Time Series Analysis

Computer Lab B: ARIMA models-4
Model selection, Forecasting

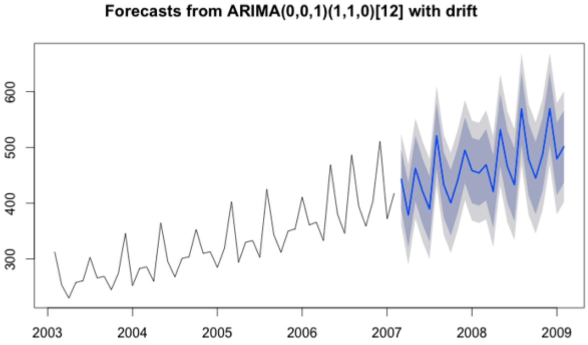
Tohid Ardeshiri

Linköping University
Division of Statistics and Machine Learning

September 23, 2019

Forecasting

- We have our series $x_1 \ldots x_n$
- Use series to predict m steps ahead: x^n_{n+m}
- The prediction should be based on our observed data $x^n_{n+m} = g(x_1, \ldots, x_n)$



Forecasting

- Assume $g(x_1, ..., x_n) = \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k$
  - Best linear predictors
- How to find α's?

$$\min E\left[(x_{n+m} - g(x_1, ...x_n))^2\right]$$

- Prediction equations
  - Find α's by solving ($x_0 = 1$)

$$E\left[(x_{n+m} - x_{n+m}^n)x_k\right] = 0, k = 0, ..., n$$

- Note: n+1 equations, n+1 unknowns

One-step-ahead

- Denote $x_{n+1}^n = \phi_{n1} x_n + ... \phi_{nn} x_1$
- Prediction equations give

$$\Gamma_n \phi_n = \gamma_n$$

$$\Gamma_n = \begin{pmatrix} \gamma(1-1) & \gamma(2-1) & \dots & \gamma(n-1) \\ \gamma(2-1) & \gamma(2-2) & \dots & \gamma(n-2) \\ \dots & \dots & \dots & \dots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(n-n) \end{pmatrix}$$

$$\phi_n = \begin{pmatrix} \phi_{n1} \\ \dots \\ \phi_{nn} \end{pmatrix} \qquad \gamma_n = \begin{pmatrix} \gamma_1 \\ \dots \\ \gamma_n \end{pmatrix}$$

- Note: for ARMA models Γ_n is positive def $\rightarrow$ unique solution

One-step-ahead

- Causal AR(p): for $n \geq p$ best linear prediction is

$$x_{n+1}^{n} = \phi_1 x_n + \ldots + \phi_p x_{n-p+1}$$

- In general, solve system of equations $\rightarrow O(n^3)$ operations

- Much faster algorithms exist
  - Durbin-Levinson algorithm
  - Innovations algorithm
- Property: PACF of a stationary process can be obtained as ϕ_{nn} by solving $\Gamma_n \phi_n = \gamma_n$

One-step-ahead

- Mean square prediction error (MSPE)

$$P_{n+1}^{n} = E\left[(x_{n+1} - x_{n+1}^{n})^{2}\right] = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n$$

- Confidence intervals for x_{n+1}

$$x_{n+1}^{n} \pm \alpha \sqrt{P_{n+1}^{n}}$$

- m-step ahead in general? Prediction equations
  - Difficult in general

m-step-ahead for ARMA

- Assume causal and invertible ARMA(p,q)
- Finite past prediction

$$x_{n+1}^n = E(x_{n+1}|x_n, ...x_1)$$

- Infinite past prediction

$$\tilde{x}_{n+m}^n = E(x_{n+m}|x_n, ...x_1, x_0, x_{-1}, ...)$$

- m-step-ahead forecast for infinite past
  - Compute recursively

$$\tilde{x}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \hat{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j \tilde{x}_{n+m-j}, m = 1, 2, ...$$

- m-step ahead prediction error: $P_{n+m}^n = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$

Long-range forecasts

- What if $m \to \infty$?

