Introduction to Time Series Econometrics

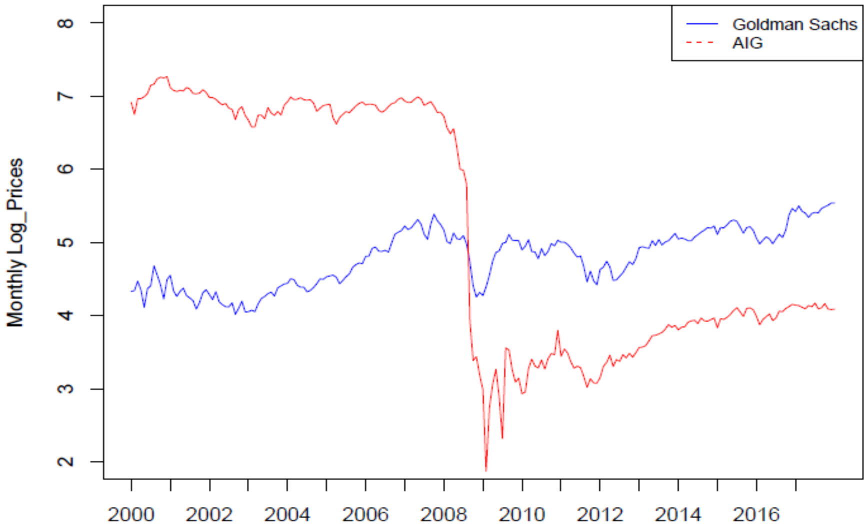
Econ 147

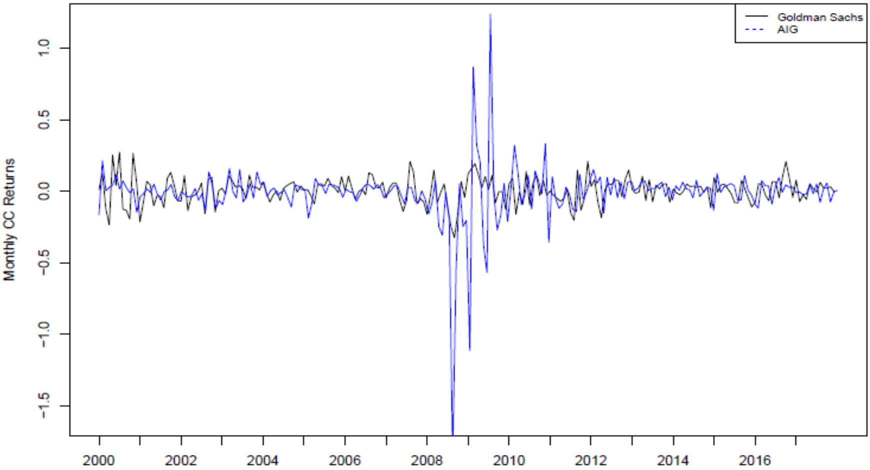
UCLA

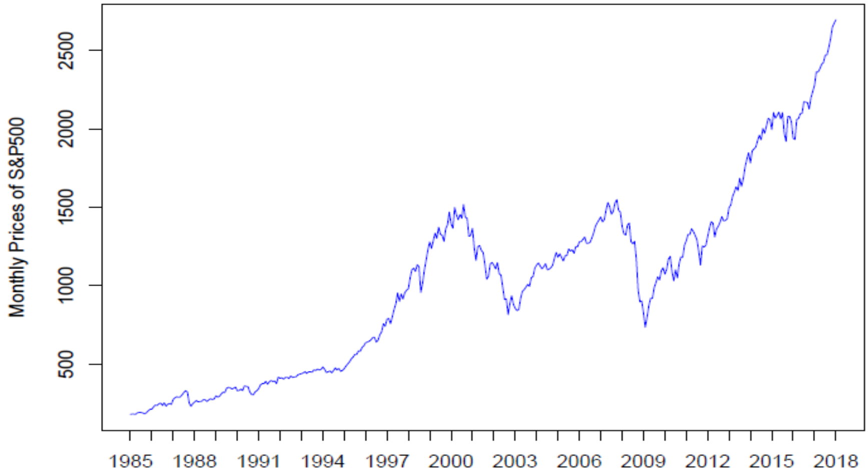
Version 1.1

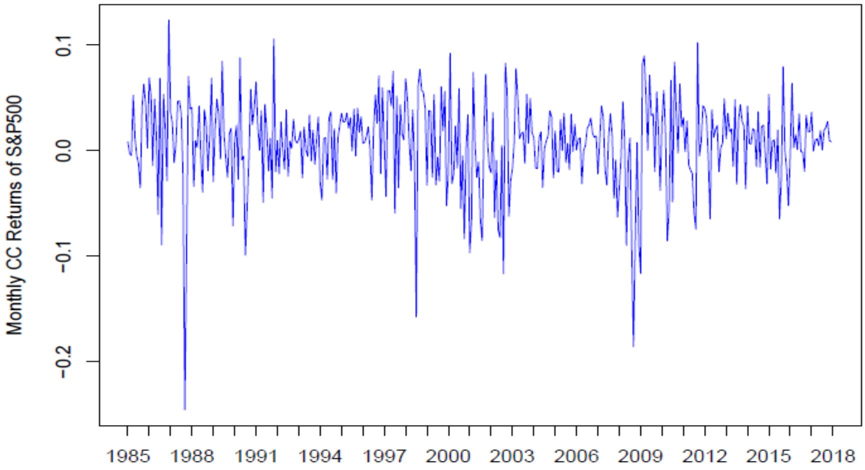
Reading

- Course slides
- Eric Zivot's book chapter on **Time Series Concepts**
- Eric Zivot's book chapter on **Descriptive Statistics**
- Optional: Chapter 2 (**Linear Time Series Models**) in Fan and Yao's book
- Optional: Chapter 4 (**Exploratory Data Analysis**), Chapter 9 (**Time Series Models: Basics**) and Chapter 10 (**Time Series Models: Further Topics**) in Ruppert's book


Goldman Sachs
AIG
Monthly Log_Prices
8
7
6
5
4
3
2
2000
2002
2004
2006
2008
2010
2012
2014
2016


Goldman Sachs
AIG
Monthly CC Returns
1.0
0.5
0.0
−0.5
−1.0
−1.5
2000
2002
2004
2006
2008
2010
2012
2014
2016


Monthly Prices of S&P500
500
1000
1500
2000
2500
1985
1988
1991
1994
1997
2000
2003
2006
2009
2012
2015
2018


Monthly CC Returns of S&P500
0.1
0.0
−0.1
−0.2
1985
1988
1991
1994
1997
2000
2003
2006
2009
2012
2015
2018

Stylized Patterns of Time Series Data

- Quickly mean-reverting (noisy) Vs Various degrees of persistence (trending or memory properties)
- time invariant Vs time-varying moments
- Transformation (e.g., 1st log-difference) do affect
- Linearity Vs Nonlinearity

Time Series Process

- Stochastic (Random) Process

$$\{\ldots, Y_1, Y_2, \ldots, Y_t, Y_{t+1}, \ldots\} = \{Y_t\}_{t=-\infty}^{\infty}$$

  - sequence of random variables indexed by time
- Observed time series of length T

$$\{Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T\} = \{y_t\}_{t=1}^{T}$$

Information Set

- $\mathcal{F}_t$ is called an *information set*, when it contains all the information or stochastic process up to time t.
  - e.g., if we are interested in X_t, Y_t,

$$\begin{aligned}\{\ldots, Y_1, Y_2, \ldots, Y_t\} &\subset \mathcal{F}_t \text{ and} \\ \{\ldots, X_1, X_2, \ldots, X_t\} &\subset \mathcal{F}_t.\end{aligned}$$

Conditional Expectation given Information Set

- We define the following conditional expectation of (continuous random variable) Y_t based on the information available up to time $s < t$,

$$E\left[Y_t|\mathcal{F}_s\right] = \int y f_{Y_t}(y|\mathcal{F}_s)dy,$$

where $f_{Y_t}(y|\mathcal{F}_s)$ denotes a conditional pdf of Y_t given $\mathcal{F}_s$.

