Times Series and Forecasting (VIII)

Chapter 8. Model Diagnostics

Jianhua Hu

hu.jianhua@mail.shufe.edu.cn

Spring 2013, SHUFE

8.1. Introduction

Until now, we have discussed the following topics:

- Model specification (model selection). This deals with specifying the values of p, d, and q that are most consistent with the observed (or possibly transformed) data. This was the topic of Chapter 6.
- Model fitting (parameter estimation). This deals with estimating model parameters in the ARIMA(p, d, q) class. This was the topic of Chapter 7.

Preview

- In this chapter, we are now concerned with model diagnostics, which, quite generally, refers to "checking the fit of the model".
- We were exposed to this topic in Section 3.5 of Chapter 3, where we encountered deterministic trend models of the form

$$Y_t = \mu_t + X_t,$$

where $\mathsf{E}(X_t) = 0$.
- We will apply many of the same techniques we used before to our situation now, diagnosing the fit of ARIMA(p, d, q) models.

8.2. Residual analysis

- Residuals are random quantities which describe the part of the variation in $\{Y_t\}$ that is not explained by the fitted model.
- In general, we have the general relationship (not just in time series models, but in nearly all statistical models):

$$\text{Residual}_t = \text{Observed} \;\; Y_t - \text{Predicted} \;\; Y_t.$$

- Calculating residuals for an ARIMA(p, d, q) model fit, based on an observed sample $Y_1, Y_2, \cdots, Y_n$, can be a difficult task. It is most straightforward with purely AR models, so we start there first.

Residuals of AR models

- Consider the stationary AR(p) model

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + e_t,$$

- As we know, this model can be reparameterized as

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t,$$

  where $\theta_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$.

- For this model, the residual at time t is

$$\widehat{e}_t = Y_t - \widehat{Y}_t = Y_t - \widehat{\theta}_0 - \widehat{\phi}_1 Y_{t-1} - \widehat{\phi}_2 Y_{t-2} - \cdots - \widehat{\phi}_p Y_{t-p},$$

  where $\widehat{\phi}_j$ is an estimator of ϕ_j for $j = 1, 2, \cdots, p$, and where $\widehat{\theta}_0 = \widehat{\mu}\left(1 - \widehat{\phi}_1 - \widehat{\phi}_2 - \cdots - \widehat{\phi}_p\right)$.

- Therefore, once we observe the values of $Y_1, Y_2, \cdots, Y_n$ in our sample, we can compute the n residuals.

Subtlety

- The first p residuals must be computed using backcasting, which is a mathematical technique used to "reverse predict" the unseen values of $Y_0, Y_{-1}, \cdots, Y_{1-p}$, that is, the p values of the process $\{Y_t\}$ before time $t = 1$.

- We need backcasting, a technique of backcasting the past value. For example, in the AR(p) model, the past values $Y_0, Y_{-1}, \cdots, Y_{1-p}$ can be obtained via backcasting.
- If a time series satisfies with

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t,$$

it can be shown that the time series also satisfy with

$$Y_t = \phi_1 Y_{t+1} + \phi_2 Y_{t+2} + \cdots + \phi_p Y_{t+p} + e_t.$$

Residuals of ARMA models

- To define residuals for an invertible ARMA model containing MA terms, we exploit the fact that the model can be written as an inverted AR process.
- To be specific, recall that any invertible ARMA(p, q) model can be written as

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \pi_3 Y_{t-3} + \cdots + e_t,$$

  where the π coefficients are functions of the ϕ and θ parameters in the specific ARMA(p, q) model.
- Residuals are of the form

$$\widehat{e}_t = Y_t - \widehat{\pi}_1 Y_{t-1} - \widehat{\pi}_2 Y_{t-2} - \widehat{\pi}_3 Y_{t-3} - \cdots ,$$

  where $\widehat{\pi}_j$ is an estimator for π_j, for $j = 1, 2, \cdots ,$.

Backcasting

- We need the backcasting technique in the ARMA(p,q) model.
- If a time series satisfies with

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t + \theta_{t-1} e_{t-1} + \cdots + \theta_{t-q} e_{t-q},$$

it can be shown that the time series also satisfy with

$$Y_t = \phi_1 Y_{t+1} + \phi_2 Y_{t+2} + \cdots + \phi_p Y_{t+p} + e_t + \theta_{t+1} e_{t+1} + \cdots + \theta_{t+q} e_{t+q}.$$

- Both equations results in time series the same mean, variance and autocorrelation function.
- In the same way that the first equation can be used to forecast the future value, the second equation can be used to backcast into the past value.

