Financial Econometrics

Junye Li

Department of Finance
ESSEC Business School

What is Financial Econometrics

- Financial econometrics is about the statistical tools that are needed to analyze and address the specific types of questions that come up in finance.
- Nowhere in economics is the availability of data better than in financial markets.
  - Data recording enforced by regulations.
  - Computerized markets facilitate recording.

- The starting point of any financial model is uncertainty. Without uncertainty, finance would simply be applied microeconomics.
- Finance is all about risk and return tradeoffs.
  - Which risks are worth taking?
  - How much should I be compensated for taking a given risk?
- Risk due to uncertainty is therefore of central importance in finance, both in the models we build and in the object of econometric interest.

General Set-Up

- Let p_{it} denote the value of asset *i* at time *t*.
- The simple asset return can be calculated by

$$r_{it} = (p_{it} - p_{it-1})/p_{it-1}$$

where we assume zero dividend. If the dividend is non-zero,

$$r_{it} = (p_{it} + d_{it} - p_{it-1})/p_{it-1}$$

- However, continuously compounded return is computed by

$$r_{it} = log(p_{it} / p_{it-1})$$

- Denote the vector of returns on *k* assets as

$$R_t = [r_{1t}, r_{2t}, \ldots, r_{kt}]'$$

- At the time the investment decision is made, we don’t know what the outcomes will be in the future for the vector of returns.
- Nevertheless, the investment decision can be reached without knowing the outcomes but with knowledge about how likely the different outcomes are --- the joint distribution.
- Let F_t denote an information set available at time t:
  - For example, it can be all past returns: $R_{t-1}, R_{t-2}, R_{t-3}, \ldots$
  - More generally, it can be any available information that might be useful in predicting returns

- A central object of interest is the joint distribution of returns in the *t+1* given information F_t.
- If *f(.)* is the joint density function, then this conditional distribution is denoted by $f(R_{t+1}|F_t)$ and is generally the object of interest
- Basic modeling question is how to link F_t to the distribution of future returns R_{t+1}.

Examples

- Estimates of the condition mean return

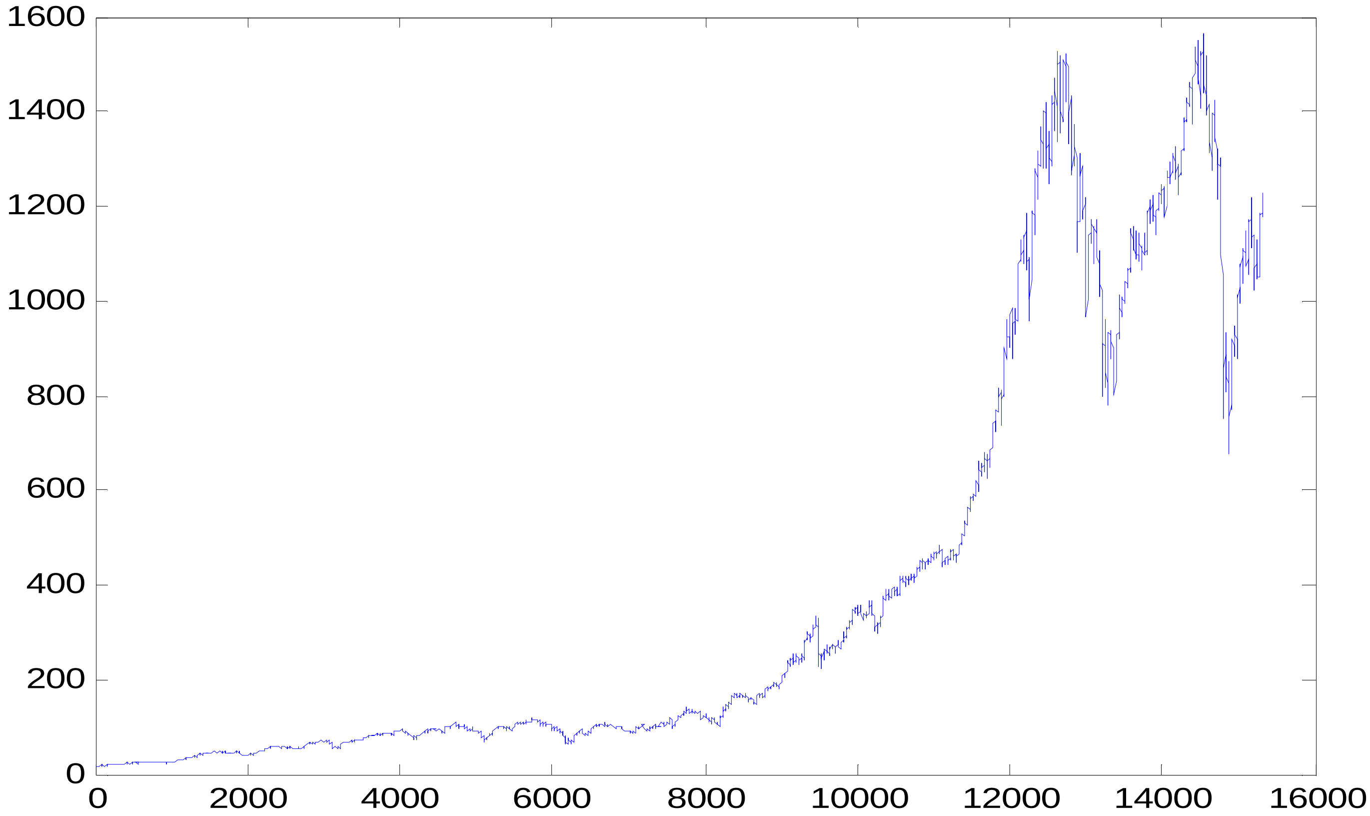
$$u_{t+1} = E[R_{t+1} \mid F_t]$$

are used to test market efficiency and build optimal portfolios.

- In Black-Scholes model, option pricing depends on volatility of the underlying asset, which can be estimated by

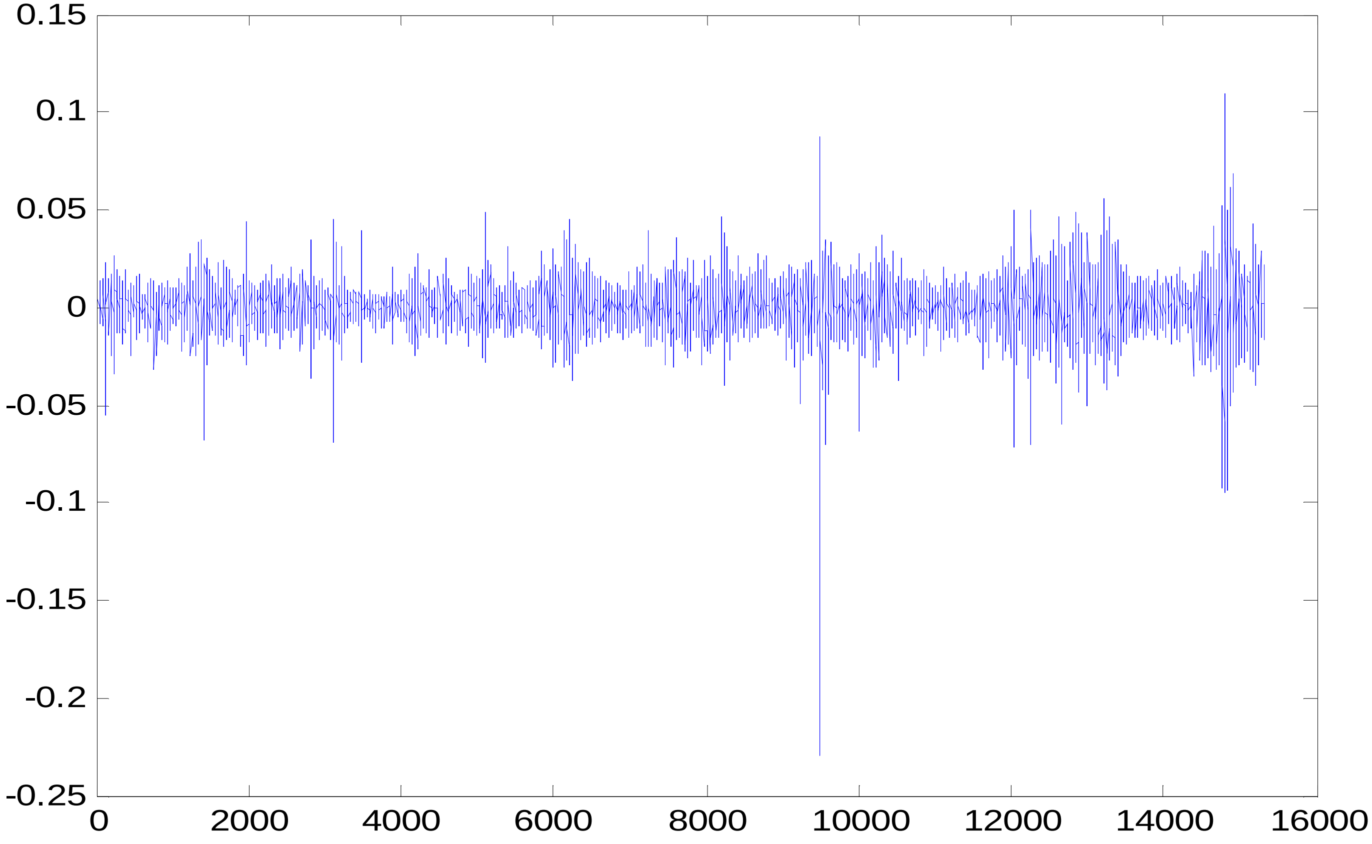
$$\sigma_{t+1} = \mathrm{Var}(r_{t+1} \mid F_t).$$

S&P 500 Index



1600
1400
1200
1000
800
600
400
200
0
0
2000
4000
6000
8000
10000
12000
14000
16000

S&P 500 Index



0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
0
2000
4000
6000
8000
10000
12000
14000
16000

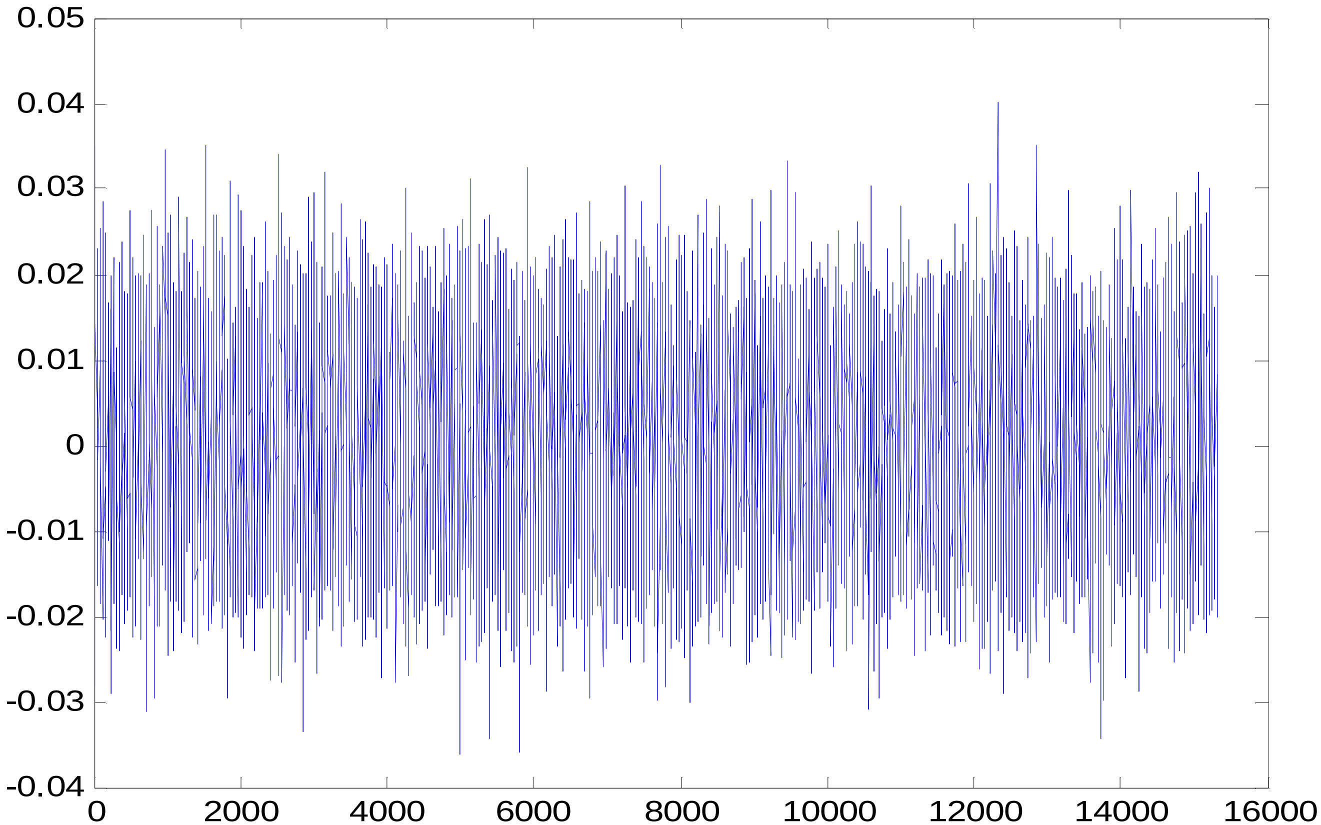
Summary Statistics

- **Mean / Std. Dev. / Skewness / Kurtosis**

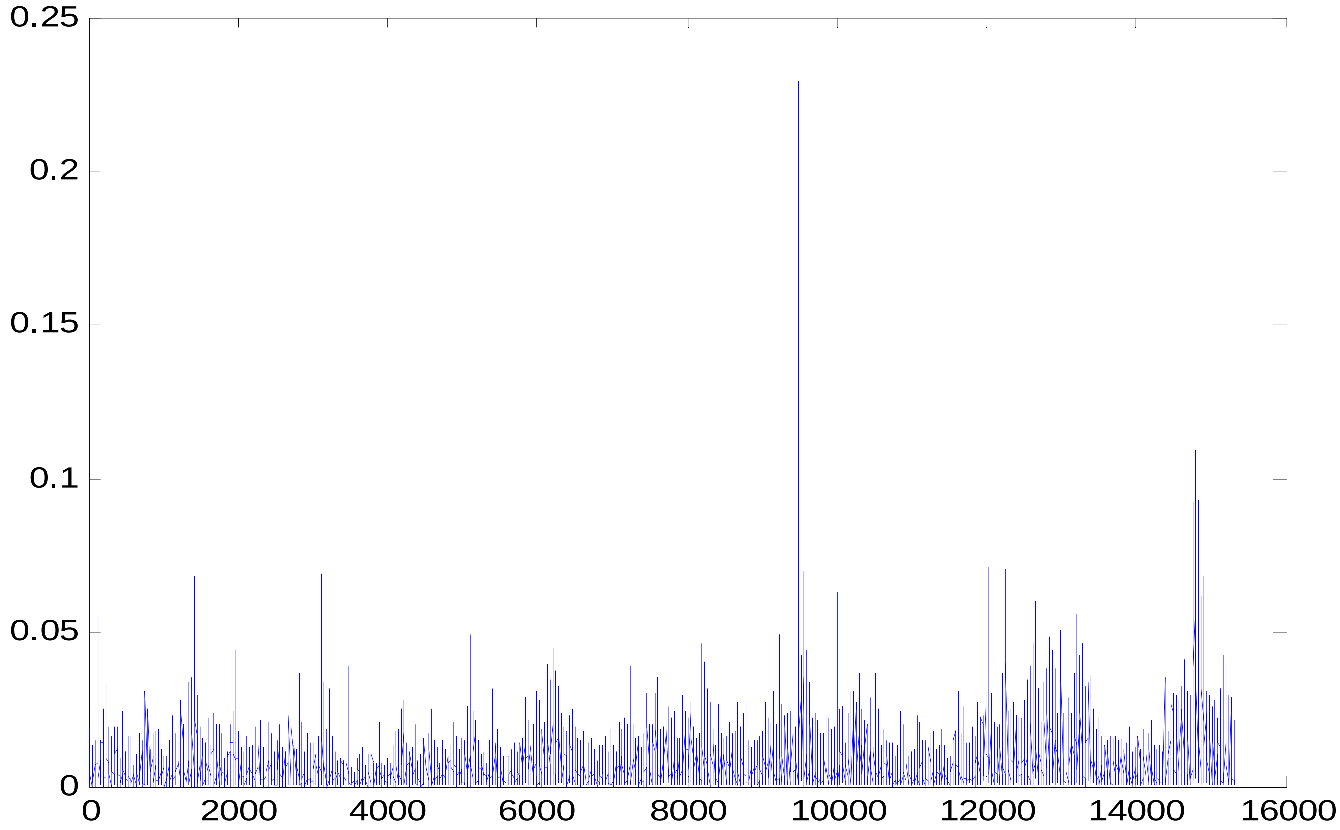
Mean	Std. Dev.	Skewness	Kurtosis
0.07	**0.15**	**-1.06**	**32.11**

- Skewness is different from zero;
- Kurtosis is far away from 3;
- Both indicate that S&P 500 returns are not normally distributed.
- What do they look like if they are normal with mean 0.07 and standard deviation 0.15.