  - if the information up to time s is equal to the information of

$$\{\ldots, X_1, X_2, \ldots, X_s\}$$

then $f_{Y_t}(y|\mathcal{F}_s) = f_{Y_t}(y|X_s, \ldots X_2, X_1, \ldots)$ and hence

$$E\left[Y_t|\mathcal{F}_s\right] = E\left[Y_t|X_s, \ldots X_2, X_1, \ldots\right]$$

Conditional Expectation given Information Set

- The following Law of Iterated Expectations (LIE) holds

$$E\left[E\left[Y_t|\mathcal{F}_s\right]\right] = E\left[Y_t\right].$$

- The usual properties also hold for all $s < t$:
  - $E\left[aY_t + bZ_t|\mathcal{F}_s\right] = aE\left[Y_t|\mathcal{F}_s\right] + bE\left[Z_t|\mathcal{F}_s\right]$ (Linearity)
  - $E\left[Y_t Z_s|\mathcal{F}_s\right] = Z_s E\left[Y_t|\mathcal{F}_s\right]$

Conditional Expectation and Forecasting

- We are often interested in "forecasting" based on information set $\mathcal{F}_{t-1}$
- **Claim:** $E\left[Y_t|\mathcal{F}_{t-1}\right]$ is the best prediction of Y_t given $\mathcal{F}_{t-1}$
- If $\mathcal{F}_{t-1}$ contains too many random variables, $E\left[Y_t|\mathcal{F}_{t-1}\right]$ will be difficult to analyze
- To model $E\left[Y_t|\mathcal{F}_{t-1}\right]$, we need determine
  - which (conditioning) variables should be included in $\mathcal{F}_{t-1}$
  - what is the function form given these variables

- **Claim:** $E\left[Y_t|\mathcal{F}_{t-1}\right]$ is the best prediction of Y_t given $\mathcal{F}_{t-1}$
- Consider any alternative predictor U of Y_t based on the information set $\mathcal{F}_{t-1}$
- U is not random given $\mathcal{F}_{t-1}$ since its information is in $\mathcal{F}_{t-1}$
- Consider the MSE of the prediction:

$$
\begin{aligned}
E[|U-Y_t|^2] &= E[|U-E\left[Y_t|\mathcal{F}_{t-1}\right]+E\left[Y_t|\mathcal{F}_{t-1}\right]-Y_t|^2] \\
&= E[|U-E\left[Y_t|\mathcal{F}_{t-1}\right]|^2]+E[|Y_t-E\left[Y_t|\mathcal{F}_{t-1}\right]|^2]
\end{aligned}
$$

where the second equality is by

$$
\begin{aligned}
& E\left[(U-E\left[Y_t|\mathcal{F}_{t-1}\right])(Y_t-E\left[Y_t|\mathcal{F}_{t-1}\right])|\mathcal{F}_{t-1}\right] \\
= & (U-E\left[Y_t|\mathcal{F}_{t-1}\right])E[Y_t-E\left[Y_t|\mathcal{F}_{t-1}\right]|\mathcal{F}_{t-1}] \\
= & (U-E\left[Y_t|\mathcal{F}_{t-1}\right])(E\left[Y_t|\mathcal{F}_{t-1}\right]-E\left[Y_t|\mathcal{F}_{t-1}\right])=0
\end{aligned}
$$

Conditional Expectation and Forecasting

- As an **illustration**, one may assume that

$$E\left[Y_t|\mathcal{F}_{t-1}\right] = \beta_0 + \beta_1 X_{t-1}$$

- Two assumptions are imposed in the above equation
  - $\mathcal{F}_{t-1}$ is identical to all the information about X_{t-1}, therefore

$$E\left[Y_t|\mathcal{F}_{t-1}\right] = E\left[Y_t|X_{t-1}\right]$$

  - $E\left[Y_t|X_{t-1}\right] = \int y f_{Y_t}(y|X_{t-1})dy = \beta_0 + \beta_1 X_{t-1}$
- There are more flexible ways to model $E\left[Y_t|\mathcal{F}_{t-1}\right]$

Stationary Processes

- $\{Y_t\}$ is stationary if some aspects of its behavior are unchanged by shifts in time
- A stochastic process $\{Y_t\}_{t=1}^{\infty}$ is *strictly stationary* if, for any set of subscripts $t_1, t_2, \ldots, t_r$ with any given finite integer r,the joint distribution of

$$(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_r})$$

is the same as the joint distribution of

$$(Y_{t_1+k}, Y_{t_2+k}, \ldots, Y_{t_r+k})$$

with any time shift k.

Stationary Processes

- For example, the distribution of (Y_1, Y_5) is the same as the distribution of (Y_{12}, Y_{16}).
- For a strictly stationary process, Y_t has the same mean, variance and higher moments for all t, *if they exist*.
  - strict stationarity however *does not* require the existence of *any* moment
  - iid $T(1)$ is strict stationary but no finite moment exist
  - some popular process in finance (e.g., IGARCH) could be strict stationary without a finite second moment
- Any function/transformation $g(\cdot)$ of a strictly stationary process, $\{g(Y_t)\}$ is also strictly stationary. E.g., if $\{Y_t\}$ is strictly stationary then $\{Y_t^2\}$ is strictly stationary.

Stationary Processes

- Strict stationarity is convenient, but sometimes hard to justify.
  - may be implausible for financial/macro time series data
- Joint distributions are also difficult to work with.
- It is therefore useful to introduce a simpler concept of stationarity which only characterizes the first two moments of the process.

Autocovariance and Autocorrelation

- Let $\{Y_t\}_{t=-\infty}^{\infty}$ be such that $E[|Y_t|^2] < \infty$ for all t
- $\mathrm{Cov}(Y_t, Y_{t-j}) = E\left[(Y_t - E(Y_t))(Y_{t-j} - E(Y_{t-j}))\right] := \gamma_{t,j}$ is called the j-lag *autocovariance* and measures the direction of linear time dependence
- $\mathrm{Corr}(Y_t, Y_{t-j}) := \rho_{t,j}$ is called the j-lag *autocorrelation* and measures the direction and strength of linear time dependence

$$\mathrm{Corr}(Y_t, Y_{t-j}) = \rho_{t,j} = \frac{\mathrm{Cov}(Y_t, Y_{t-j})}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_{t-j})}}$$

Covariance (Weak) Stationary Processes

- A stochastic process $\{Y_t\}_{t=1}^{\infty}$ is *covariance stationary* if,
  - $E[|Y_t|^2] < \infty$ for all t
  - $E[Y_t] = \mu$ and $\mathrm{Var}(Y_t) = \sigma^2$ for all t (time-invariant mean and variance)
  - $\mathrm{Cov}(Y_t, Y_{t-j}) = \gamma_j$ depends on j and not on t (time-invariant covariance)

Covariance Stationary Processes

- Both stationarity concepts are used in practice.
- A strictly stationary process is covariance stationary if $E[|Y_t|^2] < \infty$ for all t.
- However, if $E[|Y_t|^2]$ does not exist, we cannot use covariance stationarity concept.
- Strict stationarity can be defined without any finite moment condition.
  - Some financial time series may not have mean or variance.

