Time Series Analysis

Lecture 5

Vector Autoregressive (VAR) Models

datascience@berkeley

Regression With Multiple Trending Time Series: An Introduction

Introduction

- Classical linear regression models assume that the stochastic errors are uncorrelated.

- In the context of time series analysis, however, this assumption is often violated.

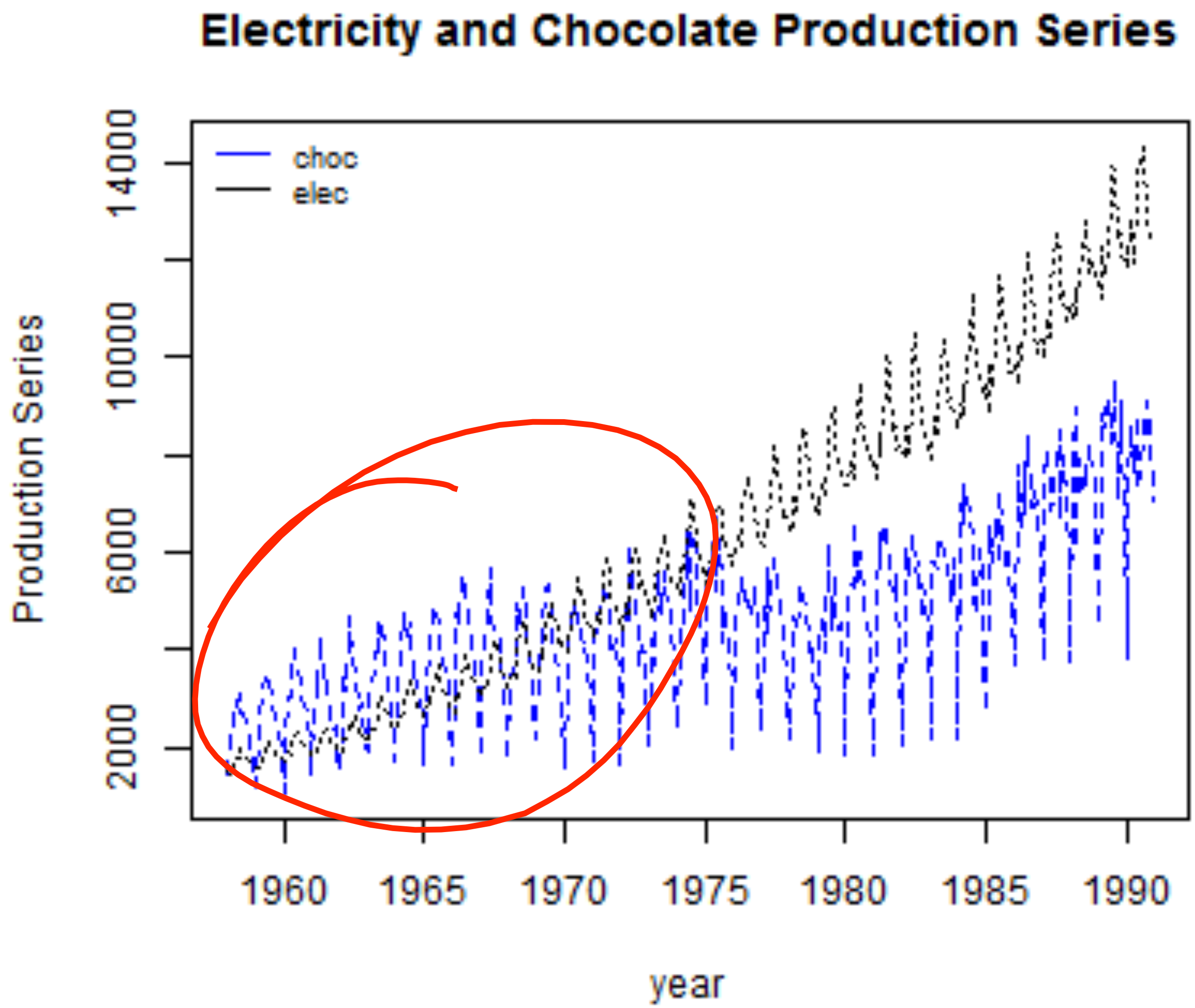
- More troublesome is that two independent time series could appear related to each other and have very high "correlation" when they are in fact independent of each other.

- We will study examples of this phenomenon and introduce statistical tests relevant for this situation.

- After discussing concepts such as spurious correlation, **cointegration**, and testing for **unit roots**, we will discuss **vector autoregressive models** that can be used to model multiple time series.

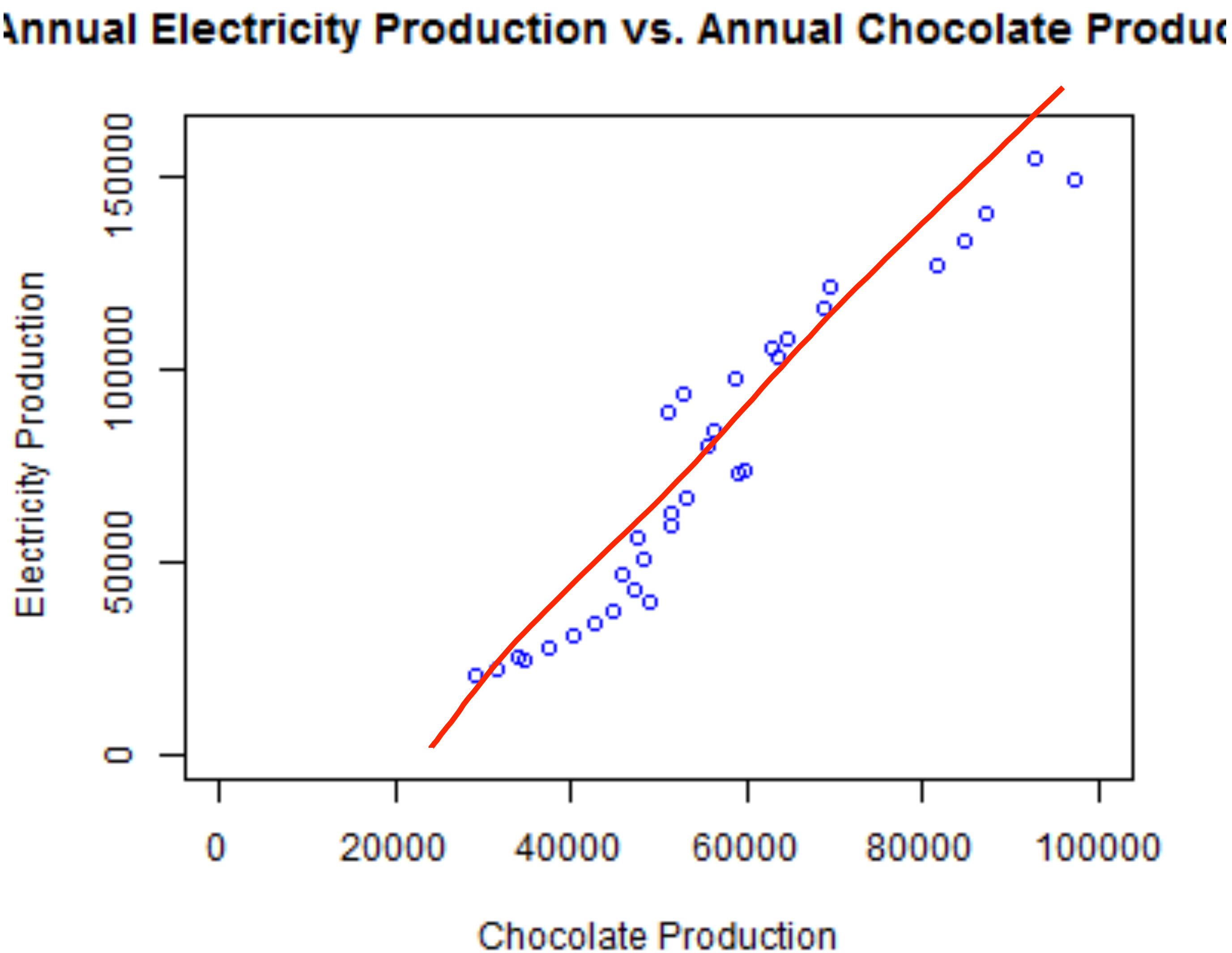
An Example: Electricity and Chocolate Production