Remarks for Residuals of ARMA models

- If the model is correctly specified and our estimates are 'reasonably close" to the true parameters, then the residuals should behave roughly like an iid normal white noise process.
- If the model is not correctly specified, then the residuals will not behave roughly like an iid normal white noise process.
- Furthermore, examining the residuals carefully may help us identify a better model.

Standardized residuals

- Let
$$\widehat{e}_t^* = \frac{\widehat{e}_t}{\widehat{\sigma}_e},$$
where $\widehat{\sigma}_e^2$ is an estimate of the white noise error variance σ_e^2.
- $\widehat{e}_t^*$ is call these standardized residuals.
- If the model is correctly specified, nearly all of the standardized residuals should fall between -3 and 3.
- Those that fall outside this range could correspond to observations which are "outlying" in some sense.

Remarks for standardized residuals

- Histograms and qq plots of the residuals can be used to assess the normality assumption visually.
- Time series plots of the residuals can also be helpful to detect "patterns" which violate the independence assumption.
- We can also apply the tests for normality (Shapiro-Wilk) and independence (runs test) with the standardized residuals, just as we did in Chapter 3 with the deterministic trend models.

Example 8.1. Lake Huron elevation data with a mixed fit of AR(2) and linear trend

Example

We examined the Lake Huron elevation data and we used the following process to model them.

$$Y_t = 50.5109 - 0.0216t + X_t$$

with

$$X_t = e_t + 1.0048X_{t-1} - 0.2913X_{t-2}$$

Therefore, the residual process is

$$\widehat{e}_t = X_t - 1.0048X_{t-1} + 0.2913X_{t-2}$$

where $X_t = Y_t - 50.5109 + 0.0216$ with $\hat{\sigma}^2 = 0.47605$.

Lake Huron elevation data

- Figures in sas output display the time series plot, the histogram, and the qq plot of the standardized residuals.
- The Shapiro-Wilk normality test: W =.990759, p=.7499.
- Run test for independence: $p-$value=0.3099; observed.runs=45; expected.runs=50.
- The (standardized) residuals from the AR(2) fit with a linear trend look to satisfy the normality and independence assumptions.

Large-sample results

- In Chapter 6, we discovered that for a white noise process, the kth sample autocorrelation satisfies

$$r_k \sim AN\left(0, \frac{1}{n}\right),$$

  when n is large, for all k.
- The sample autocorrelations r_j and r_k, for $j \neq k$, are approximately uncorrelated.

ACF of the residuals

- To further check the adequacy of a fitted model, it is a good idea to examine the sample ACF of the residuals, which we denote by $\widehat{r}_k$, for $k = 1, 2, \cdots,$.

  > "If the model is correctly specified and our estimates are reasonably close to the true parameters, then the residuals should behave roughly like an iid normal white noise process."

- We say "roughly", because, even if the correct model is fit, the sample ACF of the residuals, $\widehat{r}_k$, have sampling distributions that are a little different than that of white noise;
- In addition, $\widehat{r}_j$ and $\widehat{r}_k$, for $j \neq k$, are correlated, again, notably so at early lags.

Some large-sample results

In what follows, we suppose that the correct model is fit.

- AR(1).

$$\text{var}(\widehat{r}_1) \approx \frac{\phi^2}{n}, \quad \text{var}(\widehat{r}_k) \approx \frac{1-(1-\phi^2)\phi^{2k-2}}{n}, \quad \text{for } k > 1$$

$$\text{corr}(\widehat{r}_1, \widehat{r}_k) \approx -\text{sign}(\phi)\left[\frac{(1-\phi^2)\phi^{k-2}}{1-(1-\phi^2)\phi^{2k-2}}\right], \quad \text{for } k > 1,$$

  where $\text{sign}(\phi) = 1$, if $\phi > 0$ and $\text{sign}(\phi) = -1$, if $\phi < 0$.

