TIME SERIES ANALYSIS LECTURE 1

datascience@berkeley

Introduction to Time Series Analysis

How Is Univariate Time Series Analysis (TSA) Different From Classical Linear Regression Analysis?

- **Focus:**
  - In univariate TSA, we focus on one series.
    - Example: the monthly unemployment rates in the United States between 1950 and 2014.
  - In regression analysis, we focus on using a set of independent variables to explain the dependent variable. Put differently, we are trying to characterize the conditional mean function of the dependent variable, conditional on the independent variables.
    - Example: Student and family characteristics are used to "explain" student test scores.
- **Dependency among observations**
  - In univariate TSA, the observations are very likely to be dependent of each other. In fact, it is the dependency structure of the series that we want to characterize.
  - In CLR, the observations within each of the variables are assumed to be uncorrelated with each other. (Remember the random sampling assumption (i.e., iid) when we studied the Classical Linear Regression models?)
- **Data:**
  - In univariate TSA, the data consist only of **T** observations of the same variable; the dataset takes the form of a **Tx1** vector.
  - In classical linear regression analysis, the data consist of **n** observations of the dependent variable and a set of **k** independent variables, each of which also consists of **n** observations. Therefore, the dataset takes the form of a **n x k** matrix.

Objectives and Applications of Time Series Analysis

Time series analysis has many applications:

1. Capture the key pattern observed in the data, such as the long-term trend in climate, global average surface temperature in the last 120 years, or an annual growth rate in national health care expenditure in the United States.
2. Predict future values of a variable of interest, such as house price index, consumer spending, population size, sales, and exchange rates (e.g., USD/EUR).
3. Separate (or filter) "signals" from "noise," such as removing the seasonal component when projecting the long-term trend of red wine sales.
4. Test hypotheses using historical data, such as the hypothesis related to global warning.
5. Simulation: The water level of a reservoir depends heavily on daily water input to the system. If the input is modeled using a time series model, once the model is estimated, it can be used to "simulate" a large number of independent sequences of daily inputs. Combining the size and mode of operation of a particular reservoir with the simulated inputs, one can determine under which input conditions and at which timing the reservoir will run out of water.

Time Series Analysis Plays an Important Role in Forecasting

Time series analysis plays an important role in forecasting :

1. Governments forecast tax revenue, retail sales, unemployment claims, etc.
2. Companies forecast sales, consumers' demand, expenditure, and labor cost for strategic planning.
3. National school enrollments projection, which can be used for facilities usage and faculty recruitment planning.
4. Natural gas suppliers use TSA to forecast demand to determine the number of orders to place from the offshore fields.
5. Airlines forecast future capacity for fleet-expansion decision.
6. Retail stores, such as Walmart, Target, and Home Depot, forecast demand during holiday seasons to plan for inventory level and staffing.

Example 1: Government Budget and Key Economic Indictor Projections

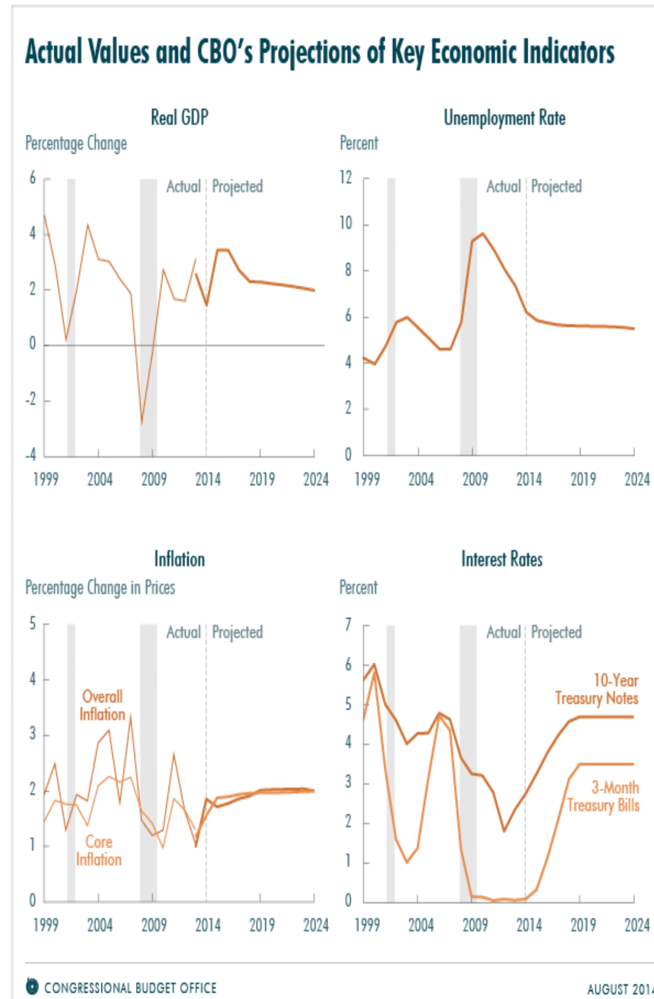
Deficits Projected in CBO's Baseline

	Actual, 2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Total 2015-2019	Total 2015-2024
Revenues	2,775	3,006	3,281	3,423	3,605	3,748	3,908	4,083	4,257	4,446	4,644	4,850	17,965	40,243
Outlays	3,455	3,512	3,750	3,979	4,135	4,308	4,569	4,820	5,076	5,391	5,601	5,810	20,741	47,439
Total Deficit	**-680**	**-506**	**-469**	**-556**	**-530**	**-560**	**-661**	**-737**	**-820**	**-946**	**-957**	**-960**	**-2,777**	**-7,196**
Net Interest	221	231	251	287	340	412	492	566	627	687	746	799	1,782	5,207
Primary Deficit[a]	-459	-275	-218	-269	-190	-148	-169	-170	-193	-259	-211	-161	-995	-1,989
Memorandum (As a percentage of GDP):														
Total Deficit	-4.1	-2.9	-2.6	-2.9	-2.7	-2.7	-3.0	-3.3	-3.5	-3.8	-3.7	-3.6	-2.8	-3.2
Primary Deficit[a]	-2.8	-1.6	-1.2	-1.4	-1.0	-0.7	-0.8	-0.8	-0.8	-1.1	-0.8	-0.6	-1.0	-0.9
Debt Held by the Public at the End of the Year	72.0	74.4	74.0	73.6	73.0	72.8	73.1	73.6	74.3	75.4	76.4	77.2	n.a.	n.a.

Source: Congressional Budget Office.

Note: GDP = gross domestic product; n.a. = not applicable.

a. Excludes net interest.



