Times Series Analysis (XIII) – Model Diagnostic (I)

Jianyong Sun
School of Mathematics and Statistics
Xi'an Jiaotong University

24th Oct., 2017

- We have now discussed methods for specifying models and for efficiently estimating the parameters in those models.
- Model diagnostics, or model criticism, is concerned with testing the goodness of fit of a model and, if the fit is poor, suggesting appropriate modifications.
- We shall present two complementary approaches:
    - analysis of residuals from the fitted model and
    - analysis of overparameterized models;
        - that is, models that are more general than the proposed model but that contain the proposed model as a special case.

Residual Analysis

Consider in particular an AR(2) model with a constant term:

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

Having estimated ϕ_1, ϕ_2 and θ_0, the residuals are defined as

$$\hat{e}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \hat{\theta}_0$$

Residual Analysis

For general ARMA models containing moving average terms, we use the inverted, infinite autoregressive form of the model to define residuals. For simplicity, we assume that θ_0 is zero. From the inverted form of the model, we have

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

so that the residuals are defined as

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

Here the π's are not estimated directly but rather implicitly as functions of the ϕ's and θ's.

In fact, the residuals are not calculated using this equation but as a by-product of the estimation of the ϕ's and θs.

In the next week, we shall argue that

$$Y_t = \hat{\pi}_1 Y_{t-1} + \hat{\pi}_2 Y_{t-2} + \hat{\pi}_3 Y_{t-3} - \cdots$$

is the best forecast of Y_t based on $Y_{t-1}, Y_{t-2}, \cdots$. Thus, we have

residual = actual - predicted

in direct analogy with regression models.

Residual Analysis

- If the model is correctly specified and the parameter estimates are reasonably close to the true values, then the residuals should have nearly the properties of white noise.
- They should behave roughly like independent, identically distributed normal variables with zero means and common standard deviations.
- Deviations from these properties can help us discover a more appropriate model.

Plots of the Residuals

Our first diagnostic check is to inspect a plot of the residuals over time.

If the model is adequate, we expect the plot to suggest a rectangular scatter around a zero horizontal level with no trends whatsoever.

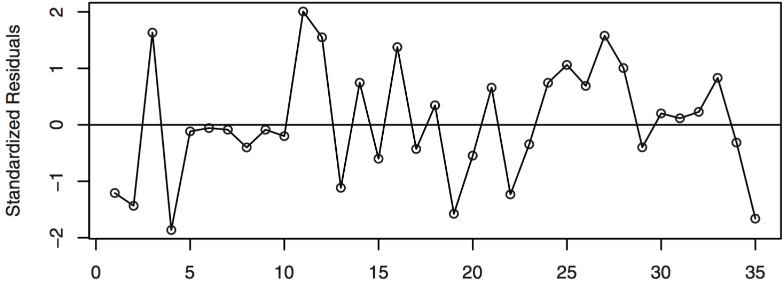


Figure: Standardized Residuals from AR(1) Model of Color.

Plots of the Residuals

- This shows such a plot for the standardized residuals from the AR(1) model fitted to the industrial color property series.

$$\text{standardized residuals} = \frac{\text{residual}}{\sqrt{Var(\text{predicted})}}$$

- Standardization allows us to see residuals of unusual size much more easily.
- The parameters were estimated using maximum likelihood. This plot supports the model, as no trends are present.

Plots of the Residuals

As a second example, we consider the Canadian hare abundance series. We estimate a subset AR(3) model with ϕ_2 set to zero. The estimated model is

$$\sqrt{Y_t} = 3.483 + 0.919\sqrt{Y_{t-1}} - 0.5313\sqrt{Y_{t-3}} + e_t$$

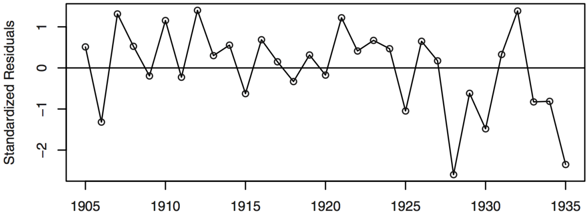


Figure: Standardized Residuals from AR(3) Model for Sqrt(Hare).

