MATH4425 (T1A) – Tutorial 9

Kazovskaia Anastasiia

April, 18

Important information

- T1A: **Thursday 19:00 - 19:50** (Rm 1033, LSK Bldg)
- Office hours: **Wednesday 14:00 - 14:50** (Math support center, 3rd floor, Lift 3)
- Any questions to be addressed to **akazovskaia@connect.ust.hk**

1 Parameter Estimation, Diagnostic Checking and Model Selection. Properties of the Parameter Estimates

Let $\hat{\alpha} := (\hat{\phi}, \hat{\mu}, \hat{\theta})$ be the CLS, ULS, or ML estimator of $\alpha_0 := (\phi_0, \mu_0, \theta_0)$.

1.1 Consistency

Consistency means that $\hat{\alpha} \longrightarrow \alpha_0$ in some sense. **Weak consistency** corresponds to *convergence in probability*. **Strong consistency** corresponds to *almost sure convergence*.

It can be proved that, under some assumptions, $\hat{\alpha}$ is a **strongly consistent estimator**, i.e. $\hat{\alpha} \xrightarrow{\text{a.s.}} \alpha_0$.

1.2 Asymptotic Normality

Also, it can be proved that, under some assumptions,

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \xrightarrow{\text{d}} \mathcal{N}(0, V(\hat{a})),$$

where

$$V(\hat{\alpha}) = \sigma_a^2 (\bar{X}_{\hat{\alpha}}^T \bar{X}_{\hat{\alpha}})^{-1}$$

$V(\hat{\alpha})$ can be estimated as

$$\hat{V}(\hat{\alpha}) = \hat{\sigma}_a^2 (\bar{X}_{\hat{\alpha}}^T \bar{X}_{\hat{\alpha}})^{-1} =: (\hat{\sigma}^2_{\hat{\alpha}_i \hat{\alpha}_j})$$

Note: Here, $\bar{X}_{\hat{\alpha}}$ is defined by the optimization method, model, and dataset.

1.3 Hypothesis Testing

With this said, we can now test the hypothesis of the form

$$H_0: \quad \alpha_{i0} = c_i$$

using the following **t-statistics**:

$$t = \frac{\hat{\alpha}_i - c_i}{\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_i}}$$

with $n - (p + q + 1)$ degrees of freedom for ARMA(p, q) model.

Note: For n big enough (say, $n > 30$) t-distribution becomes insignificantly different from standard normal distribution.

1.4 Overparametrization Detection

The estimated correlation matrix of the estimates is

$$R(\hat{\alpha}) = (\hat{\rho}_{\hat{\alpha}_i \hat{\alpha}_j}),$$

where

$$\hat{\rho}_{\hat{\alpha}_i \hat{\alpha}_j} = \frac{\hat{\sigma}^2_{\hat{\alpha}_i \hat{\alpha}_j}}{\hat{\sigma}_{\hat{\alpha}_i \hat{\alpha}_i} \hat{\sigma}_{\hat{\alpha}_j \hat{\alpha}_j}}$$

A high correlation between estimates indicates **overparameterization**.

2 Parameter Estimation, Diagnostic Checking and Model Selection. Diagnostic Checking

Given data $Z_1, \dots, Z_n$. Assume the model considered is ARMA(p, q) (possibly with drift):

$$Z_t = \mu + \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

First of all, we find the estimates $\hat{\phi}, \hat{\mu}, \hat{\theta}$ to build the model.

2.1 Model Assumptions

Once we finalize the model, we have to assess its adequacy by checking whether the model assumptions are satisfied. The basic assumptions is that a_t are **white noise**.

The **residuals** $\hat{a}_t$ defined as

$$\hat{a}_t := Z_t - \hat{\mu} - \hat{\phi}_1 Z_{t-1} - \cdots - \hat{\phi}_p Z_{t-p} + \hat{\theta}_1 \hat{a}_{t-1} + \cdots + \hat{\theta}_q \hat{a}_{t-q}, \quad \forall 1 \le t \le n$$

are **estimates** of these unobserved white noise a_t. To calculate those, we might need to introduce initial values Z_*, a_*.