Source: https://www.cbo.gov/publication/45653

Example 2: Companies Forecast Sales

Nestle forecasts 2015 sales near low end of long-term target

http://www.bloomberg.com/news/articles/2015-02-19/nestle-reports-slowest-annual-sales-growth-in-five-years

---

Walmart reported decent holiday quarter results but its forecasts suggest it knows it has to raise its game a few notches.
http://fortune.com/2015/02/19/walmarts-holiday-results/

Caterpillar warns of 2015 sales hit from falling oil price … cut its 2015 profit outlook and warned the plunge in oil prices would hurt its energy equipment business.
http://www.reuters.com/article/2015/01/27/us-caterpillar-results-idUSKBN0L01E420150127

Joseph Hinrichs, Ford Motor Co.'s president of the Americas, said today the company is predicting industrywide U.S. auto sales will top 17 million next year, the most since 2001.
http://www.bloomberg.com/news/articles/2014-09-25/ford-s-hinrichs-sees-auto-sales-at-14-year-high-in-2015

Example 3: NCES Projections of Education Statistics

- National Center for Education Statistics makes projections for enrollment, graduates, teachers, and expenditures in both public and private schools.
- Multiple methods, such as single and double exponential smoothing and linear regressions, were used.

Source: http://nces.ed.gov/pubs2013/2013008.pdf

An excerpt from NCSE Projections of Education Statistics to 2021, pp. 81:

When using single exponential smoothing for a time series, P_t, a smoothed series, $\hat{P}$, is computed recursively by evaluating

$$\hat{P}_t = \propto P_t + (1 - \propto)\hat{P}_{t-1}$$

where $0 < \propto \leq 1$ is the smoothing constant.

By repeated substitution, we can rewrite the equation as

$$\hat{P}_t = \propto \sum_{s=0}^{t-1} (1 - \propto)^s P_{t-s}$$

where time, *s*, goes from the first period in the time series, 0, to time period *t-1*.

The forecasts are constant for all years in the forecast period. The constant equals

$$\hat{P}_{T+k} = \hat{P}_T$$

where T is the last year of actual data and k is the kth year in the forecast period where $k > 0$.

These equations illustrate that the projection is a weighted average based on exponentially decreasing weights. For higher smoothing constants, weights for earlier observations decrease more rapidly than for lower smoothing constants.

For each of the approximately 1,200 single exponential smoothing equations in this edition of *Projections of Education Statistics*, a smoothing constant was individually chosen to minimize the sum of squared forecast errors for that equation. The smoothing constants used to produce the projections in this report ranged from 0.001 to 0.999.

Example 4: CMS Projection on National Health Expenditure

- The Centers for Medicare & Medicaid Services (CMS) produces short-term (11 years) projections of health care spending for categories in the National Health Expenditure Accounts annually.

- Its baseline projection relies on an econometric model estimated using time series data.

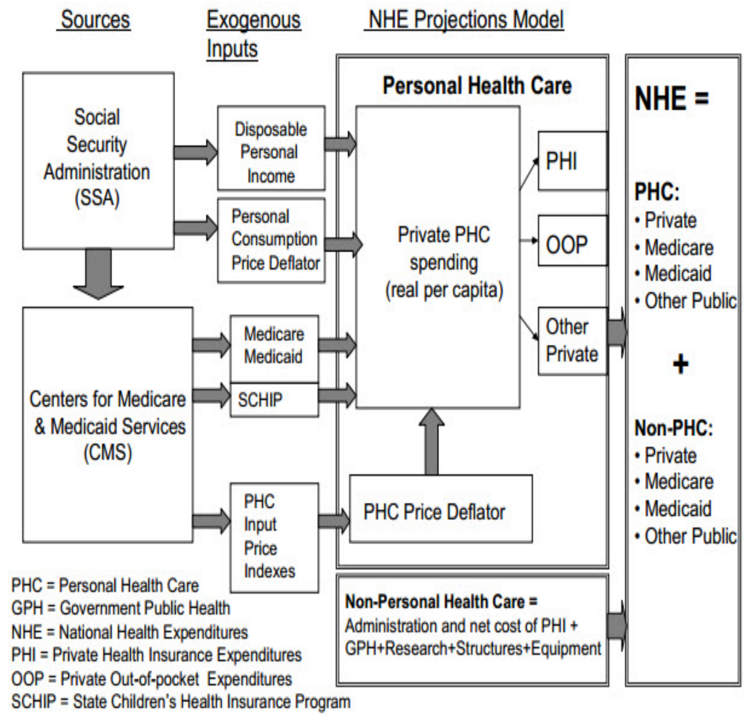
- As is common in many statistical and econometric applications, the empirical model is built on under theories (and in this case, economic theory and health economics literature).

- Importantly, it relies on other inputs (such as macroeconomic forecast) that are themselves projection from other econometric models.

a. Personal health care (PHC) spending

The Baseline NHE Projection Model for health spending by all private sources of funds is an econometric model that is estimated based on time series data from the historical National Health Expenditures. The structure and parameters of the model draw on standard economic theory and the health economics literature. The model is reestimated annually following the release of updated data for the NHEA. The fit and appropriateness of model specifications for individual series are reviewed at this time.

The diagram below provides a schematic view of the aggregate health sector within the Baseline NHE Projections Model and shows the linkages among the data sources, exogenous data, the personal health care (PHC) model, the non-PHC output, and the aggregate baseline NHE projections.



Source: Projections of National Health Expenditures (7/28/2011)

Basic Terminology of Time Series Analysis

An Introduction to Concepts of Time Series

We will discuss the following concepts used in time series analysis:

- Stochastic process
  - Continuous-time stochastic process
  - Discrete-time stochastic process
- Time series

- Note: As in any statistical modeling, we will have to capture only the essential features of the problem being studied. Specifically, we will put some structures around the process by assuming that it follow certain probabilistic laws and not necessarily assuming that it follows a deterministic model.

Notion of Time Series

- A **time series** is a set of observations generated sequentially in time.
- If the set is continuous, such as $[0, f\pi]$ where f is some frequency, the time series is said to be a *continuous* time series.
- If the set is discrete, the time series is said to be a *discrete* time series.
- In theory, a time series begins in the infinite past and continues into the infinite future. This notion plays an important role when deriving important properties of the time series models we will study in this course.
- In practice, we often encounter time series with a specific origin and end in a finite number of time periods.
- With a discrete time series, the time index can be denoted as $t_1, t_2, \cdots, t_N, \cdots$ and the observations of the series can be denoted as $z(t_1), z(t_2), \cdots, z(t_N), \cdots$.
- For time series observed in equidistant for N successive periods, we can simply express the series as $z_1, z_2, \cdots, z_N$.

Notion of Stochastic Processes (1)

- A **stochastic process** is a statistical phenomenon that evolves in time according to some probability laws.
- We will use the terms **stochastic processes** and **processes** interchangeably.
- A time series is a particular realization of an underlying stochastic process (or an underlying probability mechanism). Statisticians and econometricians often call this probability mechanism a "data generating process".
- A discrete time stochastic process is a sequence of random variables $\{\cdots, z_{-1}, z_0, z_1, \cdots\}$
- A finite subset of this realization, $\{z_1, z_2, \cdots, z_N\}$ of the process is called a **sample path**, which can be considered a collection of observations.

Notion of Stochastic Processes (2)

– This marks one of the biggest differences with the cross-section analysis in the first half of the course. A cross-section of observations is consider many "realizations" or "draws" from the underlying probability distribution. In practice, many time series (such as the unemployment rate in the U.S.) is considered only "one draw" or "one realization" in the context of time series analysis.