- These are two series that we have worked with before and are provided by the book.



An Example: Electricity and Chocolate Production

- Aggregate the series into an annual series.
- According to the scatter plot, the correlation with the two series seems to be very high.
- In fact, the correlation is **0.96!**



An Example: Electricity and Chocolate Production

- It may be very tempting to fit a regression from one trending series on another and report high R^2. Don't!
- Fitting a regression of one variable as a linear function of the other, with added random variation, can often lead to *spurious* regression.
- The term *spurious regression* is also used when underlying stochastic trends in both series happen to be "coincident."

Correlation of Time Series With Trends

Stochastic Trends of Two Independent Series

- Stochastic trends are an important feature of ARIMA process with unit root.
- We can use two independent random walk series, whose correlation is -0.88, to illustrate the concept of spurious regression.

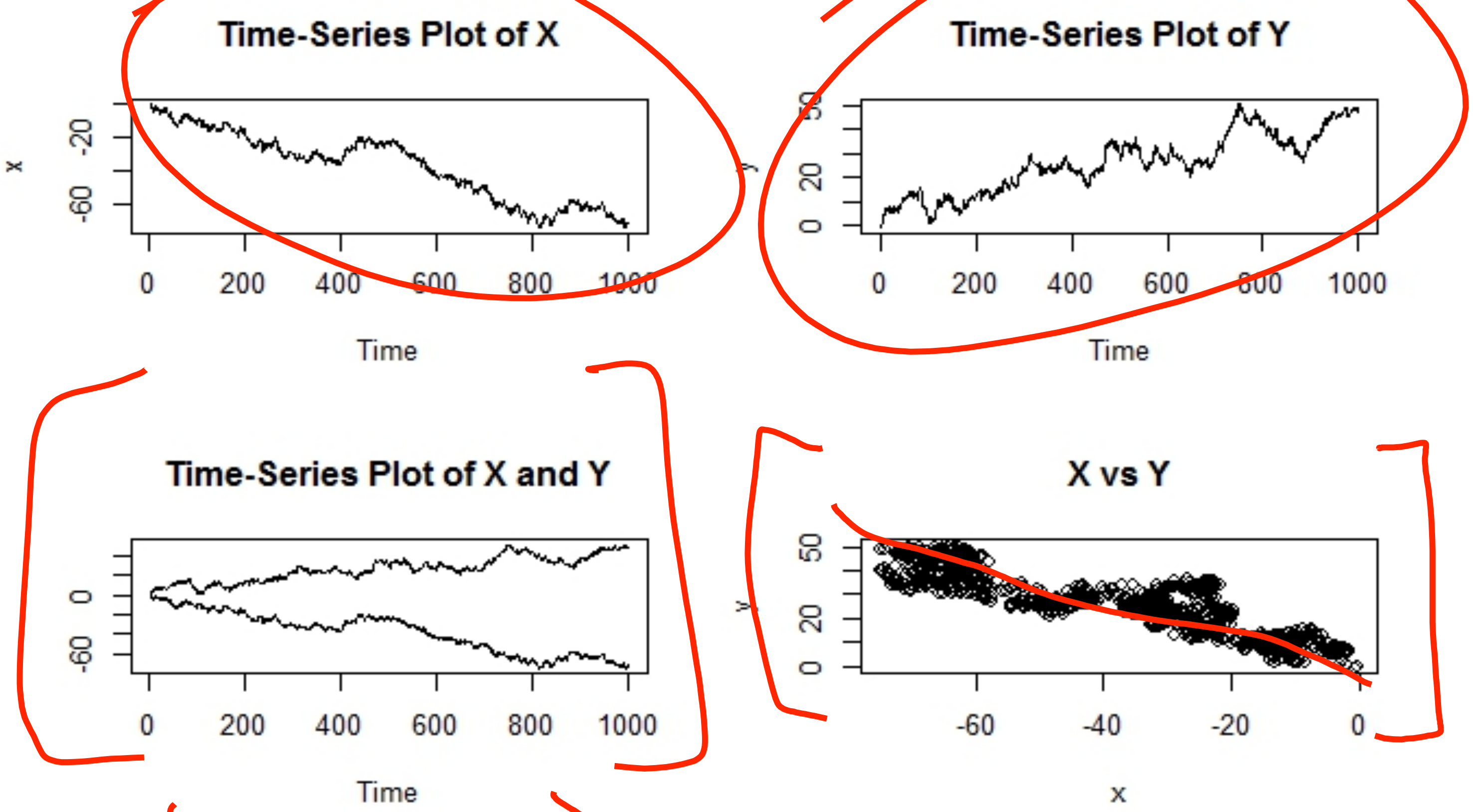
```
set.seed(14) # this seed is chosen (by trial and error) to produce
             # a high dependence between the two spurious regression
  #Create two independent white noise series
  x <- y <- rnorm(1000)
    cor(x,y)
    head(cbind(x,y))
    mean(x-y)
    sd(x-y)
  y <- rnorm(1000)

  # Create two independent random walks
  for (i in 2:1000) {
    x[i] <- x[i-1] + rnorm(1)
    y[i] <- y[i-1] + rnorm(1)
  }
```

