

stationary_test_experiment

liang kuang

March 21, 2017

Introduction

In mathematics and statistics, a stationary process (a.k.a. a strict(ly) stationary process or strong(ly) stationary process) is a stochastic process whose joint probability distribution does not change when shifted in time. Consequently, parameters such as mean and variance, if they are present, also do not change over time.

Definition:

$$F_X(x_{t_1+\tau}, \dots, x_{t_k+\tau}) = F_X(x_{t_1}, \dots, x_{t_k})$$

Since τ does not affect $F_X(\cdot)$, F_X is not a function of time.

The standard ADF test with a trend estimates the following regression:

$$y_t = \alpha + c * trend + \rho y_{t-1} + \sum_{j=1}^{pmax} \Delta y_{t-j} + \epsilon_t$$

This is an example to test the results of different stationary algorithm on the oil price from 2001, which obtained using 'Quandl' API.

Testing whether time series is stationary is very important since for many model (e.g. ARIMA, VAR) and tests, the prerequisite is that the time series are stationary. There are many tools to check stationary. In this case, the packages 'fpp' and 'forest' are used.

```
# load the required libraries
library('zoo')
```

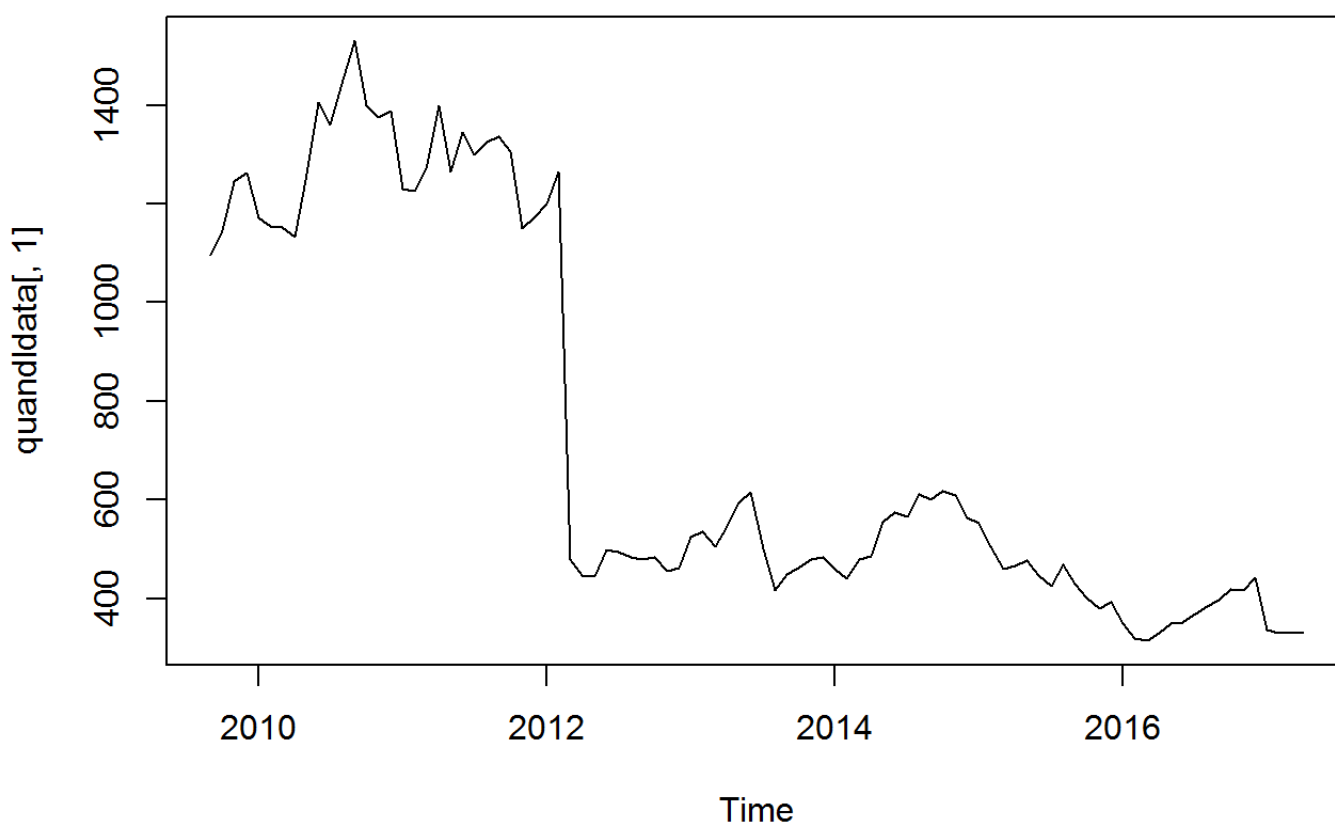
```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

