Time Series Analysis

Lecture 4: ARIMA models-1, Estimation

Tohid Ardeshiri

Linköping University
Division of Statistics and Machine Learning

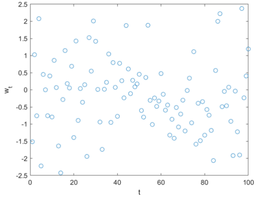
September 13, 2019

White noise

Simplest and most random time series: white noise

- w_t uncorrelated $E(w_t w_{t-h}) = 0$ for all $h \neq 0$
$$w_t \sim wn(0, \sigma_w^2)$$
- w_t white independent noise: independent and identically distributed
independence: $f(w_t, w_{t-h}) = f(w_t)f(w_{t-h})$
$$w_t \sim iid(0, \sigma_w^2)$$
- w_t white normal noise: independent and identically normal distributed

$$w_t \sim iidN(0, \sigma_w^2)$$



Autocovariance and ACF

- Autocovariance function

$$\gamma(s,t) = \mathrm{cov}(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]$$

  Note $\mathrm{var}(x_t) = \gamma(t,t)$

- Autocorrelation function (ACF)

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

Useful fact: If $U = \sum_{j=1}^{m} a_j x_j$ and

$$V = \sum_{k=1}^{r} b_k y_k$$

$$\mathrm{cov}(U,V) = \sum_{j=1}^{m} \sum_{k=1}^{r} a_j b_k \, \mathrm{cov}(x_j, y_k)$$

Stationarity

- Time series x_t is weakly stationary (stationary) if
  - $Ex_t = const$
  - $\gamma(s, t) = \gamma(|s - t|)$
  - $\text{var}(x_t) < \infty$
- $\gamma(t, t + h) = \gamma(|t + h - t|) = \gamma(h)$
  - Autocovariance depends on lag only!
- Autocovariance for stationary process $\gamma(h) = \text{cov}(x_t, x_{t+h})$
- ACF for stationary process $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$

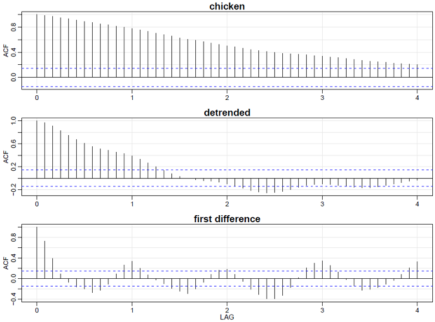
Sample ACF

Theorem: Under weak conditions, if x_t is white noise and $n \to \infty$ then $\hat{\rho}(h)$ is approximately $N(0, \frac{1}{n})$

Consequence: If some $|\hat{\rho}(h)| > \frac{2}{\sqrt{n}}$ then the time series is not a white noise (with approximately 95 % confidence).

Typical modeling strategy:

- Propose a model
- Fit a model
- Compute residuals
- Check ACF within $\pm\frac{2}{\sqrt{n}}$



Moving average models

- **Moving average model of order q, MA(q)**

$$\mathbf{1}\ x_t = \mathbf{1}\ w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

$$x_t = \sum_{j=0}^{q} \theta_j w_{t-j}$$

  - $w_t \sim wn(0, \sigma_w^2)$
  - $\theta_1, \ldots \theta_q$ constants, $\theta_q \neq 0$ and $\theta_0 = 1$

- **Moving average operator**

$$\theta(B) = \sum_{j=0}^{q} \theta_j B^j$$

- **MA(q):**

$$\boxed{x_t = \theta(B) w_t}$$

Autoregressive models

- **Autoregressive model of order p, $AR(p)$**

$$\mathbf{1}\ x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \mathbf{1}\ w_t$$

$$x_t - \sum_{j=1}^{p} \phi_j x_{t-j} = w_t$$

  - x_t is stationary if x_0 is sampled from the stationary distribution
  - $w_t \sim wn(0, \sigma_w^2)$
  - $\phi_1, \ldots \phi_p$ constants, $\phi_p \neq 0$
  - $Ex_t = 0$ if $Ex_0 = 0$

- **Autoregressive operator**

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$$

- **AR(p) model**

$$\boxed{\phi(B) x_t = w_t}$$

ARMA models

- **Autoregressive moving average ARMA(p,q)**

$$\mathbf{1}\ x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \mathbf{1}\ w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