Stochastic Trends of Two Independent Series

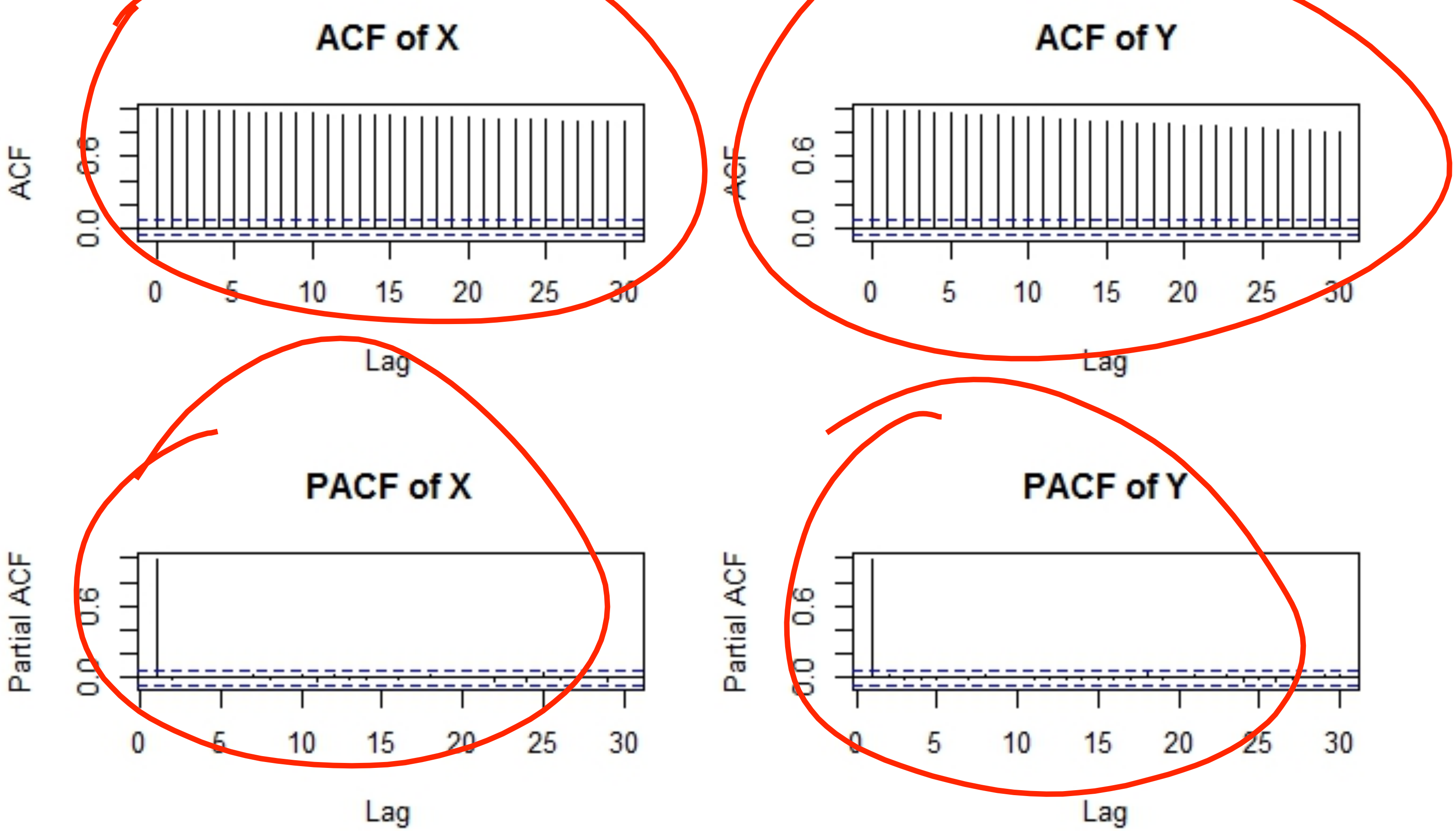
- Stochastic trends are an important feature of ARIMA process with unit root.

```
> cor(x,y)
[1] -0.88
> head(cbind(x,y), 10)
          x     y
 [1,] -0.66 -0.95
 [2,] -0.33 -0.24
 [3,] -2.09  0.77
 [4,] -3.00  0.80
 [5,] -2.96  2.46
 [6,] -3.40  4.75
 [7,] -5.52  4.24
 [8,] -4.61  4.87
 [9,] -3.21  5.96
[10,] -3.26  6.96
```



ACF of the Two Independent Series

- The ACF of the two series confirmed that both of the series have a strong trend.



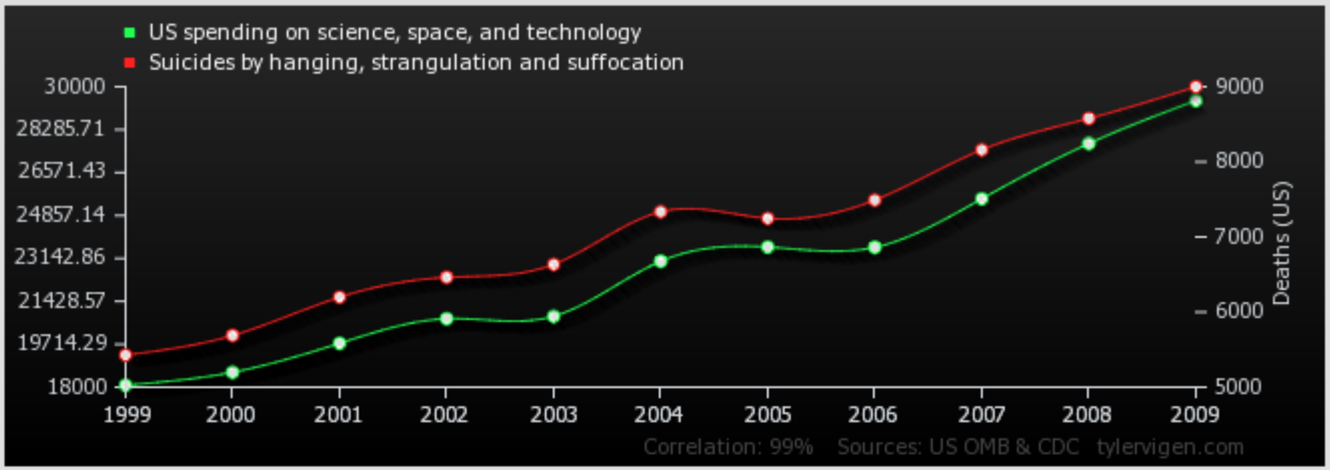
Spurious Correlation

Spurious Correlation: Example 1

US spending on science, space, and technology

correlates with

Suicides by hanging, strangulation and suffocation



	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
US spending on science, space, and technology *Millions of todays dollars (US OMB)*	18,079	18,594	19,753	20,734	20,831	23,029	23,597	23,584	25,525	27,731	29,449
Suicides by hanging, strangulation and suffocation *Deaths (US) (CDC)*	5,427	5,688	6,198	6,462	6,635	7,336	7,248	7,491	8,161	8,578	9,000

