Forecasting with ARIMA Models: Part I

Video 33

STAT 485/685, Fall 2020, SFU

Sonja Isberg

Our Roadmap

1. **Key Ideas:** Fundamental concepts (Ch. 1-2), Estimating trends (Ch. 3), Models for stationary time series (Ch. 4), Models for non-stationary time series (Ch. 5)
2. **Building a Model:**
   - **Model specification (Ch. 6):** How do we choose between the different models that we know?
   - **Parameter estimation (Ch. 7):** Now that we've chosen a model, there will be parameters whose values are unknown. How do we estimate these parameter values?
   - **Model diagnostics (Ch. 8):** How good is our chosen model? Should we be using a different model?
3. **Forecasting (Ch. 9):** How can we forecast the values of a time series at future time points? Can we assess the precision of these forecasts?
4. Other topics, as time permits.

Video 33 Learning Objectives

By the end of this video, we should be able to:

- Obtain the equation for a forecast at a future time value, using an ARIMA model
- Calculate the forecast error variance for an ARIMA model
- Understand how the behaviour of the forecast error variance differs between stationary and non-stationary models

Forecasting: Introduction

Forecasting with time series involves several primary goals:

- Given a time series dataset $Y_1, Y_2, \ldots, Y_t$ (up to some time point t), forecast the values at some future time(s) $Y_{t+\ell}$ (where $\ell > 0$)
- Assess the precision of these forecasts, by obtaining information about the *forecast error*
- Use this information to construct *prediction intervals* for the forecasts

Forecasting: Introduction

Forecasting with time series involves several primary goals:

- Given a time series dataset $Y_1, Y_2, \dots, Y_t$ (up to some time point t), forecast the values at some future time(s) $Y_{t+\ell}$ (where $\ell > 0$)
- Assess the precision of these forecasts, by obtaining information about the *forecast error*
- Use this information to construct *prediction intervals* for the forecasts

Several important comments:

- Note the notation we will use for our dataset: $Y_1, Y_2, \dots, Y_t$ (i.e., up to t, not "n")
- Throughout this chapter, we will assume that all of the parameters $\{\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \mu, \sigma_e^2, \gamma_0, \text{etc.}\}$ are known
  - In practice, we replace these by their estimates; this doesn't change the results much

Minimum Mean Square Error Forecasting

We are interested in using our time series data $Y_1, Y_2, \ldots, Y_t$, to forecast the value of $Y_{t+\ell}$, which will occur ℓ time units into the future.

We refer to t as the **forecast origin**, and ℓ as the **lead time**.

We denote the forecast by $\hat{Y}_t(\ell)$.

Minimum Mean Square Error Forecasting

We are interested in using our time series data $Y_1, Y_2, \dots, Y_t$, to forecast the value of $Y_{t+\ell}$, which will occur ℓ time units into the future.

We refer to t as the **forecast origin**, and ℓ as the **lead time**.

We denote the forecast by $\hat{Y}_t(\ell)$.

The **minimum mean square error (MSE) predictor** of a variable Y is the predictor $\hat{Y}$ that minimizes

$$E[(Y - \hat{Y})^2]$$

(Appendices E & F:) If a random variable Y is to be predicted by a function of the random variables $X_1, X_2, \dots, X_n$, then the minimum MSE predictor is

$$\hat{Y} = E(Y|X_1, X_2, \dots, X_n)$$

Minimum Mean Square Error Forecasting

We are interested in using our time series data $Y_1, Y_2, \ldots, Y_t$, to forecast the value of $Y_{t+\ell}$, which will occur ℓ time units into the future.

We refer to t as the **forecast origin**, and ℓ as the **lead time**.

We denote the forecast by $\hat{Y}_t(\ell)$.

The **minimum mean square error (MSE) predictor** of a variable Y is the predictor $\hat{Y}$ that minimizes

$$E[(Y - \hat{Y})^2]$$

(Appendices E & F:) If a random variable Y is to be predicted by a function of the random variables $X_1, X_2, \ldots, X_n$, then the minimum MSE predictor is

$$\hat{Y} = E(Y | X_1, X_2, \ldots, X_n)$$

Therefore, the minimum MSE forecast of $Y_{t+\ell}$ is

$$\hat{Y}_t(\ell) = E(Y_{t+\ell} | Y_1, Y_2, \ldots, Y_t)$$

Forecasting for ARMA(p,q)

Recall the ARMA(p,q) model with non-zero constant mean:

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Replace t by $t + \ell$:

$$\begin{aligned} Y_{t+\ell} = \theta_0 + \phi_1 Y_{t+\ell-1} + \phi_2 Y_{t+\ell-2} + \cdots + \phi_p Y_{t+\ell-p} + e_t \\ - \theta_1 e_{t+\ell-1} - \theta_2 e_{t+\ell-2} - \cdots - \theta_q e_{t+\ell-q} \end{aligned}$$

Forecasting for ARMA(p,q)

Recall the ARMA(p,q) model with non-zero constant mean:

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Replace t by $t+\ell$:

$$\begin{aligned} Y_{t+\ell} = \theta_0 + \phi_1 Y_{t+\ell-1} + \phi_2 Y_{t+\ell-2} + \cdots + \phi_p Y_{t+\ell-p} + e_t \\ - \theta_1 e_{t+\ell-1} - \theta_2 e_{t+\ell-2} - \cdots - \theta_q e_{t+\ell-q} \end{aligned}$$

Take $E(\cdot | Y_1, Y_2, \ldots, Y_t)$ of both sides:

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

Forecasting for ARMA(p,q) (cont'd)

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

Let's first address the Y's:

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

where, actually,

$$\hat{Y}_t(j) = \begin{cases} \text{the actual forecast } \hat{Y}_t(j) & \text{if } j > 0 \\ \text{the observed value } Y_{t+j} & \text{if } j \leq 0 \end{cases}$$

Forecasting for ARMA(p,q) (cont'd)

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

Recall: $e_{t+\ell} \perp\!\!\!\perp \{Y_1, \ldots, Y_t\}$ since $\ell > 0$.

Therefore: $E(e_{t+\ell}|Y_1, Y_2, \ldots, Y_t) = E(e_{t+\ell}) = 0$.

Forecasting for ARMA(p,q) (cont'd)

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

Recall: $e_{t+\ell} \perp\!\!\!\perp \{Y_1, \ldots, Y_t\}$ since $\ell > 0$.

Therefore: $E(e_{t+\ell}|Y_1, Y_2, \ldots, Y_t) = E(e_{t+\ell}) = 0$.

So:

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

Forecasting for ARMA(p,q) (cont'd)

By the same reasoning, for all $j > 0$: $E(e_{t+j}|Y_1, Y_2, \ldots, Y_t) = E(e_{t+j}) = 0$.

However, for $j \leq 0$: $E(e_{t+j}|Y_1, Y_2, \ldots, Y_t) = e_{t+j}$.

Forecasting for ARMA(p,q) (cont'd)

By the same reasoning, for all $j > 0$: $E(e_{t+j}|Y_1, Y_2, \dots, Y_t) = E(e_{t+j}) = 0$.

However, for $j \leq 0$: $E(e_{t+j}|Y_1, Y_2, \dots, Y_t) = e_{t+j}$.

Therefore, our final forecast equation is:

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

where

$$\hat{Y}_t(j) = \begin{cases} \text{the actual forecast } \hat{Y}_t(j) & \text{if } j > 0 \\ \text{the observed value } Y_{t+j} & \text{if } j \leq 0 \end{cases}$$

and

$$E(e_{t+j}|Y_1, Y_2, \dots, Y_t) = \begin{cases} 0 & \text{if } j > 0 \\ e_{t+j} & \text{if } j \leq 0 \end{cases}$$

Alternative formulation: Remove θ_0, and replace all $\hat{Y}_t(j)$'s by $(\hat{Y}_t(j) - \mu)$.

Forecasting for ARMA(p,q): Forecast Error

The **forecast error** is given by

$$e_t(\ell) = Y_{t+\ell} - \hat{Y}_t(\ell)$$

Forecasting for ARMA(p,q): Forecast Error

The **forecast error** is given by

$$e_t(\ell) = Y_{t+\ell} - \hat{Y}_t(\ell)$$

It can be shown (pg. 201) that this is equal to

$$e_t(\ell) = \sum_{j=0}^{\ell-1} \psi_j e_{t+\ell-j}$$

where $\psi_0 = 1$, and $\psi_1, \psi_2, \ldots, \psi_{\ell-1}$ are the first $\ell - 1$ coefficients in the general linear process formulation of the ARMA model.

Therefore (due to independence):

$$Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2$$

Forecasting for ARMA(p,q): Forecast Error

The **forecast error** is given by

$$e_t(\ell) = Y_{t+\ell} - \hat{Y}_t(\ell)$$

It can be shown (pg. 201) that this is equal to

$$e_t(\ell) = \sum_{j=0}^{\ell-1} \psi_j e_{t+\ell-j}$$

where $\psi_0 = 1$, and $\psi_1, \psi_2, \ldots, \psi_{\ell-1}$ are the first $\ell - 1$ coefficients in the general linear process formulation of the ARMA model.

