Oregon State University, Department of Statistics

Tutorial on performing a portmanteau lack of fit test

Example of learning/teaching a new topic

1 INTRODUCTION

After a time series model has been fitted, as part of the diagnostic checks, the portmanteau lack of fit (LOF) test, first appearing in Box et al. (1994, section 8.2.2), evaluates the **overall appropriateness** of the fit of the model. The test is "portmanteau", or compound, in the sense that there are many ways the model may depart from the white noise (the null model) and we are not being specific as to which way we are most interested in checking. Therefore, in practice, the portmanteau LOF test often serves as a first step in residual analysis, followed by closer examinations at residuals and/or updating of original model.

In this tutorial, we will firstly present the form of the LOF test statistic, along with its asymptotic distribution. In section 3, we will list the steps of performing such a test by hand in much detail. In section 4, we will introduce the R function `Box.test`, which is specifically designed for this procedure. In the last section, we will apply the LOF test to a dataset acquired from the `TSA` package in a step-by-step fashion.

2 THEORETICAL RESULTS

Suppose x_t is a stationary time series of length n, which in practice, could be a de-trended and/or de-seasonalized series. Assume that we fit an ARMA(p,q) model to this series, which is:

$$\phi(B)x_t = \theta(B)w_t \tag{1}$$

where w_t is white noise.

Now, let $\hat{\boldsymbol{w}} = (\hat{w}_1, ..., \hat{w}_n)'$ be the residuals calculated from the sample series using model (1) with all parameters of $\phi(B)$ and $\theta(B)$ substituted by their least-squares (or maximum likelihood or method of moments) estimates. Let $\hat{\boldsymbol{r}} = (\hat{r}_1, ..., \hat{r}_m)'$ be the first m autocorrelation coefficients of $\{\hat{w}_i\}_{i=1}^n$.

Box and Pierce (1970) show that the statistic

$$Q(\hat{r}) = n\sum_{k=1}^{m}\hat{r}_k^2$$

follows asymptotically a χ^2 distribution with degrees of freedom $m-p-q$. This statistic is known as the **Box-Pierce** version of the portmanteau LOF test statistic.

However, Ljung & Box (1978) points out that the distribution of $Q(\hat{r})$ can deviate greatly from χ^2_{m-p-q} unless n is large relative to m. Instead, a modified test statistic is proposed:

$$\tilde{Q}(\hat{r}) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{r}_k^2$$

and it can be shown that the distribution of $\tilde{Q}(\hat{r})$ can be approximated well by χ^2_{m-p-q} even when n is relatively small. A LOF test using $\tilde{Q}(\hat{r})$ is called the **Ljung-Box** version of the portmanteau test.

2.1 Box-Pierce or Ljung-Box?

A natural question is which of the two versions of the test statistics above one should use. Box & Pierce (1970) noticed that due to over-approximation of standard errors of $\hat{r}$'s, the test statistic $Q(\hat{r})$ is underestimated, resulting in power loss. Ljung & Box (1987) fixed this problem by including the original expression for standard errors in their test statistic. Therefore, the Ljung-Box version is more sensitive to lack of fit than Box-Pierce, and is more useful in practice because our goal is to find the best possible model.

3 Steps of the Test

(1) Difference (and/or seasonally difference) the series until it is stationary. Denote the length of the stationary series n.

(2) Fit a candidate ARMA(p, q) model to the stationary series, where p and q are pre-determined by, for instance, examining the ACF/PACF of the series. Obtain the residuals $\hat{w}_1, ..., \hat{w}_n$.

(3) Choose the number of lags to test, m. Robert J. Hyndman [4] suggests that:

- For non-seasonal time series, use $m = \min(10, n/5)$;
- For seasonal time series, use $m = \min(2s, n/5)$, where s is seasonal period.

(4) Calculate the estimated sample ACF for each lag $k \in \{1, ..., m\}$, using the following formula:

$$\hat{r}_k = \frac{\sum_{t=k+1}^{n} \hat{w}_t \hat{w}_{t-k}}{\sum_{t=1}^{n} \hat{w}_t^2}$$

(5) Calculate the test statistic (here only use Ljung-Box version):

$$\tilde{Q}(\hat{r}) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{r}_k^2$$

(6) For a level α test, compare the test statistic $\tilde{Q}(\hat{r})$ with the upper αth quantile of χ^2 distribution with $m-p-q$ degrees of freedom. Reject the null hypothesis if $\tilde{Q}(\hat{r}) > \chi^2_{m-p-q}(\alpha)$.

4 R IMPLEMENTATION

Both the Box-Pierce and Ljung-Box versions of the portmanteau test have been conveniently implemented in the R function `Box.test()`. This is a function in the `stats` package, which should have been installed along with base-R, and no additional packages are required.

The `Box.test` function has the following form:

```
Box.test(x, lag = 1, type = c("Box-Pierce", "Ljung-Box"), fitdf = 0)
```

Here `x` is the stationary series, `lag=` is the maximum number of lags considered, `type=` specifies the version of test to use, and `fitdf` is $p+q$ as from the proposed model ARMA(p,q).

Section 5 describes though an example how to use this function on real datasets.