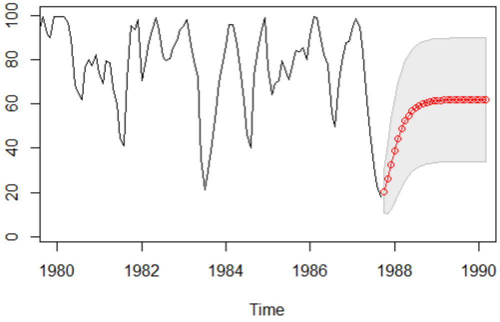
$$\tilde{x}_{n+m} \to 0(\mathrm{or}\mu)$$

$$P^n_{n+m} \to \sigma^2_x$$

m-step-ahead

- Recruitment, AR(2)

$$x_{n+m}^{n} \pm 2\sqrt{P_{n+m}^{n}}$$



Truncated prediction

- Ignore non-positive j in x_j

$$\tilde{x}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j}, m = 1, 2, \ldots$$

- For ARMA, truncated prediction formula:
  - Recursive computation, explicit

$$\tilde{x}^n_{n+m} = \phi_1 \tilde{x}^n_{n+m-1} + \ldots \phi_p \tilde{x}^n_{n+m-p} + \theta_1 \tilde{w}^n_{n+m-1} + \ldots + \theta_q \tilde{w}^n_{n+m-q}$$
$$\hat{w}^n_t = \tilde{x}^n_t - \phi_1 \tilde{x}^n_{t-1} - \ldots - \phi_p \tilde{x}^n_{t-p} - \theta_1 \tilde{w}^n_{t-1} - \ldots - \theta_q \tilde{w}^n_{t-q}$$

- Boundary conditions: $\tilde{x}^n_t = x_n, 1 \leq t \leq n, \tilde{x}^n_t = 0, t \leq 0$

$$\tilde{w}^n_t = 0, t \leq 0 \quad \text{or } t > n$$

Model selection

- ARIMA models

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

- What is p, d, q in ARIMA(p,d,q)?

Step 1: Check ACF, PACF and EACF to define a few tentative models

Model selection

- Step 2: Fit the tentative models, compare them
  - Analytical measures: AIC, BIC
    - ★ Penalize models with many paramters $\rightarrow$ simpler models
  - Residual analysis
- Akaike Information Criterion (AIC)

$$AIC = -2\log(L) + 2k$$

$$k = p + q \text{ or } k = p + q - 1 \text{ (intercept)}$$

- Corrected Akaike Information Criterion (AICc)

$$AIC_c = AIC + \frac{2(k+1)(k+2)}{n-k-2}$$

- Bayesian information criterion (BIC)

$$BIC = -2\log(L) + k\log(n)$$

Model selection

- Example: GNP data
  - Fitting ARIMA(1,1,0) to log(gmp)
  - Write down equation of the model

```
Coefficients:
         ar1  constant
      0.3467    0.0083
s.e.  0.0627    0.0010

sigma^2 estimated as 9.03e-05:  log likelihood = 718.61,  aic = -1431.22

$degrees_of_freedom
[1] 221

$ttable
         Estimate     SE t.value p.value
ar1        0.3467 0.0627  5.5255       0
constant   0.0083 0.0010  8.5398       0

$AIC
[1] -8.294483

$AICc
[1] -8.285023

$BIC
[1] -9.263925
```

Model selection

- Example: GNP data
  - Fitting ARIMA(0,1,2) to log(gmp)
  - Write down equation of the model

```
Coefficients:
         ma1     ma2  constant
      0.3028  0.2035    0.0083
s.e.  0.0654  0.0644    0.0010

sigma^2 estimated as 8.919e-05:  log likelihood = 719.96,  aic = -1431.93

$degrees_of_freedom
[1] 220

$ttable
         Estimate     SE t.value p.value
ma1        0.3028 0.0654  4.6272  0.0000
ma2        0.2035 0.0644  3.1593  0.0018
constant   0.0083 0.0010  8.7177  0.0000

$AIC
[1] -8.297814

$AICc
[1] -8.288023

$BIC
[1] -9.251978
```

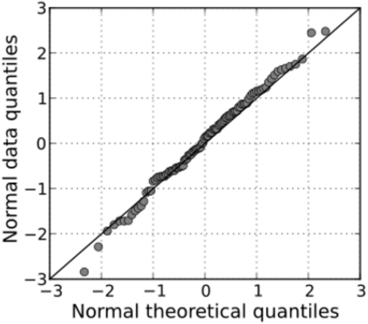
Which model is optimal according to AIC, AICc and BIC?

Residual analysis

- Residuals $e_t = x_t - \hat{x}_t^{t-1}$?they are innovations
  - Note: computed from one-step-ahead predictions!
  - Measures predictive quality of the model (compare OLS)
- Residual analysis
  - Visual inspection: stationary? Patterns?
  - Histograms, Q-Q plots
  - ACF, PACF
  - Runs test
  - Box-Ljung test

Q-Q plots

1. Sort data
2. For each x_t, compute $f_k = \frac{\#(x_i \leq x_k)}{n}$
3. For each x_k, compute $g_k = p_N^{-1}(f_k)$
4. Plot (g_k, x_k)



Wikipedia

If ECDF reminds normal, quantiles should coincide? straight line

Runs test

- Used to test independence

- H_0 : x_t values are i.i.d.
- H_a : x_t values are not i.i.d.