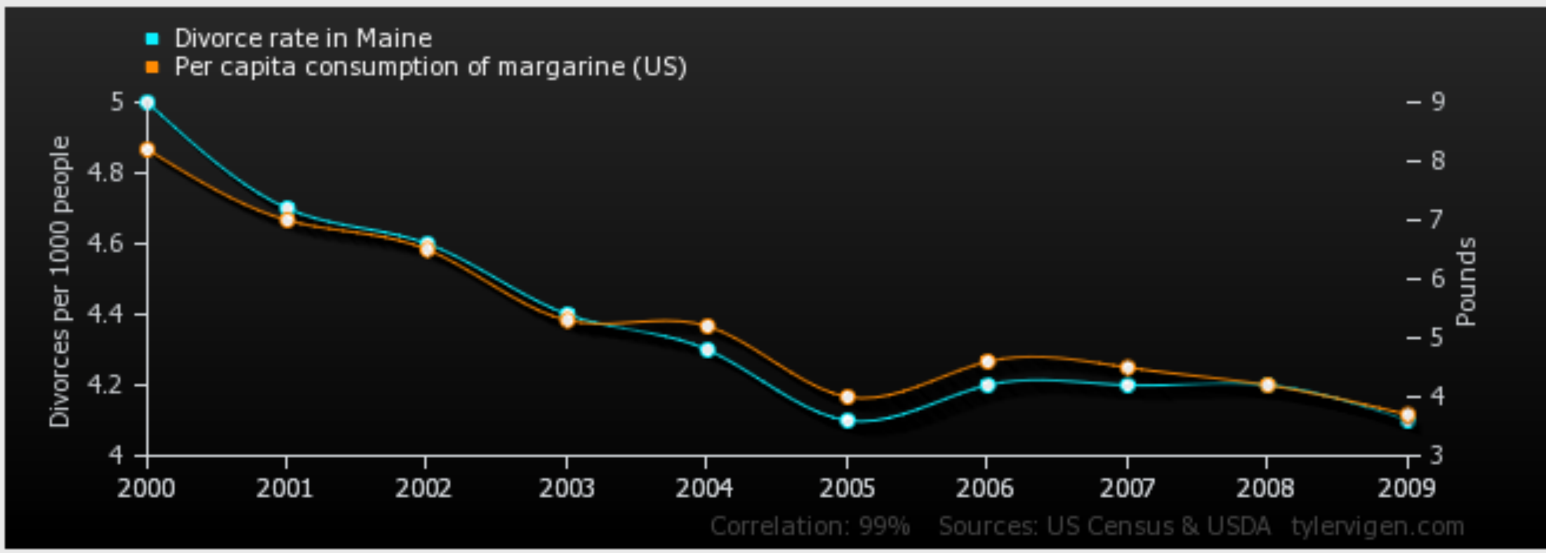
Correlation: 0.992082

Spurious Correlation: Example 2

Divorce rate in Maine

correlates with

Per capita consumption of margarine (US)



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Divorce rate in Maine *Divorces per 1000 people (US Census)*	5	4.7	4.6	4.4	4.3	4.1	4.2	4.2	4.2	4.1
Per capita consumption of margarine (US) *Pounds (USDA)*	8.2	7	6.5	5.3	5.2	4	4.6	4.5	4.2	3.7

Correlation: 0.992558

Spurious Correlation

- We just illustrated that two independent series (each with a stochastic trend) can produce high correlation!

- We also demonstrate that completely unrelated time series can generate high correlation.

- This is called "spurious correlation," which in general is used to describe a situation in which correlation between two variables is driven by some underlying common driver or that the correlation be "coincidental."

- You may find it trivial and say who would do that in practice, but you may be surprised how frequently you may encounter practitioners telling you how high the correlation is between two (trending) time, such as some revenue drivers of a company and macroeconomic time series.

Correlation Revisit: Mathematical Form

- Whenever we examine a trending time series, is correlation a good measure of the dependency of the two series?
- For that matter, is a sample mean a good measure of the "average" of a trending time series?
- Let's examine their mathematical forms of the sample estimates of mean, variance, covariance, and correlation. What is a key underlying assumption when applying these formulas?

Sample Mean: $m(\boldsymbol{x}) = \frac{1}{n}\sum^{n} x_i, \quad m(\boldsymbol{y}) = \frac{1}{n}\sum^{n} y_i$

Sample Variance: $s^2(\boldsymbol{x}) = \frac{1}{n-1}\sum_{i=1}^{n}[x_i - m(\boldsymbol{x})]^2, \quad s^2(\boldsymbol{y}) = \frac{1}{n-1}\sum_{i=1}^{n}[y_i - m(\boldsymbol{y})]^2$

Sample Covariance: $s(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{n-1}\sum_{i=1}^{n}[x_i - m(\boldsymbol{x})][y_i - m(\boldsymbol{y})]$

Sample Correlation: $r(\boldsymbol{x},\boldsymbol{y}) = \frac{s(\boldsymbol{x},\boldsymbol{y})}{s(\boldsymbol{x})s(\boldsymbol{y})}$

Correlation and Causation

- Also, as you've heard numerous times in this program, **correlation does not imply causation!**

- All of these examples clearly show that these (high) correlations really mean nothing and may even be misleading.

- However, in reality, spurious correlation between any two variables may not be as obvious. As data scientists, your bosses, clients, or colleagues may ask you to build a regression of trending time series. The concepts and techniques discussed in this lecture will be useful in dealing with these situations.

