Time Series Analysis

Forecasting

Hossein Moradi Rekabdarkolaee

South Dakota State University

email: *hossein.moradirekabdarkolaee@sdstate.edu*.

Office: CAME Building, *Room: 240*.

Introduction

- So far, we have learned
  - Different components of time series
    - Trend,
    - Seasonality,
    - Error.
  - Different type of time series
    - Stationary (weak and strong)
    - Non-stationary.
  - Different models for time series
    - Autoregressive,
    - Moving Average,
    - ARMA,
    - ARIMA.

Introduction

- Different way to find the order of time series
  - ACF,
  - PACF,
  - EACF,
  - Information Criteria-based method (AIC, BIC, AICc, etc.)
- Different approaches to estimate the parameters of the model we choose to fit the time series.
  - Method of Moment,
  - Least Squares,
  - Maximum Likelihood.

Introduction

- After fitting the model, we did
    - Model diagnostics or model criticism,
    - What does it mean
    - How to do it.
    - Do some examples.
- In this section, we are going to study the prediction at future time.

Forecasting

- One of the primary objectives of building a model for a time series is to be able to forecast the values for that series at future times.
- The assessment of the precision of those forecasts are equally importance.
- The calculation of forecasts and their properties for both deterministic trend models and ARIMA models will be considered.
- Also, the forecasts for models that combine deterministic trends with ARIMA stochastic components will be studied.

Minimum Mean Square Error Forecasting

- Based on the available history of the series up to time t, i.e. $I_t = \{Y_1, Y_2, \ldots, Y_{t-1}, Y_t\}$, we would like to forecast the value of $Y_{t+\ell}$ that will occur ℓ time units into the future.
- Define $\widehat{Y}_t(\ell)$ to be the forecast of $Y_{t+\ell}$ based on I_t where ℓ is called **lead time** for the forecast.
- Then, the forecast error is given by $\epsilon_t(\ell) = Y_{t+\ell} - \widehat{Y}_t(\ell)$.
- The mean square error of the forecast is

$$MSE(\epsilon_t(\ell)) = E[\epsilon_t(\ell)^2] = E[(Y_{t+\ell} - \widehat{Y}_t(\ell))^2]$$

- It can be shown that the minimum MSE forecast (best forecast) of $Y_{t+\ell}$ based on I_t is

$$\widehat{Y}_t(\ell) = E[Y_{t+\ell}|I_t] = E[Y_{t+\ell}|Y_1, Y_2, \ldots, Y_{t-1}, Y_t] \tag{1}$$

Deterministic Trends

- Consider a model with the deterministic trend as follows

$$Y_t = \mu_t + X_t, \qquad \text{with} \ \ E[X_t] = 0.$$

- For this section, we assume that $\{X_t\}$ is in fact white noise with variance γ_0.

- Then, the forecast can be obtained as

$$\begin{aligned} \widehat{Y}_t(\ell) &= E[\mu_{t+l} + X_{t+l} | Y_1, Y_2, \ldots, Y_{t-1}, Y_t] \\ &= \mu_{t+\ell} \end{aligned}$$

- For the linear trend case, $\mu_t = \beta_0 + \beta_1 t$, we have $\widehat{Y}_t(\ell) = \beta_0 + \beta_1(t + \ell)$.

Deterministic Trends

- For seasonal models where, say, $\mu_t = \mu_{t+12}$, the forecast is $\widehat{Y}_t(\ell) = \mu_{t+12+\ell} = \widehat{Y}_t(\ell + 12)$.

- That is, the forecast will be periodic, as desired.