Here we see possible reduced variation in the middle of the series and increased variation near the end of the series—not exactly an ideal plot of residuals.

Plots of the Residuals

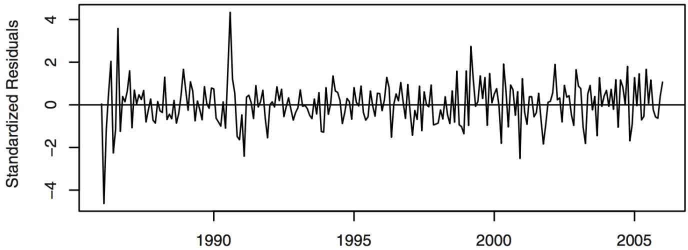
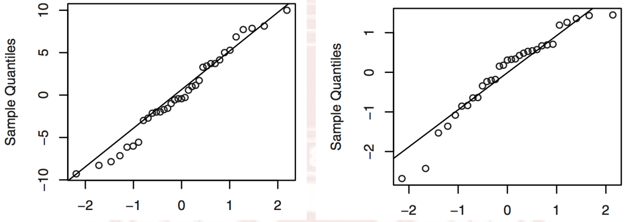


Figure: Standardized Residuals from IMA(1,1) Model for Log Oil Price.

- Parameters obtained by MLE
- There are at least two or three residuals early in the series with magnitudes larger than 3 — very unusual in a standard normal distribution.
- We should try to learn what outside factors may have influenced unusually large drops or unusually large increases in the price of oil.

Normality of the Residuals:
quantile-quantile plots are an effective tool for assessing normality.



(a) (b)

Figure: a. AR(1) Color Model Residuals; b. AR(3) Hare Model Residuals.

Plots of the Residuals

- For the color model residual, the points seem to follow the straight line fairly closely — especially the extreme values.
- the Shapiro-Wilk normality test applied to the residuals produces a test statistic of $W = 0.9754$, which corresponds to a p-value of 0.6057 $\rightarrow$ we would not reject normality based on this test.
- For the hare model residual, the extreme values look suspect. However, the sample is small ($n = 31$) and, as stated earlier, the Bonferroni criteria for outliers do not indicate cause for alarm.

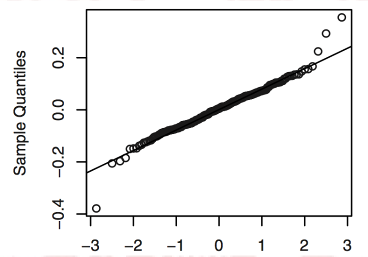


Figure: QQ plot: Residuals from IMA(1,1) Model for Oil Price.

- the IMA(1,1) model that was used to model the logarithms of the oil price series.
- Here the outliers are quite prominent ....

Autocorrelation of the Residuals

- To check on the independence of the noise terms in the model, we consider the sample autocorrelation function of the residuals, denoted $\hat{r}_k$.
- For true white noise and large n, the sample autocorrelations are approximately uncorrelated and normally distributed with zero means and variance $1/n$.
- Unfortunately, even residuals from a correctly specified model with efficiently estimated parameters have somewhat different properties.
  - for small lags k and j, the variance of $\hat{r}_k$ can be substantially less than $1/n$ and the estimates $\hat{r}_k$ and $\hat{r}_j$ can be highly correlated
  - for larger lags, the approximate variance $1/n$ does apply, and further $\hat{r}_k$ and $\hat{r}_j$ are approximately uncorrelated.

