Lecture 26

ARIMA and SARIMA models

Arnab Hazra

ARIMA models: definition (recap)

- A process X_t is said to be ARIMA(p, d, q) if $\nabla^d X_t = (1 - B)^d X_t$ is 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 $\delta = \mu(1 - \phi_1 - \ldots - \phi_p)$.

Random Walk with Drift (recap)

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

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

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

- The two-step-ahead forecast is given by

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

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

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

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 δ starting from X_T.

Regression with Autocorrelated Errors

- We discuss the modifications to the regression model when the errors are correlated.
- That is, consider the regression model

$$Y_t = \sum_{j=1}^{r} \beta_j z_{tj} + X_t$$

- Here X_t is a process with some covariance function $\gamma_X(s,t)$.
- In ordinary least squares, the assumption is that X_t is white Gaussian noise, in which case $\gamma_X(s,t) = 0$ for $s \neq t$ and $\gamma_X(t,t) = \sigma^2$, independent of t.
- If this is not the case, then weighted least squares should be used.

Regression with Autocorrelated Errors (contd.)

- Write the model in vector notation,

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{X},$$

  where $\boldsymbol{Y} = (Y_1, \ldots, Y_T)'$ and $\boldsymbol{X} = (X_1, \ldots, X_T)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)'$, and $\boldsymbol{Z}$ is the $n \times r$ matrix composed of the input variables.

- The weighted estimate of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{\Gamma}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Gamma}^{-1}\boldsymbol{Y}$, and the variance-covariance matrix of the estimator is $V(\hat{\boldsymbol{\beta}}) = (\boldsymbol{Z}'\boldsymbol{\Gamma}^{-1}\boldsymbol{Z})^{-1}$.

- In case of $\boldsymbol{X}$ being white noise, $\boldsymbol{\Gamma} = \sigma^2\boldsymbol{I}_T$ and $\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$.

Optimization

- It is possible to assume a stationary covariance structure for the error process X_t that corresponds to a linear process and try to find an ARMA representation for X_t.
- If the error process is ARMA(p, q), i.e., $\phi(B)X_t = \theta(B)W_t$, we transform by $\pi(B)X_t = W_t$, where $\pi(B) = \theta(B)^{-1}\phi(B)$.
- Multiplying the regression equation $Y_t = \sum_{j=1}^{r} \beta_j z_{tj} + X_t$ through by the transformation $\pi(B)$ yields,

$$\pi(B)Y_t = \sum_{j=1}^{r} \beta_j \pi(B) z_{tj} + \pi(B)X_t,$$

where $\pi(B)X_t = W_t$.
- In this case the error sum of squares depends on $\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}$:

$$S(\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{t=1}^{T} W_t^2 = \sum_{t=1}^{T} \left[\pi(B)Y_t - \sum_{j=1}^{r} \beta_j \pi(B) z_{tj} \right]^2.$$

Practical procedure

i. First, run an ordinary regression of Y_t on $z_{t1}, \ldots, z_{tr}$ (acting as if the errors are uncorrelated). Retain the residuals, $\widehat{X}_t = Y_t - \sum_{j=1}^{r} \beta_j z_{tj}$.

ii. Identify ARMA model(s) for the residuals $\widehat{X}_t$.

iii. Run weighted least squares (or MLE) on the regression model with autocorrelated errors using the model specified in step (ii).

iv. Inspect the residuals $\widehat{W}_t$ for whiteness, and adjust the model if necessary.

Multiplicative Seasonal ARIMA Models: motivation

- We next discuss several modifications to the ARIMA model to account for seasonal and nonstationary behavior.
- Often, the dependence on the past tends to occur most strongly at multiples of some underlying seasonal lag s.
- For example, with monthly economic data, there is a strong yearly component occurring at lags that are multiples of $s = 12$, because of the strong connections of all activity to the calendar year.
- Data taken quarterly will exhibit the yearly repetitive period at $s = 4$ quarters.
- Natural phenomena such as temperature also have strong components corresponding to seasons.
- Hence, the natural variability of many physical, biological, and economic processes tends to match with seasonal fluctuations.