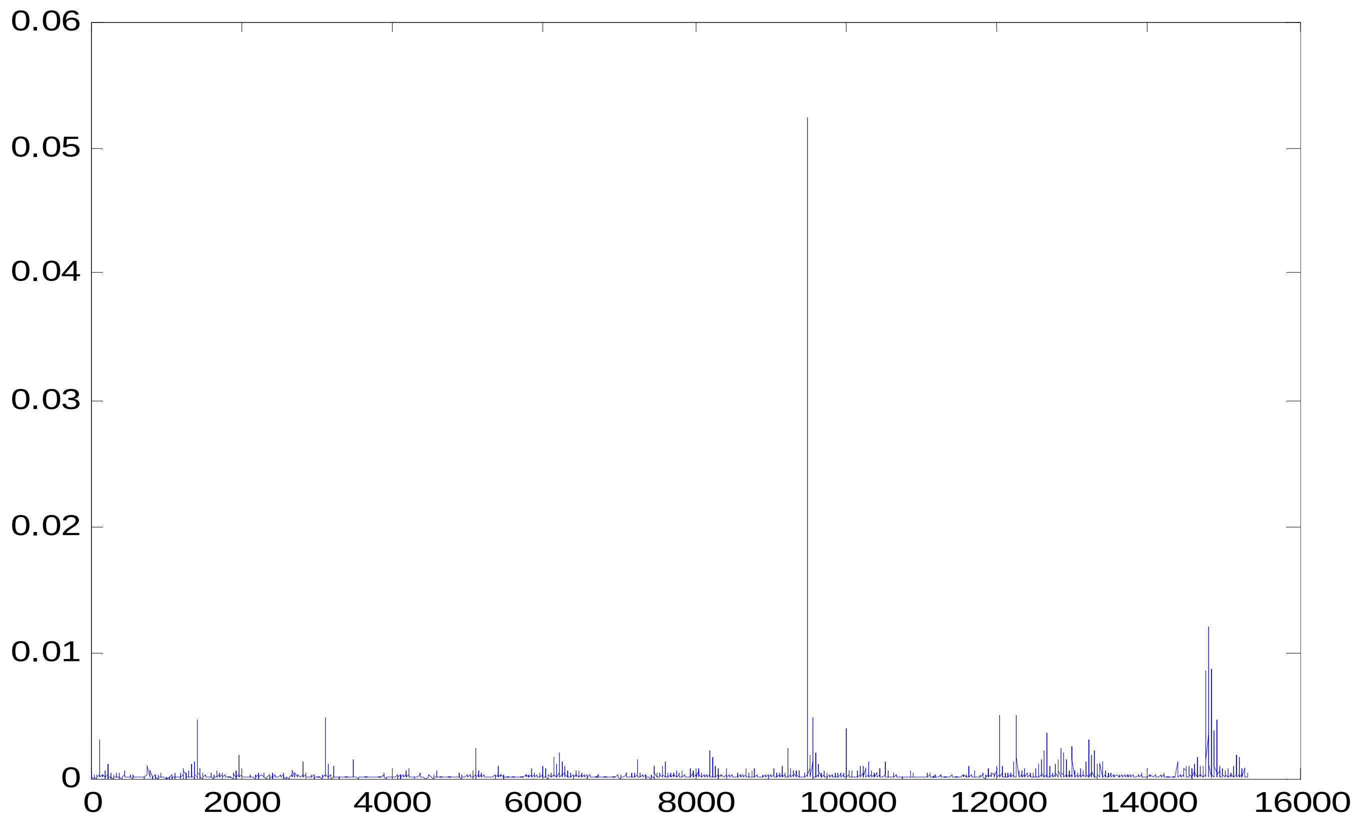
Simulated Returns



Absolute Returns



Squared Returns



Stylized Facts of Returns

- Stock returns are not normally distributed
  - Left-skewed (at least for index returns)
  - Fat-tailed
- Time-varying conditional mean
- Time-varying conditional standard deviation (volatility)

Road-Map

- We begin with dynamic time series models for the conditional mean:
  - Focus on univariate time series
  - ARMA models
- We then move to model the time-varying volatility:
  - ARCH/GARCH models (2003 Nobel Prize in Economics)
  - Extensions

Univariate Time Series Analysis

Junye Li

Finance Department
ESSEC Business School

Introduction to Time Series and Autoregressive Models

- Time-Series Data and Dependence
- Checking for Dependence
- The Autocorrelation Function
- The AR(1) Model
- The AR(p) Model
- The Partial Autocorrelation Function
- Estimate AR(p) Model

Time-Series Data

- Time-series data are simply a collection of observations gathered over time.
- For now, consider only a single variable y.
- To emphasize that the data are time-series, we index observations by *t*.
- The data: $y_1, y_2, \ldots, y_t, \ldots, y_T.$ The interval between observations can be daily, weekly, monthly and yearly.
- High-frequency data are even in second.

Stationary Time Series

- The foundation of time series analysis is stationarity.
- A time series is called *<u>Strong Stationary</u>* if

  $$f(y_t, y_{t-1}, \ldots, y_{t-j}) = f(y_{t+s}, y_{t+s-1}, \ldots, y_{t+s-j})$$

  for all *t* and *s*.
- A time series is called *<u>Weak Stationary</u>* or *<u>Covariance Stationary</u>* if

  $E[y_t] = \mu$ (constant)

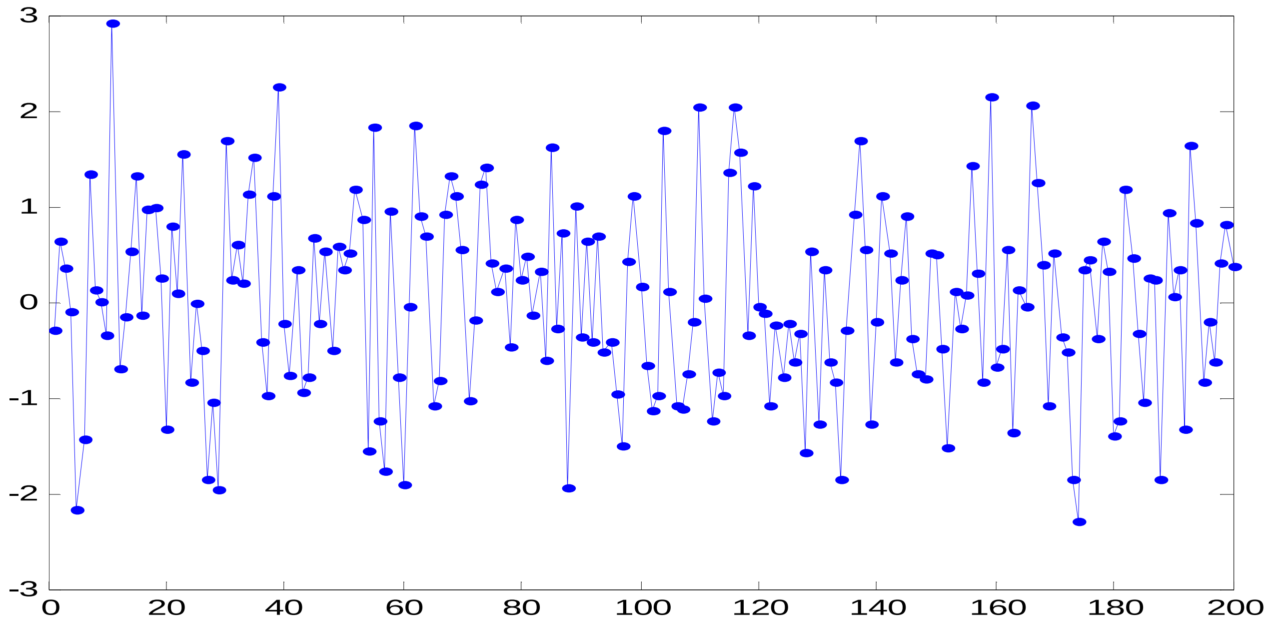
  $E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j$ (constant)

Weak Stationarity

- The weak stationarity implies that the time plot of the data would show that the T values fluctuate with constant variation around a fixed level.
- In applications, weak stationarity enables one to make inference concerning future observations, e.g. predictions.
- If y_t is strictly stationary and its first two moments are finite, it is also weakly stationary. But the converse is not true in general.

- An Example: i.i.d data:

  y_t = *random number generated from N(0, 1)*



White Noise

- The basic building block in time series is a sequence ε_1 , ε_2 , ε_3 , ..., whose elements have zero mean and constant variance

$$E[\varepsilon_t] = 0, \ Var[\varepsilon_t] = \sigma^2$$

$$\text{and } E[\varepsilon_t \ \varepsilon_s] = 0$$

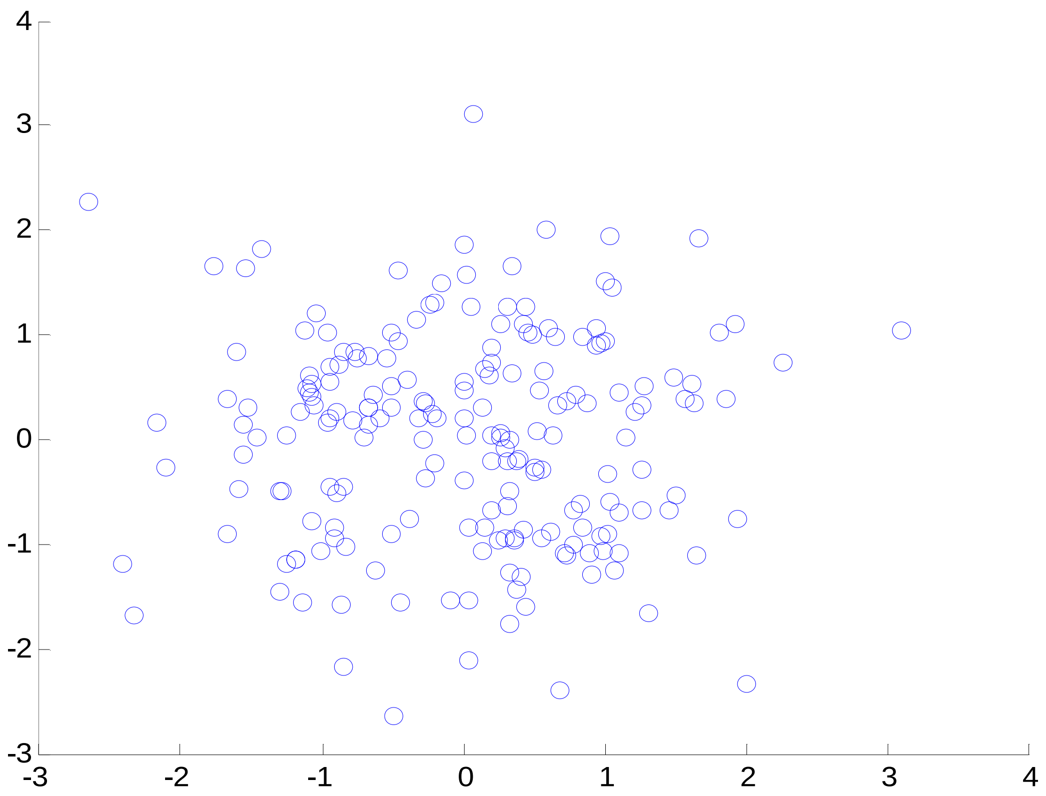
- This is called the white noise process. If we assume a normal distribution to ε_t , it is called Gaussian white noise.

- Suppose that we know the returns from time *1* to time *T* ($y_1, y_2, \ldots, y_t, \ldots, y_T$).
- What is your prediction for the next time (*T+1*) return, y_{T+1}?
- If our observations are i.i.d, the prediction would simply the unconditional mean: the average of all past observations.
- However, in practice, we find that using information about y_T can give us a more precise prediction:
  - Most of financial data are not i.i.d

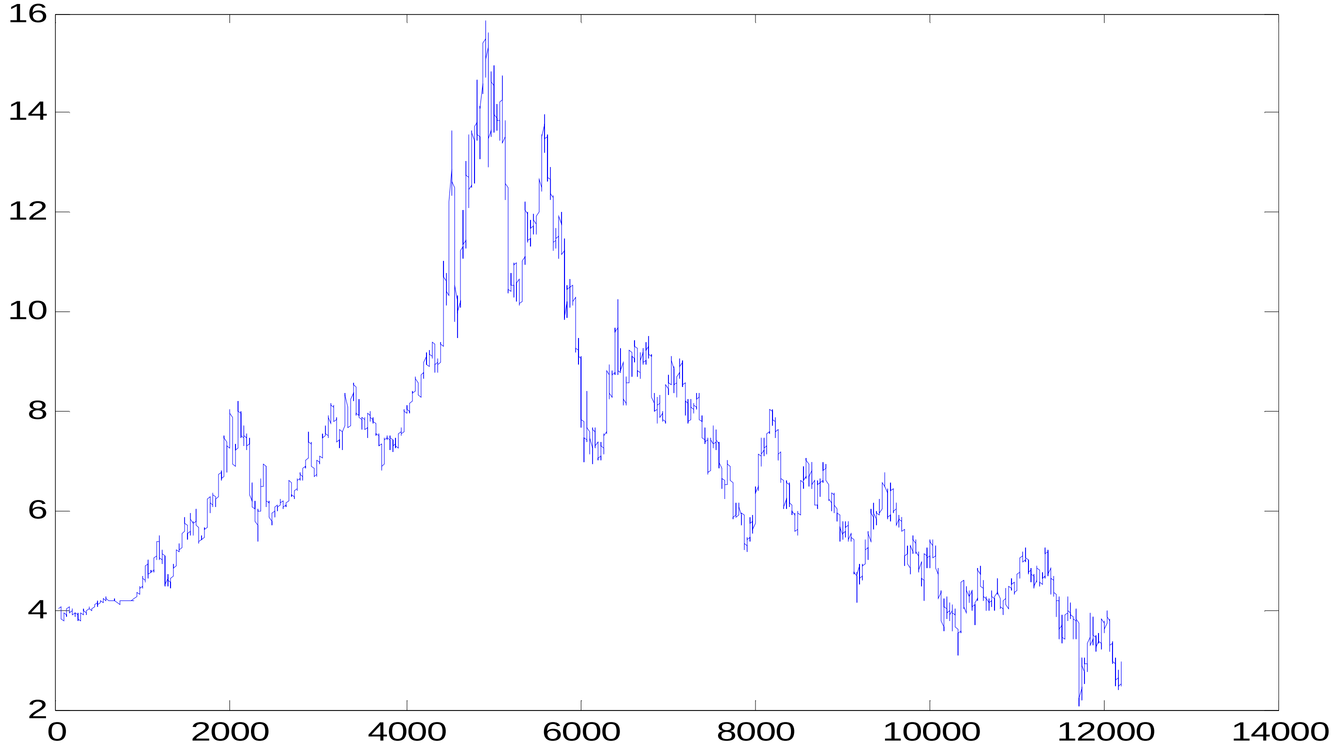
Checking for Dependence

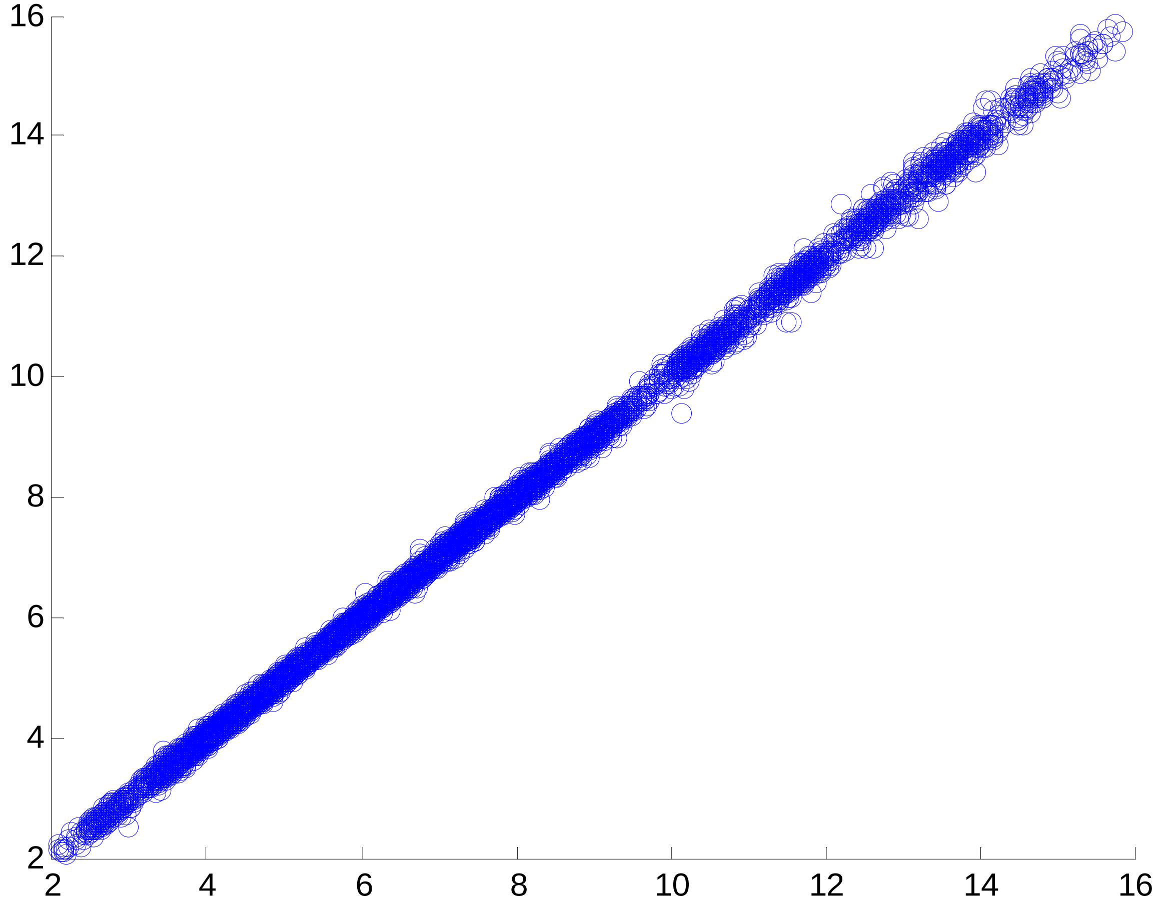
- We can have a visual assessment from the time-series plot whether the observations are independent or not.
- Independence of observations means that knowing previous values does not help us to predict the next time value.
- To see whether y_{t-1} *helps predict* y_t , we can use the scatter plot

- Take the previous generated i.i.d data as an example



- Real Data: 10-year Bond Yield: clearly not i.i.d




16
14
12
10
8
6
4
2
2
4
6
8
10
12
14
16

The Autocorrelation Function

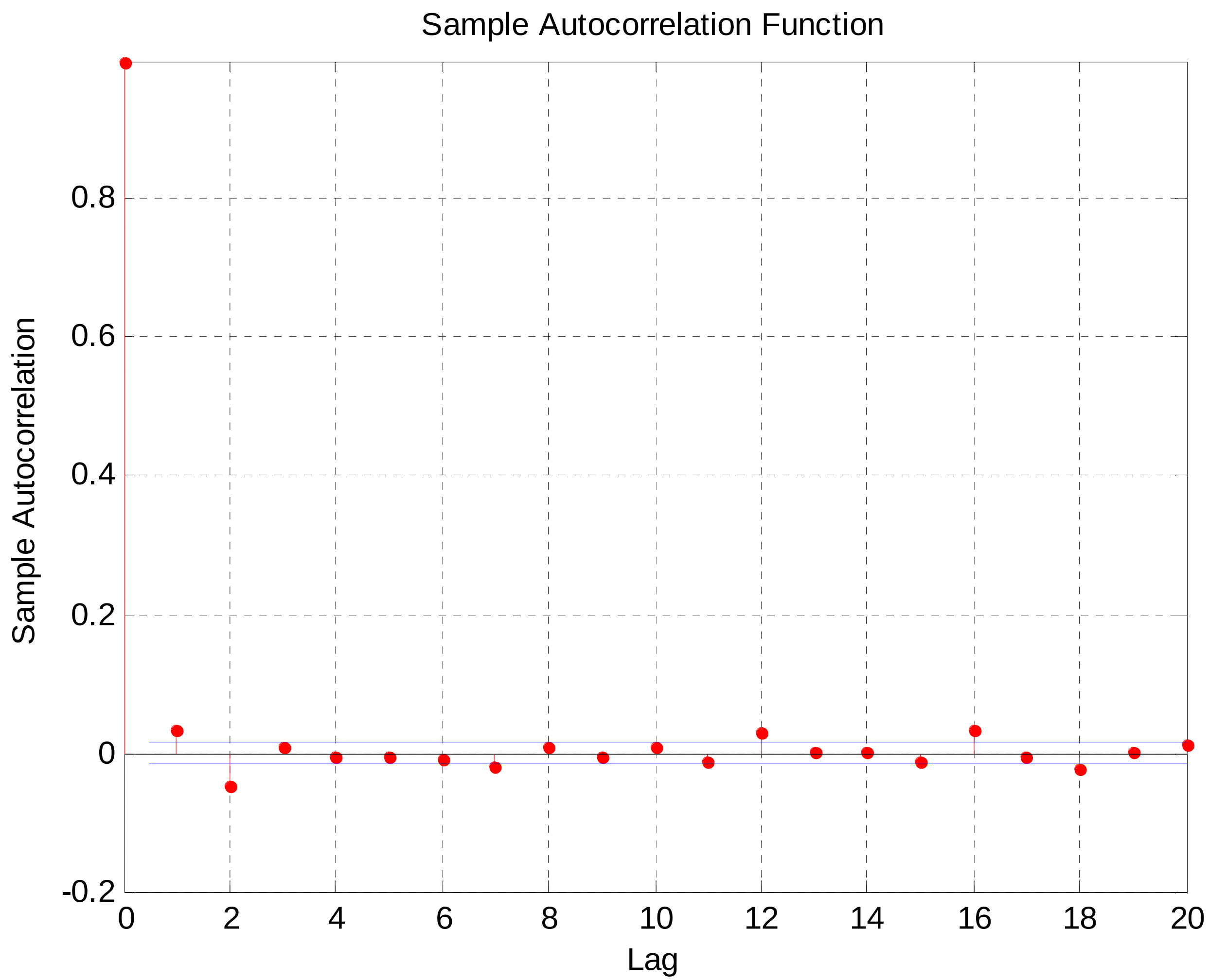
- The correlations between *y* and the lagged values of *y* are called autocorrelations.
- The autocorrelation function (*ACF*) is simply all of the autocorrelation values for all possible lags *L.*
- Consider a weakly stationary series y_t

$$\rho_l = \frac{Cov(y_t, y_{t-l})}{\sqrt{Var(y_t)Var(y_{t-l})}} = \frac{Cov(y_t, y_{t-l})}{Var(y_t)}$$

- The property $Var(y_t) = Var(y_{t-l})$ for a weakly stationary series is used;
- From the above definition, we have $\rho_0 = 1$, and $\rho_l = \rho_{-l}$;
- For a given sample, the ACF can be estimated by

$$\hat{\rho}_l = \frac{\sum_{t=2}^{T} (y_t - \bar{y})(y_{t-l} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

- Under some general conditions, this estimate is consistent.