Autocorrelation of the Residuals

As an example of these results, consider a correctly specified and efficiently estimated AR(1) model. It can be shown that, for large *n*,

$$
\begin{aligned}
Var(\hat{r}_1) &\approx \frac{\phi^2}{n} && (1)\\
Var(\hat{r}_k) &\approx \frac{1-(1-\phi^2)\phi^{2k-2}}{n} \text{ for } k>1 && (2)\\
Corr(\hat{r}_1,\hat{r}_k) &\approx -sign(\phi)\frac{(1-\phi^2)\phi^{k-2}}{1-(1-\phi^2)\phi^{2k-2}} \text{ for } k>1 && (3)
\end{aligned}
$$

where $sign(\phi)=0$ if $\phi=0$, $sign(\phi)=1$ if $\phi>0$ otherwise $sign(\phi)=-1$ if $\phi<0$

Autocorrelation of the Residuals

ϕ	0.3	0.5	0.7	0.9	ϕ	0.3	0.5	0.7	0.9
k	Standard deviation of $\hat{r}_k$ times $\sqrt{n}$					Correlation $\hat{r}_1$ with $\hat{r}_k$			
1	0.30	0.50	0.70	0.90		1.00	1.00	1.00	1.00
2	0.96	0.90	0.87	0.92		−0.95	−0.83	−0.59	−0.21
3	1.00	0.98	0.94	0.94		−0.27	−0.38	−0.38	−0.18
4	1.00	0.99	0.97	0.95		−0.08	−0.19	−0.26	−0.16
5	1.00	1.00	0.99	0.96		−0.02	−0.09	−0.18	−0.14
6	1.00	1.00	0.99	0.97		−0.01	−0.05	−0.12	−0.13
7	1.00	1.00	1.00	0.97		−0.00	−0.02	−0.09	−0.12
8	1.00	1.00	1.00	0.98		−0.00	−0.01	−0.06	−0.10
9	1.00	1.00	1.00	0.99		−0.00	−0.00	−0.03	−0.08

Figure: Approximations for Residual Autocorrelations in AR(1) Models

The table illustrates these formulas for a some ϕ and k values. $Var(\hat{r}_1) \approx 1/n$ is a reasonable approximation for $k \geq 2$ over a wide range of ϕ-values.

Autocorrelation of the Residuals

If we apply these results to the AR(1) model that was estimated for the industrial color property time series with $\hat{\phi} = 0.57$ and $n = 35$, we obtain the results

Lag k	**1**	**2**	**3**	**4**	**5**	**> 5**
$\sqrt{\hat{Var}(\hat{r}_k)}$	0.096	0.149	0.163	0.167	0.168	0.169

Figure: Approximate Standard Deviations of Residual ACF values.

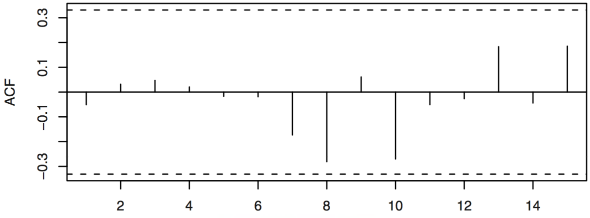


Figure: Sample ACF of Residuals from AR(1) Model for Color.

Autocorrelation of the Residuals

For an AR(2) model, it can be shown that

$$Var(\hat{r}_1) \approx \frac{\phi_2^2}{n} \qquad \text{and } Var(\hat{r}_2) \approx \frac{\phi_2^2 + \phi_1^2(1+\phi_2)^2}{n}$$

If the AR(2) parameters are not too close to the stationarity boundary, then

$$Var(\hat{r}_k) \approx \frac{1}{n}$$

for $k \geq 3$.