Steps to Analyze Time Series Data

General Approach That We Will Use to Learn Time Series Modeling

1. Learn about the mathematical formulation of a model or an important concept (such as autocorrelation function).
2. Derive some of the most important properties of the model.
3. Simulate (using R or Python) a number of realizations from the model.
4. Examine the empirical properties exhibited in the simulated realizations.
5. Apply the model or concept to a real-world datasets.

General Steps to Analyze a Time Series

1. Based on the interaction of theory, subject matter expertise, or even practical experience, consider a useful class of models.
2. Collect and cleanse the data.
3. Conduct with exploratory time series data analysis (ETSDA) by plotting the series using various graphical techniques and examine both the main patterns and atypical observations in the graphs, after collecting and “cleaning” the data :
   - Trend
   - The fluctuation around a trend
   - Sharp change in behavior (i.e., structural change or jumps)
   - Outliers
4. Examine and (statistically) test whether the series is stationary.

General Steps to Analyze a Time Series (2)

5. If the series is not stationary, transform the series to a stationary series (if a stationary time series model will be used), because the time series models covered in this course apply only to stationary or integrated times series. Common transformation techniques include trend removal (i.e., detrending), seasonality removal, logarithmic, and difference transformation.
6. Model the transformed series using a stationary or integrated time series model.
7. Examine the validity of the model's underlying assumptions.
   - This is an important step, because if the model's underlying assumptions are not satisfied, one should not proceed to conducting statistical inference and forecasting.
8. Among the valid models, choose the one that perform "best" according to some prespecified metrics or business needs (if the model is for business use).
9. Once a (statistically) valid model is chosen, conduct forecasting (or other statistical inference).

Common Empirical Time Series Patterns

Pattern 1: Trend and Fluctuation Around the Trend

Example: Passenger bookings of an airline

Data:

- The number of passengers traveling on the airline increased with time (i.e., an increasing trend).
- The bookings behaved cyclically with some fixed time periods.
- The fluctuation increased over time.

Pattern to observe:

- Upward trend—a systematic dynamic in a time series that is not periodic.
- Seasonal effect—the seasonal effect is apparent in the box plot as well.

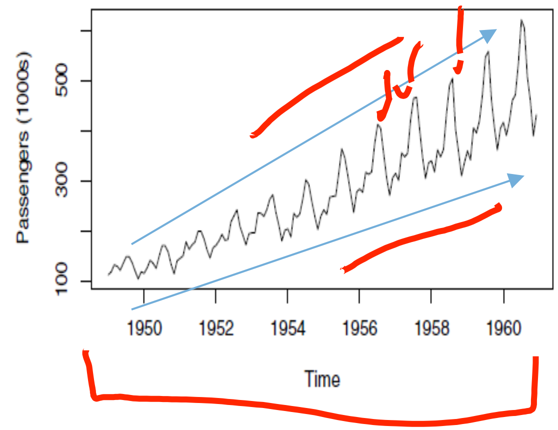
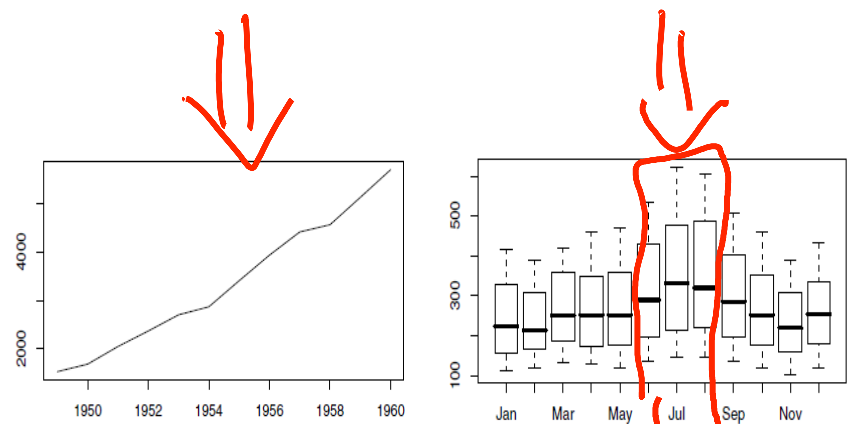


Fig. 1.1. International air passenger bookings in the United States for the period 1949–1960.



(a) Aggregated annual series

(b) Boxplot of seasonal values

Pattern 2: Change in Structure

Example: Annual average global temperature change between 1880 and 2009

Observations:

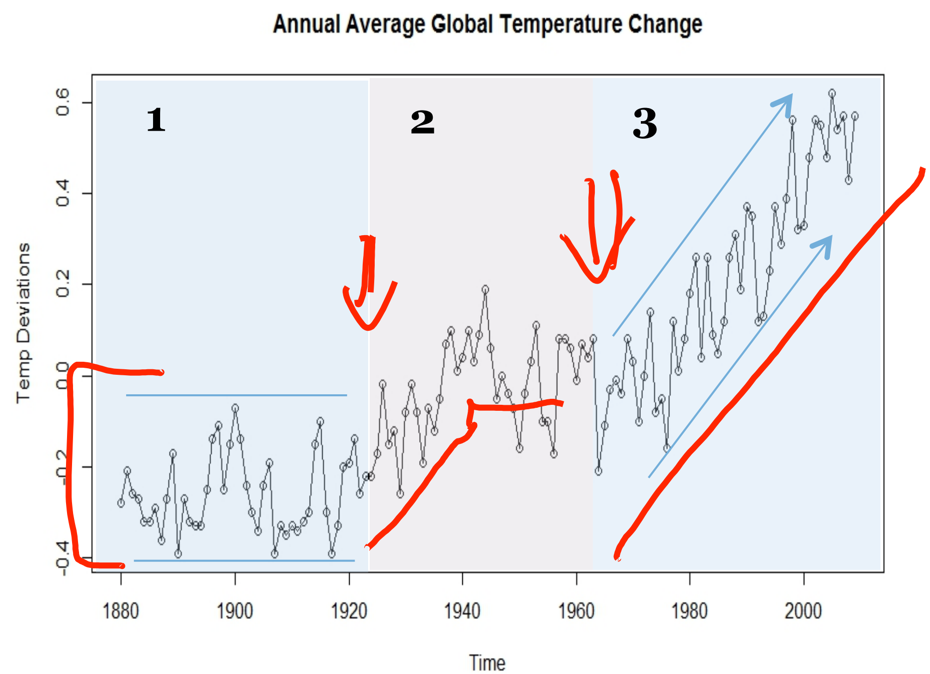
i. The temperate change was range bounded between 0 and -0.4% from 1880 to 1920.

ii. The range of temperate change started to increased to fluctuated between -0.2% to 0.2% from 1920 to 1960.

iii. The temperate change continued to trend up since 1960, sparking the debate on global warming.