Autocorrelation of Stationary Processes

- Let $\{Y_t\}_{t=-\infty}^{\infty}$ be covariance stationary process, then $\mathrm{Var}(Y_t) = \mathrm{Var}(Y_{t-j}) = \sigma^2$
- $\mathrm{Var}(Y_t) = \mathrm{Cov}(Y_t, Y_t) := \gamma_0 = \sigma^2$, so

$$\mathrm{Corr}(Y_t, Y_{t-j}) = \rho_j = \frac{\mathrm{Cov}(Y_t, Y_{t-j})}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_{t-j})}} = \frac{\gamma_j}{\sigma^2} = \frac{\gamma_j}{\gamma_0}$$

- Autocorrelation Function (ACF): ρ_j as a function of j
- The sample ACF: $\hat{\rho}_j$ as a function of j where

$$\widehat{\rho}_j = \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0} \text{ for } |j| < T,$$

$$\widehat{\gamma}_j = (T-j)^{-1} \sum_{t=1}^{T-j} (Y_t - \overline{Y})(Y_{t+j} - \overline{Y}), \ \overline{Y} = T^{-1} \sum_{t=1}^{T} Y_t$$

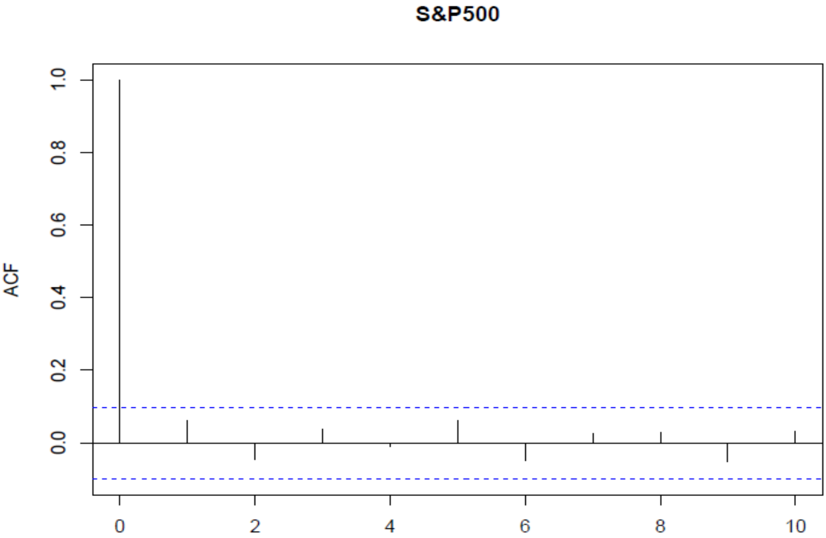
- Plot of $\hat{\rho}_j$ against j - commonly used "Descriptive Statistics"

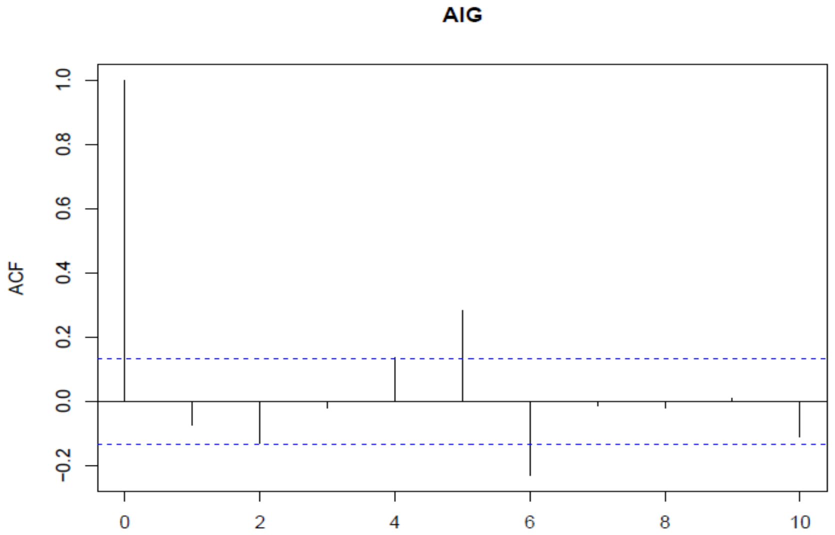
ACF in R

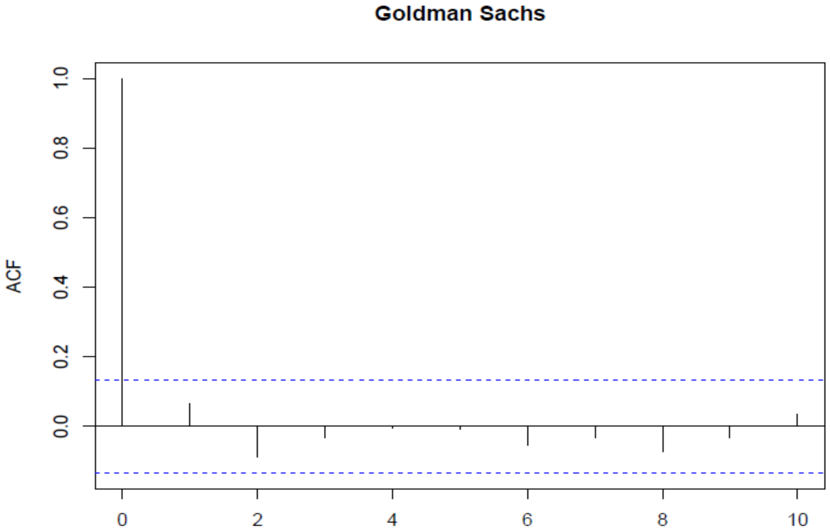
- The *R*-function to plot the sample ACF is

$$\mathrm{acf}(x,\ \mathrm{lag.max} = m)$$

where x is a data vector, m is the number of lags


S&P500
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10


AIG
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
2
4
6
8
10


Goldman Sachs
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10

Autocovariance and Autocorrelation

- Absence or presence of autocorrelation is sometimes indicated as the absence of presence of statistical dependence
  - Is this correct? No!
- In terms of strength of "linear dependence measure"
  - uncorrelated time series process: autocorrelation is zero (e.g., iid)
  - (linearly) dependent time series process: autocorrelation is not zero

Why stationarity is important

- If some distributional aspects are time invariant, e.g. $E(Y_t) = \mu$ then we can use all T observations to estimate $E(Y_t)$ using, say $\frac{1}{T}\sum_{t=1}^{T} Y_t$
- If not, for example, in unit root model below $E(X_t)$ changes over time.
  - then each observation x_t comes from a different distribution
  - we cannot use conventional statistical methods to do inference.

IID Process

- "IID" = "**I**ndependent and **I**dentically **D**istributed"
- $Y_t \sim \mathrm{IID}\ (0, \sigma^2)$ is strictly AND covariance stationary, Why?
- $\mathrm{Cov}(Y_t, Y_s) = 0$ for any $t \neq s$, so is uncorrelated.

MDS Process

- To introduce *martingale difference sequence,* we first need a concept of *Martingale.*
- A time series process $\{Y_t\}$ is called a *martingale process* (MG henceforth) if

$$E\left[Y_t|\mathcal{F}_{t-1}\right] = Y_{t-1} \text{ for all } t,$$

i.e., the current value is the best predictor of the next value of this process.

  - MG hypothesis has been traditionally assumed for stock price index (or its transformation, such as log)

MDS Process

- A time series process $\{d_t\}$ is called a *martingale difference sequence* (MDS henceforth) if it is a difference of martingale processes, i.e.,

$$d_t = Y_t - Y_{t-1} \text{ for all } t,$$

where Y_t is MG.

- Equivalently, d_t is MDS if

$$E\left[d_t | \mathcal{F}_{t-1}\right] = 0 \text{ for all } t.$$

MDS Process

- The above definition for MDS is a very general one, so we can even allow for nonstationary process as long as it satisfies $E\left[d_t|\mathcal{F}_{t-1}\right]=0$.
- For example, we can allow a non-stationarity in the second moment such as $d_t \sim mds\left(0, \sigma^2(t)\right)$ where $\sigma^2(t)$ is a time-varying function.
- Many researchers now believe time-varying (conditional or unconditional) volatility in financial market, so the MDS structure is very useful in time series/finance applications.

