Statistics 3

Hadrien Montanelli

1 Times series basics

Characteristics to consider first

- Is there a *trend*, *i.e.*, on average, the measurements tend to increase (or decrease) over time?
- Is there *seasonality*, *i.e.*, a regularly repeating pattern of highs and lows related to calendar time?
- Are there *outliers*?
- Is there a *period* unrelated to seasonality factors?
- Is there *constant variance* over time?

Weakly stationary series A series X_t is said to be *weakly stationary* if $\mathrm{E}(X_t)$, $\mathrm{var}(X_t)$, and $\mathrm{cov}(X_t, X_{t-1})$ do not depend on time t.

Autocorrelation function The *autocorrelation function* is defined by

$$\rho(t_1, t_2) = \frac{\mathrm{cov}(X_{t_1}, X_{t_2})}{\sqrt{\mathrm{var}(X_{t_1})\mathrm{var}(X_{t_2})}}.$$

For a *stationary* series, it only depends on lag h,

$$\rho_h = \frac{\mathrm{cov}(X_t, X_{t-h})}{\mathrm{var}(X_t)}.$$

Sample autocorrelation function The *sample autocorrelation function* is given by

$$\hat{\rho}_h = \frac{\displaystyle\frac{1}{T}\sum_{t=1}^{T-h}(X_{t+h} - \overline{X})(X_t - \overline{X})}{\displaystyle\frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{X})^2}, \quad \overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t.$$

Partial autocorrelation function The *partial autocorrelation function* is defined by

$$\frac{\mathrm{cov}(X_t, X_{t-h}|X_{t-h+1}, \dots, X_{t-1})}{\sqrt{\mathrm{var}(X_t|X_{t-h+1}, \dots, X_{t-1})\mathrm{var}(X_{t-h}|X_{t-h+1}, \dots, X_{t-1})}}.$$

Partial autocorrelation function The *sample partial autocorrelation function* is the last component of

$$\Gamma_h^{-1}(\hat{\rho}_1, \dots, \hat{\rho}_h)^T, \quad \text{with} \quad \Gamma_h = [\hat{\rho}_{i-j}]_{i,j=1,\dots,h}.$$

2 AR models

$\mathrm{AR}(1)$ **model** The AR(1) model,

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2),$$

has mean, variance and autocorrelation

$$\mathrm{E}(X_t) = \frac{\varphi_0}{1-\varphi_1}, \;\; \mathrm{var}(X_t) = \frac{\sigma_\epsilon^2}{1-\varphi_1^2}, \;\; \text{and} \;\; \rho_h = \varphi_1^h.$$

$\mathrm{AR}(p)$ **models**

$$X_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2)$$

Least squares for $\mathrm{AR}(p)$

$$\begin{bmatrix} X_p \\ X_{p+1} \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} 1 & X_{p-1} & X_{p-2} & \dots & X_1 \\ 1 & X_p & X_{p-1} & \dots & X_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n-1} & \dots & \dots & X_{n-p} \end{bmatrix} \begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix}$$

3 MA models

$\mathrm{MA}(1)$ **model** The MA(1) model,

$$X_t = \theta_0 + \theta_1 \epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\epsilon^2),$$

has mean, variance and autocorrelation

$$\mathrm{E}(X_t) = \theta_0, \;\; \mathrm{var}(X_t) = (1+\theta_1^2)\sigma_\epsilon^2, \;\; \text{and} \;\; \rho_1 = \frac{\theta_1}{1+\theta_1^2},$$

with $\rho_h = 0$ for $h \geq 2$.

$\mathrm{MA}(q)$ **models**

$$X_t = \theta_0 + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$$

4 ARIMA models

$\mathrm{ARIMA}(p, d, q)$ **models**

$$Z_t = (1-L)^d X_t, \quad LX_t = X_{t-1},$$

$$Z_t = \mu + \sum_{i=1}^{p} \varphi_i Z_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$$

5 Identifying and testing a model

Identifying

- ACF $\downarrow 0$, p non-zeros in the PACF: try $\mathrm{AR}(p)$.
- PACF $\downarrow 0$, q non-zeros in the ACF: try $\mathrm{MA}(q)$.
- Both ACF and PACF $\downarrow 0$: try $\mathrm{ARMA}(p, q)$.
- ACF and PACF do not $\downarrow 0$: try $\mathrm{ARIMA}(p, d, q)$.

Testing

- Look at the significance of the coefficients.
- Look at the ACF of the residuals. For a good model, all ACFs should be non-significant.
- Look at Box–Pierce (Ljung) tests for possible residual autocorrelation at various lags.
- If non-constant variance is a concern, look at a plot of residuals versus fits.