LECTURE 1. INTRODUCTION

JEN-WEN LIN, PhD, CFA

01/03/2018

COURSE INFORMATION

- Course instructor: Jen-Wen Lin, PhD, CFA
- Contact email: **jenwen@utstat.toronto.edu**
- Time/location: Thursday 3-6pm/AH100
- Office hours/location: 0200-0245pm/TBD
- Others:
  - Lecture slides will be posted on Portal
  - Tell me your answer before asking the question

REFERENCE BOOKS

1. *Time Series Analysis—Univariate and Multivariate Methods*
2. *The Analysis of Time Series: An Introduction*
3. *Applied Econometric Time Series*
4. *Time Series Analysis: Forecasting and Control*
5. *Time Series Analysis*
6. *Time Series: Theory and Methods*
7. *Time Series Analysis and Its Applications With R Examples*

WHAT TO TEST IN THE EXAM

- Interview questions (mainly financial industries)
- "Important" and "Practical" time series modeling concepts and techniques

TENTATIVE SCHEDULE

Week	Date	Topic (tentative)
1	4-Jan-18	Introduction
2	11-Jan-18	ARMA model I
3	18-Jan-18	ARMA model II
4	25-Jan-18	ARIMA model, Unit roots, and Forecasting tests
5	1-Feb-18	Transfer function noise model
6	8-Feb-18	Vector autoregression and Granger causality test
7	15-Feb-18	Midterm test
8	**22-Feb-18**	**Read Week-no classes**
9	1-Mar-18	Cointegration (return midterm test)
10	8-Mar-18	Bootstrap, bagging and boosting time series models
11	15-Mar-18	TBD*
12	22-Mar-18	TBD*
13	29-Mar-18	TBD*

* : Something related big data and ML, such as Bagging/boosting/ensemble time series forecast

MARKING SCHEME (TENTATIVE)

- **Undergraduate**

  midterm test + final exam + class participation (tentative)

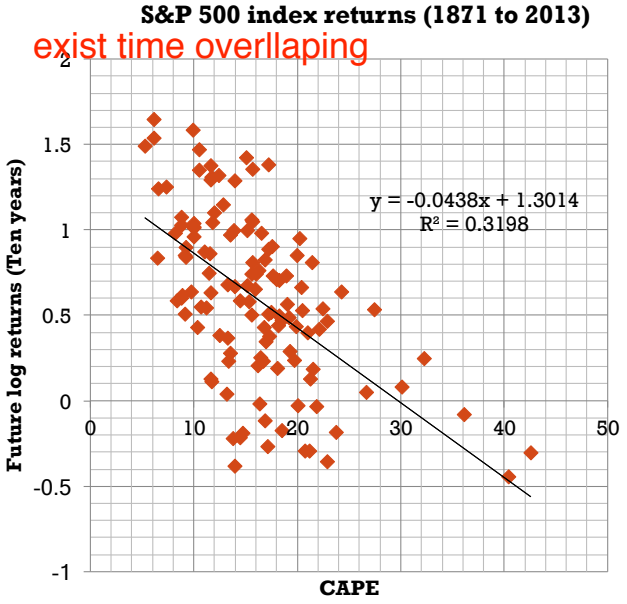
  - If students miss the midterm test with a legitimate reason, his/her weight on the midterm test will be shifted to final exam
  - Students usually have high marks in their midterm test so do your best to write the midterm test
  - No pencil is allowed in the test

- **Graduate student**
  - Research project and presentation

WHAT DID WE LEARN IN STATS 201

1. Model identification and specification
   - Use descriptive statistics to understand data: Preliminarily assess data and guess possible statistical models
   - Statistical model: e.g. normal distribution
2. Model estimation
   - Such as maximum likelihood estimation
3. Model evaluation (Hypothesis testing)
   - Goodness of fit test
   - E.g. H_0: mu=0

CAMPBELL AND SHILLER (1998, JPM)