MDS Process

- If we want to restrict our attention to "covariance stationary process", the following subclass among MDS can be employed:
- A time series process $\{d_t\}$ is called a MDS with constant variance if

$$E\left[d_t|\mathcal{F}_{t-1}\right] = 0 \text{ and}$$
$$Var(d_t) = E\left(d_t^2\right) = \sigma^2 \text{ for all } t.$$

- We use a natural expression for this process: $d_t \sim MDS\left(0, \sigma^2\right)$ and this is another example of *uncorrelated* covariance stationary process.
- See Examples in lecture note.

White Noise Processes

- White Noise Process is uncorrelated by definition:

$$Y_t \sim WN(0, \sigma^2)$$
$$E[Y_t] = 0, \ \mathrm{Var}(Y_t) = \sigma^2$$
$$\mathrm{Cov}(Y_t, Y_s) = 0 \text{ for } t \neq s$$

- Here, $Y_t \sim WN(0, \sigma^2)$ represents an uncorrelated process with mean zero and variance σ^2.
- Similarly to MDS, WN can be very general if we do not impose constant conditional variance. For example, Y_t and Y_s can exhibit non-linear dependence (e.g. Y_t^2 can be correlated with Y_s^2) but still can be WN.

Relations

- IID is stronger than MDS

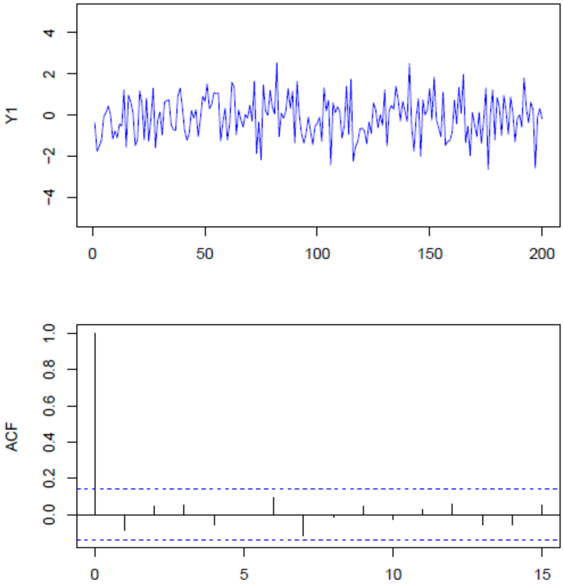
$$Y_t \sim \mathrm{IID}(0, \sigma^2) \Rightarrow Y_t \sim \mathrm{MDS}(0, \sigma^2)$$

- MDS (with constant variance) is stronger than White Noise (with constant variance)

$$Y_t \sim \mathrm{MDS}(0, \sigma^2) \Rightarrow Y_t \sim \mathrm{WN}(0, \sigma^2)$$

- These three processes are commonly used for time series data with no (or negligible) evidence of correlation

$Y_{1,t} \sim \text{IID}(0, 1)$



Testing White Noise Assumption

- H_0^*: $Y_t \sim WN(0, \sigma^2)$
- One important implication of H_0 is

$$\begin{aligned}\rho_j &= \frac{\mathrm{Cov}(Y_t, Y_{t-j})}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_{t-j})}} \\ &= \frac{\mathrm{Cov}(Y_t, Y_{t-j})}{\mathrm{Var}(Y_t)} = 0 \text{ for any } j \neq 0\end{aligned}$$

- We can test the following null hypothesis implied by H_0^*

$$H_0 : \sum_{j=1}^{m} \rho_j^2 = 0$$

where $m \geq 1$ is a prescribed integer.

- Under H_0 any consistent estimator of $\sum_{j=1}^{m} \rho_j^2$ should be close to zero
- One popular statistic for testing H_0 is the Ljung-Box statistic

Testing White Noise Assumption

- $H_0 : \sum_{j=1}^{m} \rho_j^2 = 0$
- Given data $\{Y_t\}_{t=1}^{T}$, we can estimate ρ_j as

$$\widehat{\rho}_j = \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0} \text{ for } |j| < T$$

where

$$\widehat{\gamma}_j = (T-j)^{-1} \sum_{t=1}^{T-j} (Y_t - \overline{Y})(Y_{t+j} - \overline{Y}),\ \overline{Y} = \sum_{t=1}^{T} Y_t$$

- The Ljung-Box statistic is

$$Q_m = (T+2) \sum_{j=1}^{m} (1 - j/T) \widehat{\rho}_j^2$$

- We reject H_0 for large values of Q_m

Testing White Noise Assumption

- $H_0 : \sum_{j=1}^{m} \rho_j^2 = 0$
- The Ljung-Box statistic:

$$Q_m = (T+2)\sum_{j=1}^{m}(1 - j/T)\widehat{\rho}_j^2$$

- The critical value of the Ljung-Box test depends on the theoretical distribution of Q_m under H_0
- Q_m is approximated a Chi-square random variable with degree of freedom m if $Y_t \sim IID(0, \sigma^2)$
- One popular choice is $\chi^2_{1-\alpha}(m)$: the $(1-\alpha)$-quantile of the Chi-square distribution with degree of freedom m

Ljung-Box Test in R

- The *R*-function to perform the Ljung-Box test is

$$\text{Box.test}(x, \text{ lag} = m, \text{ type} = \text{"Ljung"})$$

  where x is a data vector and m is the number of lags prescribed in the Ljung-Box statistic

Ljung-Box Test of GS, AIG and S&P500

Ljung-Box Test on Monthly CC Returns

		$m = 2$	$m = 5$	$m = 10$	$m = 20$
GS	L-B	3.2939	3.4261	5.6956	17.693
	C.V.	5.9914	11.071	18.307	31.410
AIG	L-B	2.3800	36.563	55.296	69.829
	C.V.	5.9914	11.071	18.307	31.410
S&P500	L-B	2.2937	4.4088	7.4415	11.754
	C.V.	5.9914	11.071	18.307	31.410

Autoregressive Processes

- Idea: create a stochastic process that exhibits multi-period geometrically decaying linear time dependence
- AR(1) Model (mean-adjusted form)

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t, \quad -1 < \phi < 1$$
$$\varepsilon_t \sim \text{iid } N(0, \sigma_\varepsilon^2)$$

  - AR(1) model is covariance stationary if $|\phi| < 1$ (see the supplemenatry note for derivation)
  - $\varepsilon_t \sim \text{iid } N(0, \sigma_\varepsilon^2)$ could be replaced by $\varepsilon_t \sim \text{mds } N(0, \sigma_\varepsilon^2)$ or $\varepsilon_t \sim \text{WN } N(0, \sigma_\varepsilon^2)$
  - normal distribution assumption is not necessary

Autoregressive Processes

- **Properties**

$$E[Y_t] = \mu$$
$$\mathrm{var}(Y_t) = \sigma_Y^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2}$$
$$\gamma_j = \mathrm{cov}(Y_t, Y_{t-j}) = \phi\gamma_{j-1} = \phi^j\sigma_Y^2$$
$$\rho_j = \mathrm{corr}(Y_t, Y_{t-j}) = \frac{\gamma_j}{\gamma_0} = \frac{\phi^j\sigma_Y^2}{\sigma_Y^2} = \phi^j$$

Note: Since $|\phi| < 1$

$$\lim_{j\to\infty}\rho_j = \lim_{j\to\infty}\phi^j = 0.$$

Autoregressive Processes

- **AR(1) Model (regression model form)**

$$\begin{aligned} Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t \Rightarrow \\ Y_t &= \mu - \phi\mu + \phi Y_{t-1} + \varepsilon_t \\ &= c + \phi Y_{t-1} + \varepsilon_t \end{aligned}$$

where

$$c = (1-\phi)\mu \Rightarrow \mu = \frac{c}{1-\phi}$$

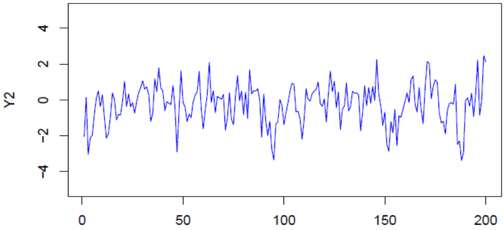
Remarks:

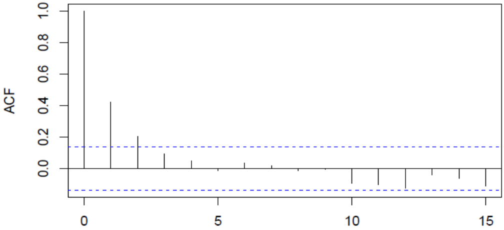
- Regression model form is convenient for estimation by linear regression.

