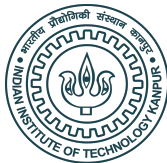


# Lecture 25

## ARIMA models

Arnab Hazra



# Integrated models

- ▶ In many situations, time series can be thought of as being composed of two components,  $X_t = \mu_t + Y_t$ , where  $\mu_t = \beta_0 + \beta_1 t$  and  $Y_t$  is stationary.
- ▶ Differencing such a process will lead to a stationary process  $\nabla X_t = X_t - X_{t-1} = \beta_1 + Y_t - Y_{t-1} = \beta_1 + \nabla Y_t$ .
- ▶ If  $\mu_t$  is a  $k$ -th order polynomial,  $\mu_t = \sum_{j=0}^k \beta_j t^j$ , then the differenced series  $\nabla^k X_t$  is stationary.

## Integrated models (contd.)

- ▶ We have seen that if  $X_t = X_{t-1} + W_t$ , then by differencing  $X_t$ , we found that  $\nabla X_t = W_t$  is stationary.
- ▶ Another model that leads to first differencing is  $\mu_t = \mu_{t-1} + V_t$ , where  $X_t = \mu_t + Y_t$  and  $V_t$  is stationary. In this case,  $\nabla X_t = V_t + \nabla Y_t$  is stationary.
- ▶ Stochastic trend models can also lead to higher order differencing. For example, suppose  $\mu_t = \mu_{t-1} + V_t$  and  $V_t = V_{t-1} + e_t$ , where  $e_t$  is stationary. Then,  $\nabla X_t = V_t + \nabla Y_t$  is not stationary, but  $\nabla^2 X_t = e_t + \nabla^2 Y_t$  is.

# ARIMA models: definition

- ▶ A process  $X_t$  is said to be  $\text{ARIMA}(p, d, q)$  if  $\nabla^d X_t = (1 - B)^d X_t$  is  $\text{ARMA}(p, q)$ .
- ▶ In general, we will write the model as  $\phi(B)(1 - B)^d X_t = \theta(B)W_t$ .
- ▶ If  $E(\nabla^d X_t) = \mu$ , we write the model as  $\phi(B)(1 - B)^d X_t = \delta + \theta(B)W_t$ , where  $\delta = \mu(1 - \phi_1 - \dots - \phi_p)$ .

# Random Walk with Drift

- ▶ We consider the model  $X_t = \delta + X_{t-1} + W_t$ , for  $t = 1, 2, \dots$ , and  $X_0 = 0$ , which we can write as an ARIMA(0, 1, 0) model.

- ▶ Given data  $X_1, \dots, X_T$ , the one-step-ahead forecast is given by

$$X_{T+1}^T = E(X_{T+1}|X_T, \dots, X_1) = E(\delta + X_T + W_{T+1}|X_T, \dots, X_1) = \delta + X_T.$$

- ▶ The two-step-ahead forecast is given by

$$X_{T+2}^T = E(X_{T+2}|X_T, \dots, X_1) = E(\delta + X_{T+1} + W_{T+2}|X_T, \dots, X_1) = 2\delta + X_T.$$

- ▶ Consequently, the  $m$ -step-ahead forecast, for  $m = 1, 2, \dots$ , is

$$X_{T+m}^T = m\delta + X_T.$$

## Random Walk with Drift (contd.)

- ▶ To obtain the forecast errors, we rewrite  $X_t = t\delta + \sum_{j=1}^T W_j$ .
- ▶ We may write

$$X_{T+m} = (T+m)\delta + \sum_{j=1}^{T+m} W_j = m\delta + \left\{ T\delta + \sum_{j=1}^T W_j \right\} + \sum_{j=T+1}^{T+m} W_j = m\delta + X_T + \sum_{j=T+1}^{T+m} W_j.$$

- ▶ From this it follows that the  $m$ -step-ahead prediction error is given by

$$P_{T+m}^T = E[(X_{T+m} - X_{T+m}^T)^2] = E \left[ \left( \sum_{j=T+1}^{T+m} W_j \right)^2 \right] = E \left[ \sum_{j=T+1}^{T+m} W_j^2 \right] = m\sigma_W^2.$$

- ▶ Hence, unlike the stationary case, as the forecast horizon grows, the prediction errors increase without bound and the forecasts follow a straight line with slope  $\delta$  starting from  $X_T$ .

# Building ARIMA Models (Lab exercise)

Basic steps to fitting ARIMA models to time series data are

- ▶ Plotting the data
- ▶ Possibly transforming the data
- ▶ Identifying the dependence orders of the model
- ▶ Parameter estimation
- ▶ Diagnostics
- ▶ Model comparison

Thank you!