Therefore (due to independence):

$$Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2$$

Recall, for a general linear process: $\gamma_0 = Var(Y_t) = \sigma_e^2 \sum_{j=0}^{\infty} \psi_j^2$ (Video 14).

Therefore, for large ℓ: $Var(e_t(\ell)) \approx \gamma_0$.

Example: Forecasting for AR(1)

AR(1) model with non-zero constant mean (alternative formulation):

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

Recall:

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

Therefore:

$$\hat{Y}_t(\ell) - \mu = \phi[\hat{Y}_t(\ell-1) - \mu]$$

$$\hat{Y}_t(\ell) = \mu + \phi[\hat{Y}_t(\ell-1) - \mu] \quad \text{for all } \ell \geq 1$$

Example: Forecasting for AR(1)

AR(1) model with non-zero constant mean (alternative formulation):

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

Recall:

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

Therefore:

$$\hat{Y}_t(\ell) - \mu = \phi[\hat{Y}_t(\ell - 1) - \mu]$$

$$\hat{Y}_t(\ell) = \mu + \phi[\hat{Y}_t(\ell - 1) - \mu] \quad \text{for all } \ell \geq 1$$

We start with (i.e., for $\ell = 1$):

$$\hat{Y}_t(1) = \mu + \phi[\hat{Y}_t(0) - \mu] = \mu + \phi(Y_t - \mu)$$

Example: Forecasting for AR(1) (cont'd)

$$\hat{Y}_t(\ell) = \mu + \phi[\hat{Y}_t(\ell - 1) - \mu] \quad \text{for all } \ell \geq 1$$

This can be repeated recursively to obtain the explicit expression:

$$\begin{aligned}
\hat{Y}_t(\ell) &= \mu + \phi[\hat{Y}_t(\ell - 1) - \mu] \\
&= \mu + \phi\Big[\mu + \phi[\hat{Y}_t(\ell - 2) - \mu] - \mu\Big] = \mu + \phi^2[\hat{Y}_t(\ell - 2) - \mu] \\
&\vdots \\
&= \mu + \phi^\ell(Y_t - \mu)
\end{aligned}$$

Note: Since, for a stationary process $|\phi| < 1$:

$$\hat{Y}_t(\ell) \approx \mu \quad \text{for large } \ell$$

Example: Forecasting for AR(1) – Forecast Error

The variance of the ℓ-steps-ahead forecast error is

$$
\begin{aligned}
Var(e_t(\ell)) &= \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 \\
&= \sigma_e^2 \sum_{j=0}^{\ell-1} (\phi^j)^2 \\
&= \sigma_e^2 \left(\frac{1-\phi^{2\ell}}{1-\phi^2} \right)
\end{aligned}
$$

Example: Forecasting for AR(1) – Forecast Error

The variance of the ℓ-steps-ahead forecast error is

$$
\begin{aligned}
Var(e_t(\ell)) &= \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 \\
&= \sigma_e^2 \sum_{j=0}^{\ell-1} (\phi^j)^2 \\
&= \sigma_e^2 \left(\frac{1-\phi^{2\ell}}{1-\phi^2} \right)
\end{aligned}
$$

Note: For large ℓ:

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

(Recall from Video 16 that this is the variance of an AR(1) process.)

Example: Forecasting for MA(1)

MA(1) model with non-zero constant mean (alternative formulation):

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

Recall:

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

Therefore:

$$\hat{Y}_t(\ell) - \mu = -\theta E(e_{t+\ell-1}|Y_1, Y_2, \ldots, Y_t)$$

$$\hat{Y}_t(\ell) = \mu - \theta E(e_{t+\ell-1}|Y_1, Y_2, \ldots, Y_t)$$

Example: Forecasting for MA(1) (cont'd)

$$\hat{Y}_t(\ell) = \mu - \theta E(e_{t+\ell-1} | Y_1, Y_2, \ldots, Y_t)$$

Also recall from the definition:

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

Therefore:

$$\hat{Y}_t(\ell) = \begin{cases} \mu - \theta e_{t+\ell-1} & \text{for } \ell \leq 1 \\ \mu & \text{for } \ell > 1 \end{cases}$$

i.e.,

$$\hat{Y}_t(\ell) = \begin{cases} \mu - \theta e_t & \text{for } \ell = 1 \\ \mu & \text{for } \ell > 1 \end{cases}$$

Problem: e_t is not observed! We will cover calculation of e_t in Video 34.