Autoregressive Processes

- The AR(1) model is a good description for the following time series
- Interest rates on U.S. Treasury securities, dividend yields, unemployment
- Growth rate of macroeconomic variables
  - Real GDP, industrial production, productivity
  - Money, velocity, consumer prices
  - Real and nominal wages

$$Y_{2,t} = 0.5Y_{2,t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim \text{IID}(0,1)$$





Moving Average Processes

- Idea: Create a stochastic process that only exhibits one period linear time dependence
- MA(1) Model

$$\begin{aligned} Y_t &= \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \quad -\infty < \theta < \infty \\ \varepsilon_t &\sim iid\ N(0, \sigma_\varepsilon^2) \end{aligned}$$

- Properties

$$\begin{aligned} E[Y_t] &= \mu + E[\varepsilon_t] + \theta E[\varepsilon_{t-1}] \\ &= \mu + 0 + 0 = \mu \end{aligned}$$

- The normal assumption is not necessary. The iid assumption can be replaced by MDS or WN.

Moving Average Processes

- Properties

$$
\begin{aligned}
\mathrm{var}(Y_t) = \sigma^2 = E[(Y_t - \mu)^2] \\
&= E[(\varepsilon_t + \theta\varepsilon_{t-1})^2] \\
&= E[\varepsilon_t^2] + 2\theta E[\varepsilon_t\varepsilon_{t-1}] + \theta^2 E[\varepsilon_{t-1}^2] \\
&= \sigma_\varepsilon^2 + 0 + \theta^2\sigma_\varepsilon^2 = \sigma_\varepsilon^2(1+\theta^2)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{cov}(Y_t, Y_{t-1}) = \gamma_1 &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] \\
&= E[\varepsilon_t\varepsilon_{t-1}] + \theta E[\varepsilon_t\varepsilon_{t-2}] \\
&+ \theta E[\varepsilon_{t-1}^2] + \theta^2 E[\varepsilon_{t-1}\varepsilon_{t-2}] \\
&= 0 + 0 + \theta\sigma_\varepsilon^2 + 0 = \theta\sigma_\varepsilon^2
\end{aligned}
$$

Moving Average Processes

- Properties

$$\begin{aligned}
\operatorname{cov}(Y_t, Y_{t-2}) = \gamma_2 = E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-2} + \theta\varepsilon_{t-3})] \\
= E[\varepsilon_t\varepsilon_{t-2}] + \theta E[\varepsilon_t\varepsilon_{t-3}] \\
+ \theta E[\varepsilon_{t-1}\varepsilon_{t-2}] + \theta^2 E[\varepsilon_{t-1}\varepsilon_{t-3}] \\
= 0 + 0 + 0 + 0 = 0
\end{aligned}$$

Similar calculation show that

$$\operatorname{cov}(Y_t, Y_{t-j}) = \gamma_j = 0 \text{ for } j > 1.$$

Moving Average Processes

- MA(1) Model

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \quad -\infty < \theta < \infty$$
$$\varepsilon_t \sim iid\ N(0, \sigma_\varepsilon^2)$$

- **Properties**

$$E[Y_t] = \mu$$
$$\mathrm{var}(Y_t) = \sigma_Y^2 = \sigma_\varepsilon^2(1+\theta^2)$$
$$\gamma_j = \mathrm{cov}(Y_t, Y_{t-j}) = \begin{cases} \theta\sigma_\varepsilon^2, & |j| = 1 \\ 0, & |j| > 1 \end{cases}$$
$$\rho_j = \mathrm{corr}(Y_t, Y_{t-j}) = \begin{cases} \frac{\theta}{1+\theta^2}, & |j| = 1 \\ 0, & |j| > 1 \end{cases}$$

- Result: MA(1) is covariance stationary for any value of θ

Moving Average Processes

- Example: MA(1) model for overlapping returns

Let r_t denote the 1−month cc return and assume that

$$r_t \sim \text{iid } N(\mu_r, \sigma_r^2)$$

Consider creating a time series of 2−month cc returns using

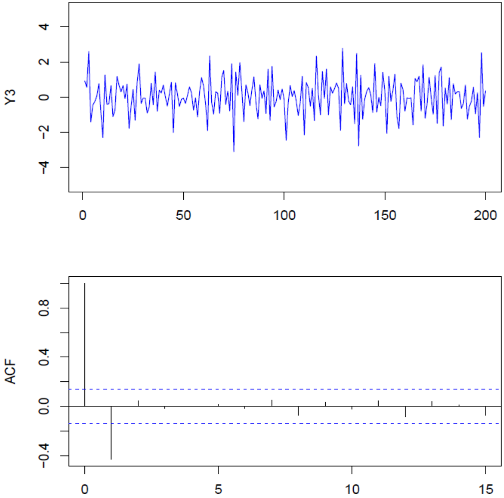
$$r_t(2) = r_t + r_{t-1}$$

These 2−month returns observed monthly overlap by 1 month

$$\begin{aligned} r_t(2) &= r_t + r_{t-1} \\ r_{t-1}(2) &= r_{t-1} + r_{t-2} \\ r_{t-2}(2) &= r_{t-2} + r_{t-3} \\ &\vdots \end{aligned}$$

Claim: The stochastic process $\{r_t(2)\}$ follows a MA(1) process

$$Y_{3,t} = \varepsilon_t - 0.5\varepsilon_{t-1} \text{ where } \varepsilon_t \sim \text{IID}(0,1)$$



ARMA Processes

- ARMA (1,1) model

$$
\begin{aligned}
Y_t - \mu &= \phi(Y_{t-1} - \mu) + \varepsilon_t + \theta\varepsilon_{t-1}, \;\; |\phi| < 1 \\
\varepsilon_t &\sim MDS\ (0, \sigma_\varepsilon^2)
\end{aligned}
$$

- If $\theta = 0$, then ARMA(1,1) becomes AR(1)
- If $\phi = 0$, then ARMA(1,1) becomes MA(1)
- Therefore, ARMA(1,1) is a more general time series model

- ARMA (1,1) model

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_t \sim MDS\ (0, \sigma_\varepsilon^2),\ \ |\phi| < 1$$

- Properties

$$E[Y_t] = \mu$$

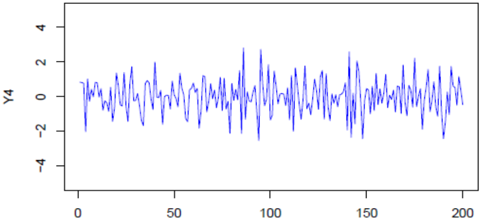
$$\mathrm{var}(Y_t) = \sigma_Y^2 = \frac{1 + \theta^2 + 2\phi\theta}{1 - \phi^2}\sigma_\varepsilon^2$$

