.68\epsilon_{2008} + \epsilon_{2009} \mid I_{2007}) = 0$$

The variances of the errors forecast are: $\hat{Var}(e_{2008}) = 1.04$ and $\hat{Var}(e_{2009}) = \hat{Var}(\epsilon_{2009} - 0.68\epsilon_{2008}) = 1.04(1 + 0.68^2)$

Forecasting with ARMA (1,1)

- Consider the simple ARMA(1,1) model:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}.$$

- We can show that:

$$\hat{y}_{t+h} = \phi_0(1 + \phi_1 + \phi_1^2 + ... + \phi_1^{h-1}) + \phi_1^h y_t + \phi_1^{h-1}\theta_1 e_t.$$

- Derive this result for $h = 1$ and $h = 2$

Forecasting with ARMA (1,1)

• This means that the asymptotic behavior of the predictor is determined by the AR component:

$$\lim_{h \to \infty} E_t(Y_{t+h}) = E(Y_t)$$

$$\lim_{h \to \infty} Var(e_{t+h}) = Var(Y_t).$$

• Generally, in an ARMA(p,q) process, when $h > q$, the prediction is driven by the AR component of the process.

Comparing predictive accuracy

- In the empirical application it is often the case that two or more time series models are available for forecasting a particular variable of interest.

- Suppose two forecasting procedures are available that produce

$$\{\widehat{y}_{1,T+h}; h = 1, ..., H\} \text{ and } \{\widehat{y}_{2,T+h}; h = 1, ..., H\}.$$

- **Are these forecasts equally good?**

- Suppose that the criterion of the evaluation is MSE. Then the method that produces the lowest MSE value over the same period is considered superior. This is the hypothesis that has to be tested

Comparing predictive accuracy

- Define the forecast errors as

$$e_{i,T+h} = \widehat{y}_{i,T+h} - y_{i,T+h}, \text{ for } i = 1,2$$

and let $f(\cdot)$ define a loss function (e.g., $f(e_{i,T+h}) = e^2_{i,T+h}$ or $f(e_{i,T+h}) = |e_{i,T+h}|$)

- We also define the loss differential between the two forecasts by

$$d_h = f(e_{1,T+h}) - f(e_{2,T+h}),$$

and say that the two forecasts have equal accuracy if and only if the loss differential has zero expectation for all h.

Comparing predictive accuracy

- The null hypothesis is formulated as

$$H_0 : \mathbb{E}(d_t) = 0 \text{ for } t$$

versus the alternative $H_1 : \mathbb{E}(d_t) \neq 0$.

In other words the null hypothesis checks if the two forecasts on avarage have the same accuracy.

Comparing predictive accuracy

Theorem (Diebold-Mariano Test)

If $\{d_h; h = 1, ..., H\}$ *is covariance stationary then*

$$\sqrt{H}(\overline{d} - \mu) \to N(0, ''Var''),$$

where $\overline{d} = \frac{1}{H}\sum_{h=1}^{H} d_h$, $\mu = \mathbb{E}(d_t)$ *and* $''Var''$ *is based on the spectral density at 0.*

Remark

Using this theorem Diebold and Mariano suggested to use the following test to check if two forecasting methods produce the same forecasts

$$DM = \frac{\overline{d}}{\sqrt{\frac{\widehat{Var}}{H}}}$$

Combination of the forecasts

- When more than one forecast is available the combination of those forecasts has the advantage of reducing the forecast error.

- Assume that two forecasts are available. Denore them by f_1 and f_2. Then the combined forecast (f_c) can be obtained as a weighted average of both forecasts

$$f_c = w_1 f_1 + w_2 f_2, \text{ for } w_1 + w_2 = 1.$$

Combination of the forecasts

- One way to select the weights w_1 and w_2 is by looking at the past performance of both methods and assigning the larger weight to that methods that perform the best, for instance based on the value of the MSE. The weight can be proportional to the inverse of the MSE, can be calculated in the following way:

$$w_1 = \frac{\frac{1}{MSE_1}}{\frac{1}{MSE_1} + \frac{1}{MSE_2}} = \frac{MSE_2}{MSE_1 + MSE_2}$$

and

$$w_2 = 1 - w_1 == \frac{MSE_1}{MSE_1 + MSE_2}$$