Sample Autocorrelation Function
Sample Autocorrelation
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20
Lag

- There seems to exist dependence of *y* on past values.
- The ACF gets smaller as the lag gets larger:
  - y_t is more strongly related to y_{t-1} than y_{t-2}
- How to evaluate significance of an autocorrelation:

$$|autocorrelation| > 2/sqrt(T)$$

- If $T = 100$, the cutoff is 0.20.

- We can test individual autocorrelation for significance, but which one should we take a look at?
- We may be interested in jointly testing whether the first k autocorrelations are all significant or not!!!
- The null hypothesis:

  H_0: $\rho_1 = \rho_2 = \ldots = \rho_k = 0$

  – Q-tests: Box-Pierce test and Ljung-Box test

Q-Tests

- Box-Pierce Test:

$$Q = T \sum \rho^2{}_j$$

  - It has a Chi-square distribution with degree of freedom *k*

- Ljung-Box Test:

$$Q = T(T+2) \sum (\rho^2{}_j / T-j)$$

  - It has better finite sample properties

A Financial Application

- In finance literature, a version of the CAPM is that the return of an asset is not predictable.
- This indicates that there should not exist autocorrelations.
- Testing for zero autocorrelations has been used as a tool to check the efficient market assumption.

The AR(1) Model

- If there is dependence in returns, we would like to have a model that can allow us to predict future returns from the past outcomes.
- Since there is information about y_t contained in the lagged values y_{t-1}, y_{t-2}, ..., the obvious way to try is a regression of y_t on its lags.
- The simplest model is the so-called AR(1) or the autoregressive model of order 1.

- The model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$$

where ε_t = *i.i.d* $N(0, \sigma^2)$

  - y_t consists of two parts: the part that depends on the past observation $\beta_0 + \beta_1 y_{t-1}$
  - and the part this is not predictable from the past ε_t
  - The model is in the same form as the well-known linear regression models
  - Interest rate and/or stochastic volatility models

- Conditional mean and variance

$$E[y_t \mid y_{t-1}] = \beta_0 + \beta_1 y_{t-1} \ ;$$

$$Var[y_t \mid y_{t-1}] = Var[\varepsilon_t\,] = \sigma^2;$$

- The AR(1) model says that y_t depends on the past only through y_{t-1}
- What does it mean:
  - knowing the past values y_{t-2}, y_{t-3}, ... does not help predict y_t if we already know the value of y_{t-1}.
  - In probability

$$f(y_t \mid y_{t-1}, y_{t-2}, \ldots) = f(y_t \mid y_{t-1})$$

$$= N(\beta_0 + \beta_1 y_{t-1} \,, \sigma^2)$$

- Unconditional mean of AR(1)

$$E[y_t] = \beta_0 + \beta_1 E[y_{t-1}] + E[\varepsilon_t]$$

$$\Rightarrow \mu = \beta_0 + \beta_1 \mu + 0$$

$$\Rightarrow \mu = \beta_0 / (1-\beta_1)$$

  - The mean exists if β_1 is different from one;
  - The mean is zeros if and only if β_0 is zero;
- Using the above result, AR(1) model can be rewritten as

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$$

$$\Rightarrow y_t = \mu(1-\beta_1) + \beta_1 y_{t-1} + \varepsilon_t$$

$$\Rightarrow y_t - \mu = \beta_1 (y_{t-1} - \mu) + \varepsilon_t$$

- The unconditional variance

$$y_t - \mu = \beta_1(y_{t-1} - \mu) + \varepsilon_t$$

$$\Rightarrow E(y_t - \mu)^2 = \beta_1^2 E(y_{t-1} - \mu)^2 + 2\beta_1 E[(y_{t-1} - \mu)\varepsilon_t] + E[\varepsilon_t^2]$$

$$\Rightarrow \gamma_0 = \sigma^2/(1 - \beta_1^2)$$

  – β_1^2 needs to be less than 1;
  – The weak stationarity implies $-1 < \beta_1 < 1$;
  – β_1 measures the persistence of the dynamic dependence of an AR(1) time series.

ACF of AR(1)

- The ACF:

  $\gamma_l = \beta_1 \gamma_{l-1}$ and $\gamma_l = \gamma_{-l}$

  $\rho_l = \beta_1 \rho_{l-1} = \beta_1^{\,l}$ because $\rho_0 = 1$

  – The starting value of the ACF of a weakly stationary AR(1) process is one;

  – It then decays exponentially with the rate β_1

- In the AR(1) model, the value of β_1 plays very important role in capturing the nature of the data.

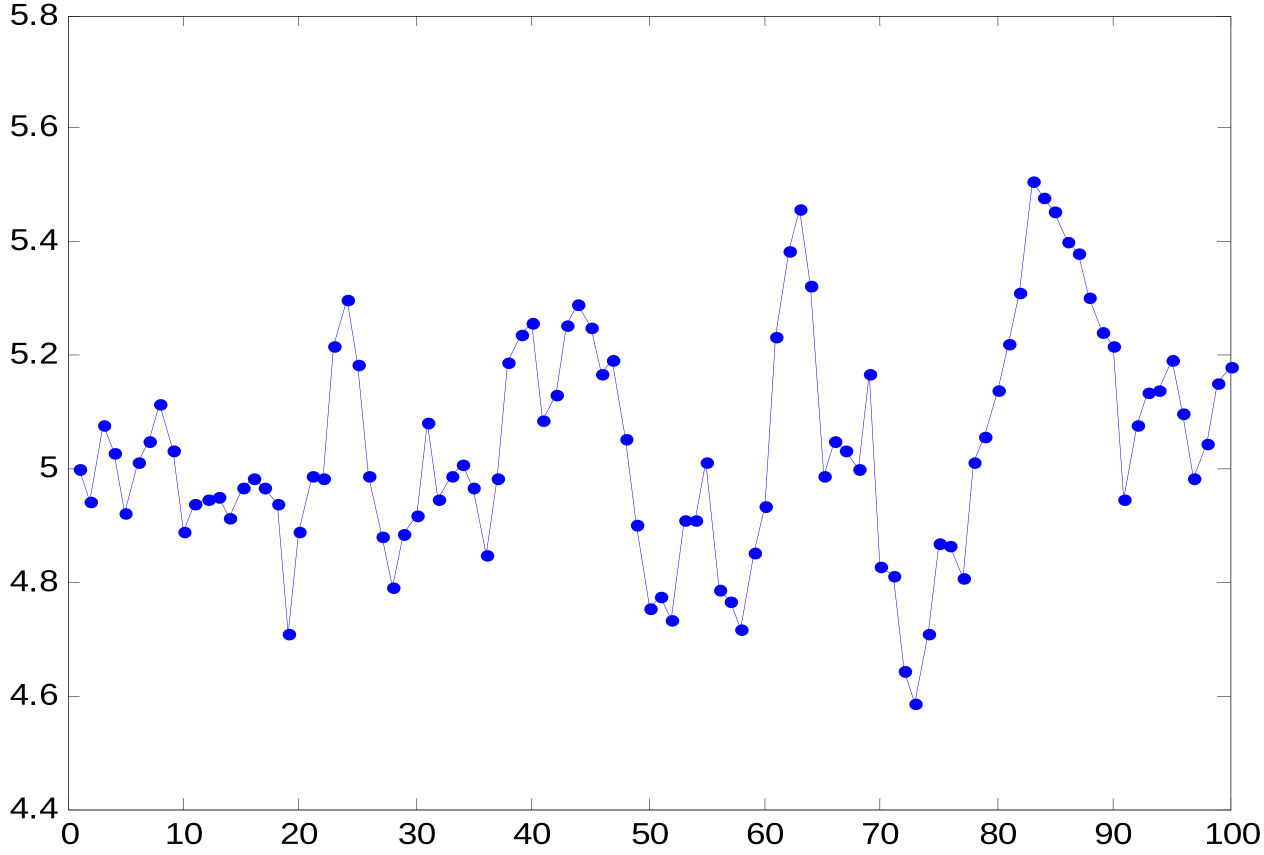
- We take a look at the simulated data for different values of β_1.

- Each data set is simulated from the AR(1) model with parameter sets given as follows.

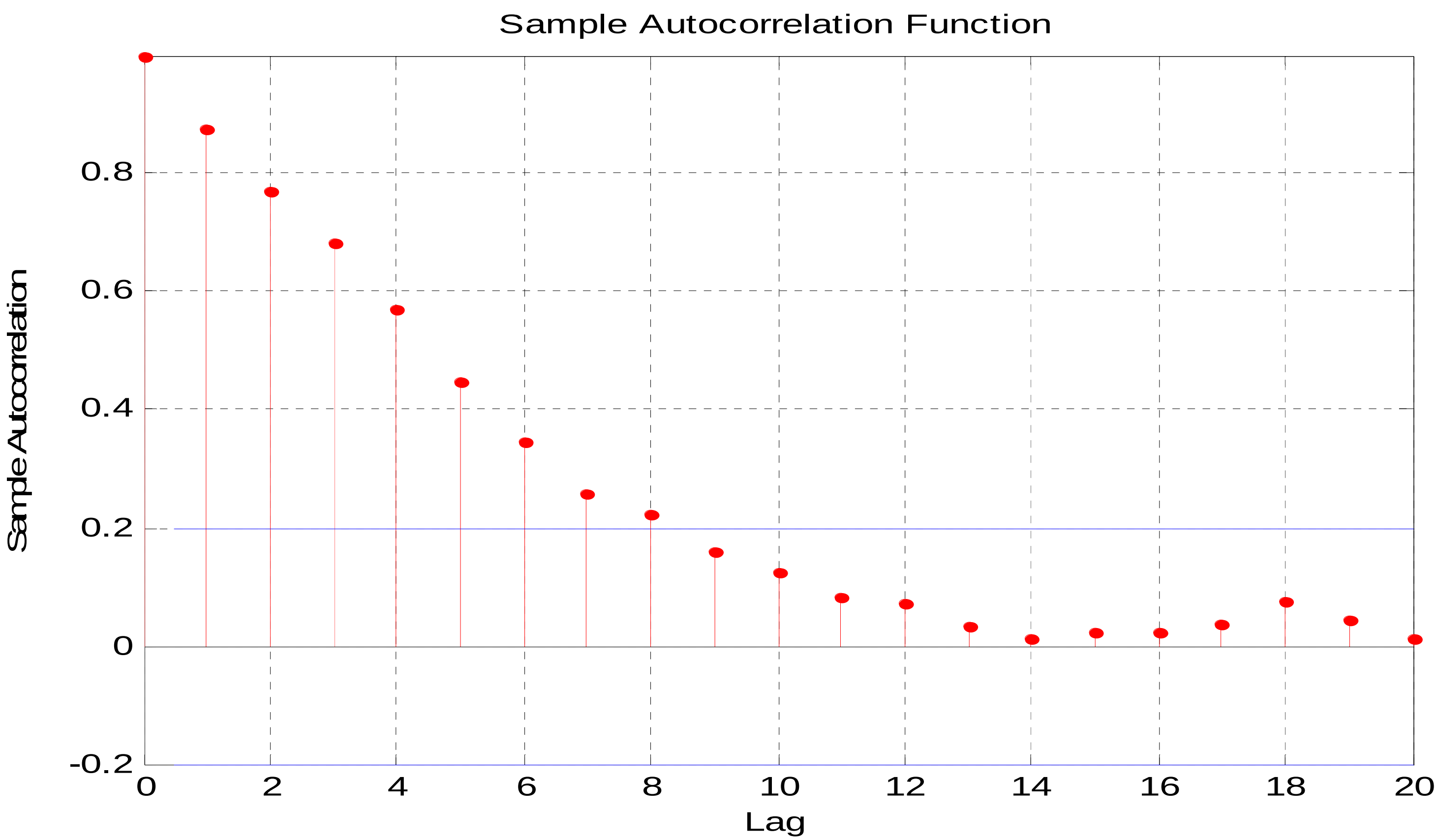
- The simulated data:
  - Series 1: $\beta_0 = 1$, $\beta_1 = 0.8$, $\sigma = 0.1$
  - Series 2: $\beta_0 = 1$, $\beta_1 = -0.8$, $\sigma = 0.1$
  - Series 3: $\beta_0 = 0.1$, $\beta_1 = 1$, $\sigma = 0.5$
  - Series 4: $\beta_0 = 0.1$, $\beta_1 = 1.1$, $\sigma = 0.5$

- Here the value of β_1 is the most interesting thing. Other parameters are just arbitrarily given.

- Series 1: Simulated Data

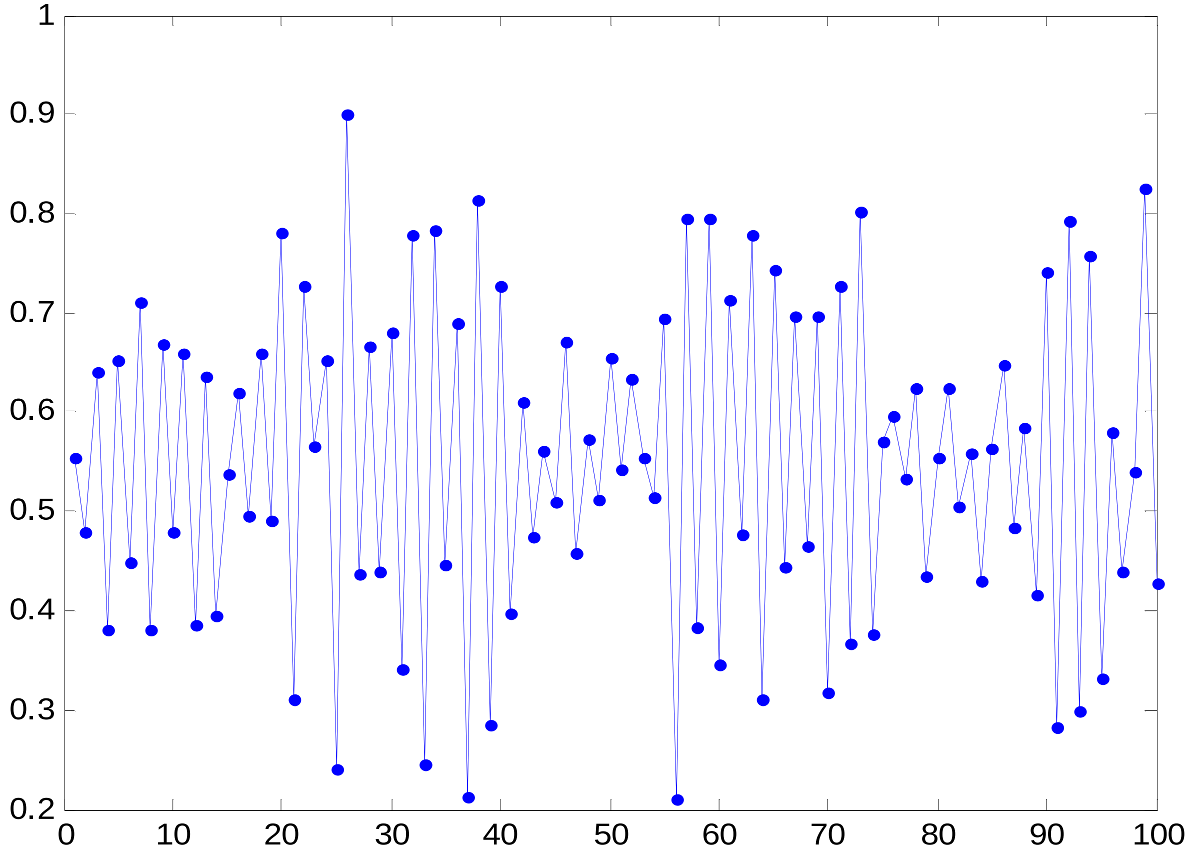


- Series 1: ACF

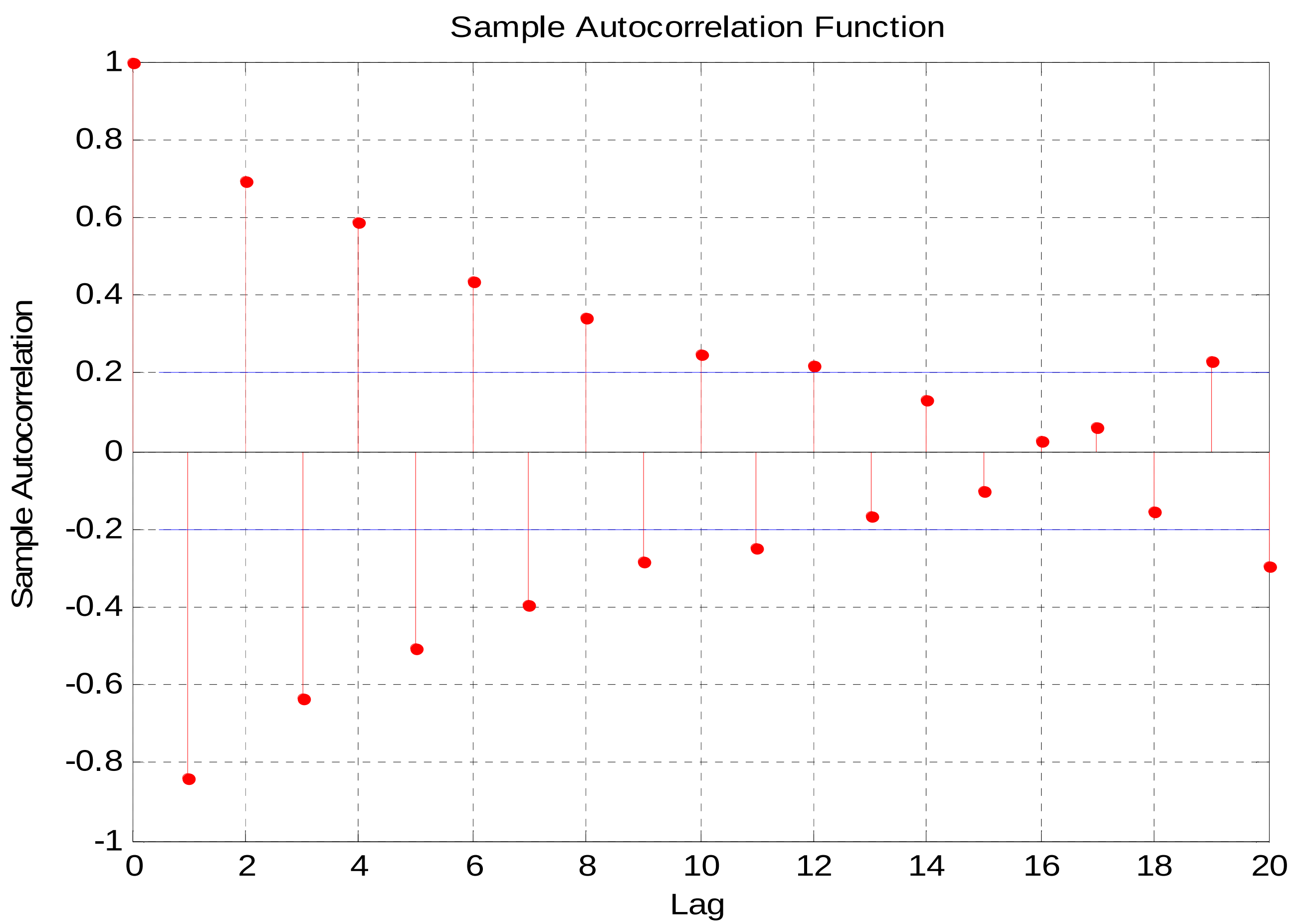


Sample Autocorrelation Function
Sample Autocorrelation
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20
Lag