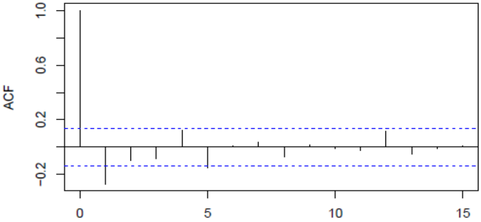
$$\gamma_j = \mathrm{cov}(Y_t, Y_{t-j}) = \begin{cases} \frac{(1+\phi\theta)(\phi+\theta)\sigma_\varepsilon^2}{1-\phi^2}, & |j| = 1 \\ \phi\gamma_{j-1}, & j \geq 2 \\ \gamma_{-j}, & j \leq -2 \end{cases}$$

$$\rho_j = \mathrm{corr}(Y_t, Y_{t-j}) = \begin{cases} \frac{(1+\phi\theta)(\phi+\theta)}{1+\theta^2+2\phi\theta}, & |j| = 1 \\ \phi\rho_{j-1}, & j \geq 2 \\ \rho_{-j}, & j \leq -2 \end{cases}$$

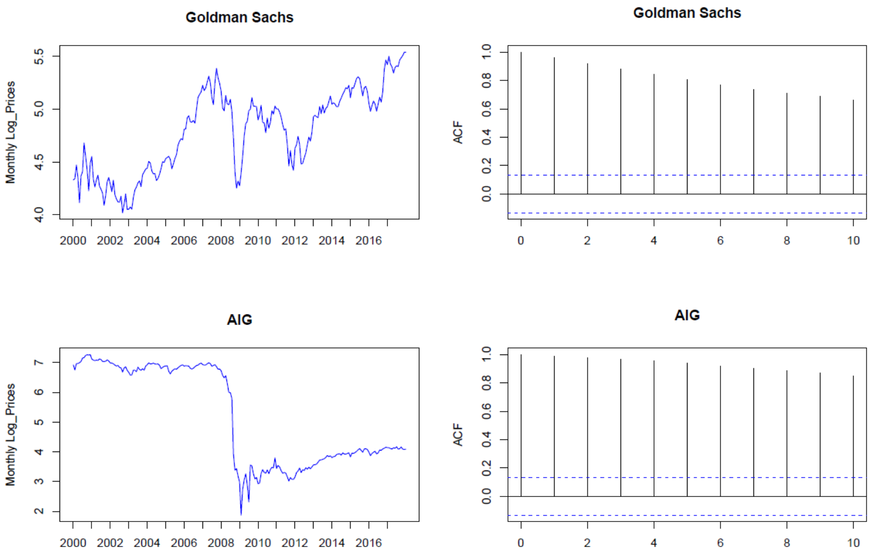
- Proof of the above results are in the supplemental notes

$$Y_{4,t} = 0.5Y_{4,t-1} + \varepsilon_t - 0.5\varepsilon_{t-1} \text{ where } \varepsilon_t \sim \mathrm{IID}(0,1)$$

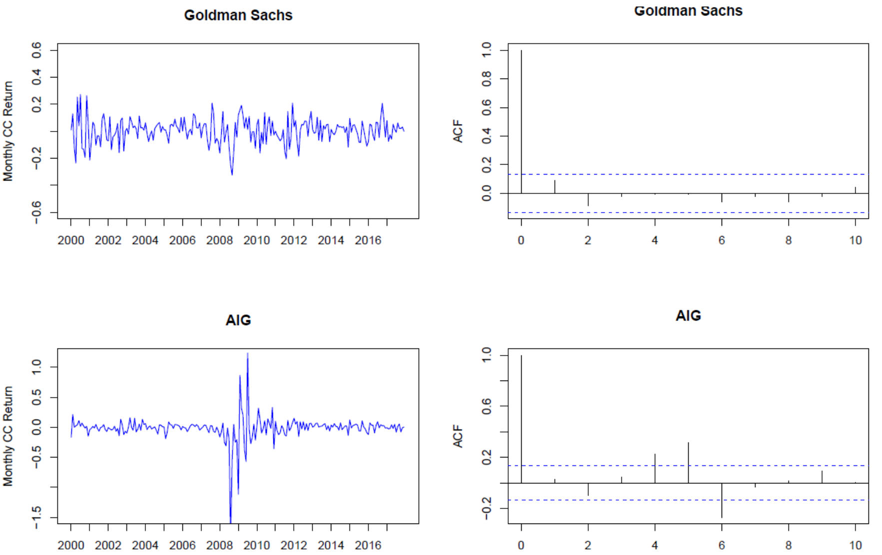




Monthly Log_Prices of Goldman Sachs and AIG (2010.1 – 2017.11)



Monthly CC Returns of Goldman Sachs and AIG (2010.1 – 2017.11)



Nonstationary Processes

- $\{Y_t\}$ is nonstationary if it is not stationary
  - nonstationary in mean (1st moment): $E(Y_t) = \mu(t) = \mu_t$
    - time-varying or structural break
  - nonstationary in variance (2nd moment): $\mathrm{var}(Y_t) = \sigma^2(t) = \sigma_t^2$
- "level" of many economic time series, such as stock price, GDP, CPI, consumption, etc. are nonstationary.
  - often take the first difference.
  - cc return is the first difference of log price, so may be (mean) stationary.

Nonstationary Processes

- Deterministically trending process

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t,\ \ \varepsilon_t \sim MDS\ (0, \sigma_\varepsilon^2)$$
$$E[Y_t] = \beta_0 + \beta_1 t \ \text{ depends on } t$$

- A detrending transformation may yield a stationary process:

$$X_t = Y_t - \beta_0 - \beta_1 t = \varepsilon_t$$

Nonstationary Processes

- Unit Root (Random Walk) Model

$$\begin{aligned} Y_t &= Y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim mds(0, \sigma_\varepsilon^2), \\ &= Y_0 + \sum_{j=1}^{t} \varepsilon_j, \ \text{if assume } Y_0 = 0 \\ &\Rightarrow \mathrm{var}(Y_t) = \sigma_\varepsilon^2 \times t \ \ \text{depends on } t \end{aligned}$$

- A differencing transformation may yield a stationary process

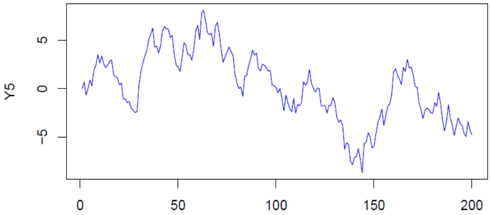
$$\Delta Y_t = Y_t - Y_{t-1} = \varepsilon_t$$

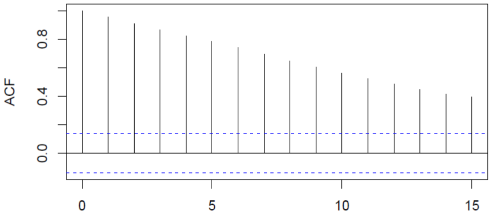
- Prototypical model for log of stock price and cc return

Nonstationary and Stationary AR Processes

- Unit Root (Random Walk) Model: AR(1) with $\phi = 1$
  - "a news or shock" impact does not decay at all
- Stationary AR(1) Model: AR(1) with $|\phi| < 1$
  - "a news or shock" impact decay exponentially
- Testing $H_0 : \phi = 1$ Versus $H_0 : \phi < 1$ is VERY IMPORTANT
  - "Unit Root Test": well developed

$$Y_{5,t} = Y_{5,t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim \text{IID}(0,1)$$





Observed Sample:

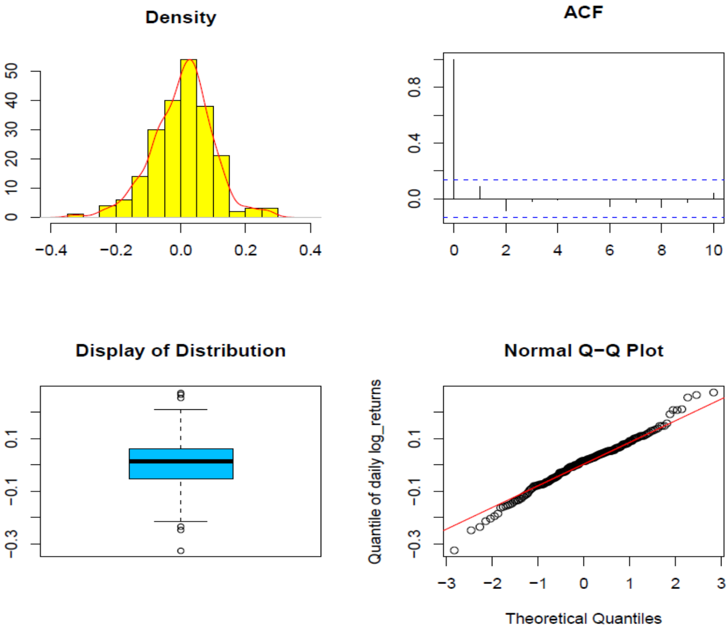
$$\{X_1 = x_1, \ldots, X_T = x_T\} = \{x_t\}_{t=1}^{T}$$

are observations generated by the stochastic process

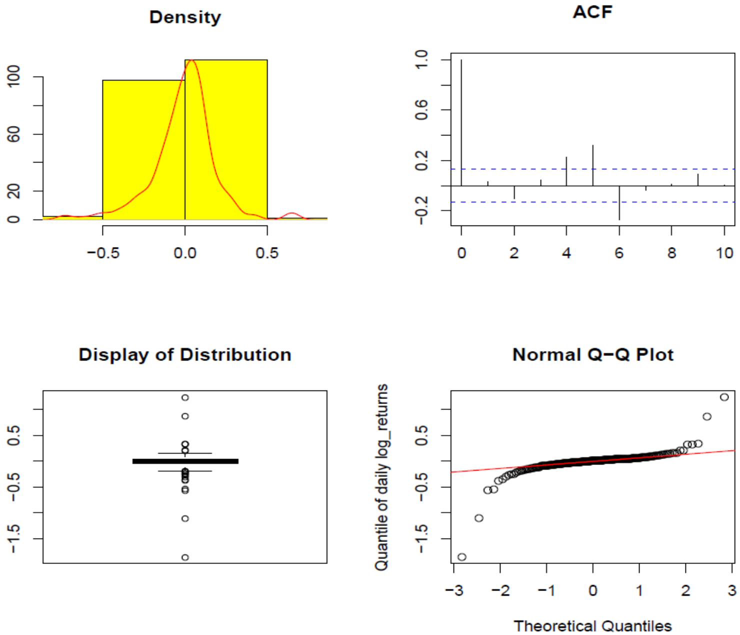
Descriptive Statistics

Data summaries (statistics) to describe certain features of the data, to learn about the unknown pdf, $f(x)$, and to capture the observed dependencies in the data

Monthly CC Returns of Goldman Sachs (2000.1 – 2017.11)



Monthly CC Returns of AIG (2000.1 – 2017.11)



Histograms

Goal: Describe the shape of the distribution of the data $\{x_t\}_{t=1}^{T}$

Histogram Construction:

1. Order data from smallest to largest values
2. Divide range into N equally spaced bins

$$[-|-|-|\cdots|-|-|-]$$

3. Count number of observations in each bin
4. Create bar chart (optionally normalize area to equal 1)

R Functions

Function	Description
`hist()`	compute histogram
`density()`	compute smoothed histogram

Note: The `density()` function computes a smoothed (kernel density) estimate of the unknown pdf at the point x using the formula

$$\hat{f}(x) = \frac{1}{Tb}\sum_{t=1}^{T} k\left(\frac{x - x_t}{b}\right)$$

$$k(\cdot) = \text{kernel function}$$

$$b = \text{ bandwidth (smoothing) parameter}$$

where $k(\cdot)$ is a pdf symmetric about zero (typically the standard normal distribution).

Empirical Quantiles/Percentiles

Percentiles:

For $\alpha \in [0,1]$, the $100 \times \alpha^{th}$ percentile (empirical quantile) of a sample of data is the data value $\hat{q}_\alpha$ such that $\alpha \cdot 100\%$ of the data are less than $\hat{q}_\alpha$.

Quartiles

$$\begin{aligned} \hat{q}_{.25} &= \text{first quartile} \\ \hat{q}_{.50} &= \text{second quartile (median)} \\ \hat{q}_{.75} &= \text{third quartile} \\ \hat{q}_{.75} - \hat{q}_{.25} &= \text{interquartile range (IQR)} \end{aligned}$$

R functions

Function	Description
`sort()`	sort elements of data vector
`min()`	compute minimum value of data vector
`max()`	compute maximum value of data vector
`range()`	compute min and max of a data vector
`quantile()`	compute empirical quantiles
`median()`	compute median
`summary()`	compute summary statistics

Sample Statistics

Plug-In Principle: Estimate population quantities using sample statistics

Sample Average (Mean)

$$\frac{1}{T}\sum_{t=1}^{T} x_t = \bar{x} = \hat{\mu}_x$$

Sample Variance

$$\frac{1}{T-1}\sum_{t=1}^{T} (x_t - \bar{x})^2 = s_x^2 = \hat{\sigma}_x^2$$

Sample Standard Deviation

$$\sqrt{s_x^2} = s_x = \hat{\sigma}_x$$

Sample Skewness

$$\frac{1}{T-1}\sum_{t=1}^{T}(x_t-\bar{x})^3/s_x^3=\widehat{skew}$$

Sample Kurtosis

$$\frac{1}{T-1}\sum_{t=1}^{T}(x_t-\bar{x})^4/s_x^4=\widehat{kurt}$$

Sample Excess Kurtosis

$$\widehat{kurt}-3$$

R Functions

Function	Package	Description (compute)
`mean()`	base	sample mean
`colMeans()`	base	column means of matrix
`var()`	stats	sample variance
`sd()`	stats	sample standard deviation
`skewness()`	PerformanceAnalytics	sample skewness
`kurtosis()`	PerformanceAnalytics	sample excess kurtosis

Note: Use the R function `apply()`, to apply functions over rows or columns of a matrix or data.frame

Empirical Cumulative Distribution Function

- Recall, the CDF of a random variable X is

$$F_X(x) = \Pr(X \leq x)$$

- The empirical CDF of a random sample is

$$\begin{aligned}\hat{F}_X(x) &= \frac{1}{n}(\#x_i \leq x) \\ &= \frac{\text{number of } x_i \text{ values } \leq x}{\text{sample size}}\end{aligned}$$

- Use the R function `ecdf()`

Comparing Empirical CDF to Normal Distribution

Question: Does observed data come from a normal distribution?