- Idea:
  - Count amount of segments (runs) where $x_t > \mathrm{median}(x_t)$
  - If the amount of segments large $\rightarrow$ negative dependence
  - If the amount of segments small $\rightarrow$ positive dependence
  - Medium?$\rightarrow$ independence

Box-Ljung test

- Do we observe a white noise if each ACF value below threshold?
  - Many of them just below threshold?

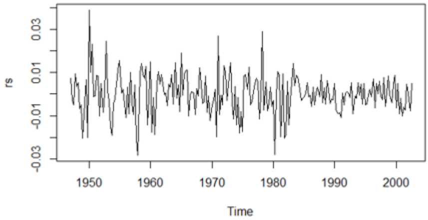
- H_0 : data are independent
- H_a : data are not independent

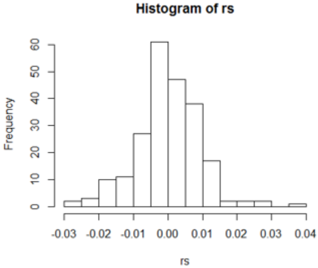
$$Q = n(n+2)\sum_{h=1}^{H} \frac{\rho_e^2(h)}{n-h}$$

- Test with different H $\rightarrow$ almost all Q-values are large when reject

Residual analysis

- Example: GNP data
  - Fitting ARIMA(1,1,0) to log(gmp)
  - Histogram and visual inspection

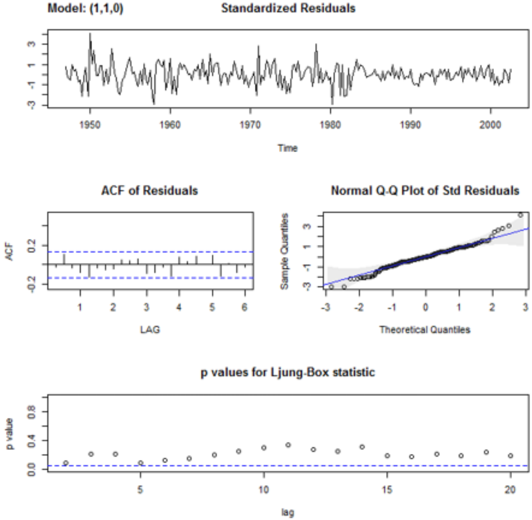




Residual analysis

- Example: GNP data

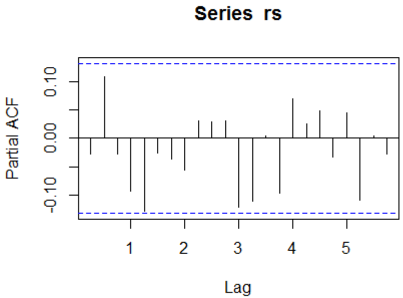
Conclusions?



Residual analysis

- Example: GNP data

Conclusions?




```
> TSA::runs(rs)
$pvalue
[1] 0.416
```

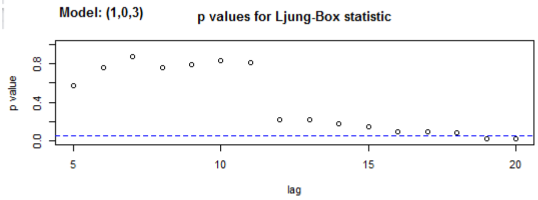
Overfitting

- Occams razor: among equally good models, choose the simplest one
- Overfitting: taking too complex models leads to bad predictions

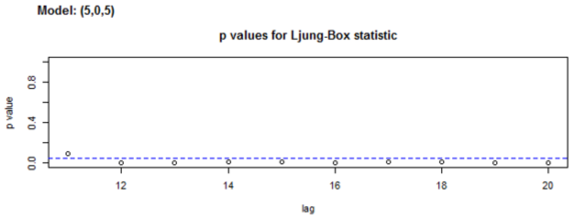
- If ARIMA(p, d, q) has almost the same predictive quality as ARIMA(p', d', q') , take the one with less parameters

Overfitting

- Example: Recruitment series
  - Fit ARIMA(1,0,3) and ARIMA(5,0,5)



- Conclusions?



Read home

- Ch 3.7-3.9
- R code: sarima, sarima.for, runs