Time Series Analysis

Lecture 3

Autoregressive Models and Moving Average Models

datascience@berkeley

Moving Average Models

Modeling Using Real-World Data: The Data and Descriptive Statistics

The Data Used in This Example

- Instead of using the data (British pounds vs. New Zealand dollar (NZD)) series provided by the textbook, I downloaded the monthly US dollar–NSZ dollar exchange rate series from the Federal Reserve website:

http://research.stlouisfed.org/fred2/series/EXUSNZ/downloaddata

The Data Used in This Example (2)

- The advantage of obtaining the data series from the official source is that you have complete control over modification and transformation made to the series.

- I will still use the data series provided by the book later in this lecture because I want to extend the example used in the book.

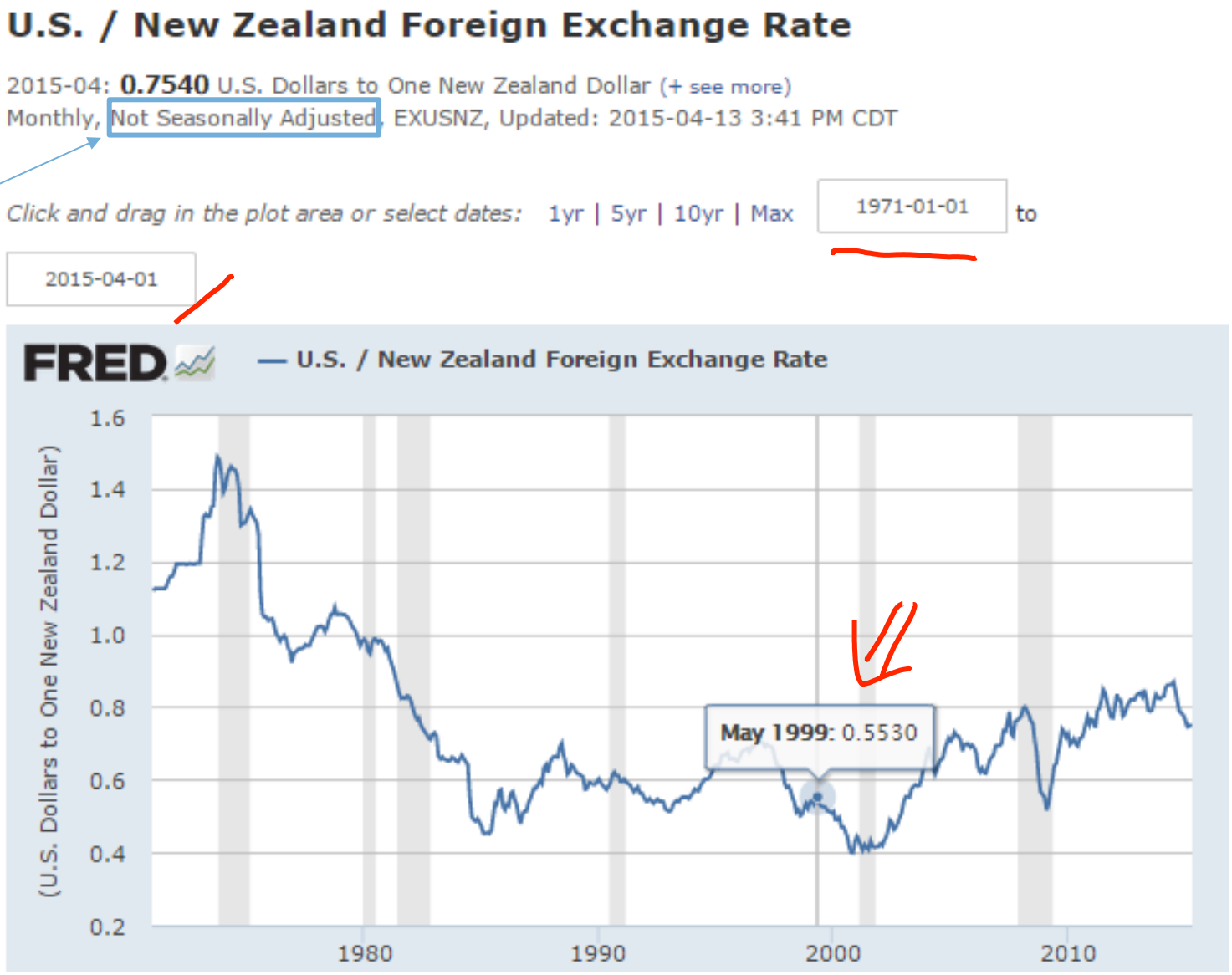
- Alternatively, one could use function in the R library "**quantmod**" to stream the data directly, which I will do in the next lecture.

- **Quantmod** is especially designed for financial time series analysis, and I will not go into the details of this library.

The Data Used in This Example (3)

The data series is not seasonally adjusted.