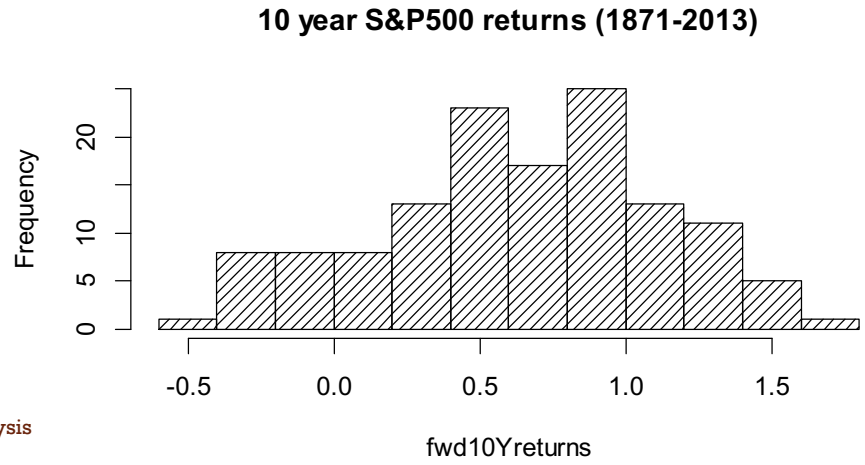
Source: http://www.econ.yale.edu/~shiller/data.htm.

- When a valuation ratio* (e.g. price to earnings ratio) is at an extreme level, either the numerator or the denominator of the valuation ratio must move in a direction that restores the ratio to a more normal level.
- Campbell and Shiller (1998) conclude that it is the stock price (denominator) that has moved to restore the ratio to its mean level.
- Campbell and Shiller (1998) tested this hypothesis by regressing forward ten year returns against Cyclically Adjusted PE ratio (CAPE).

WHAT DID WE LEARN IN STATS 201

—MODEL IDENTIFICATION

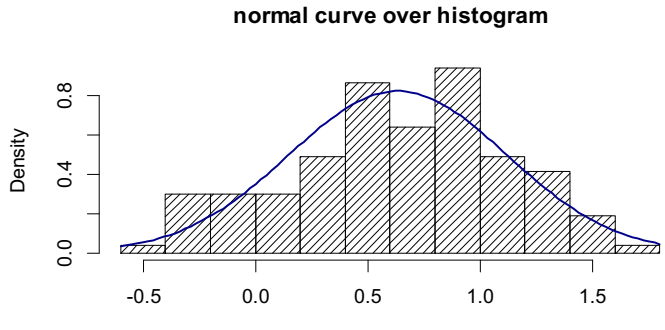
- **summary(fwd10Yreturns) #compute descriptive statistics**
  - `Min. 1st Qu. Median Mean 3rd Qu. Max. NA's`
  - `-0.4438 0.3437 0.6593 0.6341 0.9805 1.6440 10`
- **hist(fwd10Yreturns)#plot histogram**



WHAT DID WE LEARN IN STATS 201

—MODEL IDENTIFICATION AND SPECIFICATION

- Assume that stock returns follow a normal distribution which is characterized by mean and variance.
  - `mu<-mean(fwd10Yreturns,na.rm = TRUE)  #MLE estimate`
  - `sg<-sqrt(var(fwd10Yreturns, na.rm = TRUE))#MLE estimate`
- Histogram again

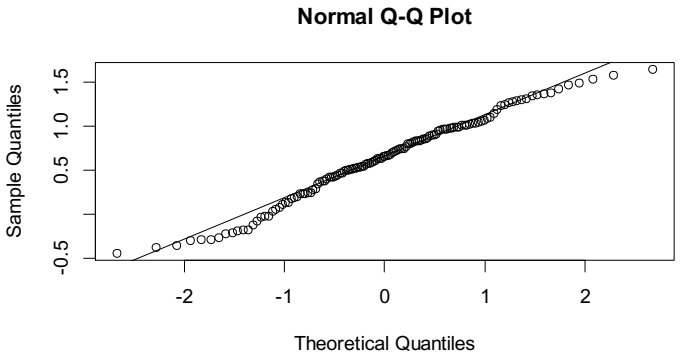



```
#R code
hist(fwd10Yreturns,
prob=TRUE,
density=20,main="no
rmal curve over
histogram")

curve(dnorm(x,
mean=mu,
sd=sg),col="darkblu
e", lwd=2,
add=TRUE, yaxt="n")
```