Pattern to observe:

- Possible structural change: A consistent upward trend with variation around the trend after a long period of range-bounded fluctuation (i.e., from 1880–1920)



Pattern 3: Variation Around a Stable Mean

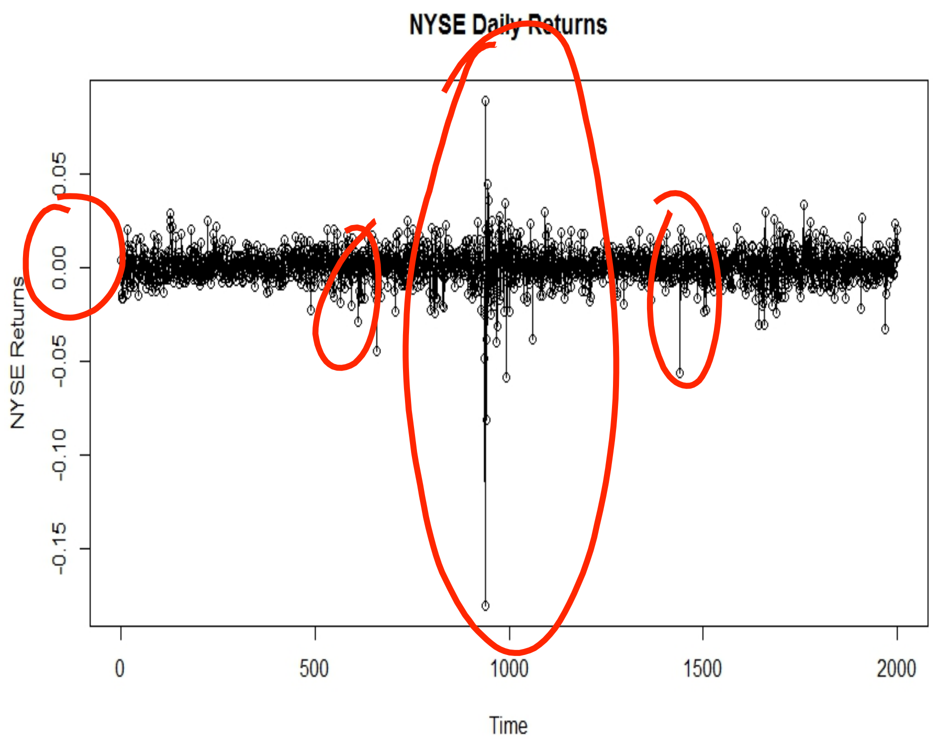
Example: New York Stock Exchange (NYSE) daily returns

Observations:

- The NYSE daily returns (in the observation window) fluctuated around zero.
- There were spikes in the fluctuations from time to time, so-called volatility clustering.

Pattern to observed:

- Stable mean—the mean of the series looks stable.
- Time-varying volatility—the volatility of the series varies over time.



Pattern 4: Periodicity

Example: Functional magnetic resonance imaging (fMRI) in different locations of the brain

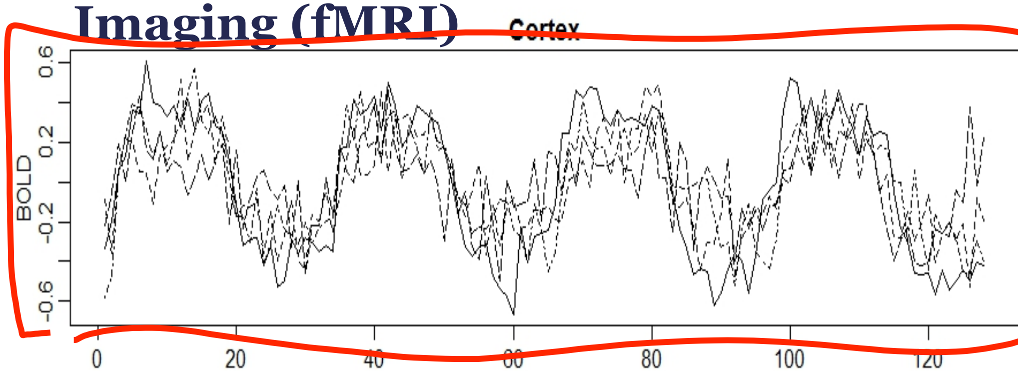
Observations:

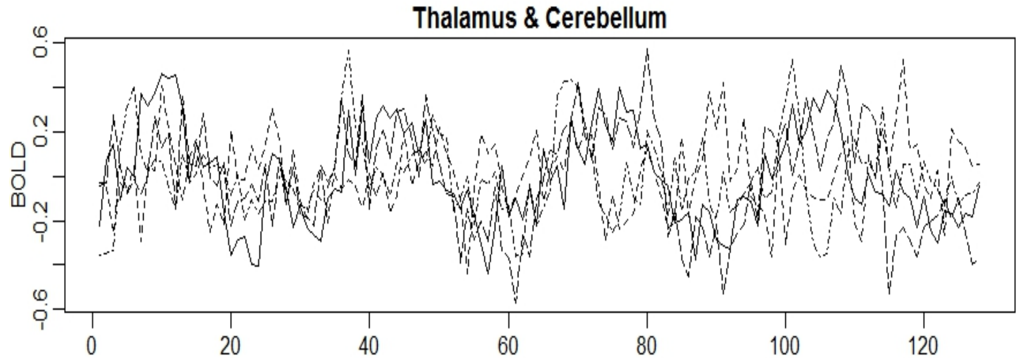
- Periodicity appear strongly in the motor cortex series.
- Periodicity appear much weaker in the thalamus and cerebellum series.

Pattern to observed:

- Periodicity—cycles occurred in fixed frequency.

Functional Magnetic Resonance Imaging (fMRI)



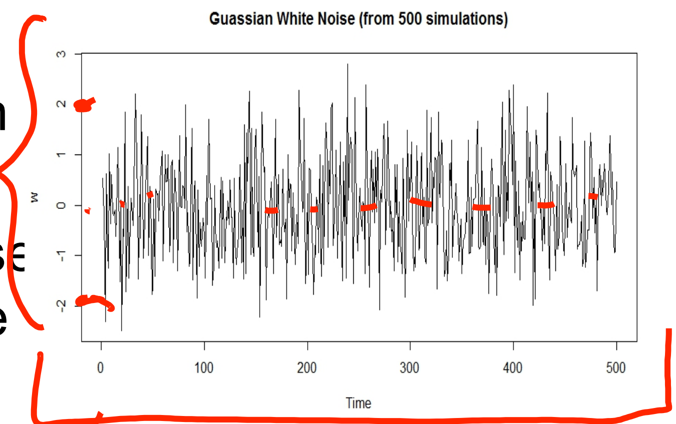


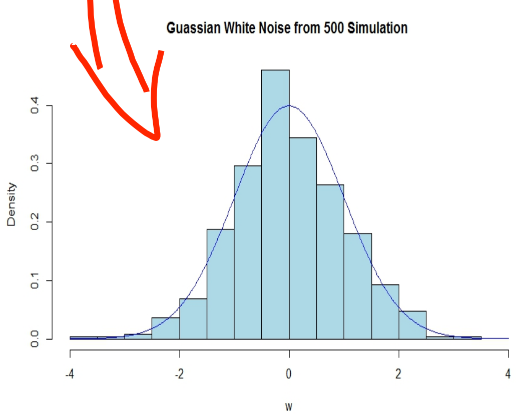
Simple Time Series Model Examples

Model 1: White Noise

White Noise:

- The most general form of a **white noise** is a collection of uncorrelated random variables with mean **0** and variance
- To put more structure on the process, we impose assumptions on the underlying distribution. One popular set of assumptions is that each of the random variables is independently and identically distributed (**iid**) with a normal distribution with mean 0 and variance 1, leading to Gaussian WN.