Autocorrelation of the Residuals

If we fit an AR(2) model by maximum likelihood to the square root of the hare abundance series, we find that $\hat{\phi}_1 = 1.351$ and $\hat{\phi}_2 = -0.776$. Thus we have

$$\begin{aligned}
\sqrt{Var(\hat{r}_1)} &\approx \frac{|-0.776|}{\sqrt{35}} = 0.131 \\
\sqrt{Var(\hat{r}_2)} &\approx \sqrt{\frac{(-0.776)^2 + (1.351)^2(1+(-0.776)^2)^2}{35}} = 0.141 \\
\sqrt{Var(\hat{r}_k)} &\approx 1/\sqrt{35} = 0.169 \text{ for } \; k \geq 3.
\end{aligned}$$

Autocorrelation of the Residuals

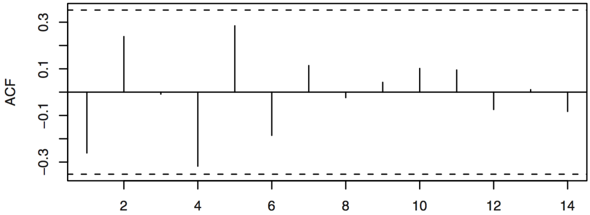


Figure: Sample ACF of Residuals from AR(2) Model for Hare

- This figure displays the sample ACF of the residuals from the AR(2) model of the square root of the hare abundance.
- The lag 1 autocorrelation here equals -0.261, which is close to 2 standard errors below zero but not quite.
- The lag 4 autocorrelation equals -0.318, but its standard error is 0.169.
- We conclude that the graph does not show statistically significant evidence of nonzero autocorrelation in the residuals.

Autocorrelation of the Residuals

- With monthly data, we would pay special attention to possible excessive autocorrelation in the residuals at lags 12, 24, and so forth. With quarterly series, lags 4, 8, and so forth would merit special attention.
- It can be shown that results analogous to those for AR models hold for MA models. In particular, replacing ϕ by θ in Eq. (1), (2) and (3) gives the results for the MA(1) case.
- Similarly, results for the MA(2) case can be stated by replacing ϕ_1 and ϕ_2 by θ_1 and θ_2, respectively, in Eqs (4), (5) and (6).
- Results for general ARMA models may be found in Box and Pierce (1970) and McLeod (1978).

- In addition to looking at residual correlations at individual lags, it is useful to have a test that takes into account their magnitudes as a group.
- For example, it may be that most of the residual autocorrelations are moderate, some even close to their critical values, but, taken together, they seem excessive.
- Box and Pierce (1970) proposed the statistic

$$Q = n(\hat{r}_1^2 + \hat{r}_2^2 + \cdots + \hat{r}_K^2)$$

to address this possibility.

The Ljung-Box Test

- They showed that if the correct ARMA(p, q) model is estimated, then, for large n, Q has an approximate chi-square distribution with $K - p - q$ degrees of freedom.
- Fitting an erroneous model would tend to inflate Q.
- Thus, a general "portmanteau" test would reject the ARMA(p, q) model if the observed value of Q exceeded an appropriate critical value in a chi-square distribution with $K - p - q$ degrees of freedom.

The Ljung-Box Test

- The chi-square distribution for Q is based on a limit theorem as $n \to \infty$, but Ljung and Box (1978) subsequently discovered that even for $n = 100$, the approximation is not satisfactory.
- By modifying the Q statistic slightly, they defined a test statistic whose null distribution is much closer to chi-square for typical sample sizes.
- The Ljung-Box statistics is given by

$$Q_* = n(n+2)\left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \cdots + \frac{\hat{r}_K^2}{n-K}\right)$$

The Ljung-Box Test

Notice that since $(n+2)/(n-k) > 1$ for every $k \geq 1$, we have $Q_* > Q$, whichpartly explains why the original statistic Q tended to overlook inadequate models.

More details on the exact distributions of Q_* and Q for finite samples can be found in Ljung and Box (1978), see also Davies, Triggs, and Newbold (1977).