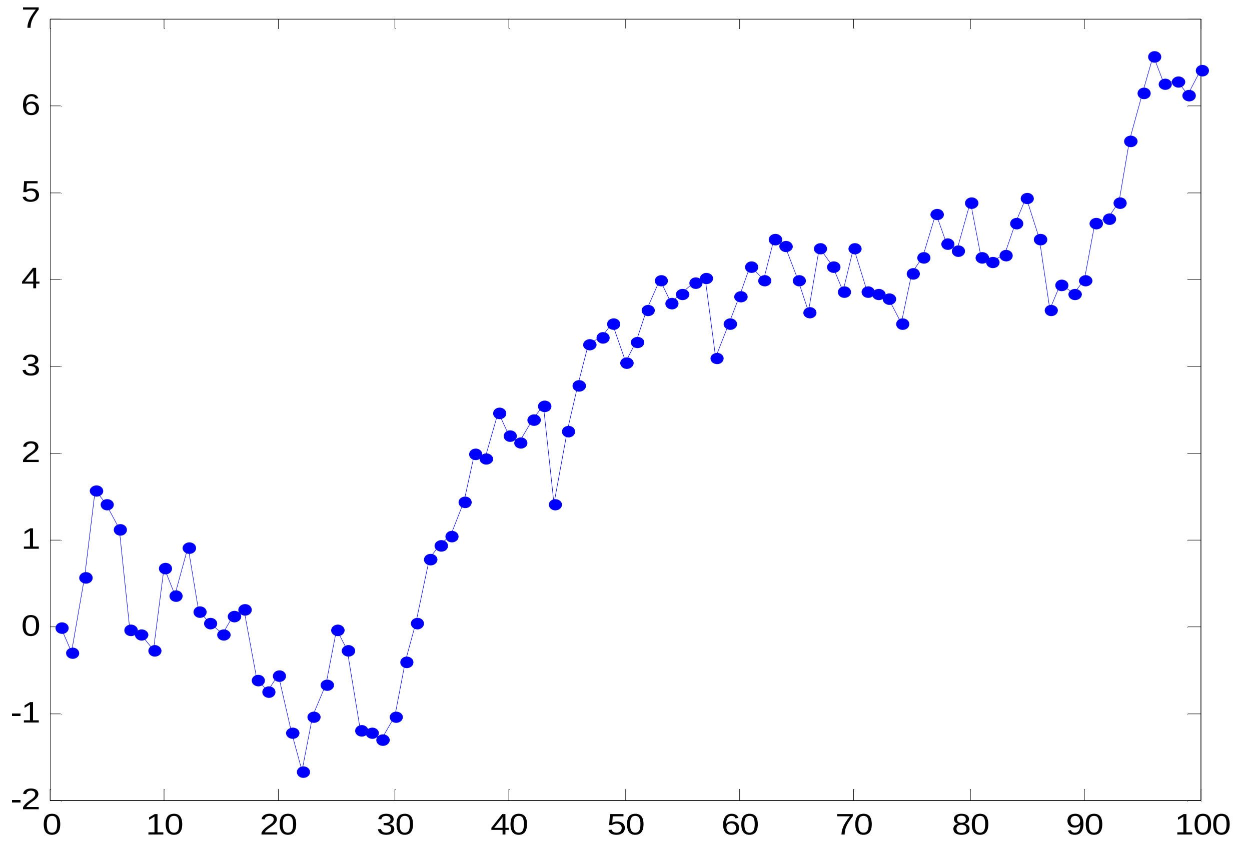
- Series 2: Simulated Data



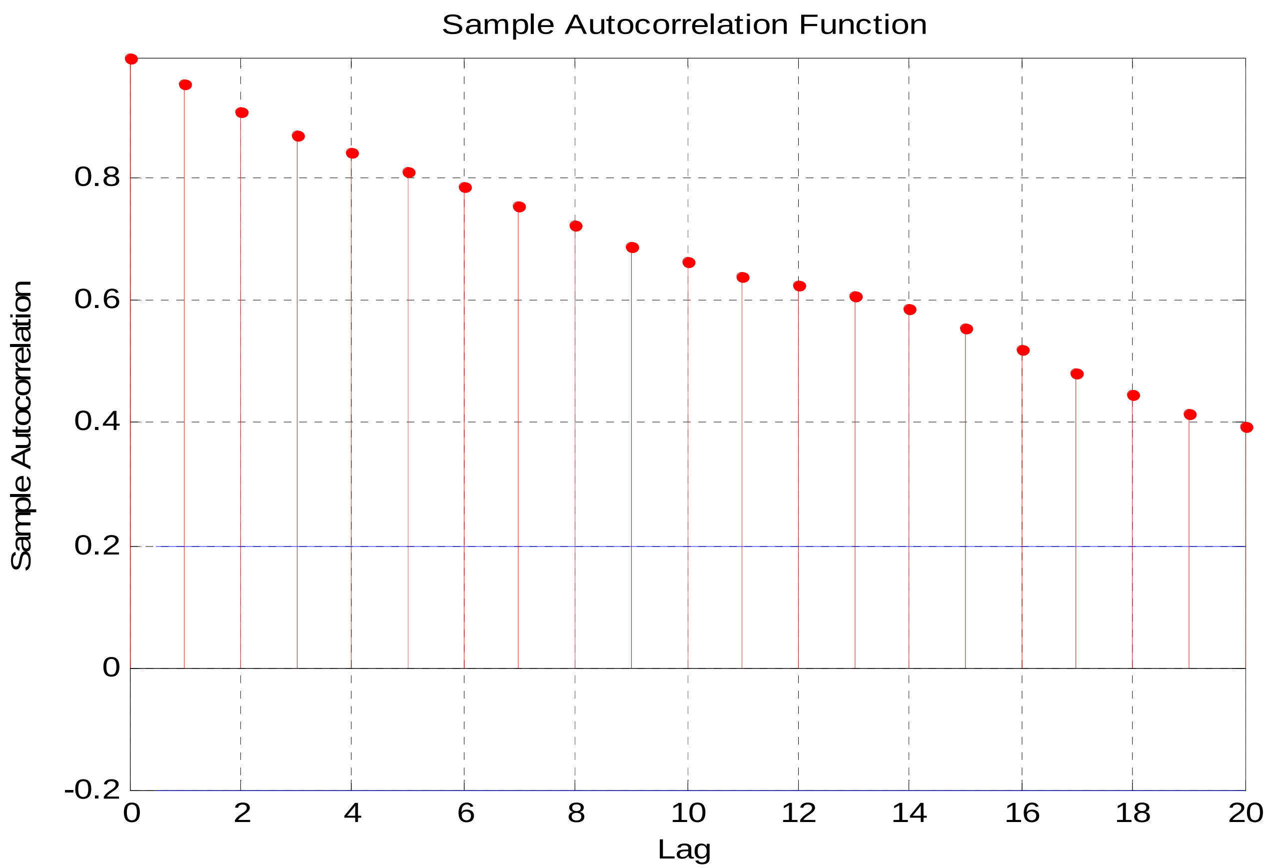
- Series 2: ACF


Sample Autocorrelation Function
Sample Autocorrelation
Lag
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
2
4
6
8
10
12
14
16
18
20

- Series 3: Simulated Data



- Series 3: ACF



Sample Autocorrelation Function
Sample Autocorrelation
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20
Lag

The Random Walk Model

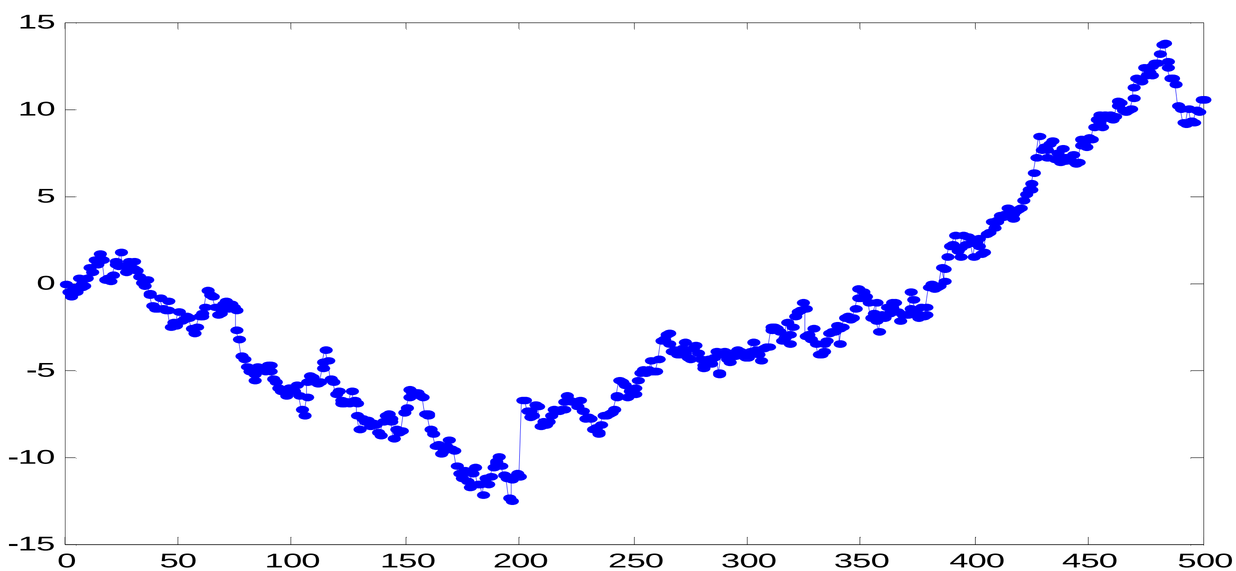
- When β_1 = 1, the AR(1) model is known as the random walk model

$$y_t = \beta_0 + y_{t-1} + \varepsilon_t$$

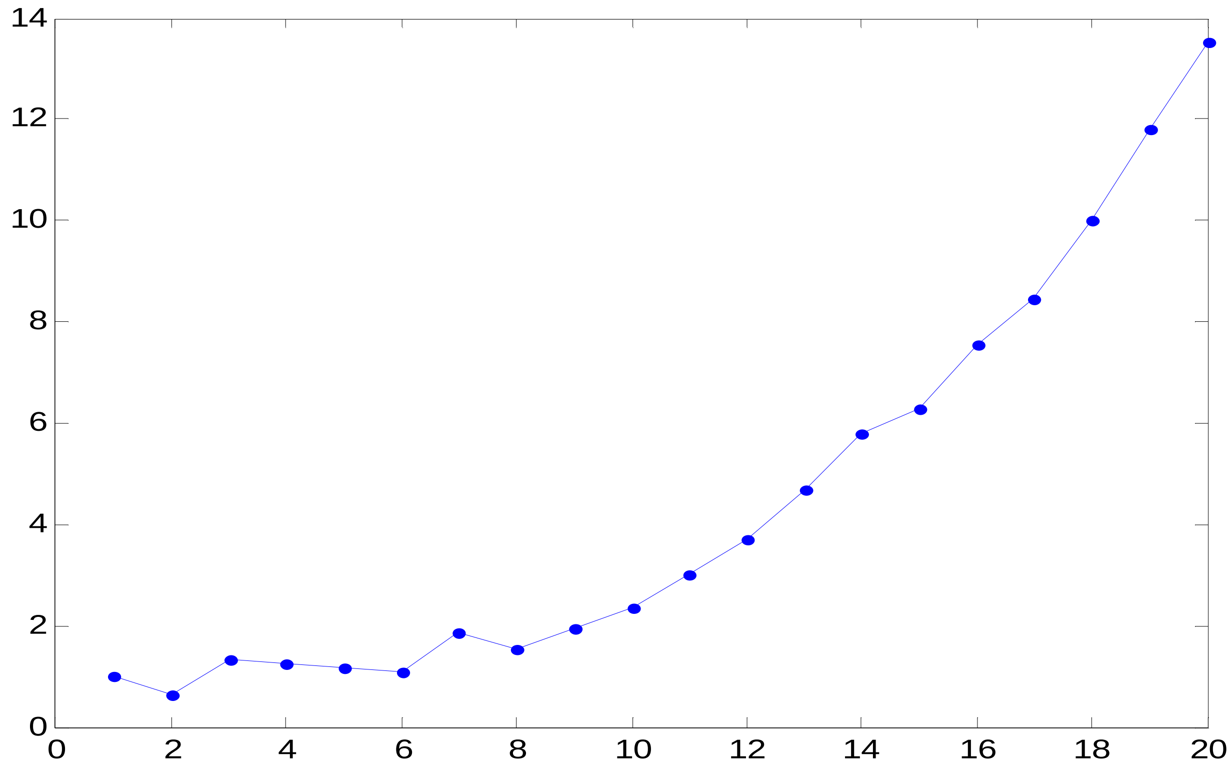
Or $$y_t - y_{t-1} = \beta_0 + \varepsilon_t = N(\beta_0 , \sigma^2)$$

- In the random walk model, the differenced series is i.i.d.

- The β_0 is called the drift:
  -- Positive drift: wandering upward;
  -- Negative drift: wandering downward;
  -- Zeros drift: the series meander around its starting value with no particular trend, but it can take very long time away its starting value.



- Series 4: Simulated Data



Summary

- $|\beta_1| < 1$:
  - The series has a mean level to which it reverts;
  - For the positive value, the series tends to wander above or below the mean level for a while;
  - For the negative value, the series tends to be above and below the mean level alternately;
  - The series is stationary: mean and variance do not change over time.

Summary

- $\beta_1 = 1$:
  - The series has no mean level, and thus is non-stationary;

- $|\beta_1| > 1$:
  - The series is explosive, and also non-stationary.

Forecasting

- Forecasting is an important application of time series analysis;
- Suppose we are at time t, and are interested in forecasting y_{t+h}
  - One-step-ahead forecast:

$$y_{t+1} = \beta_0 + \beta_1 y_t + \varepsilon_{t+1}$$

$$\hat{y}_{t+1} = E[y_{t+1} \mid F_t] = \beta_0 + \beta_1 y_t$$

- Suppose that the AR(1) model accurately describes the data and that the parameters are given by $\beta_0 = 1$, $\beta_1 = 0.8$, $\sigma^2 = 0.5$.
- If $y_T = 6$, our prediction for y_{T+1} would be obtained by plugging the parameters into the model

$$y_{T+1} = 1 + 0.8*(y_T = 6) + \varepsilon_{T+1}$$
$$= 5.8 + \varepsilon_{T+1}$$

and $$E[y_{T+1}|F_t] = 5.8$$

- Two-step-ahead forecast:

$$y_{t+2} = \beta_0 + \beta_1 y_{t+1} + \varepsilon_{t+2}$$
$$\Rightarrow \hat{y}_{t+2} = E[y_{t+2} \mid F_t] = \beta_0 + \beta_1 E[y_{t+1} \mid F_t]$$
$$\Rightarrow \hat{y}_{t+2} = \beta_0 + \beta_1(\beta_0 + \beta_1 y_t)$$

- Multi-step-ahead forecast:

$$y_{t+h} = \beta_0 + \beta_1 y_{t+h-1} + \varepsilon_{t+h}$$
$$\Rightarrow \hat{y}_{t+h} = E[y_{t+h} \mid F_t] = \beta_0 + \beta_1 E[y_{t+h-1} \mid F_t]$$
$$\Rightarrow \hat{y}_{t+h} = \beta_0 + \beta_1 \hat{y}_{t+h-1}$$

The General AR(p) Model

- We can generalize the AR(1) model to let y_t depend on p lags of y:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \varepsilon_t$$

- If p = 2, we have AR(2) model.
  - Conditional mean and variance
  - Unconditional mean and variance
  - The autocorrelation function

PACF

- When we have a time series of data, how can we determine the order *p*;
- Of course, we can use ACF. But note that the ACF of AR(p) decays slowly and can not provide precise information on *p*;
- The partial autocorrelation function (PACF) is another useful tool for selection of *p*;

PACF

- The first partial autocorrelation is simply the estimated coefficient b_1 from the regression

$$y_t = b_0 + b_1 y_{t-1}$$

- The second partial autocorrelation is the estimate b_2 from the regression

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2}$$

- In general, the *j-th* partial autocorrelation is the *j-th* estimate in a regression of y_t on y_{t-1}, …, y_{t-j}.

PACF

- An AR(p) model has a slowly decaying autocorrelation function.
- But it has non-zero first p partial autocorrelations and zero ones thereafter.
- In practice, whenever we have a time series, we usually need both ACF and PACF, since you will see later on MA(q) model has very different ACF and PACF.

How to Estimate AR(p) Model

- It is in linear form. Can we use OLS?
  - OLS assumptions: AR(1) with zeros constant

$$
\begin{aligned}
E[y_t\, \varepsilon_t] &= E[(\beta_1 y_{t-1} + \varepsilon_t)\, \varepsilon_t] \\
&= \beta_1 E[y_{t-1}\, \varepsilon_t] + E[\varepsilon_t\, \varepsilon_t] \\
&= E[\varepsilon_t\, \varepsilon_t] \neq 0
\end{aligned}
$$

  - Finite-sample properties based on the strict exogeneity do not hold;
  - However, the OLS estimator in AR(p) models has good large-sample properties.