Gaussian White Noise:

- Note that as the underlying distribution is a normal distribution, specifying the mean and variance completely characterize the distribution.
- The graph shows a simulated Gaussian white noise process.

Model 1: White Noise (2)

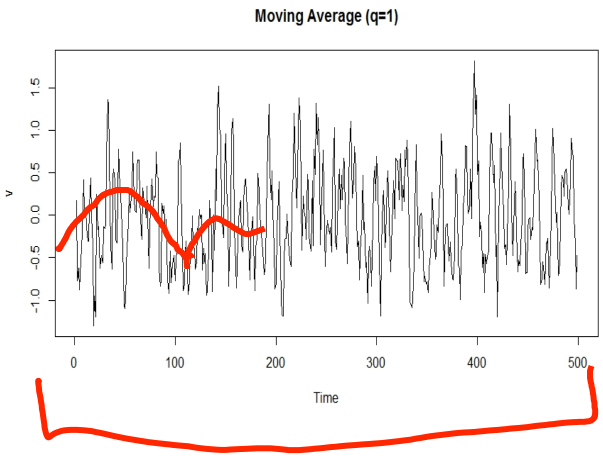
- White noise forms the most fundamental component in time series modeling and assumption testing.
- All of the time series models introduced in this course will have white noise as its stochastic components.
- As an example, a deterministic dynamic model, such as a population growth model, can be turned into a stochastic dynamic model using the addition of the white noise components.
- At this point, we will focus only on the dynamic pattern of the white noise using a time series plot and the distribution using techniques such as histogram or nonparametric kernel density.
- It is very important to note that representing a series using a histogram or density takes the time element away. It does not show the dependence of the series. And because in time series analysis, the main object of study in dependency among observations, representing a series using a histogram or density alone does not capture all of the important patterns embedded in the series.
- In later lectures, we will use other techniques to capture the dependency of a series.

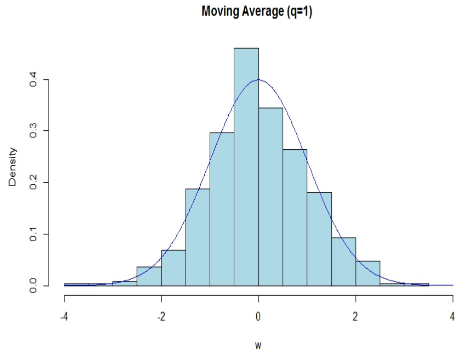
Model 2: Symmetric, Equal-Weight Moving Average Model

A **symmetric, equal-weight moving average model** takes the following general form:

The graph shows a series simulated using a centered-moving average model with

- Moving average is a common way to "smooth" (out the volatility of) a series where are Gaussian white noises introduced above.
- Moving average can be used to generate dependency among observations.

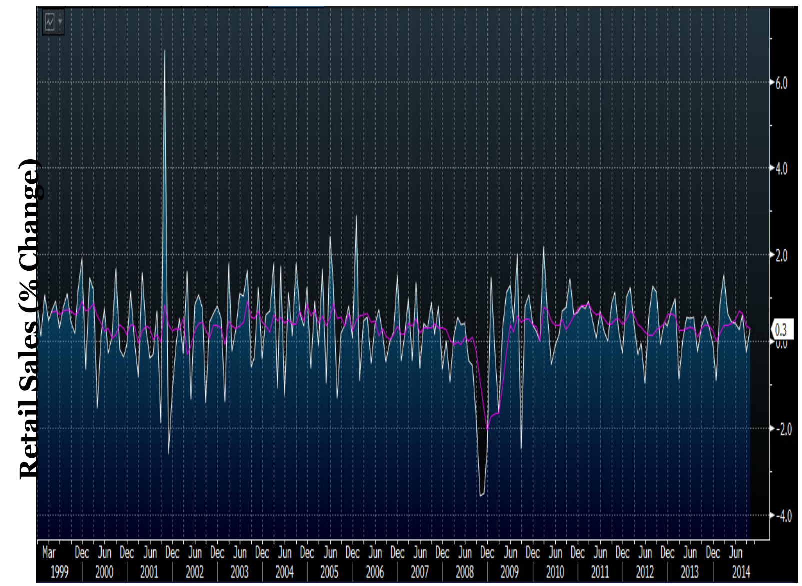




Model 2: Moving Averages: An Application

- The graph shows both the historical U.S. retail sales, measured in total monthly % change, from January 1999 to November 2014, and its six-month moving averages.
- Centered moving average is an example of a *smoothing* technique applied to the past values of the series.
- By "smoothing" out some of the variations and even seasonal effect, the underlying trend may emerge.
- In the case of adjusting for seasonality (i.e., "average-out" the seasonal effects), the length of the moving average is chosen so as to eliminate the seasonal pattern. (More on this point later.)

Retail Sales (Total Monthly % Change)



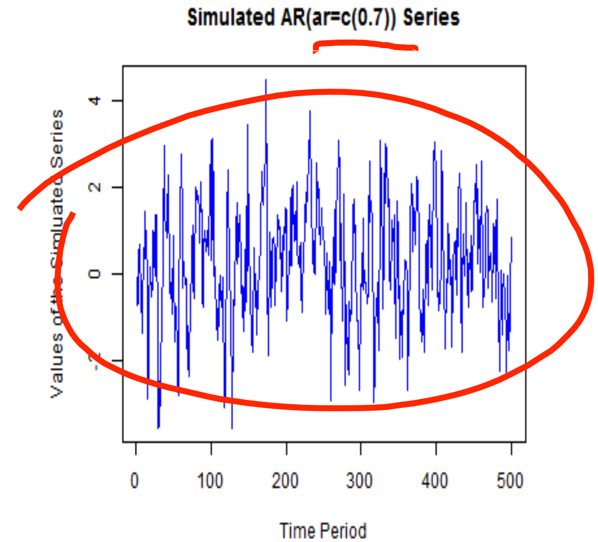
Source: Bloomberg

Model 3: Autoregressive Models

An autoregressive model with order p=2:

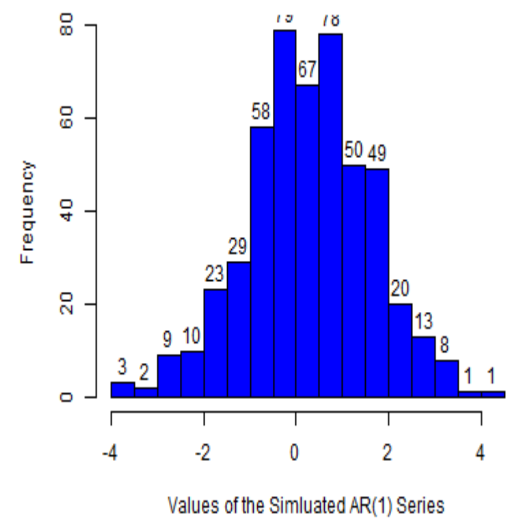
where is white noise (WN) and the are the coefficients of the model. We generally consider the series has its mean subtracted. Formally, we would write $\{x_t\}$.