WHAT DID WE LEARN IN STATS 201 —MODEL EVALUATION (HYPOTHESIS TESTING)

- Goodness of fit




```
#R code
qqnorm(fwd10Yreturns)

qqline(fwd10Yreturns)
```

- H_0: mu=0
- **`t.test(fwd10Yreturns) # One Sample t-test`**
  - `t = 15.1072, df = 132, p-value < 2.2e-16`
  - `alternative hypothesis: true mean is not equal to 0`
  - `sample estimates: mean of x = 0.6340763`

WHAT STATISTICAL MODELS DID WE LEARN IN STATS 201

- Statistical models
  1. Univariate model: such as Chi-squared, F, normal and Student t distributions
     - To infer/predict the characteristics of data itself
  2. Multivariate model: such as linear regression and bivariate normal distribution
     - To infer/predict the characteristics of data using other variables
- Note: conventional statistics usually assume all observations are identically and independently distributed (IID)

WHAT IS TIME SERIES

- A time series is a set of observations measured sequentially through time.
  - These measurements may be made continuously through time or be taken at a discrete set of time points.
    - Chris Chatfield (2000), "Time-Series Forecasting", Chapman & Hall/CRC

DESCRIPTIVE STATISTICS

1. Main descriptive statistics for time series data include
    1) Sample mean
    2) Sample autocorrelation (autocovariance) function and partial autocorrelation function

- Like conventional statistics, these sample estimates are used as our preliminary estimates of parameters of statistical models and may be used for model identification

CHARACTERISTICS IN TIME SERIES

- Visualization of characteristics of time series data
  - "*Time series data and visualization*"
- In addition to the presence of trend, seasonal and cyclical components, time series data are usually dependent with past observations
  - Autocorrelation or serial correlation (with its own past observations)
  - Cross-correlation (with past observations of other variables)
- Unstable first two moments over time
  - Fail weak stationarity

AUTO/CROSS-COVARIANCE FUNCTION

- If $\{X_t, t \in T\}$ is a process such that $var(X_t) < \infty$ for each $t \in T$, then the **autocovariance function** $\gamma_X(\cdot,\cdot)$ of $\{X_t\}$ is defined by

$$\gamma_X(r,s) = cov(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)], \qquad \forall r, s,$$

and the **autocorrelation function** is given by

$$\rho(r,s) = \gamma_X(r,s)/\sqrt{\gamma_X(r,r)\gamma_X(s,s)}\,.$$

- The **cross-covariance/correlation** between $\{X_t\}$ and $\{Y_t\}$ are

$$\gamma_{XY}(r,s) = E[(X_r - EX_r)(Y_s - EY_s)],$$

$$\rho_{XY}(r,s) = \gamma_{XY}(r,s)/\sqrt{\gamma_X(r,r)\gamma_Y(s,s)}, \forall r, s,$$

WEAKLY STATIONARY TIME SERIES

- **Definition.** The time series $\{X_t, t \in Z\}$, with index set $Z = \{0, \pm 1, \pm 2, \dots\}$, is said to be stationary if

1. $E|X_t|^2 < \infty$ for all $t \in Z$
2. $EX_t = m$ for all $t \in Z$ (m is a constant)
3. $\gamma_X(r,s) = \gamma_X(t+r, t+s)$ for all $t, r, s \in Z$ ($\gamma_X(\cdot)$ is a constant and $r - s$ is independent of t)