There are many **assumptions** to be checked:

1) $\hat{a}_t$ are normally distributed
2) the variance is constant
3) the residuals are approximately white noise

2.2 How to Check Model Assumptions?

To check whether $\hat{a}_t$ are **normally distributed**, one can construct a histogram of $\frac{\hat{a}_t}{\hat{\sigma}_a}$ and compare it with the standard normal distribution using χ^2-*goodness-of-fit test* or use any *Normality test* such as *Shapiro-Wilk test*.

To check **homoskedasticity**, one can examine the plot of residuals or use *White test*.

To check whether the **residuals are approximately white noise**, one can test *joint null hypothesis* regarding all the residue sample ACFs (*Ljung-Box test*):

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_K = 0$$

with the test statistic

$$Q = n(n+2)\sum_{k=1}^{K} \frac{1}{n-k}\hat{\rho}_k^2,$$

where K is the lag of ACF specified by the user.

It was proved that, under null hypothesis, Q statistic approximately follows $\chi^2(K-(p+q))$.

Note: It is also possible to test $H_0 : \rho_k = 0,\ \forall 1 \le k \le K$, however, in this case one should address **Multiple comparisons problem** by introducing **Multiple testing correction** such as *Bonferroni correction*.

3 Parameter Estimation, Diagnostic Checking and Model Selection. Model Selection

3.1 Akaike's Information Criteria (AIC)

AIC is defined as

$$AIC(\text{model}) = -2\ln(\text{some ML}) + 2(\text{number of parameters})$$

For the **conditional ML**, AIC can be redefined as

$$AIC(p,q) := -2\ln L_*(\hat{\phi}, \hat{\mu}, \hat{\theta}, \hat{\sigma}_a^2) + 2(p+q)$$

For the **unconditional ML**, AIC is

$$AIC(p,q) = -2\ln L(\hat{\phi}, \hat{\mu}, \hat{\theta}, \hat{\sigma}_a^2) + 2(p+q+c) = n\ln(2\pi\hat{\sigma}_a^2) + \frac{S(\hat{\phi}, \hat{\mu}, \hat{\theta})}{\hat{\sigma}_a^2} + 2(p+q+c) =$$

$$n\ln(2\pi) + n\ln(\hat{\sigma}_a^2) + \frac{n\hat{\sigma}_a^2}{\hat{\sigma}_a^2} + 2(p+q+c) = f(n) + n\ln(\hat{\sigma}_a^2) + 2(p+q)$$

Thus, AIC can be redefined in this case as

$$AIC(p,q) := \ln(\hat{\sigma}_a^2) + \frac{2(p+q)}{n}$$

3.2 Bayesian Information Criteria (BIC)

BIC is defined as

$$BIC(\text{model}) = -2\ln(\text{some ML}) + \ln(n) \times (\text{number of parameters})$$

For the **conditional ML**, BIC can be redefined as

$$BIC(p,q) := -2\ln L_*(\hat{\phi}, \hat{\mu}, \hat{\theta}, \hat{\sigma}_a^2) + \ln(n) \times (p+q)$$

For the **unconditional ML**, BIC is

$$BIC(p,q) = -2\ln L(\hat{\phi}, \hat{\mu}, \hat{\theta}, \hat{\sigma}_a^2) + \ln(n) \times (p+q+c) = n\ln(2\pi\hat{\sigma}_a^2) + \frac{S(\hat{\phi}, \hat{\mu}, \hat{\theta})}{\hat{\sigma}_a^2} + \ln(n) \times (p+q+c) =$$

$$n\ln(2\pi) + n\ln(\hat{\sigma}_a^2) + \frac{n\hat{\sigma}_a^2}{\hat{\sigma}_a^2} + \ln(n) \times (p+q+c) = f(n) + n\ln(\hat{\sigma}_a^2) + \ln(n) \times (p+q)$$

Thus, BIC can be redefined in this case as

$$BIC(p,q) := n\ln(\hat{\sigma}_a^2) + \ln(n) \times (p+q)$$

4 Model Fitting, Model Selection, Forecast. Pipeline