- In general, the forecast error, $e_t(\ell)$, is given by

$$\begin{aligned} e_t(\ell) &= Y_{t+\ell} - \widehat{Y}_t(\ell) \\ &= \mu_{t+\ell} + X_{t+\ell} - \mu_{t+\ell} = X_{t+\ell} \end{aligned}$$

- Then, we have

$$E[e_t(\ell)] = E[X_t(\ell)] = 0,$$

$$Var(e_t(\ell)) = Var(X_{t+l}) = \gamma_0.$$

AR(1) Forecasting

- Consider an AR(1) model with a nonzero mean as follows

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$$

- Then, based on Eq (1) we have

$$\begin{aligned}
\widehat{Y}_t(1) - \mu &= E[Y_{t+1} - \mu | Y_1, Y_2, \ldots, Y_{t-1}, Y_t] \\
&= E[\phi(Y_t - \mu) + \epsilon_{t+1} | Y_1, Y_2, \ldots, Y_{t-1}, Y_t] \\
&= \phi(Y_t - \mu)
\end{aligned}$$

- That is,

$$\widehat{Y}_t(1) = \mu + \phi(Y_t - \mu)$$

AR(1) Forecasting

- Similarly for forecasting l time units into the future, we have

$$\widehat{Y}_t(\ell) = \mu + \phi(\widehat{Y}_t(\ell - 1) - \mu) \quad \text{for} \quad \ell \geq 1 \tag{2}$$

- This is because for $\ell \geq 1$ we have

$$E[Y_{t+\ell-1} | Y_1, Y_2, \ldots, Y_{t-1}, Y_t] = \widehat{Y}_t(\ell - 1),$$

- and $e_{t+\ell}$ is independent from all previous lags of Y_t.
- Then, from (2) we have

$$\begin{aligned}
\widehat{Y}_t(\ell) &= \phi(\widehat{Y}_t(\ell - 1) - \mu) + \mu \\
&= \phi\big(\phi(\widehat{Y}_t(\ell - 2) - \mu)\big) + \mu \\
&= \vdots \\
&= \phi^{\ell}(Y_t - \mu) + \mu
\end{aligned}$$

Forecast error - AR(1)

- Consider one-step-ahead forecast error i.e. $e_t(1)$

$$\begin{aligned} e_t(1) &= Y_{t+1} - \widehat{Y}_t(1) \\ &= \left[\phi(Y_t - \mu) + \mu + e_{t+1}\right] - \left[\phi(Y_t - \mu) + \mu\right] \\ &= e_{t+1} \end{aligned}$$

- Therefore, $Var(e_t(1)) = \sigma^2$.
- Recall that AR(1) can be written as $Y_t = \sum_{j=0}^{\infty} \phi^j e_{t-j}$, so

$$\begin{aligned} e_t(\ell) &= Y_{t+\ell} - \widehat{Y}_t(\ell) = Y_{t+\ell} - \mu - \phi^\ell (Y_t - \mu) \\ &= \sum_{i=0}^{\ell-1} \phi^i e_{t+\ell-i} + \sum_{i=\ell}^{\infty} \phi^i e_{t+\ell-i} - \mu - \phi^\ell \sum_{j=0}^{\infty} \phi^j e_{t-j} \\ &= \sum_{i=0}^{\ell-1} \phi^i e_{t+\ell-i} \end{aligned}$$

Forecast error - AR(1)

- So, for $e_t(\ell)$ we have

$$e_t(\ell) = Y_{t+l} - \widehat{Y}_t(\ell) = \sum_{i=0}^{\ell-1} \phi^i e_{t+\ell-i}$$

- Then, we can show that

$$Var(e_t(\ell)) = \sigma^2 \left[\frac{1-\phi^{2\ell}}{1-\phi^2} \right]$$

- For large ℓ, the variance of the forecast error for AR(1) model is given by

$$Var(e_t(\ell)) \approx \frac{\sigma^2}{1-\phi^2}$$

- That is, for large value of ℓ,

$$Var(e_t(\ell)) \approx Var(Y_t) = \gamma_0 = \frac{\sigma^2}{1-\phi^2}$$

Forecast error - MA(1)

- Consider the MA(1) case with nonzero mean

$$Y_t = \mu + e_t - \theta e_{t-1}$$

- Based on Eq (1) we have

$$\begin{aligned}
\widehat{Y}_t(1) &= E[Y_{t+1}|Y_1, Y_2, \ldots, Y_{t-1}, Y_t] \\
&= E[\mu + e_{t+1} - \theta e_t | Y_1, Y_2, \ldots, Y_{t-1}, Y_t] \\
&= \mu - 0 - \theta E[e_t | Y_1, Y_2, \ldots, Y_{t-1}, Y_t] \\
&= \mu - \theta e_t
\end{aligned}$$