Spurious Correlation: Another Example

An Example From Three Currency Exchange Rates Series

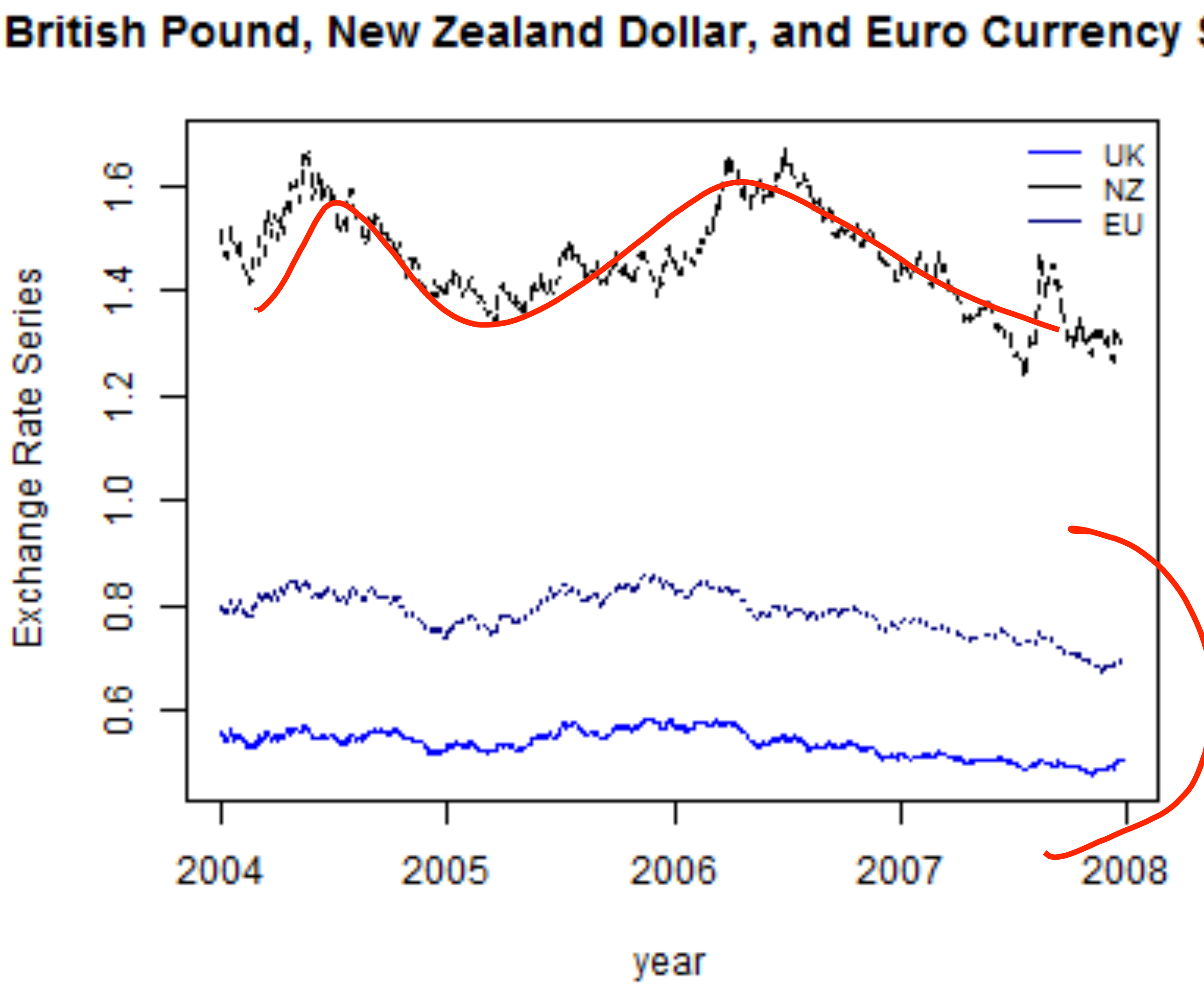
- This is a dataset we have used before, but we used only one of the three series.
- In this example, we will use all three series.

```
# First, we use the cleaned data set provided in CM2009
us_xrates <- read.table("C:/Users/K/z_Teach/MIDS_AdvStat/data/us_xrates.txt",
  str(us_xrates) # check the structure of the data
                 # 1003 observations and 3 varibles
  us_xrates[1:5,]
```

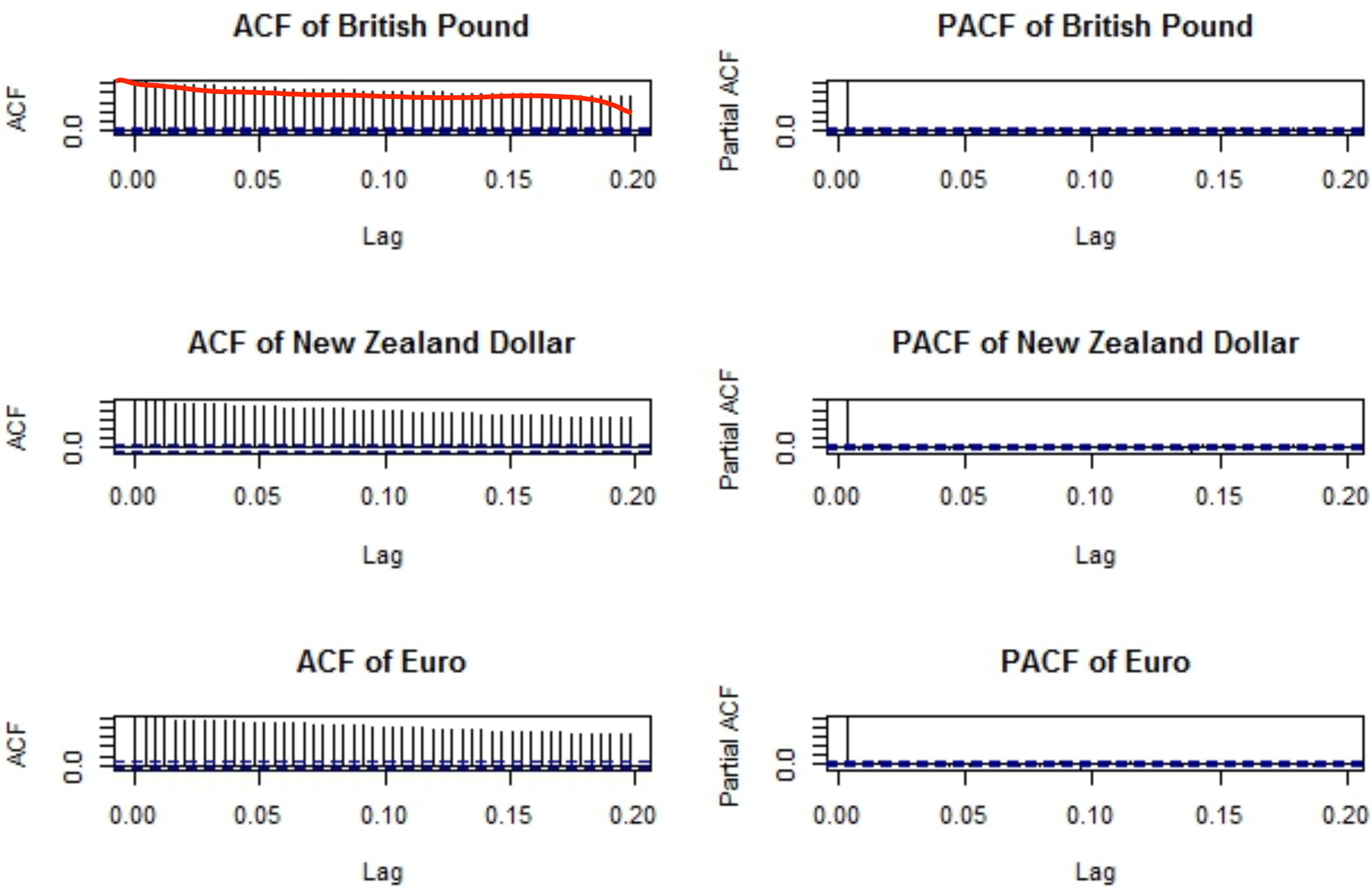
```
# Extract the currency series
UK.ts <- ts(us_xrates[,1], start=2004, freq=252)
NZ.ts <- ts(us_xrates[,2], start=2004, freq=252)
EU.ts <- ts(us_xrates[,3], start=2004, freq=252)

length(UK.ts); length(NZ.ts); length(EU.ts)
t=c(1:length(UK.ts))
```