- AR(2).

$$\text{var}(\widehat{r}_1) \approx \frac{\phi_2^2}{n}, \quad \text{var}(\widehat{r}_2) \approx \frac{\phi_2^2 + \phi_1^2(1+\phi_2)^2}{n}$$

$$\text{var}(\widehat{r}_k) \approx \frac{1}{n}, \quad \text{for } k > 2.$$

  This result for $\text{var}(\widehat{r}_k)$, for $k > 2$, may not hold if (ϕ_1, ϕ_2) is "close" to the boundary of the stationarity region for the AR(2) model.

Some large-sample results (continued)

- MA(1).

$$\begin{aligned}\text{var}(\widehat{r}_1) &\approx \frac{\theta^2}{n}\\ \text{var}(\widehat{r}_k) &\approx \frac{1-(1-\theta^2)\theta^{2k-2}}{n}, \quad \text{for } k>1\\ \text{corr}(\widehat{r}_1,\widehat{r}_k) &\approx -\text{sign}(\theta)\left[\frac{(1-\theta^2)\theta^{k-2}}{1-(1-\theta^2)\theta^{2k-2}}\right], \quad \text{for } k>1.\end{aligned}$$

- MA(2).

$$\begin{aligned}\text{var}(\widehat{r}_1) &\approx \frac{\theta_2^2}{n}\\ \text{var}(\widehat{r}_2) &\approx \frac{\theta_2^2+\theta_1^2(1+\theta_2)^2}{n}\\ \text{var}(\widehat{r}_k) &\approx \frac{1}{n}, \quad \text{for } k>2.\end{aligned}$$

Motivation of Ljung-Box test

- In addition to examining the sample ACF of the residuals individually, it is useful to consider them as a group.
- Although sample ACF may be moderate individually, as a group, the sample ACF could be "excessive", and therefore inconsistent with the fitted model.
- To address this potential occurrence, Ljung and Box (1978) developed a testing procedure to test formally whether or not a certain model in the ARMA(p, q) family was appropriate.
- Ljung and Box's procedure was built based on the sample ACF of the residuals.

Ljung-Box test

- The Ljung-Box test statistic

$$Q_* = n(n+2)\sum_{k=1}^{K}\frac{\widehat{r}_k^2}{n-k}$$

can be used to test

$$H_0: \text{ the ARMA}(p,q) \text{ model is appropriate}$$

$$\text{versus}$$

$$H_1: \text{ the ARMA}(p,q) \text{ model is not appropriate.}$$

- For a fixed K, a level α decision rule is to reject H_0 if

$$Q_* > \chi^2_{K-p-q,\alpha}.$$

How to choose K

- The sample autocorrelations $\widehat{r}_k$, for $k = 1, 2, \cdots, K$, are computed under the ARMA(p, q) model assumption in H_0.
- The value K is called the maximum lag; it's choice is somewhat arbitrary, somewhat diaphanously.
- CC recommend that K be chosen so that the Ψ_j weights of the general linear process representation of the ARMA(p, q) model (under H_0) are negligible for all $j > K$.
- Any stationary ARMA(p, q) process can be written as

$$Y_t = e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + \cdots .$$

- Typically one can simply compute Q_* for various choices of K and determine if the same decision is reached for all values of K.

LB test or ST test used in Proc arima procedure

- In Proc arima procedure, in the estimate stage, there is option

whitenoise=st | ignoremiss

- When the series contains missing values you can use this option to choose the type of test statistic that is used in the White Noise test of residuals.
- If whitenoise=ignoremiss, the standard Ljung-Box test statistic is used.
- If whitenoise=st, a modification of this statistic suggested by Stoffer and Toloi (1992) is used. Stoffer, D.S. and Toloi, C.M.C. (1992), "A note on the Ljung-Box-Pierce Portmanteau statistic with missing data", *Statistics and Probability Letters* **13**, 391-396.
- The whitenoise=st is the default.

Example 8.1. The Lake Huron elevation data

- Perform the Ljung-Box test with the Lake Huron elevation data to examine the adequacy of the AR(2) model. Here is the output from SAS using the maximum lag $K = 30$.
- Q^* = 14.94, df = 28, p-value = 0.9791
- Again, we do not have evidence against the AR(2) model for these data. Nothing in the output substantially refutes the AR(2) model.