- the last equality follows from the fact that e_t is a function of $Y_1, Y_2, \ldots, Y_t$.
- Then, the forecast error of MA(1) is obtained as

$$e_t(1) = Y_{t+1} - \widehat{Y}_t(1) = (\mu + e_{t+1} - \theta e_t) - (\mu - \theta e_t) = e_{t+1}$$

Forecast error - MA(1)

- For longer lead times, we have

$$\begin{aligned}\widehat{Y}_t(\ell) &= E[Y_{t+\ell}|Y_1, Y_2, \ldots, Y_{t-1}, Y_t] \\ &= \mu + E[e_{t+\ell}|Y_1, Y_2, \ldots, Y_t] - \theta E[e_{t+\ell-1}|Y_1, Y_2, \ldots, Y_t] = \mu.\end{aligned}$$

- This is because for $\ell > 1$, both $e_{t+\ell}$ and $e_{t+\ell-1}$ are independent of $Y_1, Y_2, \ldots, Y_t$.
- Then, the forecast error of MA(1) is obtained as

$$\begin{aligned}e_t(\ell) =& Y_{t+\ell} - \widehat{Y}_t(\ell) = (\mu + e_{t+\ell} - \theta e_{t+\ell-1}) - (\mu - \theta e_{t+\ell}) \\ =& e_{t+\ell} - \theta e_{t+\ell-1}\end{aligned}$$

- Therefore, the forecast error variance of MA(1) is given by

$$Var(e_t(\ell)) = \sigma^2(1 + \theta^2)$$

- Consider the random walk (R.W.) with drift defined by

$$Y_t = Y_{t-1} + \theta_0 + e_t$$

- Then, we can obtain the $\widehat{Y}_t(\ell)$ of the R.W. as follows

$$\begin{aligned}\widehat{Y}_t(\ell) &= E[Y_{t+\ell}|Y_1, Y_2, \ldots, Y_t] \\ &= E[Y_{t+\ell-1} + \theta_0 + e_{t+\ell}|Y_1, Y_2, \ldots, Y_t] \\ &= \widehat{Y}_t(\ell - 1) + \theta_0 = Y_t + \ell\theta_0\end{aligned}$$

- Then, the forecast error of the R. W. is obtained as follows

$$\begin{aligned}e_t(\ell) =& Y_{t+\ell} - \widehat{Y}_t(\ell) \\ =&(Y_t + \ell\theta_0 + e_{t+1} + \ldots + e_{t+\ell}) - (Y_t + \ell\theta_0) \\ =&e_{t+1} + e_{t+2} + \ldots + e_{t+\ell}\end{aligned}$$

- Therefore, the forecast error variance of the R. W. is

$$Var(e_t(\ell)) = \ell\sigma^2$$

Forecasting - ARMA(1, 1)

- Consider ARMA(1, 1)

$$Y_t = \theta_0 + \phi Y_{t-1} + e_t - \theta e_{t-1}, \qquad \theta_0 = \mu(1-\phi)$$

- Then,

$$\begin{aligned}\widehat{Y}_t(\ell) &= E[Y_{t+\ell}|Y_1, Y_2, \ldots, Y_t] \\ &= \theta_0 + \phi \widehat{Y}_t(\ell-1) + E[e_{t+\ell}|Y_1, Y_2, \ldots, Y_t] \\ &- \theta E[e_{t+\ell-1}|Y_1, Y_2, \ldots, Y_t].\end{aligned}$$

- For the first step forecast, we have

$$\begin{aligned}\widehat{Y}_t(1) &= \theta_0 + \phi Y_t - \theta e_t \\ \widehat{Y}_t(2) &= \theta_0 + \phi \widehat{Y}_t(1) \\ &\vdots \\ \widehat{Y}_t(\ell) &= \theta_0 + \phi \widehat{Y}_t(\ell-1), \quad \text{for } \ell \geq 2\end{aligned}$$

Forecasting - ARMA(1, 1)

- Therefore

$$\widehat{Y}_t(\ell) = \mu + \phi^{\ell}(Y_t - \mu) - \phi^{\ell-1}\theta e_t \text{ for } \ell \geq 1.$$