Let $k = r - s$. Weakly stationarity implies

$$\gamma(k) := cov(X_{k+1}, X_1) = cov(X_{k+2}, X_2) = \cdots,$$

and

$$\rho(k) := \frac{\gamma(k)}{\gamma(0)}, \qquad k = 0,1,2, \dots$$

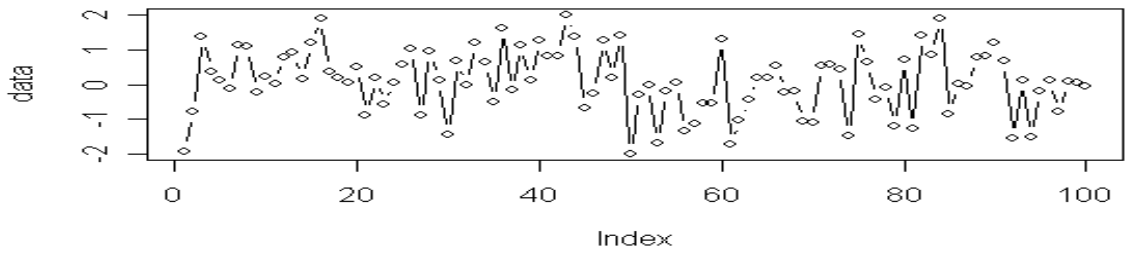
SAMPLE AUTOCORRELATION FUNCTIONS (SACF)

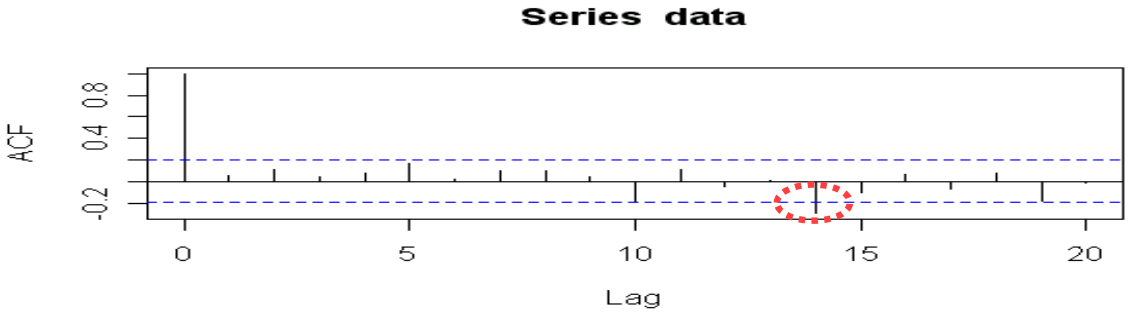
Plotting the sample autocorrelation functions (SACF) $\hat{\rho}(h)$ against the lag h for $h = 1,2, \dots, M$ are useful tool in interpreting time series, where M is usually much less than the series length.

- $\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})$ with $\hat{\gamma}(-h) = \hat{\gamma}(h)$ for $h = 0,1, \dots, n-1$

- $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$

For a random time series, $\hat{\rho}_k$ is approximately $N\left(0, \frac{1}{N}\right)$ *for* $k \neq 0$, where N is the length of the series. Thus, if a time series is random, we can expect 19 out of 20 of the values of $\hat{\rho}_k$ to lie between $\pm 2/\sqrt{N}$.

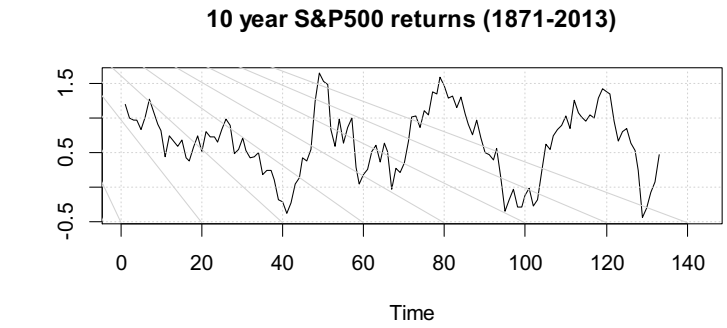


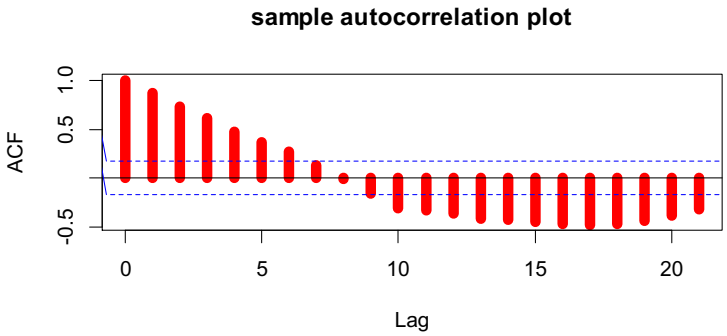


R code:

```
••data<-rnorm(100)
••par(mfrow=c(2,1))
••plot(data,type="b")
••acf(data)
```

WHAT'S WRONG WITH THE CONVENTIONAL ANALYSIS





- The fitted model and the hypothesis testing are not justified now.