The Ljung-Box Test

Lag k	1	2	3	4	5
Residual ACF	–0.051	0.032	0.047	0.021	–0.017

Figure: Residual Autocorrelation Values from AR(1) Model for Color

- This figure lists the first six autocorrelations of the residuals from the AR(1) fitted model for the color property series. Here $n = 35$.
- The Ljung-Box test statistic with $K = 6$ is equal to $Q_* \approx 0.28$.
- This is referred to a chi-square distribution with $6 - 1 = 5$ degrees of freedom. This leads to a p-value of 0.998, so we have no evidence to reject the null hypothesis that the error terms are uncorrelated.

The Ljung-Box Test

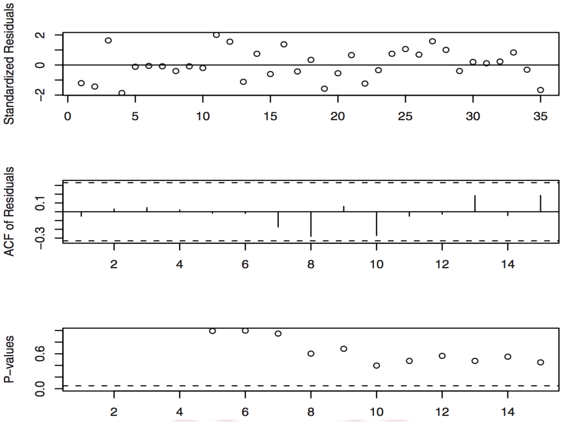


Figure: Diagnostic Display for the AR(1) Model of Color Property.

The Ljung-Box Test

- The figure shows three of our diagnostic tools in one display—a sequence plot of the standardized residuals, the sample ACF of the residuals, and p-values for the Ljung-Box test statistic for a whole range of values of K from 5 to 15.
- The horizontal dashed line at 5% helps judge the size of the p-values.
- The estimated AR(1) model seems to be capturing the dependence structure of the color property time series quite well.

The Ljung-Box Test

- the **runs test** may also be used to assess dependence in error terms via the residuals.
- Applying the test to the residuals from the AR(3) model for the Canadian hare abundance series, we obtain expected runs of 16.09677 versus observed runs of 18.
- The corresponding p-value is 0.602, so we do not have statistically significant evidence against independence of the error terms in this model.

Overfitting and Parameter Redundancy

Our second basic diagnostic tool is that of **overfitting**. After specifying and fitting what we believe to be an adequate model, we fit a slightly more general model; that is, a model “close by" that contains the original model as a special case.

For example, if an AR(2) model seems appropriate, we might overfit with an AR(3) model. The original AR(2) model would be confirmed if:

- the estimate of the additional parameter, ϕ_3, is not significantly different from zero, and
- the estimates for the parameters in common, ϕ_1 and ϕ_2, do not change significantly from their original estimates.

Overfitting and Parameter Redundancy

Take the industrial color property time series as an example. Previous studies suggested it to be an AR(1) model.

Coefficients:[†]	ar1	Intercept[‡]
	0.5705	74.3293
s.e.	0.1435	1.9151

sigma^2 estimated as 24.83: log-likelihood = -106.07, AIC = 216.15

[†] `m1.color` # R code to obtain table
[‡] Recall that the intercept here is the estimate of the process mean μ—not θ_0.

(a) AR(1)

Coefficients:	ar1	ar2	Intercept
	0.5173	0.1005	74.1551
s.e.	0.1717	0.1815	2.1463

sigma^2 estimated as 24.6: log-likelihood = -105.92, AIC = 217.84

(b) AR(2)

Figure: (a) AR(1) model results for the color property series. (b) AR(2) results

Overfitting and Parameter Redundancy

- First note that, the estimate of ϕ_2 is not statistically different from zero. This fact supports the choice of the AR(1) model.
- Secondly, we note that the two estimates of ϕ_1 are quite close, especially when we take into account the magnitude of their standard errors.
- Finally, note that while the AR(2) model has a slightly larger log-likelihood value, the AR(1) fit has a smaller AIC value. The penalty for fitting the more complex AR(2) model is sufficient to choose the simpler AR(1) model.