- For the forecasting error, for $\ell = 1$ we have

$$\begin{aligned} e_t(1) =& Y_{t+1} - \widehat{Y}_t(1) \\ =& (\theta_0 + \phi Y_t + e_{t+1} - \theta e_t) - (\theta_0 + \phi Y_t - \theta e_t) \\ =& e_{t+1} \end{aligned}$$

Forecast Error - ARMA(1, 1)

- For $\ell \geq 2$, it can be written as

$$
\begin{aligned}
e_t(\ell) =& Y_{t+\ell} - \widehat{Y}_t(\ell) \\
=& \Big[\theta_0 \sum_{i=0}^{l-1} \phi^i + \phi^\ell Y_t - \theta\phi^{\ell-1} e_t + e_{t+\ell} + \sum_{i=1}^{\ell-1} \phi^{i-1}(\phi-\theta) e_{t+\ell-i}\Big] \\
& - \Big[\theta_0 \sum_{i=0}^{\ell-1} \phi^i + \phi^\ell Y_t - \theta\phi^{\ell-1} e_t\Big] \\
=& e_{t+\ell} + \sum_{i=1}^{\ell-1} \phi^{i-1}(\phi-\theta) e_{t+\ell-i} = \sum_{i=0}^{\ell-1} \psi_i e_{t+\ell-i}
\end{aligned}
$$

- where $\psi_0 = 1$ and $\psi_j = \phi^{j-1}(\phi-\theta), j = 1, \ldots, \ell-1$

Forecasting - ARMA(p, q)

- The forecast function of ARMA(p, q) is obtained as follows

$$\begin{aligned}\widehat{Y}_t(\ell) = \theta_0 + \phi_1 \widehat{Y}_t(\ell-1) + \phi_2 \widehat{Y}_t(\ell-2) + \ldots + \phi_p \widehat{Y}_t(\ell-p)\\ - \theta_1 E[e_{t+\ell-1}|Y_1, Y_2, \ldots, Y_t] - \theta_2 E[e_{t+\ell-2}|Y_1, Y_2, \ldots, Y_t]\\ - \ldots - \theta_q E[e_{t+\ell-q}|Y_1, Y_2, \ldots, Y_t]\end{aligned}$$

- where

$$E[e_{t+j}|Y_1, Y_2, \ldots, Y_t] = \begin{cases} 0 & \text{for } j > 0 \\ e_{t+j} & \text{for } j \leq 0 \end{cases}$$

- Therefore for $\ell > q$ we have

$$\widehat{Y}_t(\ell) = \theta_0 + \phi_1 \widehat{Y}_t(\ell-1) + \phi_2 \widehat{Y}_t(\ell-2) + \ldots + \phi_p \widehat{Y}_t(\ell-p),$$

$$\begin{aligned}\widehat{Y}_t(\ell) - \mu = \phi_1(\widehat{Y}_t(\ell-1) - \mu) + \phi_2(\widehat{Y}_t(\ell-2) - \mu)\\ + \ldots + \phi_p(\widehat{Y}_t(\ell-p) - \mu)\end{aligned}$$

Forecast Error - ARMA(p, q)

- It can be proved that any ARIMA model can be written in truncated linear process form as

$$Y_{t+\ell} = C_t(\ell) + I_t(\ell) \quad \text{for} \quad \ell > 1$$

- where $C_t(\ell)$ is a certain function of the finite history of $Y_t, Y_{t-1}, \ldots, Y_1$, and $I_t(\ell)$ is given by

$$I_t(\ell) = \sum_{i=0}^{\ell-1} \psi_i e_{t+\ell-i} \quad \text{for} \quad \ell \geq 1$$

- As a result of this we have,

$$\widehat{Y}_t(\ell) = E[C_t(\ell)|Y_1, Y_2, \ldots, Y_t] + E[I_t(\ell)|Y_1, Y_2, \ldots, Y_t] = C_t(\ell)$$

Forecast Error - ARMA(p, q)