  - $\phi_p \neq 0, \theta_q \neq 0$
  - Is stationary
  - $Ex_t = 0$ if $Ex_0 = 0$
- p-autoregressive order, q-moving average order
- Alternative form

$$\boxed{\phi(B)x_t = \theta(B)w_t}$$

- Criteria for **causality** and **invertibility**
  - Check roots of the characteristic polynomials $\phi(\cdot)$ and $\theta(\cdot)$

Property: ARMA(p,q) is **causal** iff **ALL** roots $\phi(z') = 0$ are outside unit circle, i.e. $|z'| > 1$

Property: ARMA(p,q) is **invertible** iff **ALL** roots $\theta(z') = 0$ are outside unit circle, i.e. $|z'| > 1$

Linear process

For a **linear process** x_t: $x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j} = \mu + \psi(B) w_t$ where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$,

$$\gamma_x(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_{j+h} \psi_j$$

Note: $x_t = \phi^{-1}(B)\theta(B) w_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}$ But series might be non-convergent

- Coefficient matching **whiteboard**
- How to find coefficients in $\psi(B) \rightarrow$ **coefficient matching**
- Example: $x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2}$

```
> ARMAtoMA(ar=.9,ma=0.5, 6)
[1] 1.400000 1.260000 1.134000 1.020600 0.918540 0.826686
```

Differencing

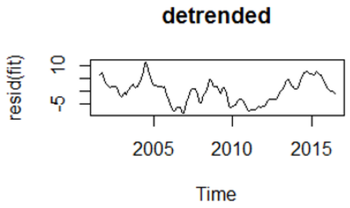
Assume $x_t = \mu_t + y_t$ and y_t stationary

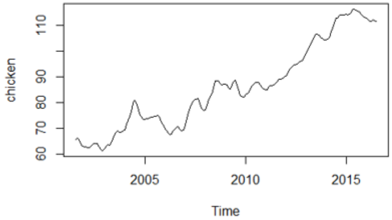
Differencing gives

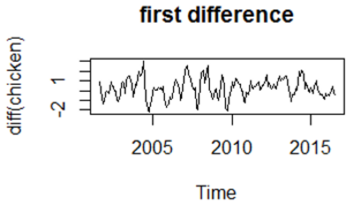
$z_t = \nabla x_t = x_t - x_{t-1}$

Also,

- $\nabla x_t = (1 - B)x_t$
- $\nabla^d = (1 - B)^d$

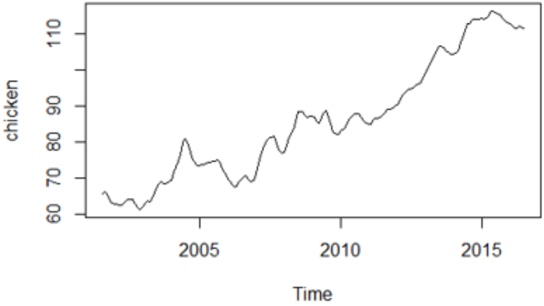






ARIMA models

- ARMA for stationary models
  - What if not stationary?



ARIMA models

- Differencing helps (lecture 2)
  - $\nabla x_t = x_t - x_{t-1}$ removes linear trend and random walk
  - $\nabla^d x_t$ removes polynomial of order d and some stochastic trends
  - $\rightarrow$ differencing is important modeling instrument!

- **Def:** x_t is **ARIMA(p,d,q)** if $\nabla^d x_t$ is ARMA(p,q), i.e.

$$\phi(B)(1-B)^d x_t = \theta(B) w_t$$

- For nonzero mean $E(\nabla^d x_t) = \mu$,

$$\phi(B)(1-B)^d x_t = \theta(B) w_t + \delta$$

$$\delta = \mu(1 - \phi_1 - \ldots - \phi_p)$$

ARIMA models

- Notation: p=0 $\rightarrow$ IMA(d,q), q=0 $\rightarrow$ ARI(p,d)
- Estimation: Differenciate + fit ARMA
- Forecasting:
  - Transform data $y_t = \nabla^d x_t$ and forecast ARMA(p,q)
  - Solve $(1 - B)^d x_t^n = y_t^n$