$$\tilde{x}_t = x_t - \mu$$



The graph shows a simulated time series using the following zero-mean AR(1) model:

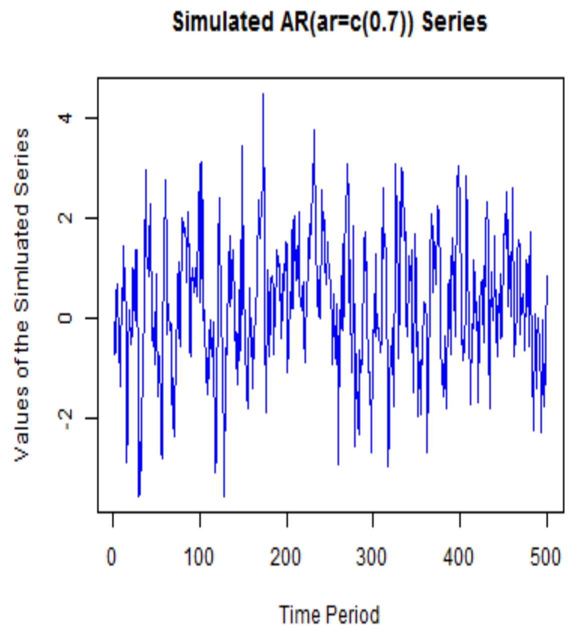
where

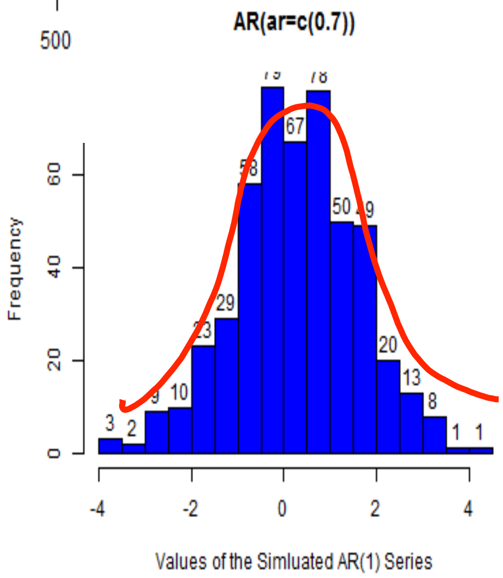


Note the simulated AR(a) series evolve around a constant level (mean = 0).

Model 3: Autoregressive Models (2)

The distribution of this simulated series looks fairly symmetric



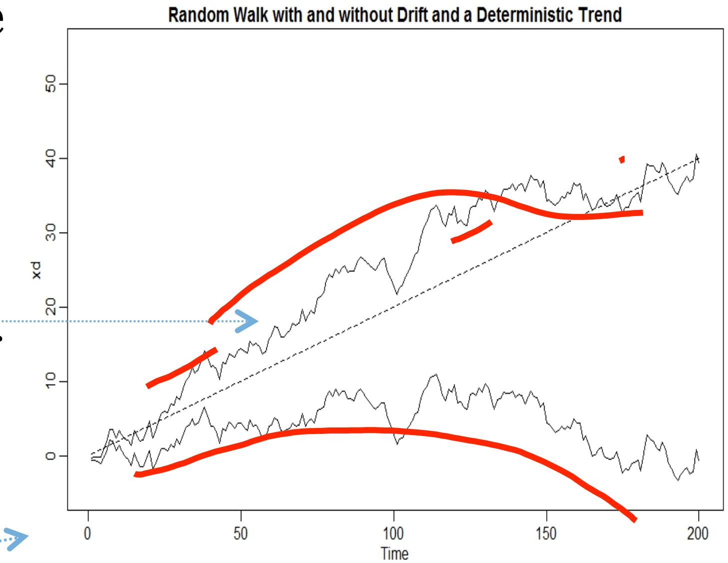


Model 4: Random Walk and Deterministic Trend

A form of **random walk (with drift)** can be expressed as + where

The graph shows three models:

1. A deterministic trend (i.e., is a deterministic function of the index . That is, knowing the value of will identify the value of



2. Random walk without drift

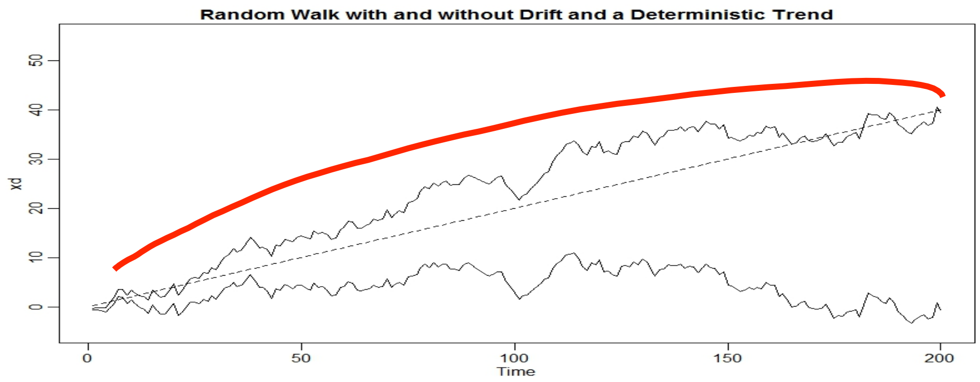
where is

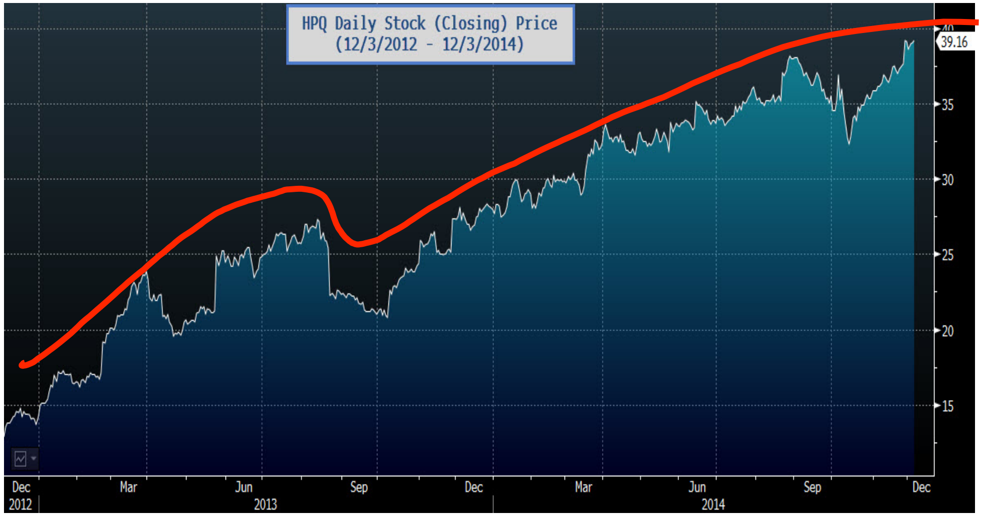
3. Random walk with drift

3.