- Therefore, the forecast error can be obtained as follows

$$\begin{aligned} e_t(\ell) =& Y_{t+\ell} - \widehat{Y}_t(\ell) \\ =& (C_t(\ell) + I_t(\ell)) - C_t(\ell) = I_t(\ell) \end{aligned}$$

- Thus, for a general invertible ARIMA process

$$E[e_t(\ell)] = E[I_t(\ell)] = 0$$

$$Var(e_t(\ell)) = \sigma^2 \sum_{j=0}^{\ell-1} \psi_j^2 \quad \text{for} \quad \ell \geq 1$$

- for large ℓ

$$Var(e_t(\ell)) \approx \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \gamma_0 \quad \text{for large} \quad \ell$$

Non-stationary Models - ARIMA(1, 1, 1)

- Similar to the random walk case, forecasting for nonstationary ARIMA models is similar to the one for stationary ARMA models, with some striking differences.
- Recall the ARIMA(1, 1, 1):

$$Y_t - Y_{t-1} = \phi(Y_{t-1} - Y_{t-2}) + \theta_0 + e_t - \theta e_{t-1}$$

- or

$$Y_t = (1 + \phi)Y_{t-1} - \phi Y_{t-2} + \theta_0 + e_t - \theta e_{t-1}$$

- Then, we have

$$\begin{aligned}
\widehat{Y}_t(1) &= (1 + \phi)Y_t - \phi Y_{t-1} + \theta_0 + e_t - \theta e_t \\
\widehat{Y}_t(2) &= (1 + \phi)\widehat{Y}_t(1) - \phi Y_t + \theta_0 \\
\widehat{Y}_t(\ell) &= (1 + \phi)\widehat{Y}_t(\ell - 1) - \phi \widehat{Y}_t(\ell - 2) + \theta_0
\end{aligned}$$

Non-stationary Models - ARIMA(1, 1, 1)

- For the general invertible ARIMA model, based on the truncated linear process representation, the forecast error is

$$e_t(\ell) = \sum_{i=0}^{\ell-1} \psi_i e_{t+\ell-i} \qquad \text{for} \quad \ell \geq 1$$

- Therefore,

$$E(e_t(\ell)) = 0, \qquad \text{for} \quad \ell \geq 1$$

- and

$$var(e_t(\ell)) = \sigma^2 \sum_{i=0}^{\ell-1} \psi_i^2 \qquad \text{for} \quad \ell \geq 1.$$

- Note that, **unlike the stationary series**, for the nonstationary time series, the ψ_j-weights do not decay to zero as j increases.

Non-stationary Models - IMA(1, 1)

- Recall the IMA(1, 1) with constant term has a form

$$Y_t = Y_{t-1} + \theta_0 + e_t - \theta e_{t-1}$$

- Then the forecasts are

$$\widehat{Y}_t(\ell) = \widehat{Y}_t(\ell - 1) + \theta_0 - \theta e_t = Y_t + \ell\theta_0 - \theta e_t,$$

- The forecasting error is

$$e_t(\ell) = e_{t+\ell} + (1-\theta)\sum_{i=1}^{\ell-1} e_{t+\ell-i} \qquad \text{for} \quad \ell \geq 1.$$

  with the variance $Var(e_t(\ell)) = \sigma^2\left[1 + (\ell - 1)(1-\theta)^2\right]$

- If $\theta_0 = 0$, we can write

$$\widehat{Y}_t(1) = (1-\theta)Y_t + (1-\theta)\theta Y_{t-1} + (1-\theta)\theta^2 Y_{t-2} + \ldots$$

Non-stationary Models - IMA(2, 2)

- Consider the IMA(2, 2) as follows

$$Y_t = 2Y_{t-1} - Y_{t-2} + \theta_0 + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

- Then the forecasts are

$$\widehat{Y}_t(1) = 2Y_t - Y_{t-1} + \theta_0 - \theta_1 e_t - \theta_2 e_{t-1}$$
$$\widehat{Y}_t(2) = 2\widehat{Y}_t(1) - Y_t + \theta_0 - \theta_2 e_t$$
$$\widehat{Y}_t(\ell) = 2\widehat{Y}_t(\ell-1) - \widehat{Y}_t(\ell-2) + \theta_0, \qquad \text{for} \quad \ell > 2$$

- This can be written as

$$\widehat{Y}_t(\ell) = A + B\ell + \frac{\theta_0}{2}\ell^2$$

where

$$A = 2\widehat{Y}_t(1) - \widehat{Y}_t(2) + \theta_0, \qquad B = \widehat{Y}_t(2) - \widehat{Y}_t(1) - \frac{3}{2}\theta_0$$

Prediction Limits

- We have the forecast, the forecast error, and the variance of the forecasting error.