How to Estimate AR(p) Model

- MLE can be always applied as long as we know the distribution of $\varepsilon_{t.}$
- For the AR(1) model:

$$
\begin{aligned}
f(y_T, \ldots, y_1) &= f(y_T \mid y_{T-1}, \ldots, y_1) f(y_{T-1}, \ldots, y_1) \\
&= f(y_T \mid y_{T-1}, \ldots, y_1) f(y_{T-1} \mid y_{T-2}, \ldots, y_1) \\
&\quad f(y_{T-2}, \ldots, y_1) \\
&\ldots \ldots \\
&= f(y_T | y_{T-1}) f(y_{T-1} | y_{T-2}) \ldots f(y_1)
\end{aligned}
$$

Information Criteria

- There are several information criteria available to determine the order p of an AR process. All of them are likelihood-based;
- If we assume Gaussian noise ε,
  - Akaike Information Criterion (AIC)

$$AIC = \ln(\hat{\sigma}^2) + \frac{2}{T} p$$

  - Schwartz-Bayesian Information Criterion (BIC)

$$BIC = \ln(\hat{\sigma}^2) + \frac{\ln(T)}{T} p$$

Selection Rule

- To use AIC and/or BIC to select an AR model, one computes AIC (BIC) for all orders from 0 to p.
- Select the order of k that has the minimum values.
- In practice, we firstly use PACF, and then use AIC or BIC.
- An example

An Example

TABLE 2.1 Sample Partial Autocorrelation Function and Some Information Criteria for the Monthly Simple Returns of CRSP Value-Weighted Index from January 1926 to December 2008

p	1	2	3	4	5	6
PACF	0.115	−0.030	−0.102	0.033	0.062	−0.050
AIC	−5.838	−5.837	−5.846	−5.845	−5.847	−5.847
BIC	−5.833	−5.827	−5.831	−5.825	−5.822	−5.818
p	7	8	9	10	11	12
PACF	0.031	0.052	0.063	0.005	−0.005	0.011
AIC	−5.846	−5.847	−5.849	−5.847	−5.845	−5.843
BIC	−5.812	−5.807	−5.805	−5.798	−5.791	−5.784

The Moving Average Models

- The White Noise model
- The MA(1) Model
- The MA(q) Model
- ACF and PACF for MA(q) models
- How to estimate MA(q) models

White Noise Revisited

- A white noise time series is simply a mean zeros series with all autocorrelations equal to zero;
- Example: $y_t = \varepsilon_{t\,,}$ where is i.i.d N(0, σ^2). This is simply the i.i.d normal time series we met before.
- We could extend this model by giving a mean: $y_t = \mu + \varepsilon_{t\,.}$

The MA(1) Model

- We can make the above model more interesting by the following extension:

$$y_t = \mu + \varepsilon_t + \vartheta \varepsilon_{t-1}$$

- The model is different from the AR model since we construct the time series for y as a combination of two random draws from a normal.

The MA(1) Model

- The expectation of y_t

$$E(y_t) = E(\mu + \varepsilon_t + \vartheta\, \varepsilon_{t-1}) = \mu$$

- The variance of y_t

$$\begin{aligned} Var(y_t) &= E(y_t - \mu)^2 \\ &= E(\varepsilon_t^2 + 2\vartheta\, \varepsilon_t \varepsilon_{t-1} + \vartheta^2 \varepsilon_{t-1}^2) \\ &= (1 + \vartheta^2)\sigma^2 \end{aligned}$$

- The first autocovariance

$$\gamma_1 = E[(y_t - \mu)\,(y_{t-1} - \mu)] = \vartheta\sigma^2$$

The MA(1) Model

- Can you see that higher autocovariances are all zero?

$$\gamma_j = \mathrm{E}(y_t - \mu)\,(y_{t-j} - \mu) = 0 \quad \textit{for } j > 1$$

- *The jth autocorrelation is defined as*

$$\rho_j = \gamma_j / \gamma_0$$

- For the *MA(1)* model, we have

$$\rho_1 = \vartheta\sigma^2 / (1 + \vartheta^2)\sigma^2 = \vartheta / (1 + \vartheta^2)$$

$$\rho_j = 0 \quad \textit{for } j > 1$$

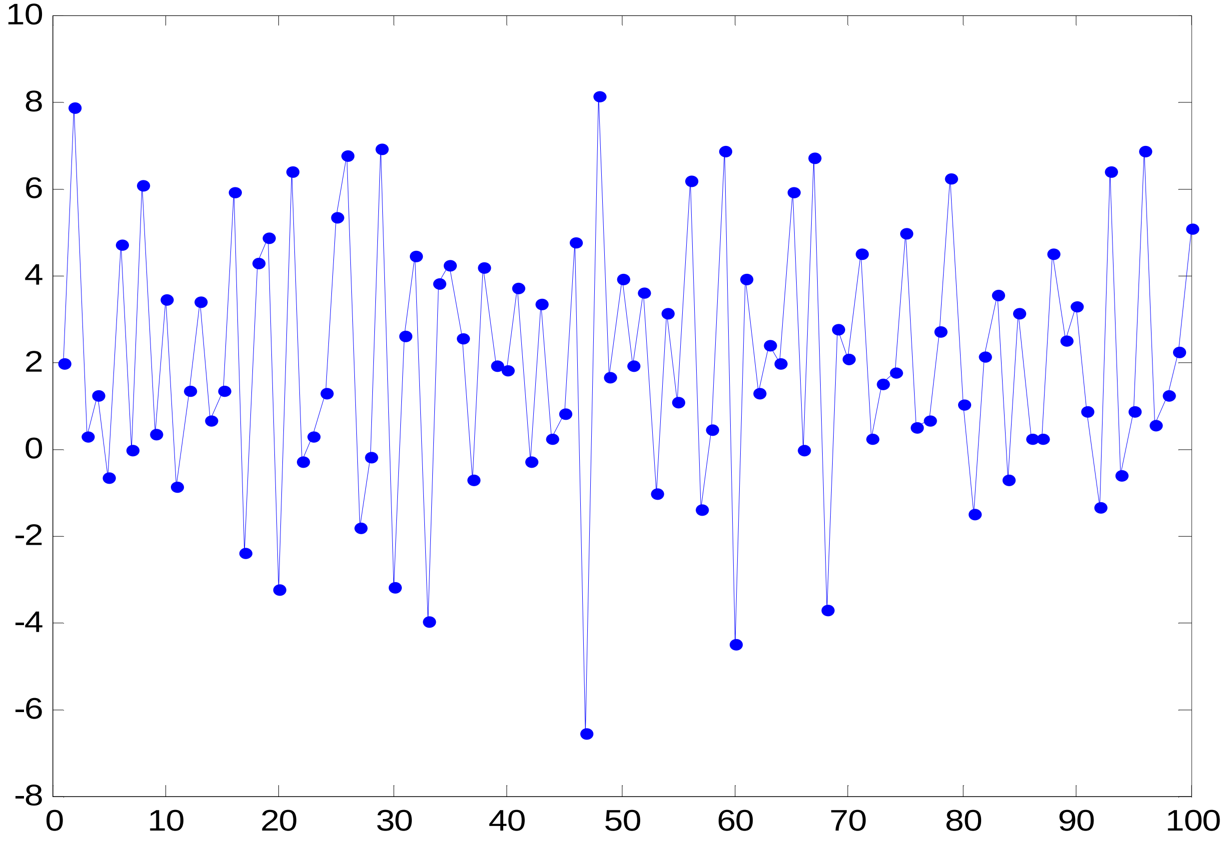
The MA(1) Model

- Notice that both y_t and y_{t-1} depend on ε_{t-1}

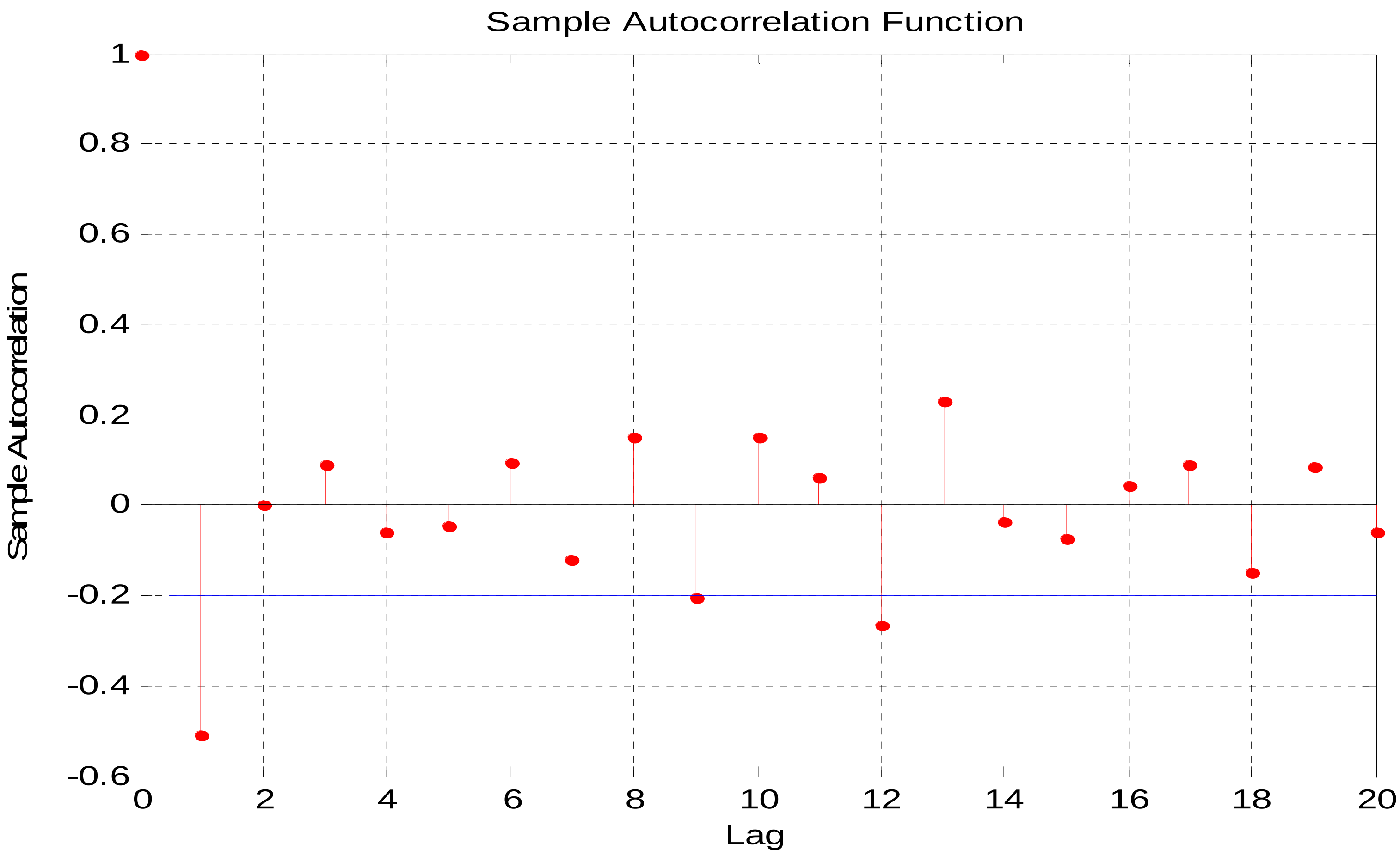
$$y_t = \mu + \varepsilon_t + \vartheta\, \varepsilon_{t-1}$$
$$y_{t-1} = \mu + \varepsilon_{t-1} + \vartheta\, \varepsilon_{t-2}$$

- Even though ε_t is i.i.d, y_t are correlated.
- For technical reasons, we will restrict $|\ \vartheta\ | \leq 1$.
- To better understand how the model works, let us take a look at the simulated data.

- Series 1: $\mu = 2, \vartheta = -0.9, \sigma = 2$

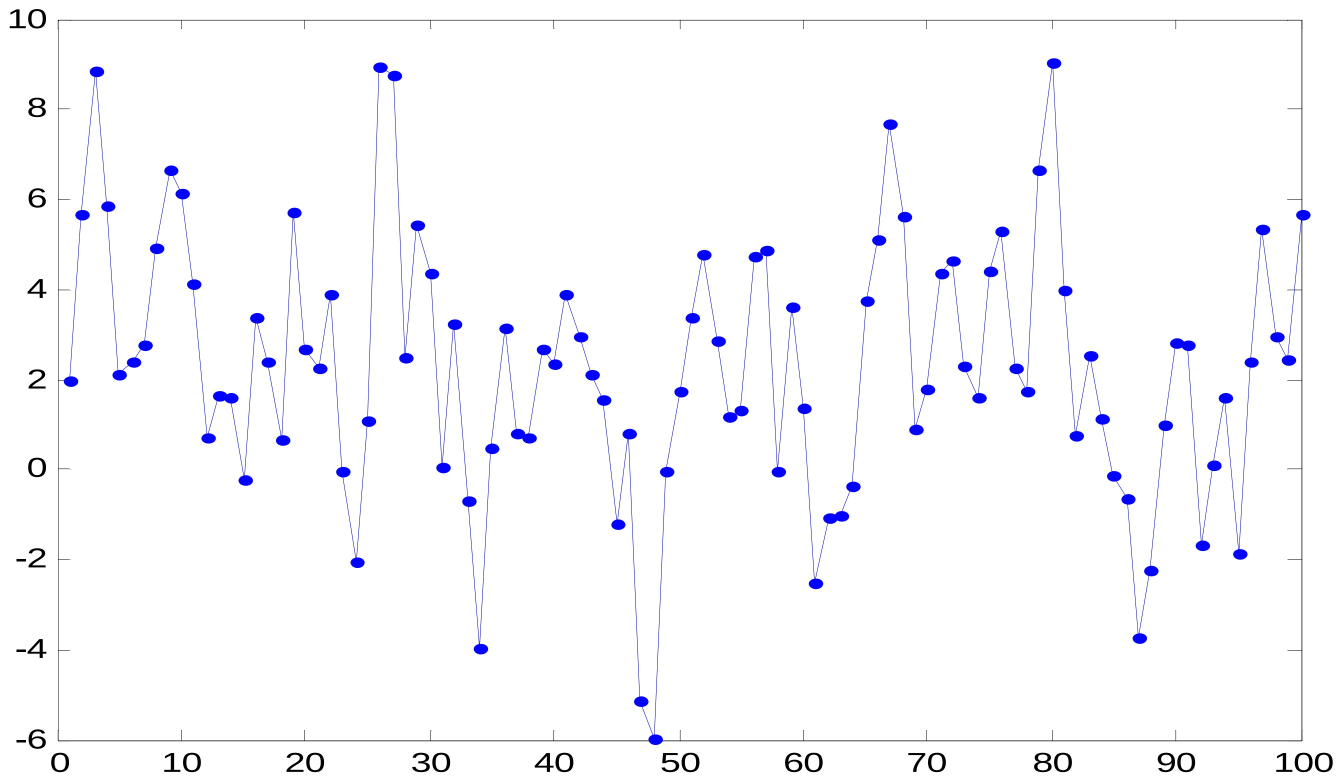


- Series 1: ACF

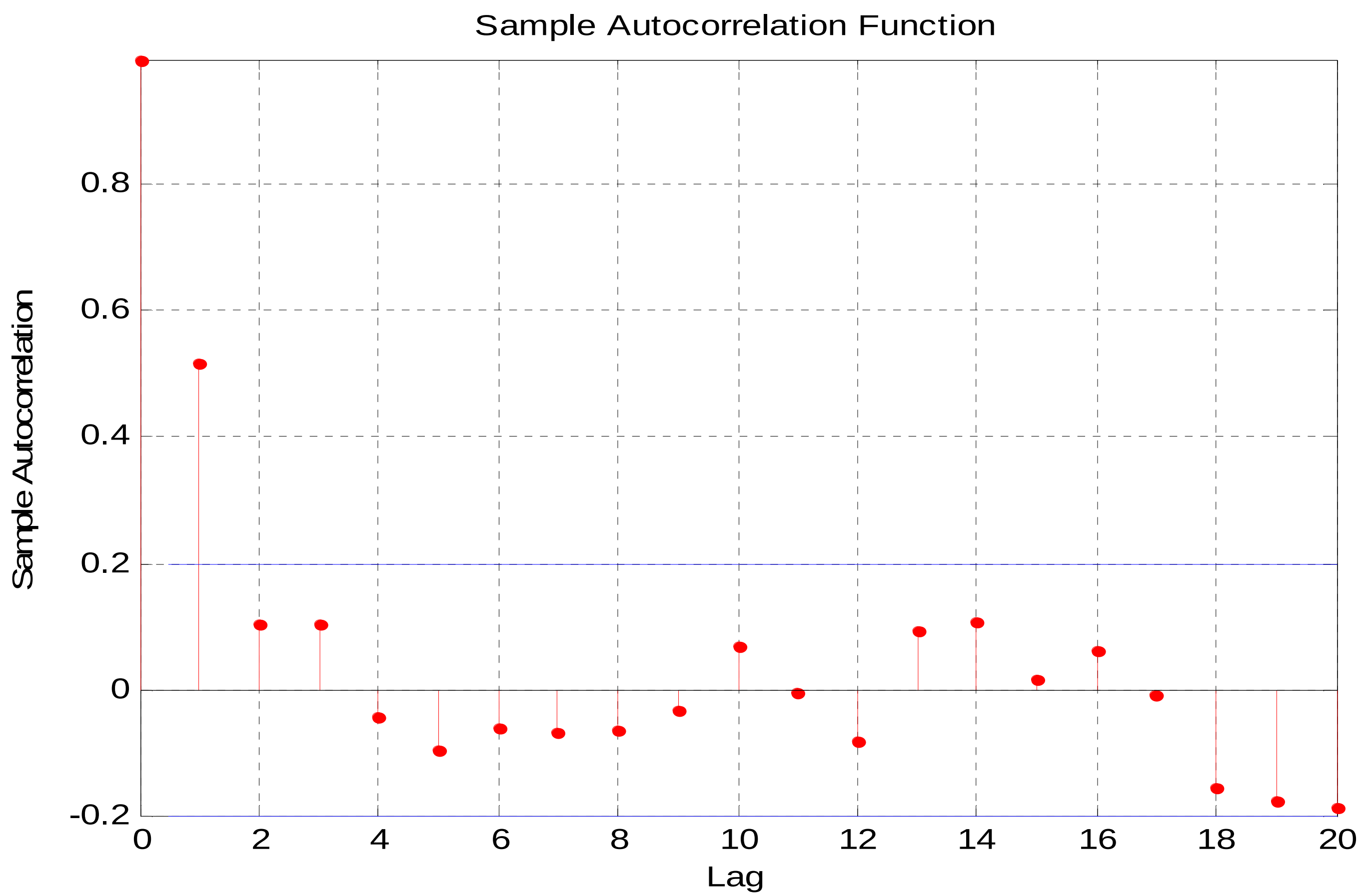


Sample Autocorrelation Function
Sample Autocorrelation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
2
4
6
8
10
12
14
16
18
20
Lag