This graph is provided in the FRED website: http://research.stlouisfed.org/fred2/series/EXUSNZ



Examine the Basic Structure of the Data

Import the data from a csv file, and name the data.frame exusnz.

```
exusnz <- read.csv("C:/Users/K/z_Teach/MIDS_AdvStat/data/EXUSNZ.csv",
                   header=TRUE, stringsAsFactors=FALSE)
```

Then, we examine the basic structure of the data:

- 532 observations and 2 variables: date and the exchange rate called "value."

```
> str(exusnz)
'data.frame': 532 obs. of  2 variables:
$ DATE : chr  "1971-01-01" "1971-02-01" "1971-03-01" "1971-04-01"
$ VALUE: chr  "1.1194" "1.1250" "1.1254" "1.1250" ...
```

- List the first and last few observations of the data.
- Always a good idea to list some observations to make sure it is what you would expect.

```
> cbind(head(exusnz), tail(exusnz))
        DATE  VALUE        DATE  VALUE
1 1971-01-01 1.1194 2014-11-01 0.7834
2 1971-02-01 1.1250 2014-12-01 0.7766
3 1971-03-01 1.1254 2015-01-01 0.7628
4 1971-04-01 1.1250 2015-02-01 0.7454
5 1971-05-01 1.1254 2015-03-01 0.7460
6 1971-06-01 1.1255 2015-04-01      .
```

This is unexpected!

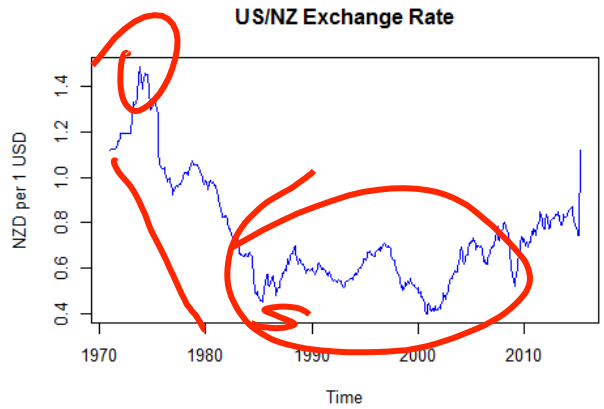
Clean Up the Data—Handling Missing Values

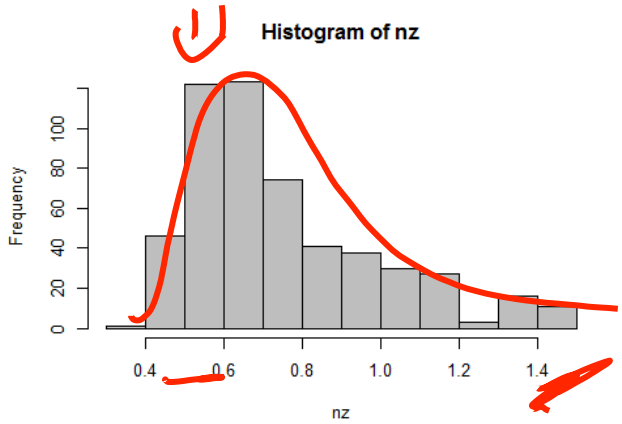
- Delete the last observation. (Note that I could have gone back to the website, downloaded the series again, and examined whether or not the last observation is really unavailable. However, this exercise is for pedagogical purpose, so I simply use the series as is, which already has over 500 observations.)
- I delete the last observation because it contains a missing value, providing no information.
- In general, the root cause of missing values should be thoroughly examined, and one should never simply impute it with mean (or any other values) without a good reason of doing so. One should also not simply delete the information without understanding the reason.

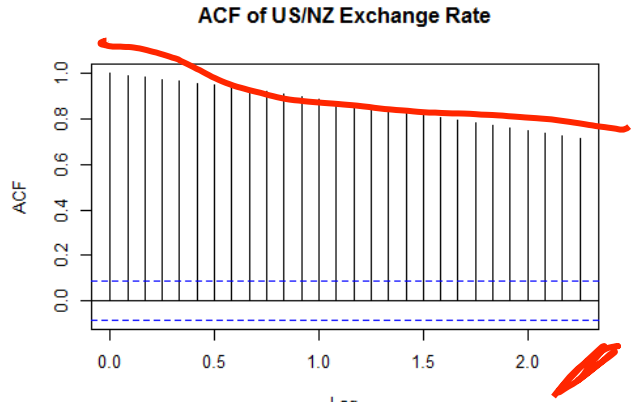
```
> exusnz2 <- exusnz[-532,] # omit the last observation because it is
> str(exusnz2)
'data.frame': 531 obs. of  2 variables:
 $ DATE : chr  "1971-01-01" "1971-02-01" "1971-03-01" "1971-04-01" .
 $ VALUE: chr  "1.1194" "1.1250" "1.1254" "1.1250" ...
```

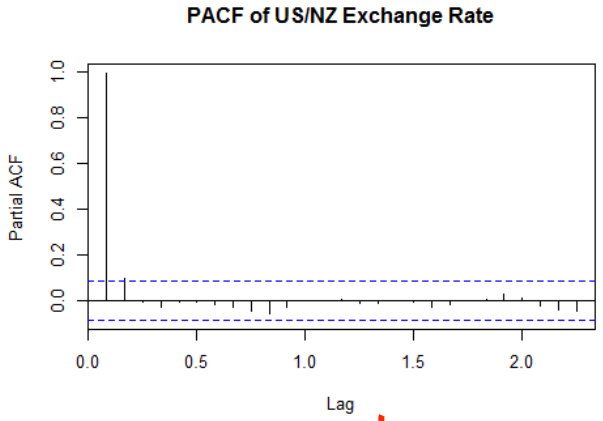
Data Visualization and Descriptive Statistics

- The visuals suggests that the series is not a sample path of a MA process.









Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.40	0.57	0.68	0.75	0.85	1.49

1%	5%	10%	25%	50%	75%	90%	95%	99%
0.41	0.46	0.51	0.57	0.68	0.85	1.12	1.30	1.45

Berkeley SCHOOL OF INFORMATION