- Similarly, we want to assess the precision of our predictions.

- For the deterministic trends, recall that

$$\hat{Y}_t(\ell) = \mu_{t+\ell},$$

and

$$var(e_t(\ell)) = var(X_{t+\ell}) = \gamma_0,$$

- If the stochastic component is normally distributed, then

$$e_t(\ell) = Y_{t+\ell} - \hat{Y}_t(\ell) = X_{t+\ell} \sim N(0, \gamma_0).$$

Prediction Limits

- A $100(1-\alpha)\%$ prediction interval for $Y_{t+\ell}$ is

$$\widehat{Y}_t(\ell) \pm z_{1-\alpha/2}\sqrt{Var(e_t(\ell))}$$

- Or equivalently

$$\widehat{Y}_t(\ell) \pm z_{1-\alpha/2}\sqrt{\gamma_0}.$$

- If we use a linear trend the corrected forecasting error is

$$\gamma_0\left[1+\frac{1}{n}+c_{n,\ell}\right],$$

where

$$c_{n,\ell}=\frac{3(n+2\ell-1)^2}{n(n^2-1)}\approx\frac{3}{n},$$

for moderate lead ℓ and large n.

Prediction Limits

- For ARIMA models, the variance of the forecasting error is

$$var(e_t(\ell)) = \sigma^2 \sum_{j=0}^{\ell-1} \psi_j^2,$$

- However, σ^2 and ψ_j are unknown since they are certain functions of the parameters of the model which are unknown.

- For large sample sizes, these estimations will have little effect on the actual prediction limits.

- For an AR(1)

$$var(e_t(\ell)) = \sigma^2 \left[\frac{1-\phi^{2\ell}}{1-\phi^2}\right].$$

Forecast Weights and Exponentially Weighted Moving Averages (EWMA)

- For ARIMA models without moving average terms, the forecasts are explicitly determined from the observed series $Y_t, Y_{t-1}, \ldots, Y_1$.
- Any invertible ARIMA process

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \pi_3 Y_{t-3} + \ldots + e_t,$$

- The one-step forecast is

$$\hat{Y}_t(1) = \pi_1 Y_t + \pi_2 Y_{t-1} + \pi_3 Y_{t-2} + \ldots$$

- For any invertible ARIMA model, π is calculated by letting $\pi_0 = -1$ and using

$$\pi_j = \begin{cases} \sum_{i=1}^{\min(j,q)} \theta_i \pi_{j-i} + \phi_j & \text{for } 1 \leq j \leq p+d \\ \sum_{i=1}^{\min(j,q)} \theta_i \pi_{j-i} & \text{for } j > p+d \end{cases}$$

Forecast Weights and EWMA

- Consider the nonstationary IMA(1,1) model

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1},$$

where $p = 0$, $d = 1$, $q = 1$, with $\phi_1 = 1$.

- Thus,

$$\begin{aligned}
\pi_1 &= \theta\pi_0 + 1 = 1 - \theta \\
\pi_2 &= \theta\pi_1 = \theta(1-\theta) \\
&\vdots \\
\pi_j &= \theta\pi_{j-1} = (1-\theta)\theta^{j-1}, \quad \text{for } j \geq 1.
\end{aligned}$$

- Thus, we can rewrite the one-step forecast as

$$\hat{Y}_t(1) = (1-\theta)Y_t + (1-\theta)\theta Y_{t-1} + (1-\theta)\theta^2 Y_{t-2} + \ldots$$

Forecast Weights and EWMA

- Since the weights decrease exponentially, and

$$\sum_{j=1}^{\infty} \pi_j = (1-\theta)\sum_{j=1}^{\infty} \theta^{j-1} = \frac{1-\theta}{1-\theta} = 1,$$

- $\hat{Y}_t(1)$ is called an exponentially weighted moving average (EWMA).
- It can be shown that

$$\hat{Y}_t(1) = (1-\theta)Y_t + \theta \hat{Y}_{t-1}(1),$$

  and

$$\hat{Y}_t(1) = \hat{Y}_{t-1}(1) + (1-\theta)[Y_t - \hat{Y}_{t-1}(1)]$$

- $1-\theta$ is often referred to as the smoothing constant in EWMA.