Currency Exchange Rates Series: T-Plots

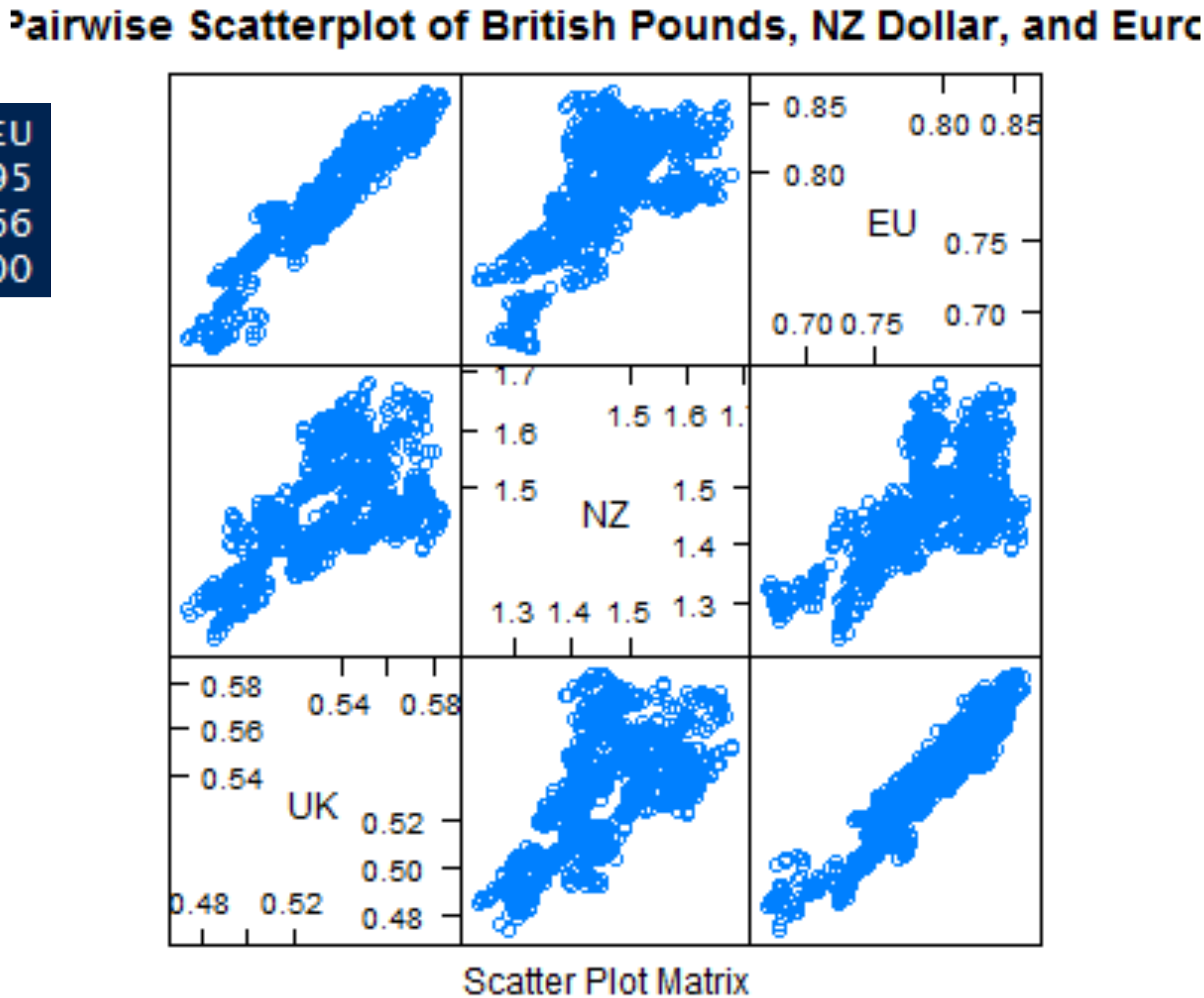


Currency Exchange Rates Series: ACF

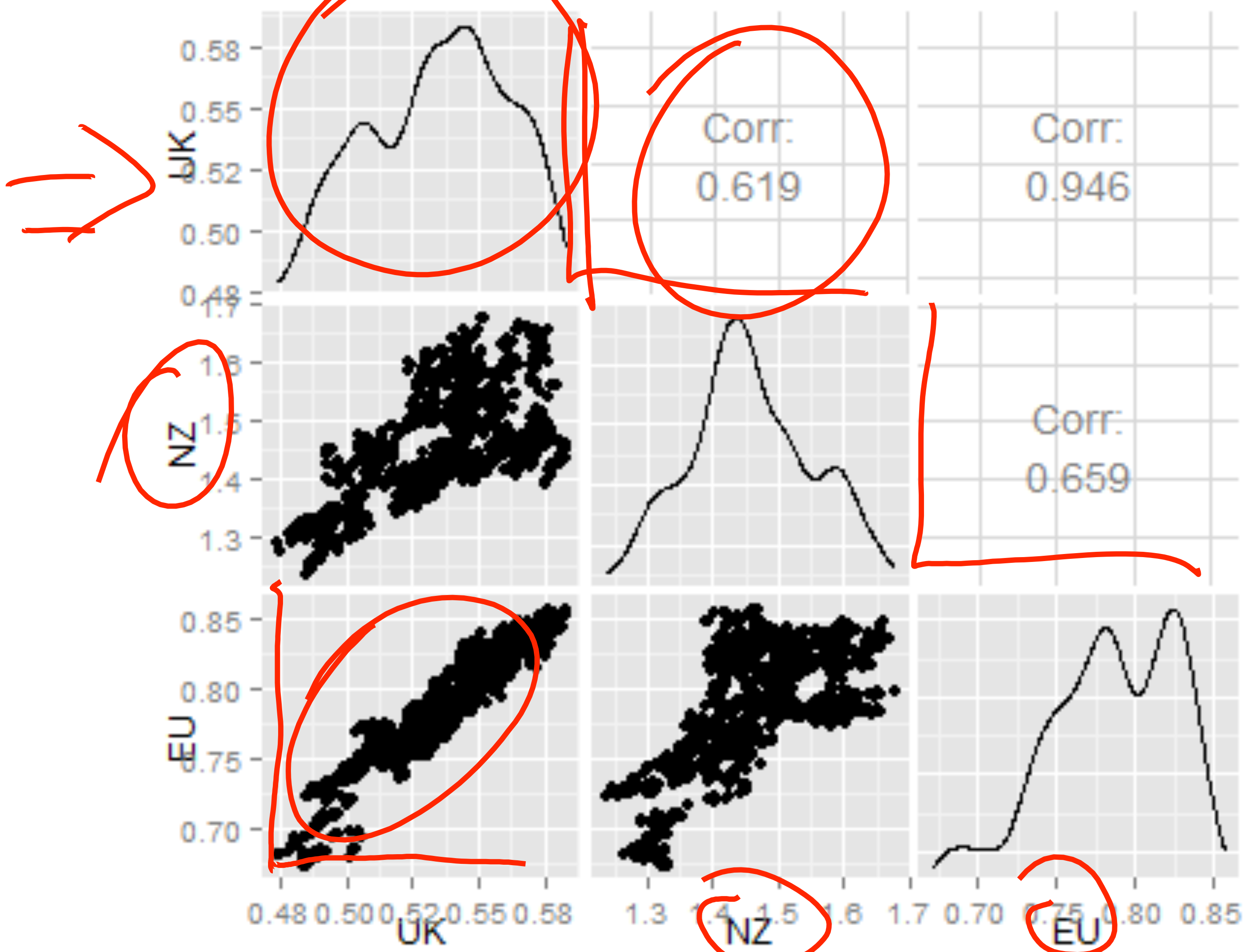


Currency Exchange Rates Series: Scatter Plot Matrix

	UK	NZ	EU
UK	1.00	0.62	0.95
NZ	0.62	1.00	0.66
EU	0.95	0.66	1.00



Currency Exchange Rates Series: Scatter Plot Matrix 2



Unit Root Nonstationarity and Dickey–Fuller Test: Definition and Intuition

Unit Roots

Recall that a stochastic process has a unit root if 1 is a root of the process's characteristic equation. Such a process is nonstationary.