Estimation

Consider **ARIMA(p,d,q)**

$$\phi(B)(1-B)^d x_t = \theta(B) w_t$$

- What are the unknowns?
  - Orders p, d and q
  - Parameters $\phi_1, \cdots, \phi_p$ and $\theta_1, \cdots, \theta_q$
  - variance σ_w^2 where $w_t \sim N(0, \sigma_w^2)$
- How to estimate these?
- Assumption: Let us assume for now that we know p, d and q
  - **Maximum likelihood (ML) estimate**
  - **Least squares**

Maximum likelihood estimation: reminder

Let $x \sim f(x|\alpha)$

- Likelihood of α given observations $x_1, \cdots, x_t$ is

$$L(\alpha) = f(x_1, \cdots, x_t|\alpha)$$

- Maximum likelihood: Optimal α

$$\widehat{\alpha} = \arg\max_{\alpha} L(\alpha)$$

- Independent observations: $x_i \overset{iid}{\sim} f(x_i|\alpha)$
- $L(\alpha) = \prod_i f(x_i|\alpha)$
- Negative log-likelihood $l(\alpha) = -\sum_i \log(f(x_i|\alpha))$
- Maximum likelihood α can be obtained from negative log-likelihood

$$\max_{\alpha} L(\alpha) = \min_{\alpha} l(\alpha)$$

Maximum likelihood estimation: reminder

Time series data are NOT independent

- Likelihood of α given observations $x_1, \cdots, x_t$ is

$$L(\alpha) = f(x_1, \cdots, x_t|\alpha)$$

- Maximum likelihood: Optimal α

$$\widehat{\alpha} = \arg\max_{\alpha} L(\alpha)$$

- Dependent data (time series): chain rule

$$L(\alpha) = f(x_1|\alpha)f(x_2|\alpha, x_1)f(x_3|\alpha, x_2, x_1)...$$

- Negative log-likelihood $l(\alpha) = -\sum_i \log(f(x_i|\alpha, x_{i-1}, \cdots))$
- Maximum likelihood: Optimal α

$$\max_{\alpha} L(\alpha) = \min_{\alpha} l(\alpha)$$

Maximum likelihood estimation: reminder

- Normal distributions: if $x_i \sim N(\mu, \sigma^2)$, iid.

$$L(\theta) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}}$$

- Maximum likelihood

$$\widehat{\mu} = \bar{x}$$
$$\widehat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$$

ML for AR(1) **Whiteboard**

- For ARMA models, assume normality of w_t!
- Negative log-likelihood

$$l(\mu, \phi, \sigma_w^2) = \frac{S(\mu, \phi)}{2\sigma_w^2} + \frac{n}{2}\log(2\pi\sigma_w^2) - \frac{1}{2}\log(1-\phi^2)$$

$$S(\mu, \phi) = (1-\phi^2)(x_1-\mu)^2 + \sum_{t=2}^{n}[(x_t-\mu) - \phi(x_{t-1}-\mu)]^2$$

- How to find optimum?
  - For σ^2 explicit

$$\widehat{\sigma}_w^2 = \frac{1}{n} S(\widehat{\mu}, \widehat{\phi})$$

  - Otherwise numerical optimization (unconstrained optimization)

Optimization methods

- Examples:
  - Steepest descent
  - Newtons Methods
  - Gauss-Newton methods
  - (least squares)
  - ...

Least squares

- Unconditional least squares
  - Estimate by numerical methods or sometimes analytically

$$\min_{\mu,\phi} S(\mu,\phi)$$

- Conditional least squares: assume x_1 given (constant)

$$min \sum_{i=1}^{t} w_i^2$$

- For AR(1), $\sum_{i=1}^{t} w_i^2 = S_c(\mu,\phi)$

$$S_c(\mu,\phi) = \sum_{t=2}^{n} [(x_t - \mu) - \phi(x_{t-1} - \mu)]^2 = \sum_{t=2}^{n} [x_t - \alpha - \phi x_{t-1}]^2$$

- Note: Minimize by doing regression $Y = x_t, X = \mathrm{lag}(x_t)$

Home reading

- Shumway and Stoffer, parts of sections 3.5, 3.6, 3.7
- R code: arima.sim, arima, polyroot, ARMAtoMA, ARMAacf