Updating ARIMA Forecasts

- Suppose we have t observations at time t .
- We have an appropriate ARMA model for Y_t that is used to obtain the forecast for Y_{t+1}, Y_{t+2}, etc.
- At $t+1$, we observe Y_{t+1}.
- Now, we want to update our forecasts using the original value of Y_{t+1} and the forecasted value of it.
- The forecast error is:

$$e_t(\ell) = Y_{t+\ell} - \widehat{Y}_t(\ell) = \sum_{j=0}^{\ell-1} \psi_j e_{t+\ell-j}.$$

Updating ARIMA Forecasts

- The error can be written as

$$
\begin{aligned}
e_{t-1}(\ell+1) &= Y_{t-1+\ell+1} - \widehat{Y}_{t-1}(\ell+1) \\
&= \sum_{i=0}^{\ell} \psi_i e_{t-1+\ell+1-i} \\
&= \sum_{i=0}^{\ell} \psi_i e_{t+\ell-i} \\
&= \sum_{j=0}^{\ell-1} \psi_j e_{t+\ell-j} + \psi_\ell e_t \\
&= e_t(\ell) + \psi_\ell e_t.
\end{aligned}
$$

Updating ARIMA Forecasts

- Therefore, we have

$$e_{t-1}(\ell+1) = e_t(\ell) + \psi_\ell e_t$$

- In other words

$$\begin{aligned}
& Y_{t+\ell} - \widehat{Y}_{t-1}(\ell+1) \\
& = Y_{t+\ell} - \widehat{Y}_t(\ell) + \psi_\ell e_t \\
\Rightarrow \quad & \widehat{Y}_t(\ell) = \widehat{Y}_{t-1}(\ell+1) + \psi_\ell e_t \\
\Rightarrow \quad & \widehat{Y}_t(\ell) = \widehat{Y}_{t-1}(\ell+1) + \psi_\ell (Y_t - \widehat{Y}_{t-1}(1)) \\
t \to t+1 \Rightarrow \quad & \widehat{Y}_{t+1}(\ell) = \widehat{Y}_t(\ell+1) + \psi_\ell (Y_{t+1} - \widehat{Y}_t(1))
\end{aligned}$$

Forecasting Transformed Series

- If we use variance stabilizing transformation, after the forecasting, we need to convert the forecasts for the original series.
- For example, if we use log-transformation, $Z = \log(Y_t)$ then,

$$E[Y_{t+\ell}|Y_1, Y_2, \ldots, Y_t] \geq \exp\left(E[Z_{t+\ell}|Z_1, Z_2, \ldots, Z_t]\right)$$

- If $X \sim N(\mu, \sigma^2)$, then $E(e^X) = \exp(\mu + \sigma^2/2)$
- The MSE forecast for the original series is:

$$\exp\left(\widehat{Z}_t(\ell) + \frac{1}{2} Var[e_t(\ell)]\right)$$

where

$$\widehat{Z}_t(\ell) = E[Z_{t+\ell}|Z_1, Z_2, \ldots, Z_t],$$
$$Var[e_t(\ell)] = Var[Z_{t+\ell}|Z_1, Z_2, \ldots, Z_t]$$

Forecasting From ARMA Model: Remarks

- In general, we need a large t to have a better estimate and it is possible to check for model stability and check the forecasting ability of the model by withholding data.
- Seasonal patterns also need large t.
- Usually, you need 4 to 5 seasons to get reasonable estimates.
- Parsimonious models are very important.
- Easier to compute and interpret models and forecasts.
- Forecasts are less sensitive to deviations between parameters and estimates.