Model 4: Random Walk and Deterministic Trend (2)

- Note the persistence of the simulated random walk with drift series (top curve in the top graph)–it gradually "drifts" upward and does not show any signs of reverting back to a the starting observed level (or a constant level).
- Observe the similarity (in the general pattern) between the simulated random walk with drift series and the Hewlett-Packard Company (HPQ) daily closing price series (from 12/3/2012 to 12/3/2014).





Source: Bloomberg

Notion and Measure of Dependency

General Description of a Stochastic Process

A complete description of a time series as a collection of *k* random variables requires a joint distribution function:

$$F(c_1, c_2, \ldots, c_n) = P(x_{t_1} \le c_1, x_{t_2} \le c_2, \ldots, x_{t_n} \le c_n)$$

- Characterizing the joint distribution function in its most general form, such as the one written in the above form, is very challenging, if not impossible.
  - A single realization of a time series does not offer enough information to describe the entire underlying joint distribution function from which the realizations are generated.
  - Doing so requires many more assumptions about the joint distribution are imposed.
- In the context of time series analysis, one of the most important probabilistic features is **the dependency structure** embedded in the joint distribution.
- We will
  - discuss how dependency is defined, measured, and estimated,
  - illustrate these concepts using the foundational time series models, and
  - the mean and variance functions of a time series.

Mean Function and Stationarity in the Mean

Before discussing the dependency structure of a time series, let's define the mean and variance functions of a stochastic process. The **mean function** is defined as

$$\boxed{\mu_x(t) = E(x_t) = \int_{-\infty}^{+\infty} x_t f_t(x_t), \mathrm{d}x_t}$$

where $E(.)$ denotes the expected value operator, and the expectation is taken over the *ensemble*, which consists of the entire population, of all possible time series that might have been produced by the underlying time series data generating process (DGP).

Note that the mean is a function of the index . If the function is constant (i.e., it does not vary with), then the underlying stochastic process is said to be **stationary in the mean**.

Variance Function and Stationarity in the Variance

The variance function of a time series model that is stationary in the mean is defined as

$$\sigma_x^2(t) = E(x_t - \mu)^t = \int_{-\infty}^{+\infty} (x_t - \mu)^2 f_t(x_t), dx_t$$

where $E(.)$ denotes the expected value operator, and the expectation is also taken over the *ensemble*. Note that the mean μ is a constant not varying with time.

- Note that the variance function is also a function of the index .
- Yet, operation-wise, it is impossible to estimate different variance for different points in time, given only a single realization of the underlying stochastic process (i.e., a single time series).
- To make the concept operational, one has to impose more structure on the underlying stochastic process.
  - One popular assumption is **stationary in variance**, which, combining with stationarity in autocovariance to be introduced in the next slide, produce a large class of models called **stationary time series models**.

Autocovariance and Autocorrelation Functions

Recall from DATASCI w203 that linear (or order) dependency is measured by covariance and correlation. In the context of time series analysis, we speak of the covariance and correlation between different random variables of the same series, and hence the term "autocovariance" and "autocorrelation".

The **autocovariance function (acvf)** is defined as

$$\gamma_x(s,t) = cov(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)] \forall s,t$$

Two natural implications are (1) $\gamma_x(s,t) = \gamma_x(t,s)$ (Exercise: verify it) and (2) $\gamma_x(s,s) = cov(x_s, x_s) = E[(x_s - \mu_s)^2]$

- A correlation of a variable with itself at different times is known as *autocorrelation*.
- If a time series is second-order stationary (i.e. stationary in both mean and variance: $\mu_t = \mu$ and $\sigma_t^2 = \sigma^2$ for all t), then an *autocovariance function* can be expressed as a function only of the time lag *k*:

$$\gamma_k = E[(x_t - \mu)(x_{t+k} - \mu)]$$

- Likewise, the **autocorrelation function (acf)** is defined as When $k = 0$, $\rho_0 = 1$ $\boxed{\rho_k = \frac{\gamma_k}{\sigma^2}}$

Autocovariance and Autocorrelation Functions (2)

$$\gamma_k = E[(x_t - \mu)(x_{t+k} - \mu)]$$

$$\boxed{\rho_k = \frac{\gamma_k}{\sigma^2}}$$

- I want to emphasize the importance of the requirement that the series is second-order stationary before defining the autocovariance and autocorrelation functions.
- As you can see in these functions, the mean and variance are both constant.
- In other words, without a constant mean and variance, these functions are not well-defined!
- In time series, it is the dependence between the values of the series that is important to measure, so at a minimum, we want to estimate autocorrelation with precision.
- It would be difficult to measure the temporal dependence if the dependence structure change at every time point.
- I'd like you to keep this notion in mind because they will have important implications in the empirical work.

Estimation of Autocovariance and Autocorrelation

Using the *moment principles*, the *acvf* and *acf* can be estimated from a time series by their sample equivalents. The sample *acvf* can be estimated using the following formula:

$$\boxed{\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} \left(x_t - \hat{x}\right)\left(x_{t+k} - \hat{x}\right)}$$

Note that the sum is divided by and not .

The sample *acf* is defi $\frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\frac{1}{T}\sum_{t=1}^{T-k}\left(x_t - \bar{x}\right)\left(x_{t+k} - \bar{x}\right)}{\frac{1}{T}\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2}$

In the next section, we will estimate and examine the autocovariance and autocorrelation of the real-world time series data and simulated time series.

Partial Autocorrelation Functions

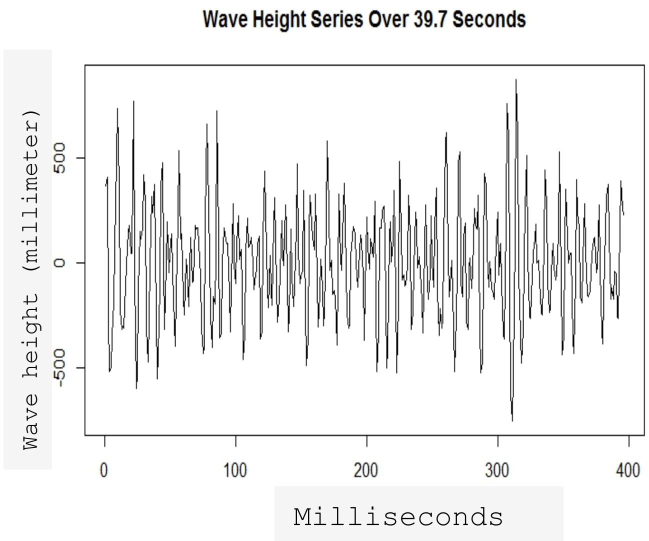
- The concept of **partial autocorrelation** in the context of time series is analogous to the **partial correlation** in the context of multiple linear regressions.
- It is a "conditional" correlation, conditional on other explanatory variables accounted for in a time series model.
- The partial autocorrelation of a process z_t at lag k, ϕ_{kk}, is the (auto)correlation between the variable z_t and z_{t-k} , adjusting for the effects from variables $z_{t-1}, z_{t-2}, \ldots, z_{t-k+1}$
- In other words, it is the coefficient of z_{t-k} in a linear regression of or z_t $z_{t-1}, z_{t-2}, \ldots, z_{t-k}$; it is called **autoregression** because the variable z_t is regressed on its own lagged values.
- The difference between autocorrelation and partial autocorrelation is similar to the difference between the coefficients in simple linear regression (i.e. regression with 1 explanatory variables) and the coefficients in a multiple linear regression.
- Like the autocorrelation, the partial autocorrelation summarizes the dynamics of a process, and as you will see in the next lecture, partial autocorrelation is a power device for identifying the order of an AR(p) model