To detect the existence of unit roots, unit root tests can be used to determine if a series should be first differenced or regressed on a deterministic function of time to render the series stationary.

Dickey–Fuller Test

An **augmented Dickey–Fuller test (ADF)** is a test for a unit root in a time series. It is an augmented version of the original Dickey–Fuller test for a larger and more complicated set of time series models.

The test procedure for the ADF test is the same as the DF test, but it is applied to the following model:

$$\Delta y_t = \boxed{\alpha + \beta t} + \boxed{\gamma y_{t-1}} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

The unit root test is then carried out under the null hypothesis $\boxed{\gamma = 0}$ against the alternative hypothesis of $\gamma < 0$.

Dickey–Fuller Test: Intuition

The intuition behind the test is that if the series is not integrated, then the lagged level of the series (y_{t-1}) will provide no relevant information in predicting the change in y_t besides the one obtained in the lagged changes (Δy_{t-k}).

In that case the $\gamma = 0$ null hypothesis is not rejected.

Dickey–Fuller Test (2)

Once a value for the test statistic

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

is computed, it can be compared to the relevant critical value for the Dickey–Fuller Test.

If the test statistic is less (this test is nonsymmetrical so we do not consider an absolute value) than the (larger negative) critical value, then the null hypothesis of $\gamma = 0$ is rejected and no unit root is present.

Unit Root Nonstationarity and Dickey–Fuller Test: An Example

Unit Root Nonstationarity and Dickey–Fuller Test

Dickey and Fuller developed a test of the null hypothesis that $\alpha = 1$ against an alternative hypothesis that $\alpha < 1$ for the model $x_t = \alpha x_{t-1} + u_t$ in which u_t is white noise. A more general test, which is known as the augmented Dickey-Fuller test (Said and Dickey, 1984), allows the differenced series u_t to be any stationary process, rather than white noise, and approximates the stationary process with an AR model. The method is implemented in R by the function `adf.test` within the `tseries` library. The null hypothesis of a unit root cannot be rejected for our simulated random walk x:

Use the library called **tseries** and the embedded series to illustrate the mechanics of conduction the ADF test in R:

```
library(tseries)
adf.test(x)
```

Even though this is a series coming with the library, it is a good idea to get familiar with the series.

Unit Root Nonstationarity and Dickey–Fuller Test

Basic structure and descriptive statistics of the series:

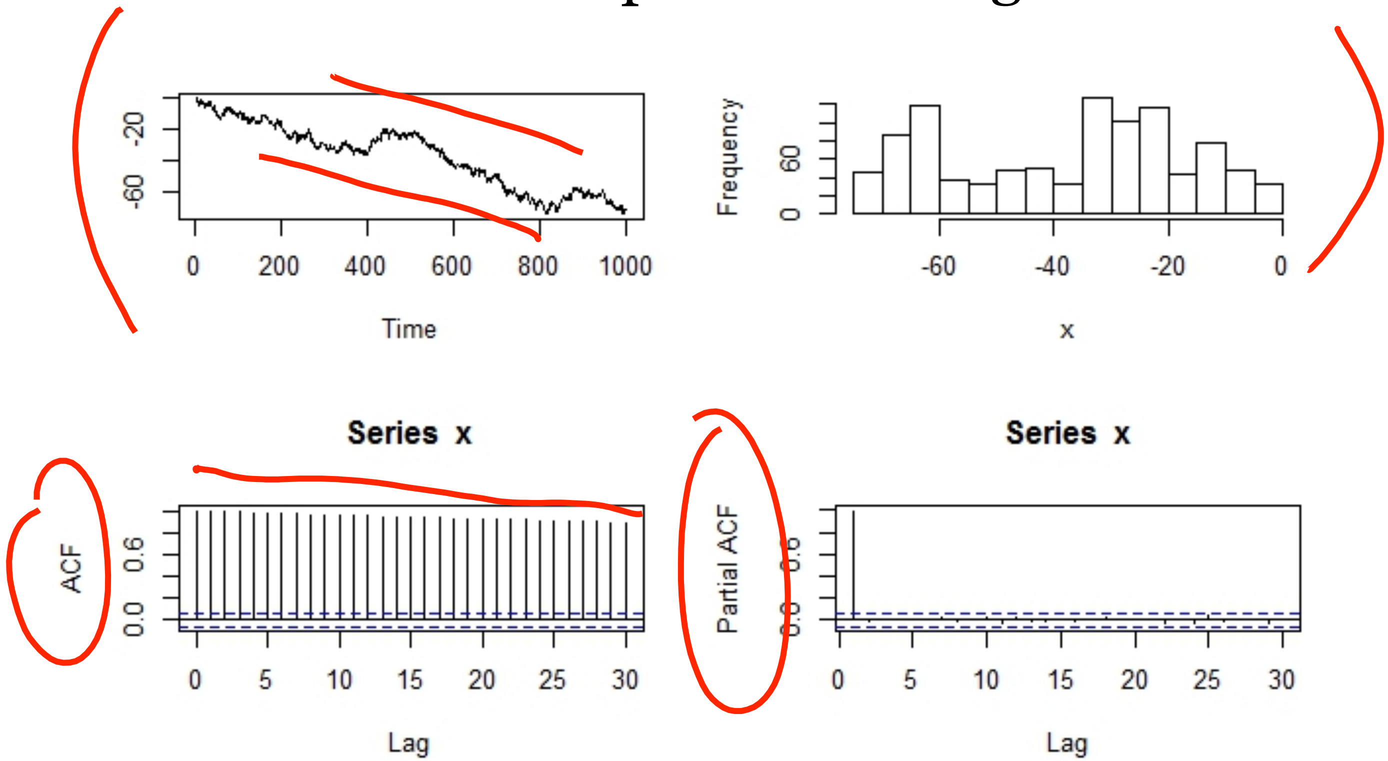
```
> str(x)
 num [1:1000] -0.662 -0.331 -2.089 -3.002 -2.962 ...
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -75     -60     -33     -38     -23       0
```

In addition to the basic structure of the data and the descriptive statistics that describe various quantiles and the mean of the distribution, we will still look at graphs to examine the dynamics of the series.