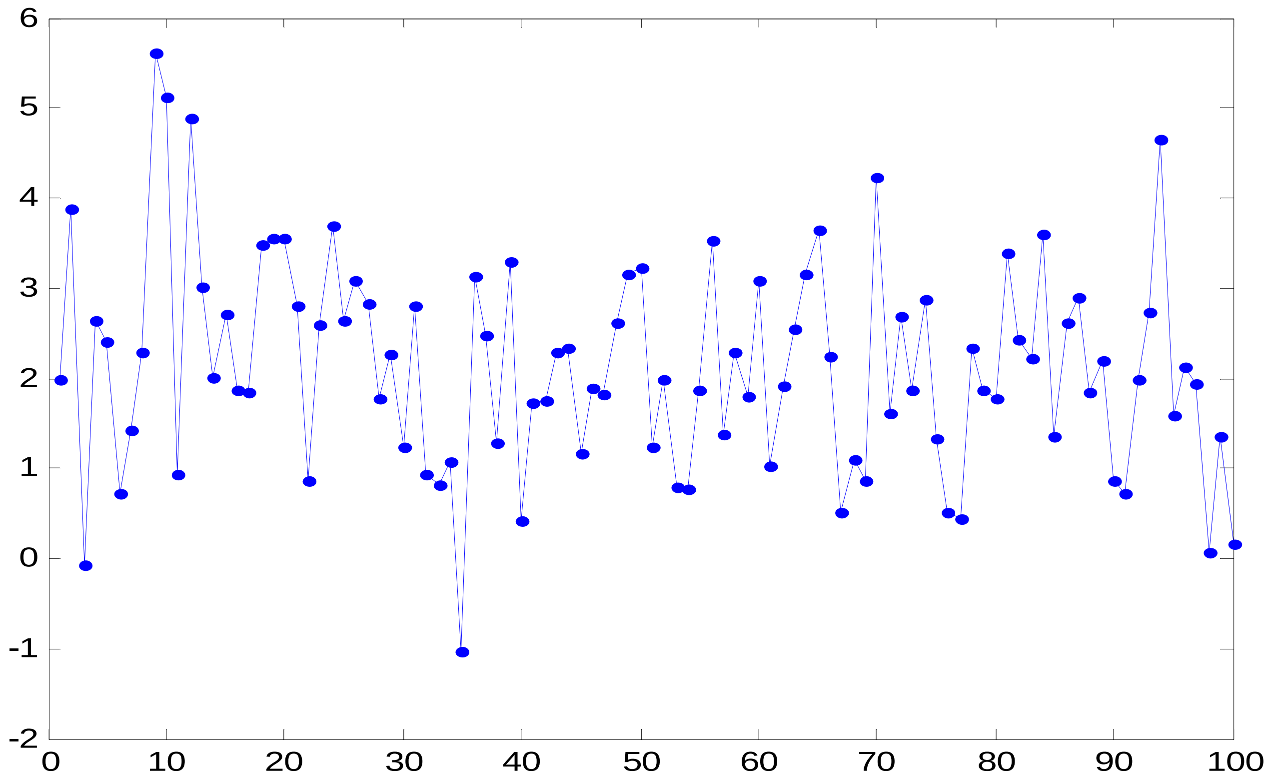
- Series 2: $\mu = 2, \vartheta = 0.9, \sigma = 2$



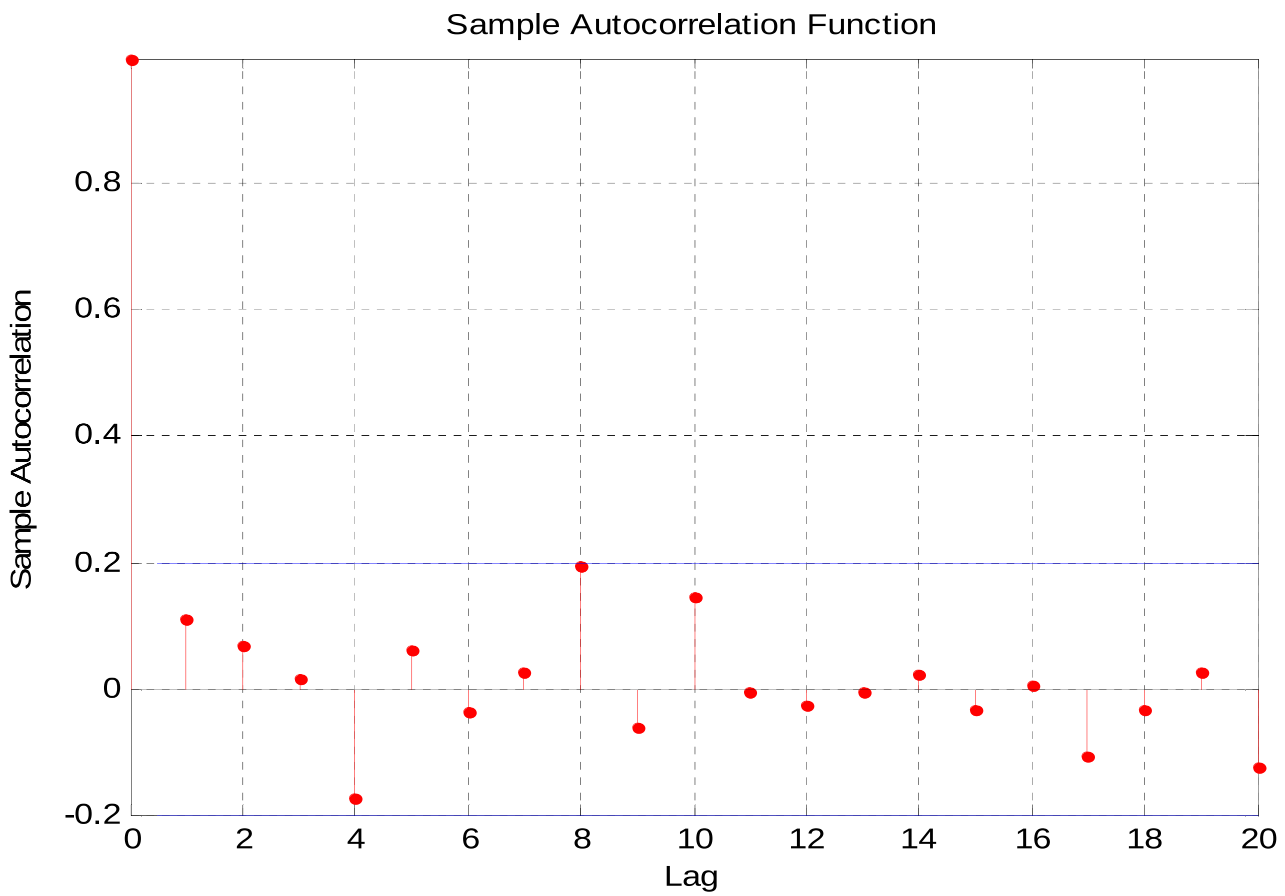
- Series 2: ACF


Sample Autocorrelation Function
Sample Autocorrelation
Lag
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20

- Series 3: $\mu = 2, \vartheta = 0.1, \sigma = 2$



- Series 3: ACF



Sample Autocorrelation Function
Sample Autocorrelation
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20
Lag

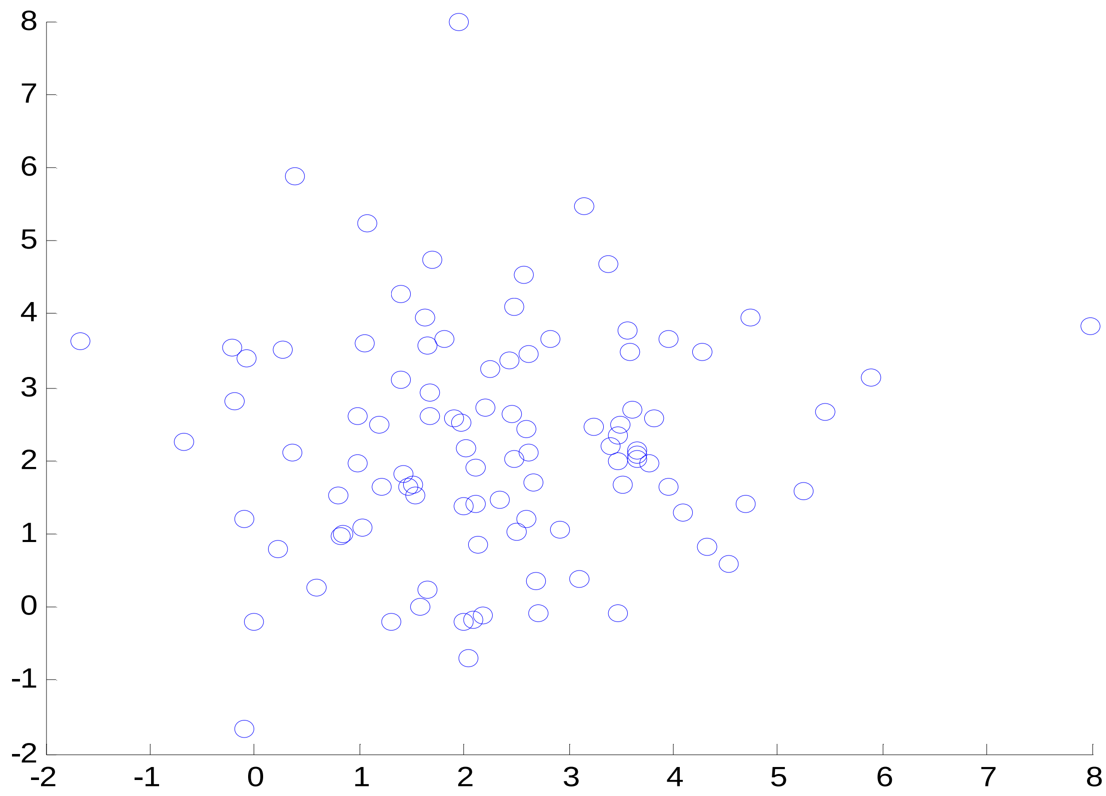
Interpreting MA(1) Model

- The MA(1) says that y_t is determined by a current shock ε_t and a lagged shock ε_{t-1};
- If ϑ is positive, an unexpected large shock yesterday will make y_t today large;
- If ϑ is negative, an unexpected large shock yesterday will make y_t today small;
- In sum, the value of y today depends on the surprise from yesterday.

The MA(q) Model

- We have seen that in MA(1) model, y_t is correlated to y_{t-1};
- What do you expect for the correlation between y_t and y_{t-2};
- We know y_t depends on ε_t and ε_{t-1};
- And y_{t-2} depends on ε_{t-2} and ε_{t-3};
- Since is ε_t i.i.d, we should expect that the correlation between y_t and y_{t-j} is zero for all $j > 1$.

- Series 2: *μ = 2, ϑ = 0.9, σ = 2*



- We can allow for richer dependence in y_t by allowing y_t to depend on more lagged values of $\varepsilon_{t.}$
- The MA(q) model says that

$$y_t = \mu + \varepsilon_t + \sum \vartheta_j \varepsilon_{t-j}$$

- So the value of y_t depends on ε_t and its q past values.

- For example, a MA(2) model looks like this:

$$y_t = \mu + \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \vartheta_2 \varepsilon_{t-2}$$

- *ε_t is i.i.d normally distributed;*
- Now y_t and y_{t-2} should be correlated since both of them depend on ε_{t-2}

$$y_{t-2} = \mu + \varepsilon_{t-2} + \vartheta_1 \varepsilon_{t-3} + \vartheta_2 \varepsilon_{t-4}$$

- The expectation of y_t in *MA(q)*

$$E(y_t) = \mu$$

- The variance of y_t

$$\mathrm{Var}(y_t) = E(y_t - \mu)^2$$
$$= (1 + \vartheta_1^2 + \vartheta_2^2 + \ldots + \vartheta_q^2)\sigma^2$$

- The autocovariances for j = 1, 2, ..., q

$$\gamma_j = E(y_t - \mu)(y_{t-j} - \mu)$$
$$= (\vartheta_j + \vartheta_{j+1}\vartheta_1 + \vartheta_{j+2}\vartheta_2 + \ldots + \vartheta_q \vartheta_{q-j})\sigma^2$$

and for j > q,

$$\gamma_j = 0$$

- For example, for an *MA(2)* model

$$\gamma_0 = (1 + \vartheta_1^2 + \vartheta_2^2)\sigma^2$$

$$\gamma_1 = (\vartheta_1 + \vartheta_2 \vartheta_1)\sigma^2$$

$$\gamma_2 = \vartheta_2 \sigma^2$$

$$\gamma_3 = \gamma_4 = \ldots = 0$$

- For any values of $\vartheta_1, \vartheta_2, \ldots \vartheta_q$, the MA(q) is covariance stationary.

- From the above discussion, we have the following result: For an MA(q) model, the first q autocorrelations will be non-zero, whereas all $q + j$ autocorrelations will be zero for $j > 0$;
- What does the PACF look like in a MA model?
- It turns out that the partial autocorrelations decay slowly.
- Do you still remember that in an AR model, what do ACF and PACF look like?

Summary of ACF and PACF

	AR(p) Model	MA(q) Model
ACF	Slowly decaying	Non-zero for the first q and zero thereafter
PACF	Non-zero for the first p and zero thereafter	Slowly decaying

Estimation of MA(q)

- Maximum likelihood estimation is commonly used to estimate MA models. There are two approaches.
- Conditional likelihood method
  - Initial shock (ε_0) is assumed to be zero;
  - A recursive procedure;
- Exact likelihood method
  - Treat initial shock as a parameter

Forecasting

- For MA(1) model

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

- Then the 1-step-ahead forecast is:

$$\hat{y}_{t+1} = E[y_{t+1} \mid F_t] = \mu + \theta\varepsilon_t$$

- And 2-step-ahead forecast

$$\hat{y}_{t+2} = E[y_{t+2} \mid F_t] = \mu$$

- What is the multi-step-ahead forecast?
- For MA(2) model?

ARMA(p, q) Models

- Combine the AR(p) model and the MA(q) model together, we obtain the so-called ARMA(p, q) model.
- And this works very well in practice.
- The simplest ARMA(1, 1) model looks like:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t + \theta\, \varepsilon_{t-1}$$

ε_{t-j} *is i.i.d normal* $(0, \sigma^2)$

Properties of ARMA(1, 1)

- Taking expectation, we have

$$E[y_t] = \beta_0 + \beta_1 E[y_{t-1}] + E[\varepsilon_t] + \theta E[\varepsilon_{t-1}]$$

$$\Rightarrow \mu = \frac{\beta_0}{1-\beta_1}$$

- Autocovariance: rewrite ARMA(1, 1)

$$y_t - \mu = \beta_1 (y_{t-1} - \mu) + \varepsilon_t + \theta \varepsilon_{t-1}$$

- Variance

$$Var(y_t) = \beta_1^2 Var(y_{t-1}) + \sigma^2 + \theta^2\sigma^2 + 2\beta_1\theta Cov(y_{t-1}\varepsilon_{t-1})$$

$$\Rightarrow \gamma_0 = \frac{(1+2\beta_1\theta+\theta^2)\sigma^2}{1-\beta_1^2}$$

- Autocovariance
  - If l = 1

$$\gamma_1 = \beta_1\gamma_0 + \theta\sigma^2 \Rightarrow \rho_1 = \beta_1 + \frac{\theta\sigma^2}{\gamma_0}$$

  - If l > 1

$$\gamma_l = \beta_1\gamma_{l-1} \Rightarrow \rho_l = \beta_1\rho_{l-1}$$

- To generalize, the ARMA(p, q) model looks like:

$$y_t = \mu + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \ldots + \vartheta_q \varepsilon_{t-q}$$

ε_{t-j} *is i.i.d normal* $(0, \sigma^2)$

- The model becomes difficult to interpret;
- And it is also difficult to determine *p* and *q* using ACF and PACF;
- However, we can use the same approach as before: try different models of AR, MA and ARMA.

Estimate the ARMA Model

- The MLE is the most convenient and neat way to estimate the ARMA model
- Then How to construct the likelihood function
  - Using an iterative approach.
- Maximize the likelihood and obtain the parameter estimates and standard deviation.
- Construct statistics to implement hypothesis analysis.

Unit-Root Non-Stationarity

- So far, we have focused on time series that are stationary.
- In some studies, interest rates, foreign exchange rates, or the price series of an asset are of interest.
- These series tend to be non-stationary.
- The best known example of unit-root non-stationary time series is the random-walk model, which we already saw.

Random Walk

- A time series y_t is a random walk if it satisfies

$$y_t = y_{t-1} + \varepsilon_t$$

  where y_0 is a real number denoting the starting value and ε_t is a white noise series.
- The random walk is a special AR(1) model with coefficient of y_{t-1} is unit.
- It is therefore not weakly stationary. We call it a unit-root non-stationary time series.

- The random walk model has widely been considered as a statistical model for the movement of the log stock prices.
- Under such a model, the stock price is not predictable or mean-reverting.
- To see this, the 1-step ahead fore cast is

$$\hat{y}_{t+1} = E[y_{t+1} \mid y_t, y_{t-1}, \ldots] = y_t$$

and the 2-step-ahead forecast is

$$\hat{y}_{t+2} = E[y_{t+2} \mid y_t, y_{t-1}, \ldots] = y_t$$

- In fact, we have

$$\hat{y}_{t+h} = E[y_{t+h} \mid y_t, y_{t-1}, \ldots] = y_t$$

- Thus, for all forecasting horizons, point forecasts of a random walk model are simply the value of the series at the forecast origin.
- The MA representation is

$$y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots$$

- Variance of h-step ahead forecast: $h\sigma^2$

- The usefulness of point forecast diminishes as h increases – unpredictable;
- Unconditional variance of y_t is unbounded, indicating that y_t can assume any values for a sufficiently large t.
- The impact of any past shock does not decay over time – permanent effect or long memory.
- The sample ACFs are all approaching one as the sample size increases.