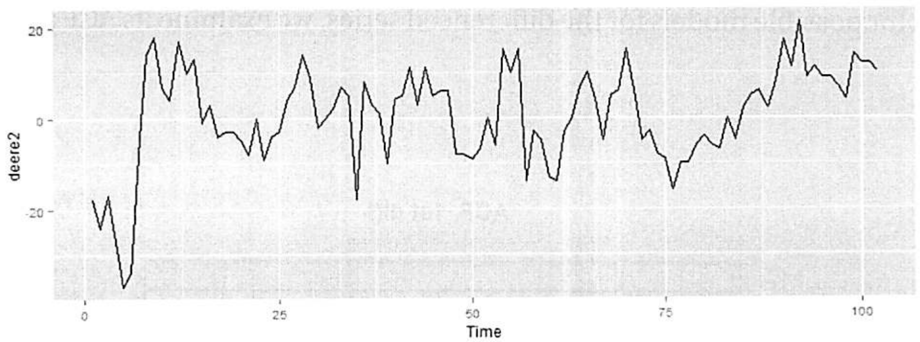
5 EXAMPLE

As an example, we use the `deere2` dataset available in the R package `TSA`:

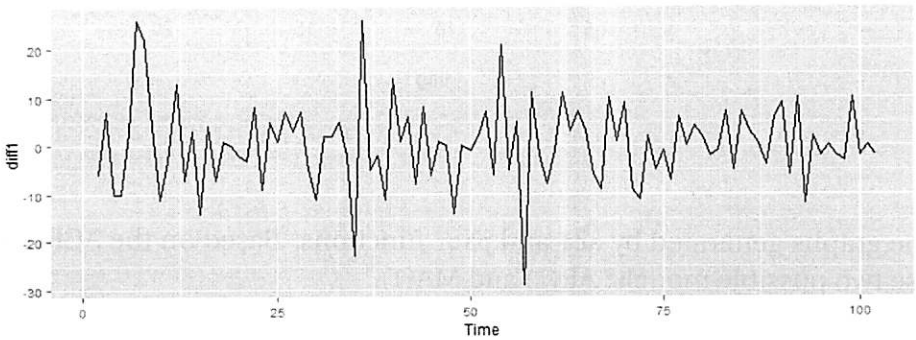
```
library(TSA)
data(deere2)
```

Our goal is to identify candidate models and check their overall appropriateness using the portmanteau lack of fit test.

The time plot of the original series shows some trend:



Taking a first difference, the series appears to be stationary:



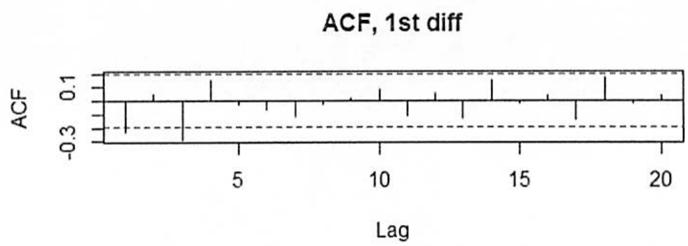
The two plots above are obtained using the following code:

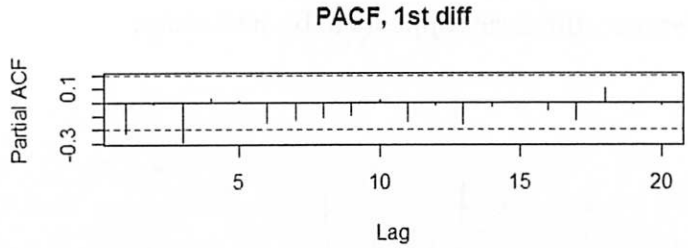
```
library(ggplot2)
n = length(deere2)

qplot(1:n,deere2,geom="line")+
  xlab("Time")

diff1 = c(NA,diff(deere2))
qplot(1:n,diff1,geom="line")+
  xlab("Time")
```

To look for possible models for the differenced series, we examine its ACF and PACF:





with the graphs generated by `acf` and `pacf` functions. Based on the ACF/PACF, we propose two possible models: AR(1) and MA(1).

Now we start to follows the steps in section 3. Firstly, we use function `arima` to fit an AR(1) and an MA(1) model on the differenced series, respectively:

```
mod.ar = arima(diff(deere2),order=c(1,0,0))
mod.ma = arima(diff(deere2),order=c(0,0,1))
```

Secondly, we acquire the residuals of the fitted models:

```
res.ar = residuals(mod.ar)
res.ma = residuals(mod.ma)
```

And finally, perform the Box-Pierce or Ljung-Box test:

```
Box.test(res.ar,lag=10,type="Box-Pierce",fitdf=1)
Box.test(res.ma,lag=10,type="Box-Pierce",fitdf=1)
### OR ###
Box.test(res.ar,lag=10,type="Ljung-Box",fitdf=1)
Box.test(res.ma,lag=10,type="Ljung-Box",fitdf=1)
```

Notice that in this example we chose the maximum number of lag m to be 10, which is a result from Hyndman's suggestion that $m = \min(10, n/5)$. In this case $n = 101$ and $n/5 = 20.5 > 10$, so we choose $m = 10$.

Here we only show results of the Ljung-Box tests:

```
##### AR(1) #####
        Box-Ljung test

data:  res.ar
X-squared = 13.9085, df = 9, p-value = 0.1256

##### MA(1) #####
        Box-Ljung test

data:  res.ma
X-squared = 13.403, df = 9, p-value = 0.1452
```

Observe that both of the tests fail to reject the null hypothesis that the error terms are truly white noise. Residual ACF/PACF and/or other model selection criteria (e.g. AIC) might be needed to select the most appropriate model for this particular dataset.

References

[1] Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis Forecasting and Control.* San Francisco: Holden-Day.

[2] Box, G. E. P. & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Statist. Assoc.* **65**, 1509-26.

[3] Ljung, G. M. & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**(2), 297-303.

[4] Hyndman, R. (January 24, 2014). Thoughts on the Ljung-Box Test. In *Hyndsight.* Retrieved March 13, 2014, from http://robjhyndman.com/hyndsight/ljung-box-test/.