- Standardize data to have zero mean and variance one

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

- Sort standardized data from smallest to largest values: $\{z_{(1)}, \ldots, z_{(n)}\}$
- Compute standard normal CDF at each sorted value: $\Phi(z_{(i)})$
- Plot $\hat{F}_Z(z_{(i)})$ and $\Phi(z_{(i)})$ against sorted data

Quantile-Quantile (QQ) Plots

A QQ plot is useful for comparing your data with the quantiles of a distribution (usually the normal distribution) that you think is appropriate for your data. You interpret the QQ plot in the following way:

- If the points fall close to a straight line, your conjectured distribution is appropriate
- If the points do not fall close to a straight line, your conjectured distribution is not appropriate and you should consider a different distribution

R functions

Function	Package	Description
`qqnorm()`	stats	QQ-plot against normal distribution
`qqline()`	stats	draw straight line on QQ-plot
`qqPlot()`	car	QQ-plot against specified distribution

Outliers

- Extremely large or small values are called "outliers"
- Outliers can greatly influence the values of common descriptive statistics. In particular, the sample mean, variance, standard deviation, skewness and kurtosis
- Percentile measures are more robust to outliers: outliers do not greatly influence these measures (e.g. median instead of mean; IQR instead of SD)

IQR (interquartile range) - outlier robust measure of spread

$$IQR = q_{.75} - q_{.25}$$

Moderate Outlier

$$\hat{q}_{.75} + 1.5 \cdot IQR < x < \hat{q}_{.75} + 3 \cdot IQR$$
$$\hat{q}_{.25} - 3 \cdot IQR < x < \hat{q}_{.25} - 1.5 \cdot IQR$$

Extreme Outlier

$$x > \hat{q}_{.75} + 3 \cdot IQR$$
$$x < \hat{q}_{.25} - 3 \cdot IQR$$

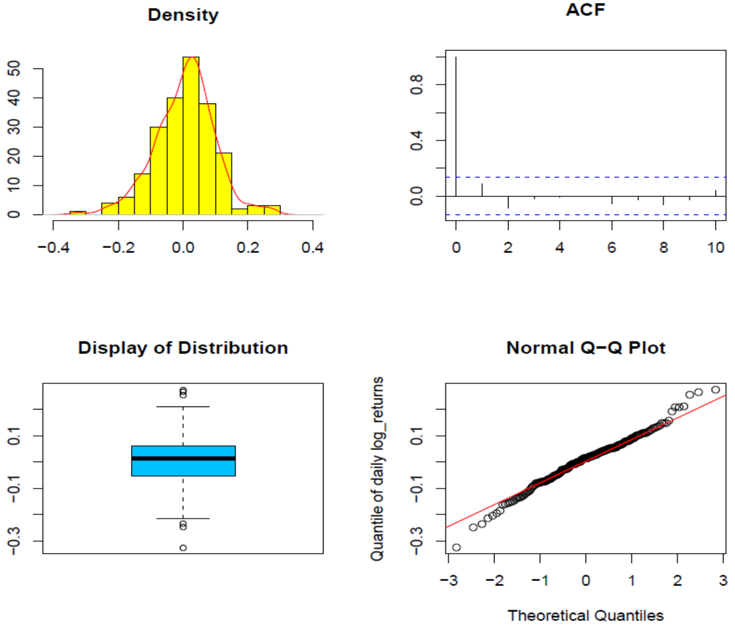
Boxplots

A box plot displays the locations of the basic features of the distribution of one-dimensional data—the median, the upper and lower quartiles, outer fences that indicate the extent of your data beyond the quartiles, and outliers, if any.

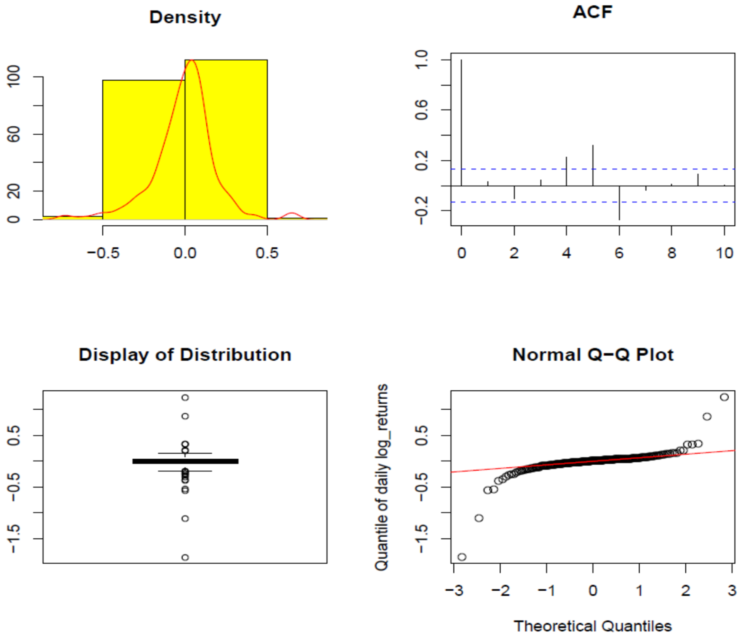
R functions

Function	Package	Description
`boxplot()`	graphics	box plot multiple series
`chart.Boxplot()`	Performance Analytics	box plot asset returns

Monthly CC Returns of Goldman Sachs (2000.1 – 2017.11)



Monthly CC Returns of AIG (2000.1 – 2017.11)



Bivariate Descriptive Statistics

$$\{\ldots, (X_1, Y_1), (X_2, Y_2), \ldots (X_T, Y_T), \ldots\} = \{(X_t, Y_t)\}$$

covariance stationary bivariate stochastic process with realized values

$$\{(x_1, y_1), (x_2, y_2), \ldots (x_T, y_T)\} = \{(x_t, y_t)\}_{t=1}^{T}$$

Scatterplot

XY plot of bivariate data

R functions: `plot()`, `pairs()`

Sample Covariance

$$\frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y}) = s_{xy} = \hat{\sigma}_{xy}$$

Sample Correlation

$$\frac{s_{xy}}{s_x s_y} = r_{xy} = \hat{\rho}_{xy}$$

R functions

Function	Package	Description
`var()`	stats	compute sample variance-covariance matrix
`cov()`	stats	compute sample variance-covariance matrix
`cor()`	stats	compute sample correlation matrix

Time Series Descriptive Statistics

Sample Autocovariance

$$\hat{\gamma}_j = \frac{1}{T-j} \sum_{t=j+1}^{T} (x_t - \bar{x})(x_{t-j} - \bar{x}), \; j = 1, 2, \ldots$$

Sample Autocorrelation

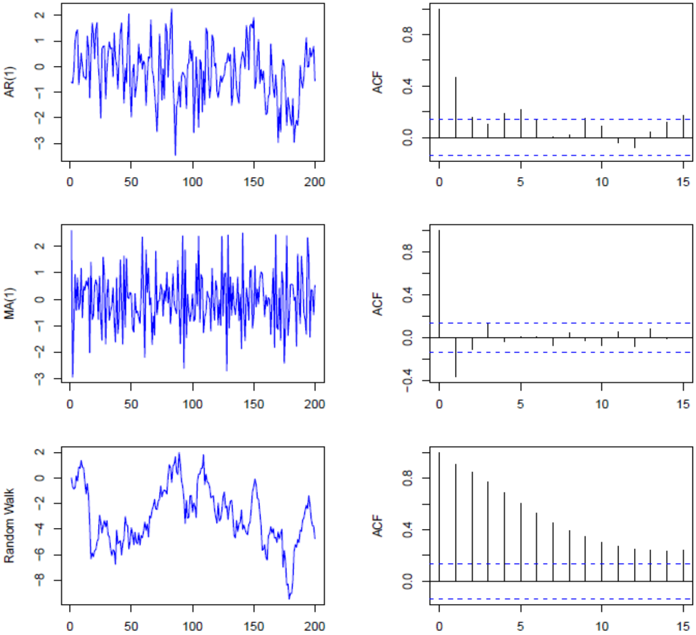
$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\sigma}^2}, \; j = 1, 2, \ldots$$

Sample Autocorrelation Function (SACF)

Plot $\hat{\rho}_j$ against j

R functions

Function	Package	Description (plot)
`acf()`	stats	sample autocorrelations
`chart.ACF()`	PerformanceAnalytics	sample autocorrelations
`chart.ACFplus()`	PerformanceAnalytics	sample autocorrelations


AR(1)
ACF
MA(1)
ACF
Random Walk
ACF

- Probability and statistics
- Essential time series concepts and models
  - stationarity
  - uncorrelated process in time series data
  - dependent processes (AR and MA models)
- Useful descriptive statistics for financial time series:
  - acf, qqplot, boxplot and histogram (density estimator)
- Explore some finance models - next topics