Random Walk with Drift

- If we introduce a drift in a random walk model, we then have

$$y_t = \mu + y_{t-1} + \varepsilon_t$$

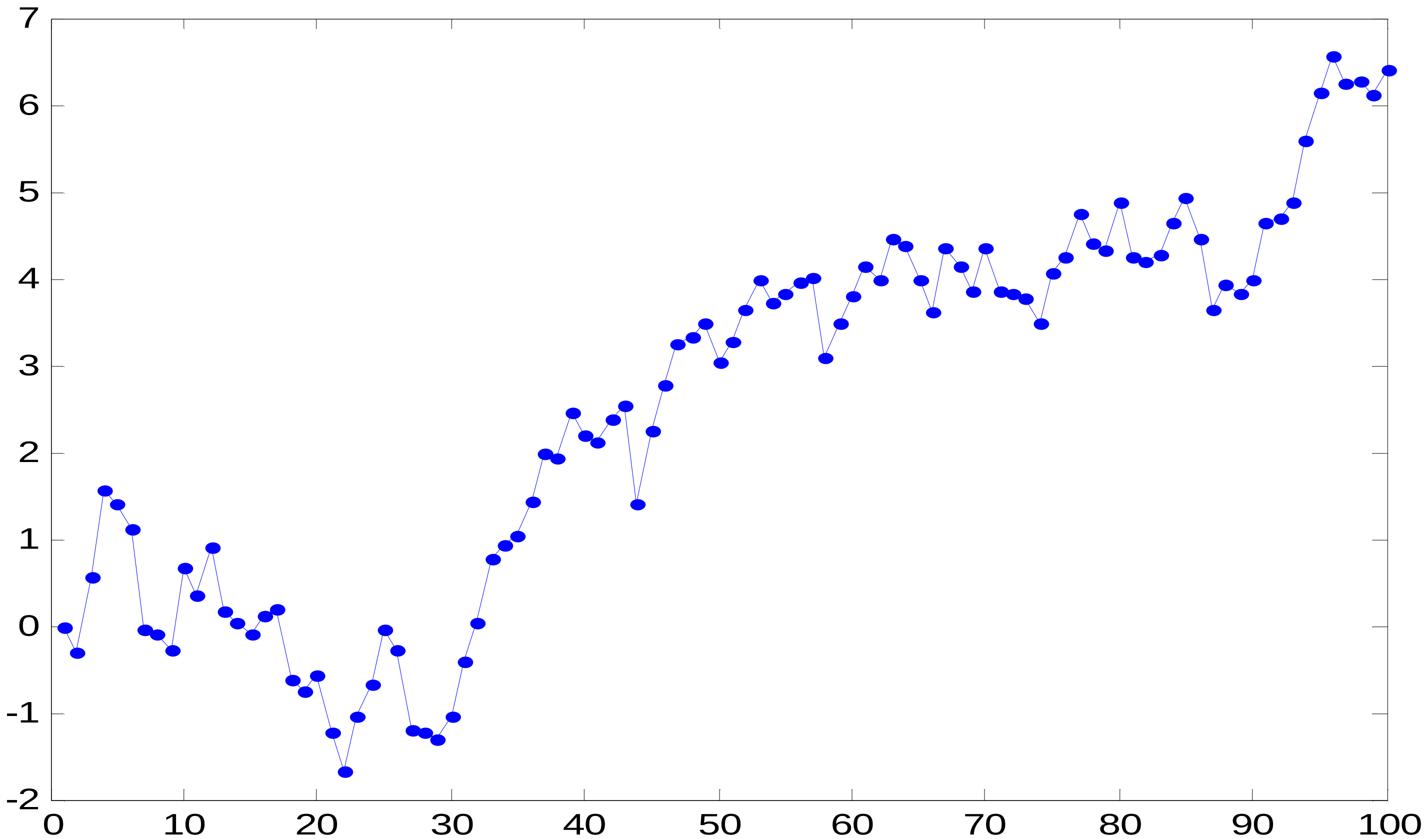
where $\mu = E[y_t - y_{t-1}]$ and ε_t is a white noise.

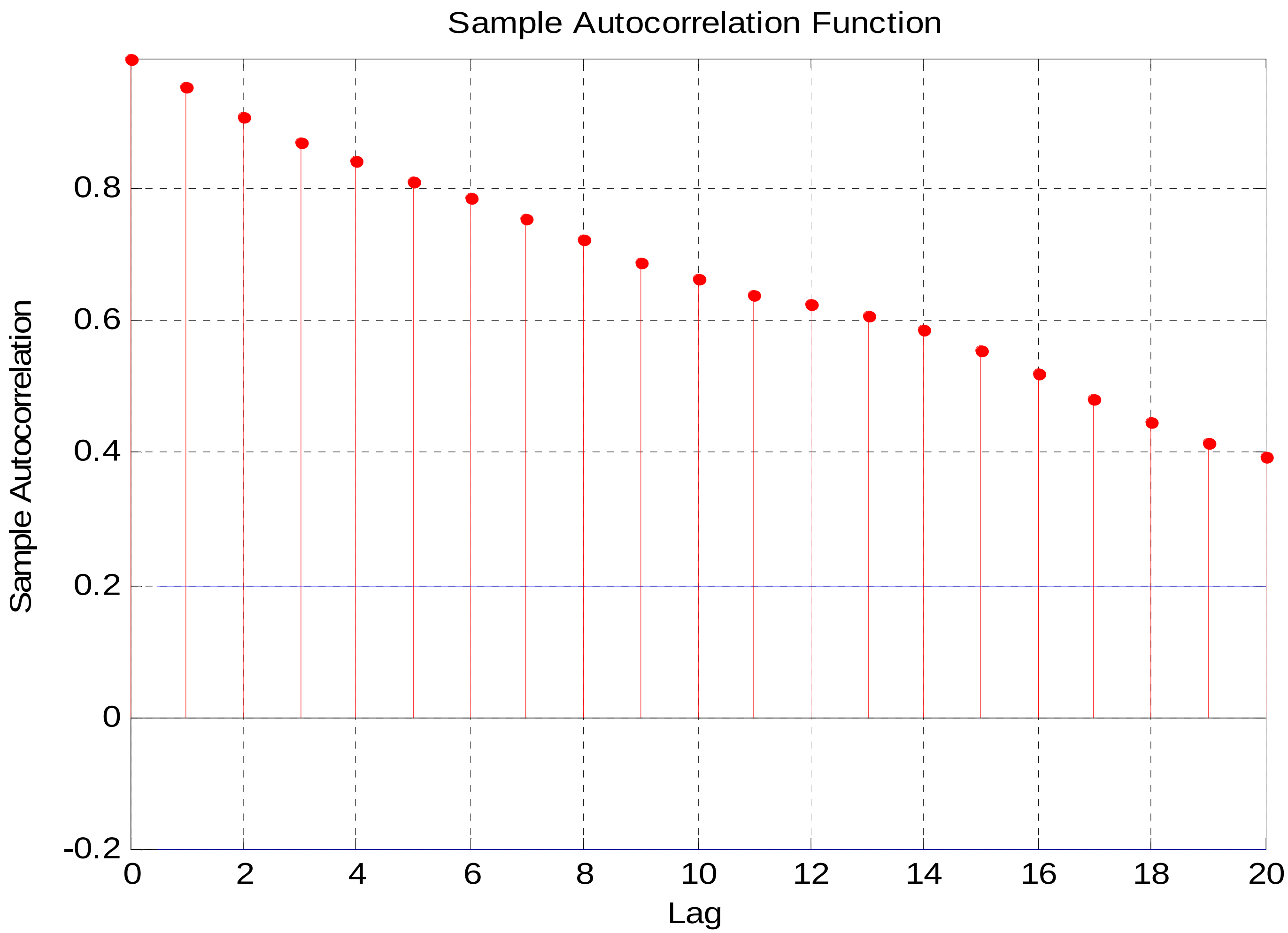
- Iterative substitution results in

$$y_t = \mu t + y_0 + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_1$$

there exists a time trend. If μ is positive, y_t eventually goes to infinity; if μ is negative, y_t would converge to negative infinity.

Random Walk with Positive Drift




Sample Autocorrelation Function
Sample Autocorrelation
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20
Lag

Trend-Stationary Time Series

- A closely related model that exhibits linear trend is the trend-stationary time series

$$y_t = \mu + \beta_1 t + r_t$$

where is r_t a stationary time series with mean zero, for example, an AR(1) process without constant.

- y_t grows linearly in time with rate β_1 and hence can exhibit behavior similar to that of a random walk with drift.

- But there is a big difference
  - Random walk with drift:

    $E[y_t] = y0 + \mu t$,

    $Var[y_t] = t\sigma^2$,

    both are time dependent
  - Trend-stationary:

    $E[y_t] = \mu + \beta_1 t$, which depends on time,

    $Var[y_t] = Var[r_t]$ which is finite and time-invariant.

General Non-Stationary Models

- For a time series y_t, if the first-order difference $\Delta y_t = y_t - y_{t-1}$ follows a stationary ARMA(p, q) process, we call yt an Autoregressive integrated moving-average (ARIMA) model, ARIMA(p, 1, q).
- In finance, price series are commonly believed to be non-stationary, but log return series, $r_t = \ln(P_t) - \ln(P_{t-1})$, is stationary.

Unit-Root Test

- To test whether a time series yt follows a random walk or a random walk with drift, we use the model

$$y_t = \beta_1 y_{t-1} + e_t$$
$$y_t = \beta_0 + \beta_1 y_{t-1} + e_t$$

- Consider the null hypothesis H_0: $\beta_1 = 1$. A convenient test is the t-test of the least-square estimate.

- This is called Dickey-Fuller unit-root test (DF).
- For the first case, the least-square gives

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y^2{}_{t-1}}, \hat{\sigma}_e = \frac{\sum_{t=1}^{T} (y_t - \hat{\beta}_1 y_{t-1})^2}{T-1}$$

- The t-ratio is

$$DF \equiv t - ratio = \frac{\hat{\beta}_1 - 1}{std(\hat{\beta}_1)} = \frac{\sum_{t=1}^{T} y_{t-1} e_t}{\hat{\sigma}_e \sqrt{\sum_{t=1}^{T} y^2{}_{t-1}}}$$

The Augmented DF Test

- Fore many economic time series, ARIMA models might be more appropriate than the simple random walk model.
- The Augmented DF test assumes that the time series follows

$$y_t = c_t + \beta y_{t-1} + \sum_{i=1}^{P-1} \phi_i \Delta y_{t-i} + e_t$$

where c_t is a deterministic function of time, and $\Delta y_j = y_j - y_{j-1}$

- Then, we have ADF test as

$$ADF-test = \frac{\hat{\beta}-1}{std(\hat{\beta})}$$

- The above model can be rewritten as

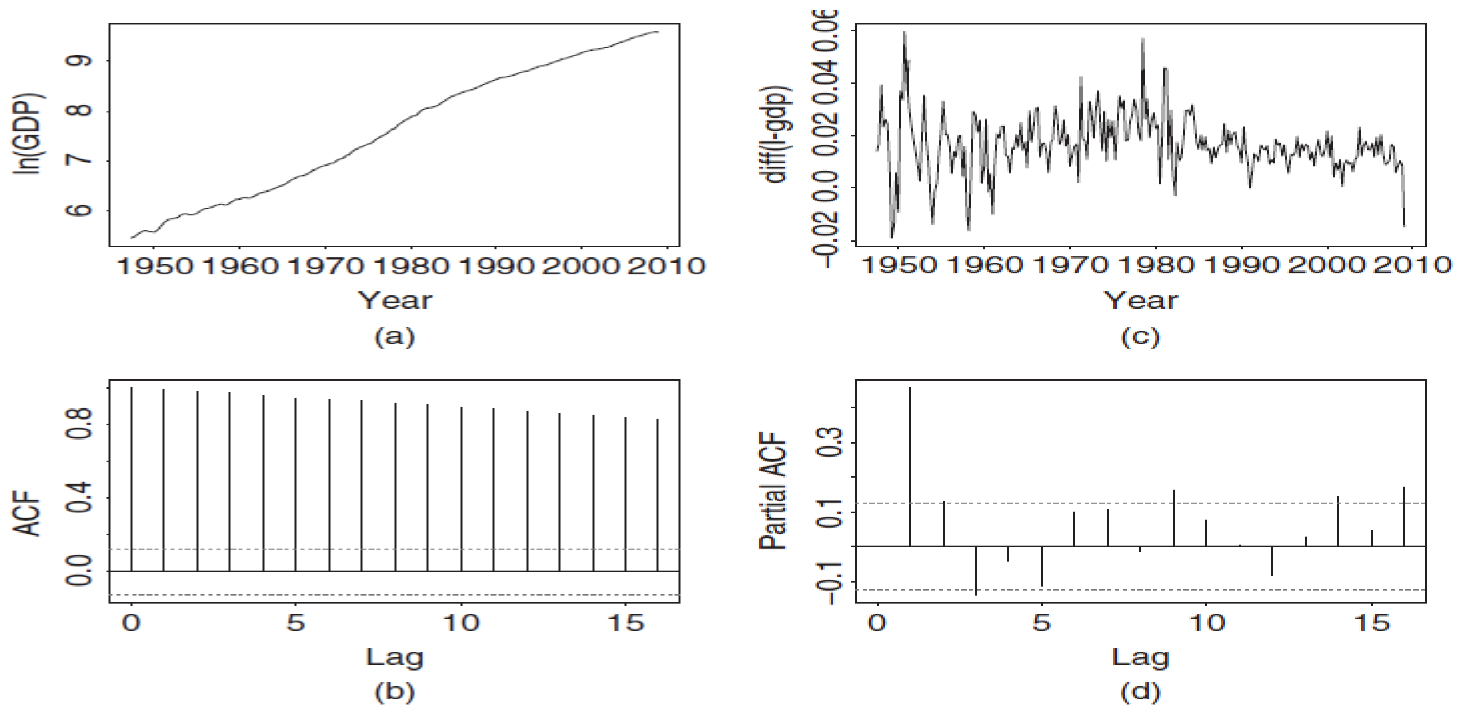
$$\Delta y_t = c_t + \beta_c y_{t-1} + \sum_{i=1}^{P-1} \phi_i \Delta y_{t-i} + e_t$$

- Now the null hypothesis becomes H_0: $\beta_c = 0$

$$ADF-test = \frac{\hat{\beta}_c}{std(\hat{\beta}_c)}$$

Example I

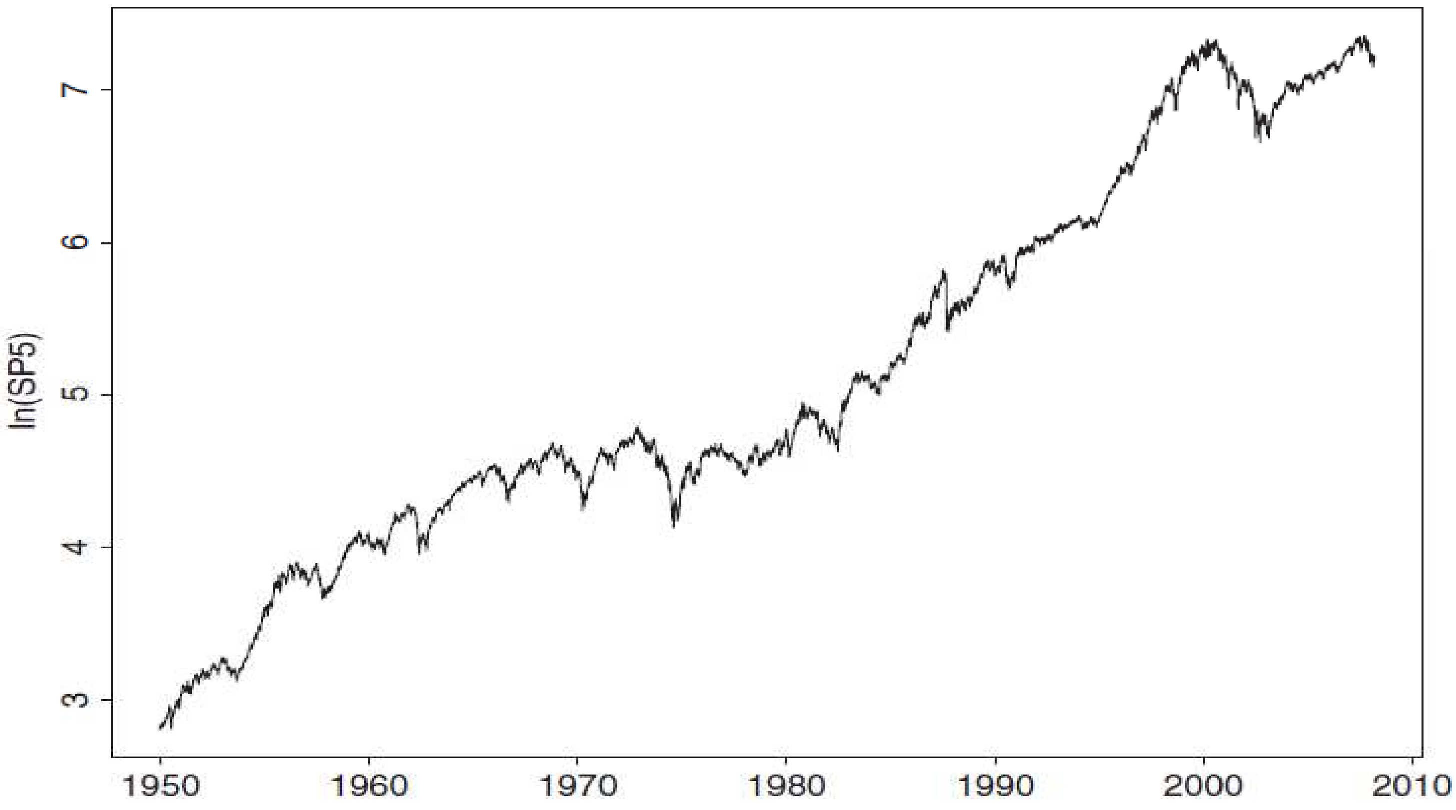
- Log series of US quarterly GDP: 1947.I – 2008.IV



```
Test for Unit Root: Augmented DF Test
Null Hypothesis: there is a unit root
   Type of Test: t-test
 Test Statistic: -1.701
        P-value: 0.4297

Coefficients:
          Value Std. Error t value Pr(>|t|)
   lag1 -0.0008  0.0005    -1.7006  0.0904
   lag2  0.3799  0.0659     5.7637  0.0000
   lag3  0.1883  0.0696     2.7047  0.0074
   ...
   lag10 0.1784  0.0637     2.8023  0.0055
constant 0.0134  0.0045     2.9636  0.0034

Regression Diagnostics:
         R-Squared 0.2877
Adjusted R-Squared 0.2564
Durbin-Watson Stat 1.9940

Residual standard error: 0.009318 on 234 degrees of freedom
```

Example II

- S&P 500 index: Jan 3, 1950 – Apr 16, 2008



```
Test for Unit Root: Augmented DF Test

 Null Hypothesis: there is a unit root
    Type of Test: t-test
  Test Statistic: -1.998
         P-value: 0.602

Coefficients:
           Value Std. Error t value Pr(>|t|)
    lag1 -0.0005  0.0003    -1.9977  0.0458
    lag2  0.0722  0.0083     8.7374  0.0000
    lag3 -0.0386  0.0083    -4.6532  0.0000
    lag4 -0.0071  0.0083    -0.8548  0.3927
    ...
   lag15  0.0133  0.0083     1.6122  0.1069
constant  0.0019  0.0008     2.3907  0.0168
    time  0.0020  0.0011     1.8507  0.0642

Regression Diagnostics:
         R-Squared 0.0081
Adjusted R-Squared 0.0070
Durbin-Watson Stat 1.9995

Residual standard error: 0.008981 on 14643 degrees of freedom
```