- We are going to learn how to conduct the right analysis for time series data in this course

SPURIOUS CORRELATION IN TIME SERIES

Yule (1926), "Why do we Sometimes get Nonsense-Correlation between Time-Series?—A Study in Sampling and the Nature of Time-Series", *Journal of the Royal Statistical Society*, vol. 89, No. 1

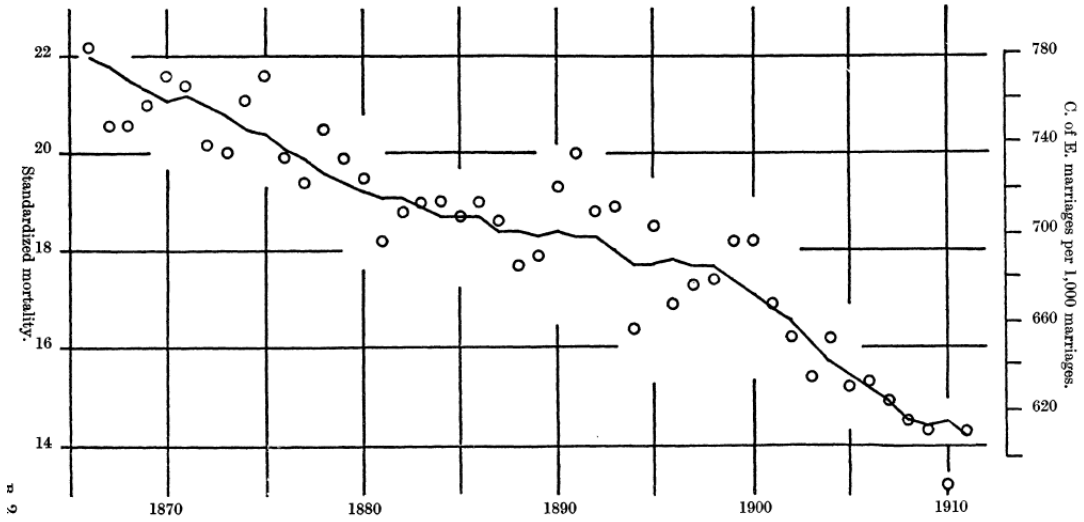


FIG. 1.—Correlation between standardized mortality per 1,000 persons in England and Wales (circles), and the proportion of Church of England marriages per 1,000 of all marriages (line), 1866–1911. $r = +0{\cdot}9512$.

CORRELATION IN SIMPLE REGRESSION

$$y_t = \alpha + \beta x_t + e_t$$

$$cov(y_t, x_t) = cov(\alpha + \beta x_t + e_t, x_t) = \beta cov(x_t, x_t) = \beta \sigma_x^2$$

$$cov(y_t, x_t) = \rho_{xy}(0) \cdot \sigma_x \sigma_y$$

$$\beta = \frac{cov(y_t, x_t)}{\sigma_x^2} = \rho_{xy}(0) \cdot \frac{\sigma_y}{\sigma_x} \propto \rho_{xy}(0)$$

- Since standard deviation is nonnegative, we may test whether beta is zero via testing whether the cross correlation function (CCF) at lag zero is statistically different from zero.
- How to test whether CCF is statistical different from zero?