Unit Root Nonstationarity and Dickey–Fuller Test

- The series clearly is not stationary.
- The autocorrelation is almost nondecreasing while the partial autocorrelation sharp drops to zero in the first lag.
- These are all evidences of a process having unit roots.



Unit Root Nonstationarity and Dickey–Fuller Test

- The mechanism of applying the augmented Dickey–Fuller test in R is very straight-forward.
- It simply call the command adf.test(x).

```
 Augmented Dickey-Fuller Test

data:  x
Dickey-Fuller = -2.2, Lag order = 9, p-value = 0.5003
alternative hypothesis: stationary
```

The ADF test result indicates that the null hypothesis (that the series has a unit root) cannot be rejected.

This result is not surprising at all, based on the data visuals that we have seen.

Cointegration Definition

Cointegration

Definition: Two non-stationary time series $\{x_t\}$ and $\{y_t\}$ are **cointegrated** if some linear combinations $ax_t + by_t$, where a and b are constants, is a stationary series.

Consider two random walks:

$$x_t = \mu_t + w_{x,t}$$

$$y_t = \mu_t + w_{y,t}$$

where $w_{x,t}$ and $w_{y,t}$ are independent white noise and the series $\{\mu_t\}$ is given by

$$\boxed{\mu_t = \mu_{t-1} + w_t}$$

with $\{w_t\}$ being a zero-mean white noise processes.

Cointegration

As they are both random walks, they are non-stationary processes.

However, their difference $\{x_t - y_t\}$

$$
\begin{aligned}
x_t - y_t &= (\mu_t + w_{x,t}) - (\mu_t + w_{y,t}) \\
&= w_{x,t} - w_{y,t}
\end{aligned}
$$

is just the difference (or a linear combination with $a = 1$ and $b = -1$ as defined above in the cointegration definition) between two independent white noise processes. This newly formed cointegrated series is covariance stationary. As an exercise, derive the mean and variance of this series.

Exercise: Derive the mean, variance, and covariance of the cointegrated series $x_t - y_t$.

Multivariate Time Series Models: An Introduction to Vector Autoregressive Models

Vector AR Models: Mathematical Formulation

- The good news of studying multivariate time series models in the form of vector autoregressive models is that the univariate autoregressive models can be easily generalized to their multivariate counterpart, although the mathematical notation is heavier.

- Consider a simple vector autoregressive process of order 1 (VAR(1)):

$$x_t = \phi_{11} x_{t-1} + \phi_{12} y_{t-1} + \omega_{x,t}$$
$$y_t = \phi_{21} x_{t-1} + \phi_{22} y_{t-1} + \omega_{y,t}$$

where $\{w_{x,t}\}$ and $\{w_{y,t}\}$ are bivariate white noise and ϕ_{ij} are the model parameters.

VAR Models in Matrix Form

Exercise: Show that if the white noise processes have mean 0 and the VAR(1) process is stationary, then both $\{x_t\}$ and $\{y_t\}$ have mean 0.

As in AR(p) models, to incorporate a mean, simply define $\{x_t\}$ and $\{y_t\}$ as deviations from their means.

To derive properties of the VAR(p) process, it is easier to use matrix notations:

ϕ $\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix}$

$$\mathbf{Z_t} = \Psi \mathbf{Z_{t-1}} + \mathbf{w_t}$$

$$\underset{2 \times 1}{\mathbf{Z_t}} = \begin{pmatrix} x_t \\ y_t \end{pmatrix} \quad \Psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix} \quad \mathbf{w_t} = \begin{pmatrix} w_{x,t} \\ w_{y,t} \end{pmatrix}$$

VAR Models: Stationarity Condition

Expressed in backward shift operator:

$$(\mathbf{I} - \Psi(B))\,\mathbf{Z_t} = \theta(B)\mathbf{Z_t} = \mathbf{w_t}$$

where θ is a matrix polynomial of order 1 and $\mathbf{I}$ is a 2 x 2 identity matrix.

The mechanics and concepts of VAR(1) process can be easily extended to the general VAR(p) process of m time series, in which case θ becomes a matrix polynomial of order p and is a m x m matrix of parameters, $\mathbf{I}$ becomes a $(m \text{ x } m)$ identity matrix, and $\mathbf{Z_t}$ is a m x 1 matrix of time series variables, and $\mathbf{w_t}$ is a multivariate white noise.

Stationarity of the VAR(p) model is similarly defined as that of the AR(p) model: **The roots of the characteristic equation all lie outside of the unity circle.**

In the case of VAR(p) model, the characteristic equation is given by the determinant of the ***theta*** matrix defined above.

Finding the Roots of the Characteristic Polynomial

Back to the VAR(1) model, the determinant is

$$\begin{vmatrix} 1-\theta_{11}x & -\theta_{12} \\ -\theta_{21}x & 1-\theta_{22}x \end{vmatrix} = (1-\theta_{11}x)(1-\theta_{11}x) - \theta_{12}\theta_{21}x^2$$

- However, although this polynomial can be easily solved using papers and pencils, we can use ***R*** functions ***polyroot*** and **Mod**.
- The function ***polyroot***, coming with the base package, is used to find zeros of a real or complex polynomial of the form $p(x) = z_1 + z_2 x + \ldots + z_n x^{n-1}$. For our purpose, it suffices to know how to use this function to find roots in ***R***.
- Interested readers can refer to the documentation https://stat.ethz.ch/R-manual/R-devel/library/base/html/polyroot.html and the reference therein.

Vector AR Models: Introduction

The function **Mod** is one of the basic functions that supports complex arithmetic in **R**. Detailed documentation can be found here `https://stat.ethz.ch/R-manual/R-patched/library/base/html/complex.html` or within **R** by typing help(Mod). We will have to use this function because roots of characteristic polynomials can be complex roots. Since our interests lie only in the absolute value of the roots, we will apply the **Mod** function to roots of the characteristic polynomials and examine if the absolute value of the roots are bigger than 1.

Ψ

As an example, if the VAR(1) model has the parameter matrix

$$\begin{array}{cc} \phi_{11} & \phi_{12} \\ \end{array}$$

$$\Psi = \begin{pmatrix} 0.4 & 0.3 \\ 0.2 & 0.1 \end{pmatrix}$$

$$\begin{array}{cc} \phi_{21} & \phi_{22} \\ \end{array}$$

and the corresponding characteristic polynomial is given by

$$\theta = \begin{vmatrix} 1-0.4x & -0.3x \\ -0.2x & 1-0.1x \end{vmatrix} = (1-0.5x-0.02x^2)$$

The absolute value of the roots of this equation can be found using ***R***: **Mod(polyroot(c($1, -0.5, -0.02$)))**.

```
> Mod(polyroot(c(1, -0.5, -0.02)))
[1] 1.9 26.9
```

Berkeley SCHOOL OF INFORMATION