

25 Spring 439/639 TSA: Lecture 17

Dr Sergey Kushnarev

Table of contents

Diagnostics	1
Residual analysis	1
Overfitting	2

Diagnostics

So far we have talked about model specification (find the model and p, q), and parameter estimation (estimate $\hat{\phi}_i, \hat{\theta}_j, \hat{\mu}, \hat{\sigma}_e^2$ for a given model). Our next step is diagnostics.

Residual analysis

In the parameter estimation step, we find the estimated parameters $\hat{\phi}_i, \hat{\theta}_j, \hat{\mu}$ using the observed data (Y_1, \dots, Y_n) . Using the estimated parameters, we can compute the predicted/fitted values \hat{Y}_t .

- For $\text{AR}(p)$, the residuals are defined as the difference between observed Y_t and fitted \hat{Y}_t : the fitted value for $\text{AR}(p)$ is just

$$\hat{Y}_t = \hat{\mu} + \hat{\phi}_1(Y_{t-1} - \hat{\mu}) + \dots + \hat{\phi}_p(Y_{t-p} - \hat{\mu}),$$

and the residual is

$$\hat{e}_t = \text{observed value} - \text{fitted/predicted value} = Y_t - \hat{Y}_t.$$

- For $\text{MA}(q)$ or $\text{ARMA}(p, q)$, we need to use the invertible representation of the model. The residuals are defined as

$$\hat{e}_t = Y_t - \hat{\pi}_1 Y_{t-1} - \hat{\pi}_2 Y_{t-2} - \dots,$$

where these $\hat{\pi}_k$ are the coefficients of the invertible representation corresponding to the estimated parameters $\{\hat{\theta}_j\}$ (and also $\{\hat{\phi}_i\}$ for ARMA), i.e.

$$\hat{\pi}_1 = \hat{\pi}_1(\hat{\phi}_i, \hat{\theta}_j), \quad \hat{\pi}_2 = \hat{\pi}_2(\hat{\phi}_i, \hat{\theta}_j), \quad \dots$$

We “want” the residuals to be: independent, normal, mean zero, constant variance.

Tool 1: sample ACF of the residuals. We can look at the sample ACF of the residuals \hat{e}_t ,

$$\hat{r}_k = \widehat{\text{corr}}(\hat{e}_t, \hat{e}_{t-k}).$$

Note: we use the notation \hat{r}_k for the sample ACF of \hat{e}_t . This is a very different thing from some similar terms we have seen in the lecture (like $r_k = \widehat{\text{corr}}(\hat{Y}_t, \hat{Y}_{t-k})$, the sample ACF of Y_t ; or $\rho_k = \text{corr}(\hat{Y}_t, \hat{Y}_{t-k})$, the theoretical ACF of Y_t).

If we think of \hat{e}_t as a white noise, then by our earlier results (see lecture 11), we should have $\hat{r}_k \approx N(0, \frac{1}{n})$ approximately. *Note:* this is incorrect, since the correct sampling distribution of \hat{e}_t is very complicated (not simply a white noise), but we just use this white noise as a heuristic.

Tool 2: plot of the standardized residuals. We can plot $\frac{\hat{e}_t}{\text{sd}(\hat{e}_t)}$. From this plot, we can check the outliers, check whether they are mean zero or have constant values.

Tool 3: QQ plot or Shapiro–Wilk test. We can look at the QQ plot or use Shapiro–Wilk test to check the normality.

Tool 4: Ljung–Box test. This is a “portmanteau” test. It tests $\text{ACF}_k(\hat{e}_t) = 0$ for all k at once. To be specific,

$$H_0 : \underbrace{\text{ACF}_k(\hat{e}_t) = 0 \text{ for all } k > 0}_{\text{suggests the fitted model is good}} \quad \text{vs.} \quad H_a : \underbrace{\text{ACF}_k(\hat{e}_t) \neq 0 \text{ for some } k > 0}_{\text{we should adjust the model}}.$$

Let’s briefly introduce this test. The Ljung–Box test uses the following test statistic

$$Q_{\text{LB}} = n(n+2) \left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \dots + \frac{\hat{r}_K^2}{n-K} \right),$$

where K is a parameter of the test. Usually, K is chosen from 5, 6, ..., 30 such that $\psi_j \approx 0$ for $j > K$ (ψ_j is the coefficient from GLP representation). Under the null hypothesis H_0 , the test statistic Q_{LB} asymptotically follows χ^2_{K-p-q} .

In this test, a critical value (c.v.) can be determined from this χ^2 distribution.

- If $Q_{\text{LB}} > \text{c.v.}$ (the p-value is small), we reject H_0 in favor of H_a . H_a suggests the \hat{e}_t are dependent, we need to update the model.
- If $Q_{\text{LB}} \leq \text{c.v.}$ (the p-value is large), we do not reject H_0 . H_0 suggests the \hat{e}_t seems to be independent, so the model is fine.

Overfitting

We already mentioned this idea in the previous lecture. Now we state it from a different perspective.

Example. Suppose AR(2) is the correct model, but we overfit with AR(3). We can still confirm AR(2) over AR(3) if

1. The parameter ϕ_3 is not significant (in AR(3) fitting).
2. The estimated $\hat{\phi}_1, \hat{\phi}_2$ in AR(3) are (almost) the same as the estimated value in AR(2) (we also need to fit an AR(2) for comparison). In addition, we can also look at their confidence intervals in both models, the corresponding CIs should overlap significantly.

Similarly, we can also overfit an ARMA(2, 1) model, and compare it to AR(2).

Caution: do not fit an ARMA(3, 1) model and compare it to AR(2).

In general, we can compare ARMA(p, q) with ARMA($p+1, q$) or ARMA($p, q+1$) for model diagnostics, but not ARMA($p+1, q+1$). Overfitting ARMA($p+1, q+1$) to an ARMA(p, q) leads to **parameter redundancy/unidentifiability**. This is because: suppose the correct model is an ARMA(p, q)

$$\underbrace{\Phi(B)}_{\text{order } p} Y_t = \underbrace{\Theta(B)}_{\text{order } q} e_t,$$

then the following holds for any c

$$\underbrace{(1-cB)\Phi(B)}_{\text{order } p+1} Y_t = \underbrace{(1-cB)\Theta(B)}_{\text{order } q+1} e_t,$$

so fitting ARMA($p+1, q+1$) for this (Y_t) is unidentifiable (parameters cannot be estimated, holds for any c).