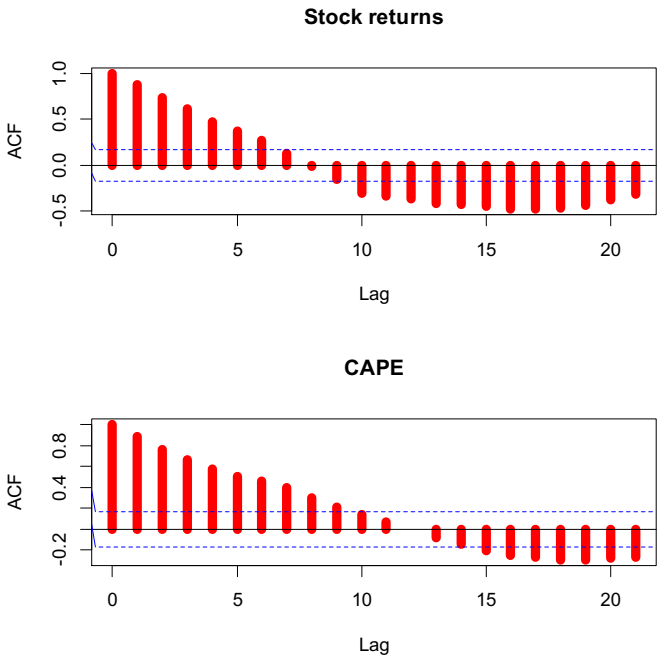
TEST ZERO CROSS-CORRELATION

- If both x_t and y_t contain no serial correlation and mutually independent, the sample CCF, $r_{xy}(k)$, is approximately normally distributed with zero mean and variance $1/n$, where n is the sample size (or the number of pairs of available).
  - The above statement is incorrect if x_t and y_t are serially correlated (even if the processes x_t and y_t are independent of each other).
- Under the assumption that both x_t and y_t are stationary and that they are independent of each other, $\sqrt{n}\, r_{XY}(k)$ is asymptotically normal with mean zero and variance

$$1 + 2\sum_{j=1}^{\infty} \rho_X(j)\rho_Y(j),$$

where rho(k) is the autocorrelation at lag k. For refinement of this asymptotic result, see Brockwell and Davis (1991, p.410).

REVISIT CAMPBELL AND SHILLER (1998, JPM)



- Both stock returns and CAPE exhibit severe serial correlation (persistent).
- Ignore the presence of serial correlation. The estimate of CCF at lag zero is -0.57 and the p-value for zero CCF is 3.6e-10.
- Take into account the serial correlation using the prewhitening technique (to discuss latter in the course). The prewhitened CCF at lag zero is -0.192 and the corresponding p-value for zero CCF is 0.033 (unable to reject H_0 at 99% significance level).
- In practice, we usually test the significance of regression coefficients based on HAC estimator, such as Newey and West (1987).

SPURIOUS CORRELATION IN TIME SERIES

- Correlation does not imply causation.
- Characteristics of time series lead to various forms of spurious correlation/regression.
  - Theoretical justification, correct model specification, and tests are useful to avoid making decision based on spurious correlation.
- In this class, we will introduce some techniques to deal with spurious correlation in time series.

APPROACHES TO TIME SERIES ANALYSIS

Time domain analysis

- Focuses on modeling some future values of a time series as a parametric function of the current and past values

Frequency domain analysis

- Assumes that the primary characteristics of interest in time series analyses relate to periodic or systematic sinusoidal variations found naturally in most data

In many cases, the two approaches may produce similar answers for long series, but the comparative performance over short samples is better done in the time domain.

DECOMPOSITION OF TIME SERIES

Statisticians usually decompose a time series into components representing

- trend
- seasonal variation
- other cyclic changes
- irregular fluctuations.

This decomposition is sometimes referred to as the classical decomposition model in time series analysis.

Seasonal variation	Cyclic variation	Trend
• Time series exhibit variation that is annual in period (or every 12 units of time). • For example, the sales of electronic companies in the second quarter are typically the lowest.	• Time series exhibit variation at a fixed period due to some other physical cause. • Examples are daily variation in temperature and business cycles.	• This may be loosely defined as 'long-term change in the mean level'.

STEPS TO TIME SERIES MODELING

1. Plot the time series and check for

- Trend, seasonal and other cyclic components, any apparent sharp changes in behavior, as well as any outlying observations

2. Remove trend and seasonal components to get residuals

3. Choose a model to fit the residuals

4. Forecasting can be carried out by forecasting residual and then inverting the transformation carried out in Step 2.

CLASSICAL DECOMPOSITION IN R

Atmospheric concentrations of CO2 are expressed in parts per million (ppm) and reported in the preliminary 1997 SIO manometric mole fraction scale.