```
library('xts')  
library('forecast');  
library('fma')  
library('expsmooth')  
library('lmtest')  
library('tseries')  
library('Quandl')  
library('fpp');  
library('urca')
```

```
quandldata = Quandl("NSE/OIL", collapse="monthly", start_date="2001-01-01", type="ts"  
)  
plot(quandldata[,1],main='Figure 1: Raw Oil Price Data')
```

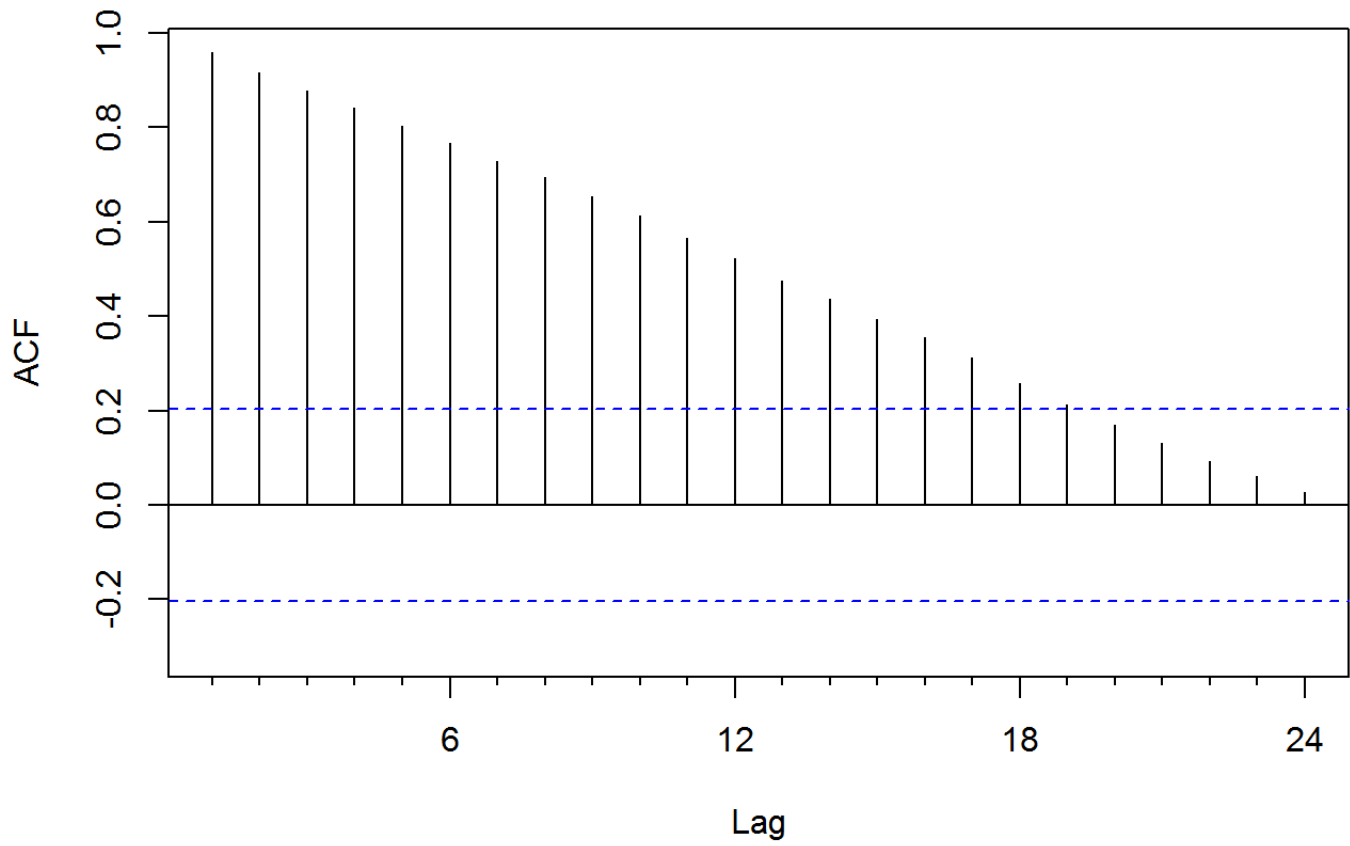
Figure 1: Raw Oil Price Data



Check using *ACF* and *PACF* graphs and check significant lags.

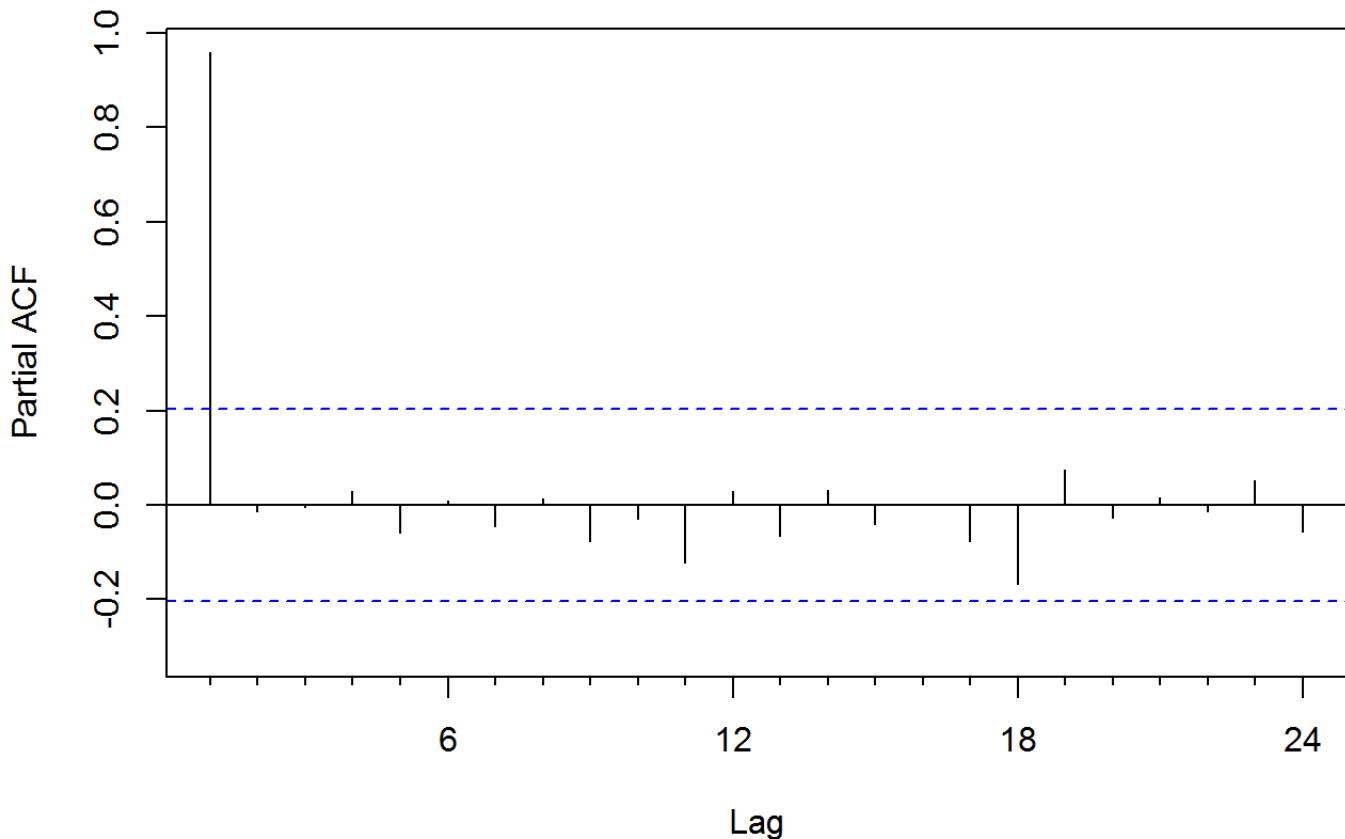
```
Acf(quandldata[,1])
```

Series quandldata[, 1]