Example: Forecasting for MA(1) – Forecast Error

The variance of the ℓ-steps-ahead forecast error is

$$Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 = \begin{cases} \sigma_e^2 & \text{for } \ell = 1 \\ \sigma_e^2(1+\theta^2) & \text{for } \ell > 1 \end{cases}$$

Example: Forecasting for MA(1) – Forecast Error

The variance of the ℓ-steps-ahead forecast error is

$$Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 = \begin{cases} \sigma_e^2 & \text{for } \ell = 1 \\ \sigma_e^2(1+\theta^2) & \text{for } \ell > 1 \end{cases}$$

Note: For large ℓ:

$$Var(e_t(\ell)) = \sigma_e^2(1+\theta^2) = \gamma_0$$

(Recall from Video 15 that this is the variance of an MA(1) process.)

Forecasting for ARIMA(p,d,q)

For an ARIMA(p,d,q) process:

1. Write out the difference equation form for the non-stationary process Y_t
2. The same equations for the forecasts $\hat{Y}_t(\ell)$ and forecast errors $e_t(\ell)$ still apply
3. However, the forecast error variance

$$Var(e_t(\ell)) = \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2$$

will grow without bound as ℓ increases

Example: Forecasting for Random Walk with Drift

Random walk with drift:

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

(This is just an "AR(1)" model with $\phi = 1$. Also: It is an ARIMA(0,1,0) model.)

Example: Forecasting for Random Walk with Drift

Random walk with drift:

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

(This is just an "AR(1)" model with $\phi = 1$. Also: It is an ARIMA(0,1,0) model.)

Recall:

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

Therefore:

$$\hat{Y}_t(\ell) = \theta_0 + \hat{Y}_t(\ell - 1) \quad \text{for all } \ell \geq 1$$

We start with (i.e., for $\ell = 1$):

$$\hat{Y}_t(1) = \theta_0 + \hat{Y}_t(0) = \theta_0 + Y_t$$

Example: Forecasting for Random Walk with Drift (cont'd)

$$\hat{Y}_t(\ell) = \theta_0 + \hat{Y}_t(\ell - 1) \quad \text{for all } \ell \geq 1$$

This can be repeated recursively to obtain the explicit expression:

$$\begin{aligned}
\hat{Y}_t(\ell) &= \theta_0 + \hat{Y}_t(\ell - 1) \\
&= \theta_0 + [\theta_0 + \hat{Y}_t(\ell - 2)] = 2\theta_0 + \hat{Y}_t(\ell - 2) \\
&\vdots \\
&= \ell\theta_0 + Y_t
\end{aligned}$$

Note: In contrast to AR(1), this forecast does *not* converge to μ for large ℓ. Instead, it follows a straight line with slope θ_0 for all ℓ.

Example: Forecasting for Random Walk with Drift – Forecast Error

The variance of the ℓ-steps-ahead forecast error is

$$
\begin{aligned}
Var(e_t(\ell)) &= \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 \\
&= \sigma_e^2 \sum_{j=0}^{\ell-1} (\phi^j)^2 \\
&= \sigma_e^2 \sum_{j=0}^{\ell-1} 1 \\
&= \ell \sigma_e^2
\end{aligned}
$$

Example: Forecasting for Random Walk with Drift – Forecast Error

The variance of the ℓ-steps-ahead forecast error is

$$
\begin{aligned}
Var(e_t(\ell)) &= \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 \\
&= \sigma_e^2 \sum_{j=0}^{\ell-1} (\phi^j)^2 \\
&= \sigma_e^2 \sum_{j=0}^{\ell-1} 1 \\
&= \ell \sigma_e^2
\end{aligned}
$$

Note: Unlike in AR(1), for large ℓ:

$$Var(e_t(\ell)) \to \infty$$

Example: Forecasting for IMA(2,2)

Suppose $\{Y_t\}$ is an IMA(2,2) model with a drift term. Then, $\{W_t\} = \{\nabla^2 Y_t\}$ is an MA(2) model:

The difference equation form for $\{Y_t\}$ is:

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

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

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

(This is just an "ARMA(2,2)" model with $\phi_1 = 2$ & $\phi_2 = -1$)

Example: Forecasting for IMA(2,2) (cont'd)

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

Recall:

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

Therefore:

$$\begin{aligned}\hat{Y}_t(\ell) = \theta_0 &+ 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) \\ &- \theta_1 E(e_{t+\ell-1}|Y_1, Y_2, \ldots, Y_t) - \theta_2 E(e_{t+\ell-2}|Y_1, Y_2, \ldots, Y_t)\end{aligned}$$