Seasonal AR operator and seasonal MA operator

- We first introduce AR and MA polynomials that identify with the seasonal lags.
- The resulting pure seasonal autoregressive moving average model, say, ARMA$(P, Q)_s$, then takes the form

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)W_t,$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}$$

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs}$$

  where the operators $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are the seasonal AR and MA operators of orders P and Q, respectively, with seasonal period s.
- ARMA$(P, Q)_s$ is causal only when the roots of $\Phi_P(z^s)$ lie outside the unit circle.
- It is invertible only when the roots of $\Theta_Q(z^s)$ lie outside the unit circle.

Definition of SARIMA models

- The multiplicative seasonal autoregressive integrated moving average model, or SARIMA model is given by

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d X_t = \delta + \Theta_Q(B^s)\theta(B)W_t,$$

  where W_t is the usual Gaussian white noise process.
- The general model is denoted as ARIMA$(p, d, q) \times (P, D, Q)_s$.
- The ordinary AR and MA components are represented by polynomials $\phi(B)$ and $\theta(B)$ of orders p and q, respectively.
- The seasonal AR and MA components are represented by $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ of orders P and Q.
- Ordinary and seasonal difference components are represented by $\nabla^d = (1 - B)^d$ and $\nabla_s^D = (1 - B^s)^D$.

Air passenger data

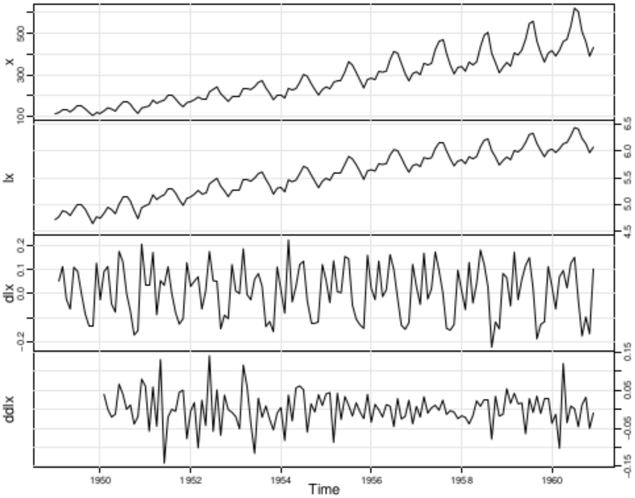

Fig. 3.22. *R data set* `AirPassengers`*, which are the monthly totals of international airline passengers* `x`*, and the transformed data:* `lx` $= \log x_t$, `dlx` $= \nabla \log x_t$, *and* `ddlx` $= \nabla_{12}\nabla \log x_t$.

ACF and PACF

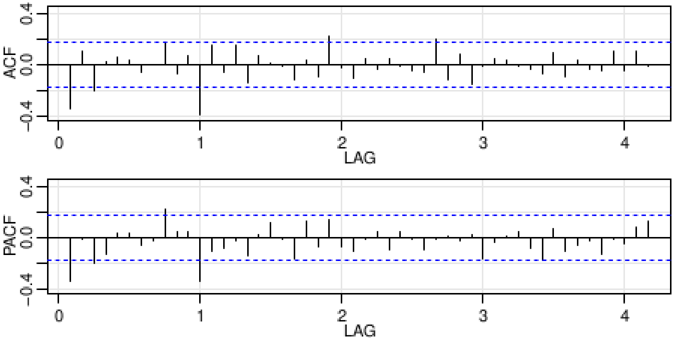

Fig. 3.23. Sample ACF and PACF of `ddlx` *($\nabla_{12}\nabla \log x_t$).*

- Here ARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ and ARIMA$(1, 1, 0) \times (0, 1, 1)_{12}$ appear to be reasonable models.

Thank you!