A different overfit for this series would be to try an ARMA(1,1) model.

Coefficients:	ar1	ma1	Intercept
	0.6721	–0.1467	74.1730
s.e.	0.2147	0.2742	2.1357

sigma^2 estimated as 24.63: log-likelihood = -105.94, AIC = 219.88

Figure: Overfit of an ARMA(1,1) Model for the Color Series.

- Notice that the standard errors of the estimated coefficients for this fit are rather larger than previous results.
- Regardless, the estimate of ϕ_1 from this fit is not significantly different from the estimate in previous.
- The estimate of the new parameter, θ, is not significantly different from zero.
- This adds further support to the AR(1) model.

Overfitting and Parameter Redundancy

any ARMA(p, q) model can be considered as a special case of a more general ARMA model with the additional parameters equal to zero.

However, when generalizing ARMA models, we must be aware of the problem of **parameter redundancy** or **lack of identifiability**.

Overfitting and Parameter Redundancy

To make these points clear, consider an ARMA(1,2) model:

$$Y_t = \phi Y_{t-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \qquad (1)$$

Now replace t by $t-1$ to obtain

$$Y_{t-1} = \phi Y_{t-2} + e_{t-1} - \theta_1 e_{t-2} - \theta_2 e_{t-3} \qquad (2)$$

If we multiply both sides of Eq. (1) by any constant c and then subtract it from Eq. (2), we obtain (after rearranging)

$$Y_t - (\phi + c) Y_{t-1} + \phi c Y_{t-2} = e_t - (\theta_1 + c) e_{t-1} - (\theta_2 - \theta_1) e_{t-2} + c\theta_2 e_{t-3}$$

Overfitting and Parameter Redundancy

This apparently defines an ARMA(2,3) process. But notice that we have the factorizations

$$1 - (\phi + c)x + \phi c x^2 = (1 - \phi x)(1 - cx)$$

and

$$1 - (\theta_1 + c)x - (\theta_2 - c\theta_1)x^2 + c\theta_2 x^3 = (1 - \theta_1 x - \theta_2 x^2)(1 - cx)$$

Thus the AR and MA characteristic polynomials in the ARMA(2,3) process have a common factor of $(1 - cx)$.
Even though Y_t does satisfy the ARMA(2,3) model, clearly the parameters in that model are not unique – the constant c is completely arbitrary. We say that we have parameter redundancy in the ARMA(2,3) model.

Overfitting and Parameter Redundancy

The implications for fitting and overfitting models are as follows:

1. Specify the original model carefully. If a simple model seems at all promising, check it out before trying a more complicated model.
2. When overfitting, do not increase the orders of both the AR and MA parts of the model simultaneously.
3. Extend the model in directions suggested by the analysis of the residuals. For example, if after fitting an MA(1) model, substantial correlation remains at lag 2 in the residuals, try an MA(2), not an ARMA(1,1).

Overfitting and Parameter Redundancy

Coefficients:	ar1	ar2	ma1	Intercept
	0.2189	0.2735	0.3036	74.1653
s.e.	2.0056	1.1376	2.0650	2.1121

sigma^2 estimated as 24.58: log-likelihood = −105.91, AIC = 219.82

Figure: Overfitted ARMA(2,1) Model for the Color Property Series.

Suppose we try an ARMA(2,1) model. Notice that even though the estimate of σ_e^2 and the log-likelihood and AIC values are not too far from their best values, the estimates of ϕ_1, ϕ_2, and θ are way off, and none would be considered different from zero statistically.

Summary

- Residual analysis
- We looked at various plots of the residuals, checking the error terms for constant variance, normality, and independence.
- The properties of the sample autocorrelation of the residuals play a significant role in these diagnostics.
- The Ljung-Box statistic portmanteau test was discussed as a summary of the autocorrelation in the residuals.
- Lastly, the ideas of overfitting and parameter redundancy were presented.

Questions?