Also recall from the definition:

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

Example: Forecasting for IMA(2,2) (cont'd)

Therefore:

$$E(e_{t+\ell-1}|Y_1, Y_2, \ldots, Y_t) = \begin{cases} e_{t+\ell-1} & \text{if } \ell \leq 1 \\ 0 & \text{if } \ell > 1 \end{cases} = \begin{cases} e_t & \text{if } \ell = 1 \\ 0 & \text{if } \ell > 1 \end{cases}$$

and

$$E(e_{t+\ell-2}|Y_1, Y_2, \ldots, Y_t) = \begin{cases} e_{t+\ell-2} & \text{if } \ell \leq 2 \\ 0 & \text{if } \ell > 2 \end{cases} = \begin{cases} e_{t-1} & \text{if } \ell = 1 \\ e_t & \text{if } \ell = 2 \\ 0 & \text{if } \ell > 2 \end{cases}$$

Example: Forecasting for IMA(2,2) (cont'd)

Therefore:

$$E(e_{t+\ell-1}|Y_1, Y_2, \ldots, Y_t) = \begin{cases} e_{t+\ell-1} & \text{if } \ell \leq 1 \\ 0 & \text{if } \ell > 1 \end{cases} = \begin{cases} e_t & \text{if } \ell = 1 \\ 0 & \text{if } \ell > 1 \end{cases}$$

and

$$E(e_{t+\ell-2}|Y_1, Y_2, \ldots, Y_t) = \begin{cases} e_{t+\ell-2} & \text{if } \ell \leq 2 \\ 0 & \text{if } \ell > 2 \end{cases} = \begin{cases} e_{t-1} & \text{if } \ell = 1 \\ e_t & \text{if } \ell = 2 \\ 0 & \text{if } \ell > 2 \end{cases}$$

So:

$$\hat{Y}_t(\ell) = \begin{cases} \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) - \theta_1 e_t - \theta_2 e_{t-1} & \text{for } \ell = 1 \\ \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) - \theta_2 e_t & \text{for } \ell = 2 \\ \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) & \text{for } \ell > 2 \end{cases}$$

Example: Forecasting for IMA(2,2) (cont'd)

$$\hat{Y}_t(\ell) = \begin{cases} \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) - \theta_1 e_t - \theta_2 e_{t-1} & \text{for } \ell = 1 \\ \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) - \theta_2 e_t & \text{for } \ell = 2 \\ \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) & \text{for } \ell > 2 \end{cases}$$

i.e.,

$$\hat{Y}_t(\ell) = \begin{cases} \theta_0 + 2Y_t - Y_{t-1} - \theta_1 e_t - \theta_2 e_{t-1} & \text{for } \ell = 1 \\ \theta_0 + 2\hat{Y}_t(1) - Y_t - \theta_2 e_t & \text{for } \ell = 2 \\ \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) & \text{for } \ell > 2 \end{cases}$$

Example: Forecasting for IMA(2,2) (cont'd)

$$\hat{Y}_t(\ell) = \begin{cases} \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) - \theta_1 e_t - \theta_2 e_{t-1} & \text{for } \ell = 1 \\ \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) - \theta_2 e_t & \text{for } \ell = 2 \\ \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) & \text{for } \ell > 2 \end{cases}$$

i.e.,

$$\hat{Y}_t(\ell) = \begin{cases} \theta_0 + 2Y_t - Y_{t-1} - \theta_1 e_t - \theta_2 e_{t-1} & \text{for } \ell = 1 \\ \theta_0 + 2\hat{Y}_t(1) - Y_t - \theta_2 e_t & \text{for } \ell = 2 \\ \theta_0 + 2\hat{Y}_t(\ell-1) - \hat{Y}_t(\ell-2) & \text{for } \ell > 2 \end{cases}$$

We won't derive the expression for $Var(e_t(\ell))$ here. It can be shown to be a cubic function of ℓ.

Final Comments

That's all for now!

In this video, we've learned how to obtain forecasts for the general ARIMA(p,d,q) model.

We've also practiced calculating these forecasts for several different ARIMA models.

Finally, we learned how to calculate forecast error variances, which will be useful to us in constructing prediction intervals.

Coming Up Next: More topics in forecasting!

Thank you!

References:

[1] Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: with applications in R.* Springer Science and Business Media.

[2] Chan, K. S., & Ripley, B. (2020). TSA: Time Series Analysis. R package version 1.2.1.