Example 8.2. The oil price price data

Example

For the oil price data in Example 7.4, we used ml to fit

$$\nabla \log Y_t = e_t - \theta e_{t-1} = e_t + 0.29372 e_{t-1},$$

an IMA(1,1) model for $\log Y_t$. To diagnose this model, we now examine the residuals using all the methods we have discussed so far in this chapter.

Analysis

- Shapiro-Wilk normality test: W = 0.9661, p-value = 0.0001.
- the Shapiro-Wilk tests strongly rejects the hypothesis of normality for the residuals from the fit

- Run test for independence: p-value=0.7273, observed.runs=130, expected.runs=132.819.

- The runs test does not refute the independence assumption on the residuals from the IMA(1,1) fit.

Figure 8.2. Oil price data with IMA(1,1) fit to $\log(Y_t)$

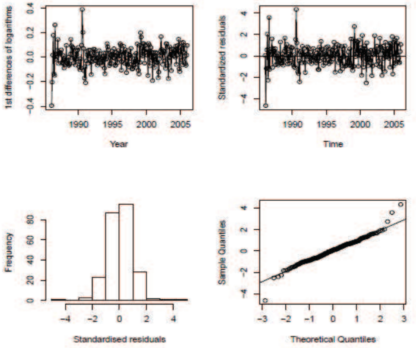

Figure 8.7: *Oil price data with IMA(1,1) fit to* $\log Y_t$. *Upper left:* $\nabla \log Y_t$ *data. Upper right: Standardized residuals with zero line added. Lower left: Histogram of the standardized residuals. Lower right: QQ plot of the standardized residuals.*

Further analysis

- The top plot displays the sresiduals from the IMA(1,1) logY_t fit with "outlier limits" at $z_{0.025/241} \approx 3.709744$, which is the upper $1 - 0.05/2(241)$ quantile of the $N(0,1)$ distribution.
- According to the Bonferroni criterion, residuals which exceed this value (3.709744) in absolute value would be classified as outliers.
- In the estimate stage of Prog arima procedure, we can use outlier to check possible outliers in data.

Further analysis

- The one around 1991 likely corresponds to the US invasion of Iraq (the first one), which had a severe effect on the price of oil.
- The sample ACF here displays no discernible patterns, and the Ljung-Box p-value plot suggests no lack of fit with the IMA(1,1) model for log-transformed process.
- The IMA(1,1) model for $\log Y_t$ appears to do a satisfactory job and explaining the variation in the oil price data. Intervention analysis (Chapter 11) could be used to adjust for the outlier observations.

8.3. Overfitting

- In addition to performing a thorough residual analysis, overfitting can be a useful diagnostic technique to further assess the validity of an assumed model.

- Basically, "overfitting" refers to the process of a fitting a model more complicated than the one under investigation and examining the significance of the additional parameter estimates and the change in the estimates from the assumed model.

Consider AR(2)

- Suppose that, after the model specification phase and residual diagnostics, we have "settled" on an AR(2) model for our $\{Y_t\}$ process, that is,

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t.$$

- To perform an overfit, we fit the following two models:
  - AR(3):

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + e_t$$

  - ARMA(2,1):

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t - \theta e_{t-1}.$$

Overfitting the AR(3) model

Compared to the AR(2) model, the AR(3) model includes one additional AR term, $\phi_3 Y_{t-3}$.

- If the additional $\widehat{\phi}_3$ is significantly different than zero, then this would be evidence that an AR(3) model is worthy of investigation.

- If, on the other hand, $\widehat{\phi}_3$ is not significantly different than zero and the estimates of ϕ_1 and ϕ_2 do not change much from their values in the AR(2) model, then this would be evidence that the AR(3) model is not needed.

Overfitting the ARMA(2,1) model

The ARMA(2,1) model includes one additional MA term, θe_{t-1}.

- If the additional $\widehat{\theta}$ is significantly different than zero, then this would be evidence that an ARMA(2,1) model is worthy of investigation.
- If, on the other hand, $\widehat{\theta}$ is not significantly different than zero and the estimates of ϕ_1 and ϕ_2 do not change much from their values in the AR(2) model, then this would be evidence that the ARMA(2,1) model is not needed.

Rule of thumb

- Rule of thumb: When overfitting an ARMA(p, q) model, we always consider two models, namely, an ARMA$(p+1, q)$ and an ARMA$(p, q+1)$.
- That is, one overfit model increases p by 1, and the other increases q by 1.

Have a nice day !