```
Pacf(quandldata[, 1])
```

Series quandldata[, 1]



As seen from the ACF graph, there are significant lags. PACF tells a slight different story.

Testing various methods

- 1. Ljung-Box test Ljung-Box test examines whether there is significant evidence for non-zero correlation at lags 1-20. **Small p-values** (i.e., less than 0.05) suggest that the series is **stationary**.

```
LB_test <- Box.test(quandldata[,1],lag=20, type='Ljung-Box')
print(LB_test)
```

```
##
## Box-Ljung test
##
## data:  quandldata[, 1]
## X-squared = 793.39, df = 20, p-value < 2.2e-16
```

While using Ljung-Box testing stationarity, it shows a very small p-value which indicates that the time series is stationary. But this is not true as we seen from the Figure 1.

As pointed out by Mihaela Solcan, LB method are used for serial correlation test. The public information on some blogger might be not right regarding stationary test.

- 2. Augmented Dickey-Fuller (ADF) t-test **Small p-values** suggest that the data is **stationary** and doesn't need to be differenced stationarity.

```
adf_test <- adf.test(quandldata[,1],alternative = 'stationary')
print(adf_test)
```

```
##
## Augmented Dickey-Fuller Test
##
## data:  quandldata[, 1]
## Dickey-Fuller = -2.0871, Lag order = 4, p-value = 0.5404
## alternative hypothesis: stationary
```

By using adf.test, it yield a big p-value which shows the data is not stationary.

- 3. Kwiatkowski-Philips-Schmidt-Shin (KPSS) test Here accepting null hypothesis means that the series is stationary, and **small p-value** suggest that the series is **NOT Stationary** and a differencing is required.

```
kpss_test <- kpss.test(quandldata[,1])
```

```
## Warning in kpss.test(quandldata[, 1]): p-value smaller than printed p-value
```

```
print(kpss_test)
```

```
##
## KPSS Test for Level Stationarity
##
## data:  quandldata[, 1]
## KPSS Level = 2.4121, Truncation lag parameter = 2, p-value = 0.01
```

KPSS shows the same results as adf.test, non-stationary data.

- 4. Based on Miheala's experience, the R *urca* package is more recommended to implement the unit root and the cointegration tests. Use `ur.df` function which allows for a lot of options, including trends, intercepts and the automatic choice of lags based on information criteria.

```
udf_test <- ur.df(quandldata[,1], type='trend', lags = 10, selectlags = "BIC")
print(udf_test)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -1.9065 1.8895 2.0946
```

In summary, `adf.test` and `kpss.test` yields consistent results regarding the stationary test. However, the Ljung-Box test shows opposite story which is not true in this case. Clearly, LB stationary test is not suitable stationary test as pointed out by Mihaela. Based on Mihaela's experience, the last approach of using `ur.df()` is recommended since it can specify trend, automatical find lag order and works well for structure data. While ADF test is weak with structure dataset.