Heteroskedastic and Autocorrelated Errors

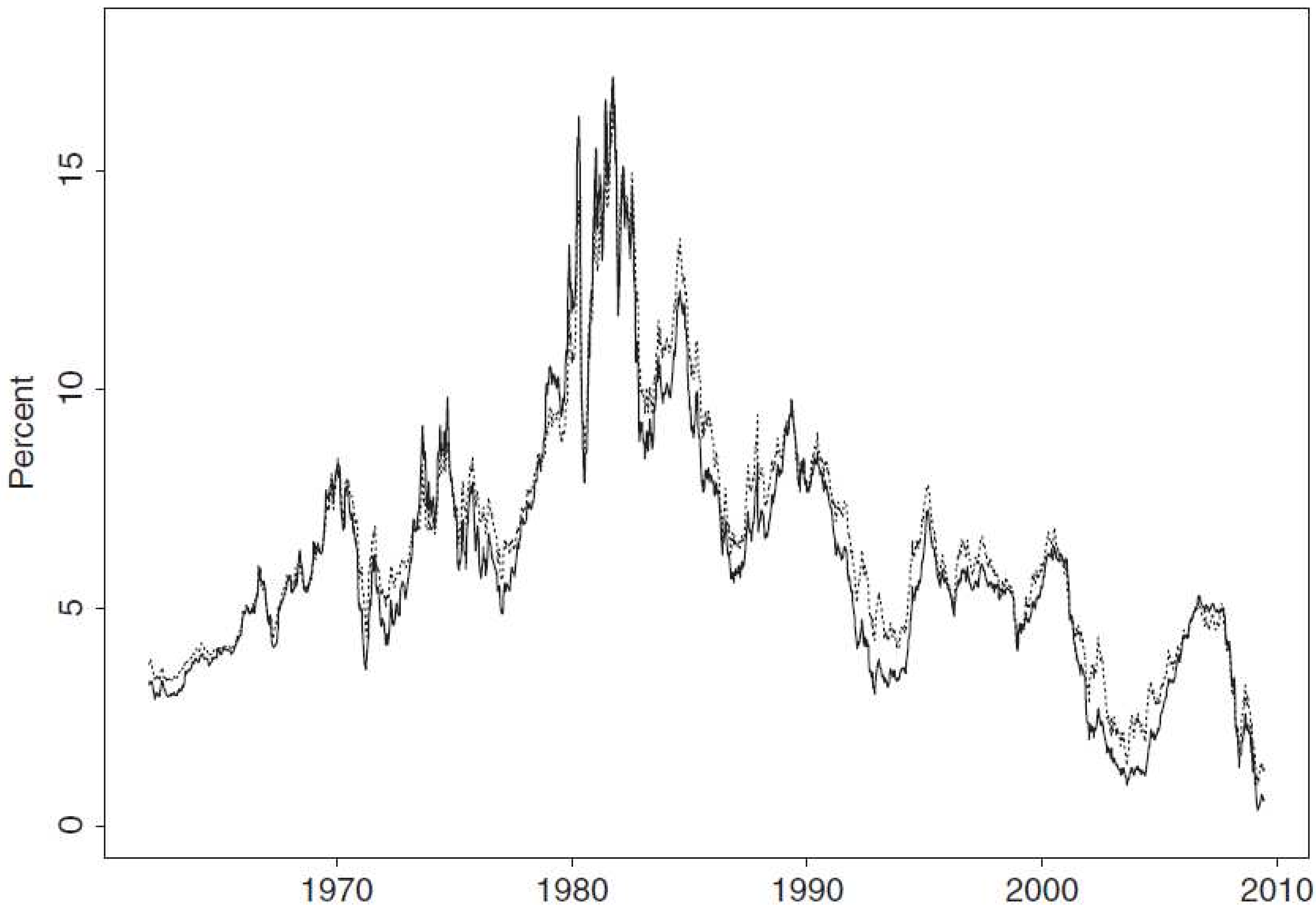
- In many applications, the relationship between two time series is of major interest.
- An obvious example is the market model in finance that relates the excess returns of an individual stock to those of a market index.
- The term structure of interest rates in another example.
- These examples lead naturally to the consideration to a linear regression in the form

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

where y_t and x_t are two time series and e_t denotes the error term.

- The least square method is often used to estimate this model
  - If et is a white noise, then the LS method produces consistent estimates;
- In practice, however, it is common that e_t is serially correlated.

- In this case, we have a regression model with time series errors, and the LS estimates may not be consistent.
- We investigate this issue by considering the relationship between two U.S. weekly interest rate series:
  - r_{1t}, the 1-year maturity treasury rate
  - r_{3t}, the 3-year maturity treasury rate
  - Both series run from Jan 5, 1962 to April 10, 2009, in total 2467 observations.

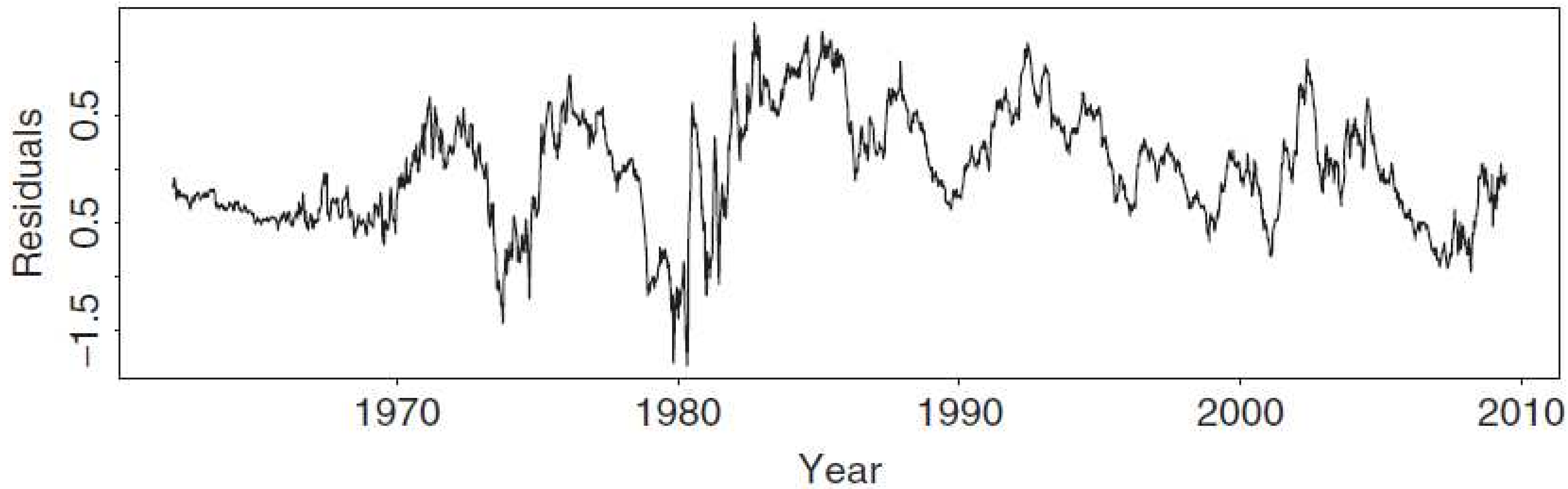

Percent
15
10
5
0
1970
1980
1990
2000
2010

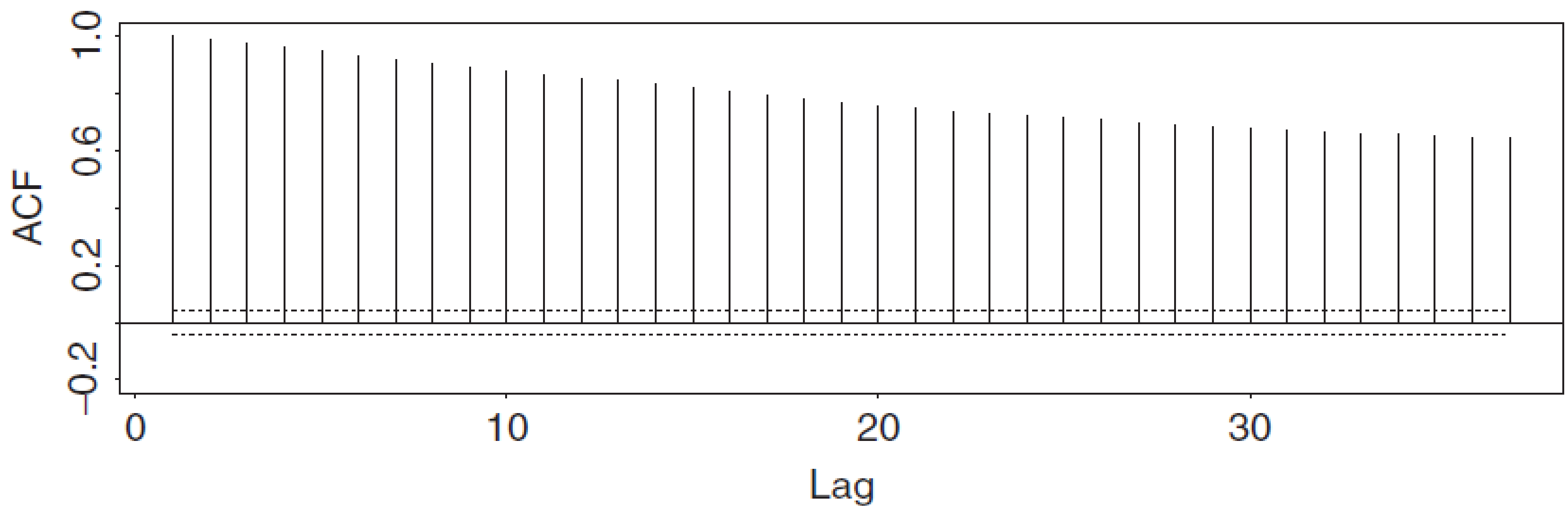
- A naïve way to investigate the relationship between these two series is to use the simple model $r_{3t} = \alpha + \beta r_{1t} + e_t$.
- The estimated model is

$$r_{3t} = 0.832 + 0.930 r_{1t} + e_t, \qquad \hat{\sigma}_e = 0.523$$

where the standard deviations of the two coefficients are 0.024 and 0.004, respectively.

- However, the residuals indicate that the model is seriously inadequate.

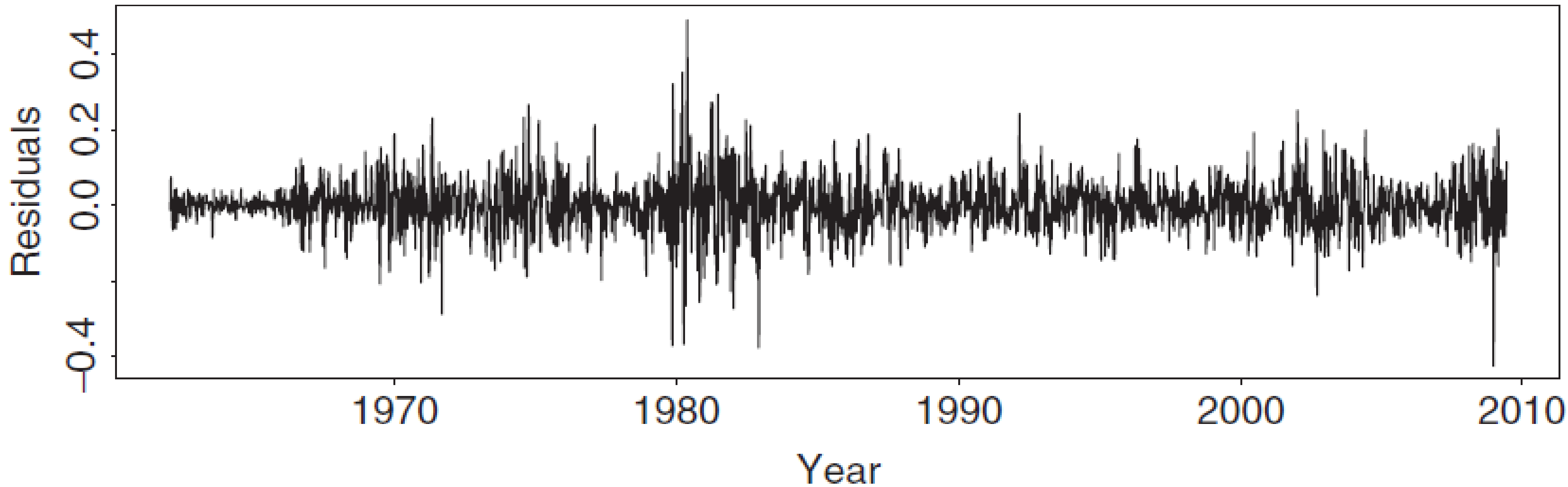

Residuals
0.5
0.5
−1.5
1970
1980
1990
2000
2010
Year
(a)


ACF
1.0
0.6
0.2
−0.2
0
10
20
30
Lag
(b)

- The residual ACF is highly significant and decays slowly, showing a pattern of a unit-root non-stationary time series.
- Instead of using levels, we use changes,
  - $c_{1t} = r_{1t} - r_{1t-1}$,
  - $c_{3t} = r_{3t} - r_{3t-1}$,

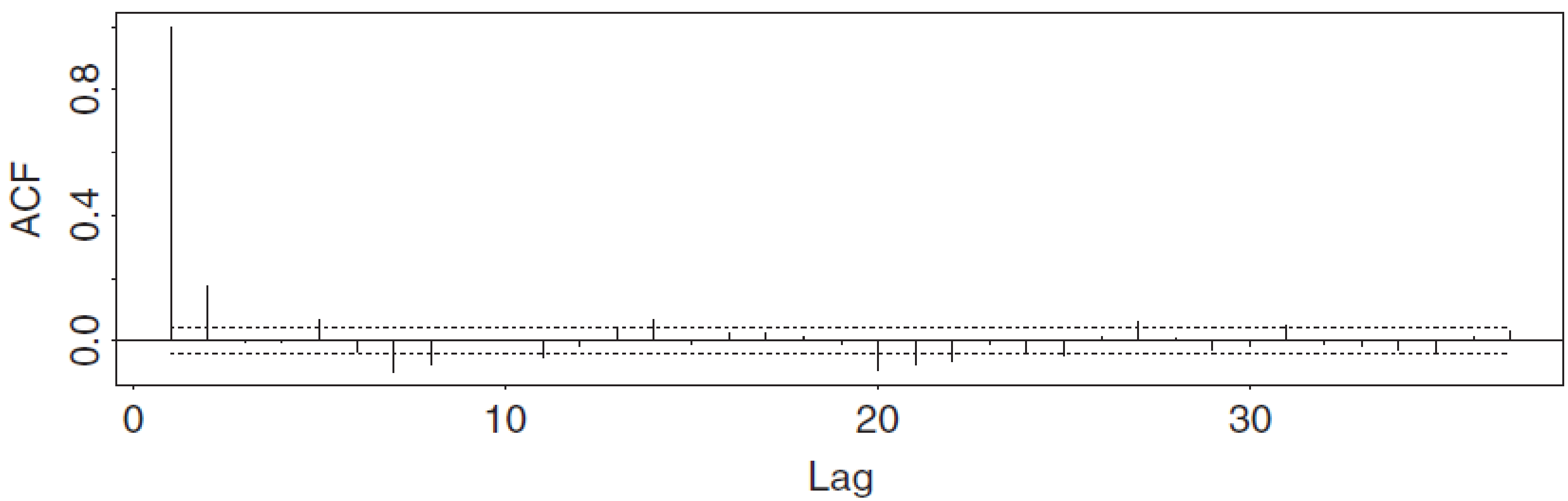
  and consider the linear regression $c_{3t} = \beta c_{1t} + e_t$
- The fitted model is given by $c_{3t} = 0.792 c_{1t} + e_t$

  the standard deviation is 0.007 and R^2 is 82.5%.



Residuals
0.4
0.2
0.0
−0.4
1970
1980
1990
2000
2010
Year

(a)



ACF
0.8
0.4
0.0
0
10
20
30
Lag

(b)

- The residual ACF again shows some significant serial correlations, but magnitudes become much smaller.
- This weak serial dependence in errors can be captured by ARMA models.
- We therefore have a linear regression model with time series errors.
- We specify a MA(1) model for the residual, and modify the model to

$$c_{3t} = \beta c_{1t} + e_t, \qquad e_t = a_t - \theta_1 a_{t-1},$$

- Now at is assumed to be a white noise.
- This model can be easily estimated using MLE. The fitted model is

$$c_{3t} = 0.794 c_{1t} + e_t, \qquad e_t = a_t + 0.1823 a_{t-1}, \qquad \hat{\sigma}_a = 0.0678,$$

```
Coefficients:
         ma1      c1
      0.1823  0.7936
s.e.  0.0196  0.0075

sigma^2 estimated as 0.0046: log likelihood=3136.62,
   aic=-6267.23
```

- And R^2 is about 83.1%

Summary

- 1. Fit a linear regression model and check serial correlations of the residuals.
- 2. If the residuals are unit-root nonstationary, take the first difference of both dependent and independent variables. Go to step 1.
- 3. If the residuals appear to be stationary, identify an ARMA model for the residuals.
- 4. Perform estimation.

Consistent Covariance Estimation

- As discussed before, there may exist situations in which the error term has serial correlations and/or conditional heteroskedasticity.
- But our main objective is to make inference concerning the regression coefficients.
- In situations under which the OLS estimates of the coefficients remain consistent, methods are available to provide consistent estimate of the covariance matrix.

- Two methods are widely used:
  - The heteroskedasticity consistent (HC) estimator (White, 1980)
  - The heteroskedasticity and autocorrelation consistent (HAC) estimator (Neway and West, 1987)
- Consider the regression model

$$y_t = x'_t \beta + e_t$$

- The OLS estimate

$$\hat{\beta} = \left[\sum_{t=1}^{T} x_t x'_t\right]^{-1} \sum_{t=1}^{T} x_t y_t, Cov(\hat{\beta}) = \sigma_e^2 \left[\sum_{t=1}^{T} x_t x'_t\right]^{-1}$$

- In the presence of serial correlations or conditional heteroskedasticity, the above covariance matrix estimate is inconsistent.
- The estimator of White (1980) is given by

$$Cov(\hat{\beta})_{HC} = \left[\sum_{t=1}^{T} x_t x'_t\right]^{-1} \left[\sum_{t=1}^{T} \hat{e}_t x_t x'_t\right] \left[\sum_{t=1}^{T} x_t x'_t\right]^{-1}$$

- The estimator of Neway and West (1987)

$$Cov(\hat{\beta})_{HAC} = \left[\sum_{t=1}^{T} x_t x'_t\right]^{-1} \hat{C}_{HAC} \left[\sum_{t=1}^{T} x_t x'_t\right]^{-1}$$

- And

$$\hat{C}_{HAC} = \sum_{t=1}^{T} \hat{e}_t x_t x'_t + \sum_{j=1}^{l} w_j \sum_{t=j+1}^{T} (x_t e_t e_{t-j} x'_{t-j} + x_{t-j} e_{t-j} e_t x'_t)$$

- Where l is a truncation parameter and w_j is a weight function.
- Neway and West (1987) suggest choosing to be the integer part of $4(T/100)^{2/9}$, and w_j can be $w_j = 1 - \frac{j}{1+l}$
- This estimator essentially uses a non-parametric method.

- A. Simple OLS

```
Coefficients:
                  Value Std. Error   t value  Pr(>|t|)
(Intercept)     -0.0001     0.0014   -0.0757    0.9397
         c1      0.7919     0.0073  107.9063    0.0000

Regression Diagnostics:
         R-Squared 0.8253
Adjusted R-Squared 0.8253
Durbin-Watson Stat 1.6456

Residual Diagnostics:
                 Stat   P-Value
Jarque-Bera 1644.6146    0.0000
  Ljung-Box  230.0477    0.0000
```

- B. White (1980)

```
Coefficients:
                Value Std. Error  t value  Pr(>|t|)
(Intercept)   -0.0001    0.0014   -0.0757   0.9396
         c1    0.7919    0.0163   48.4405   0.0000
```

- C. Neway and West (1987)

```
Coefficients:
                Value Std. Error  t value  Pr(>|t|)
(Intercept)   -0.0001   0.0016    -0.0678   0.9459
         c1    0.7919   0.0198    39.9223   0.0000
```