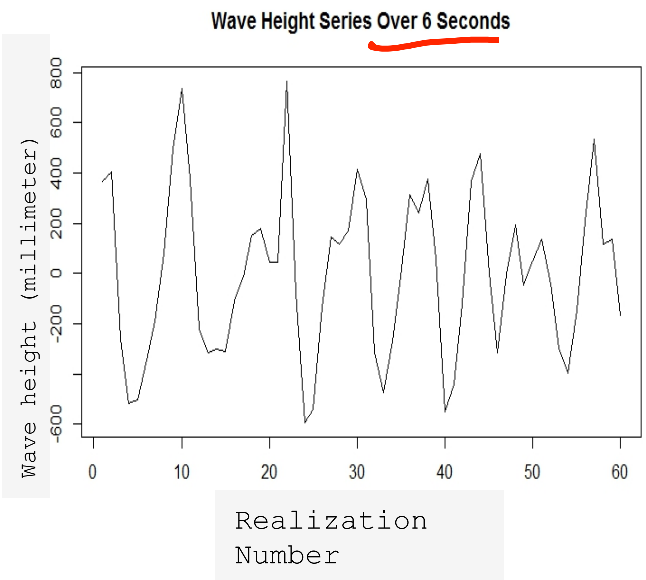
Examining Time Series Correlation—Autocorrelation Function: Example 1

Example: Wave Height

Let's go through a few examples to illustrate the concepts just covered.

- In this example, we use the wave height data provided in the textbook, *Introductory Time Series With R*.
- It is a time series of wave height, measured in millimeters (mm) relative to still water.
- When considering a time series, we need to pay attention to the sampling interval. In this example, the sampling interval is 0.1 second and the total record length is 39.7 seconds, giving almost 400 observations.
- As mentioned, we should always make a time series plot when analyzing a time series.
- The graph at the top displays the entire (sampled) series over the 39.7 seconds of recording.





Example: Wave Height (2)

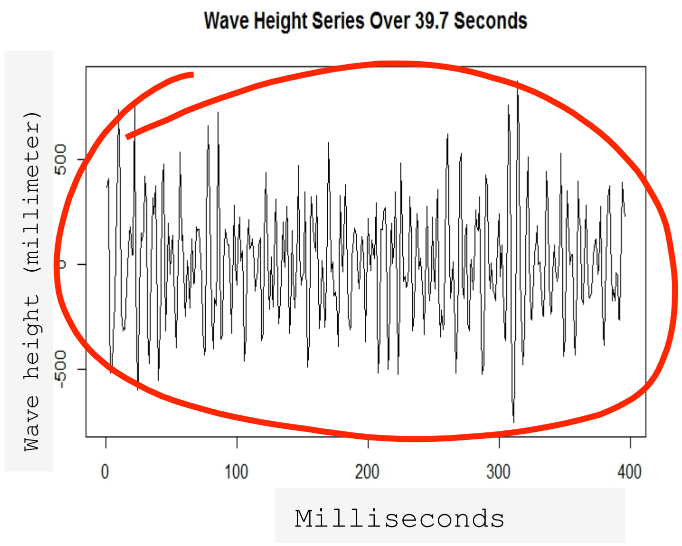
- When graphing a time series (or any other plots for that matter), ensure it includes a title and the axes are well-labeled. In a time series plot (or t-plot) the x-axis is time period (or frequency). Make sure the correct time period or interval is recorded to assist readers in understanding the graphs.

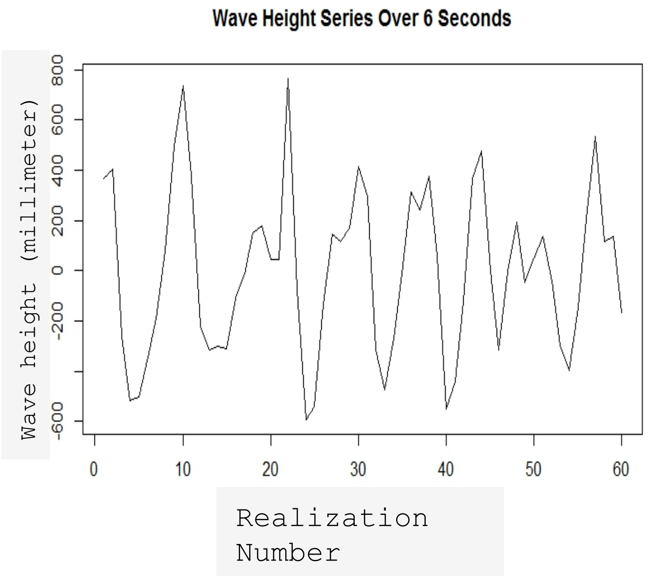
- In this example, the time period is measured in milliseconds and the unit of measurement (of wave height) is millimeters (mm).

- The wave height series does not appear to have any trends or "seasonal" components, so it is reasonable to assume that the series is a realization of a stationary process.

- In other words, a stationary process can be used to model this set of realizations.

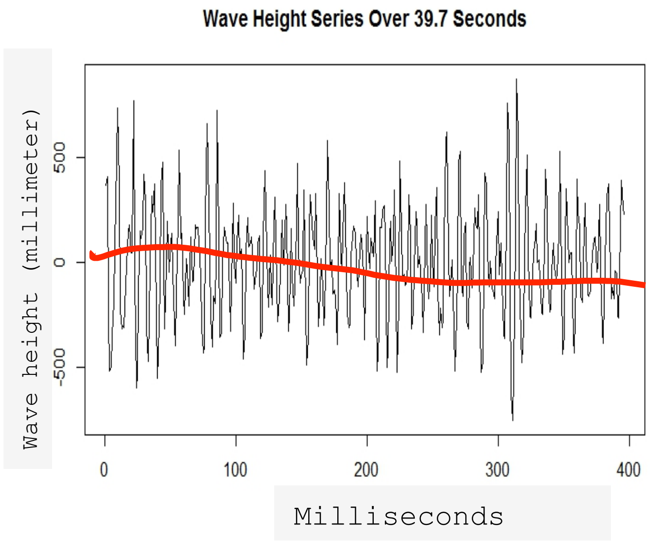
- The series also does not show any outliers.

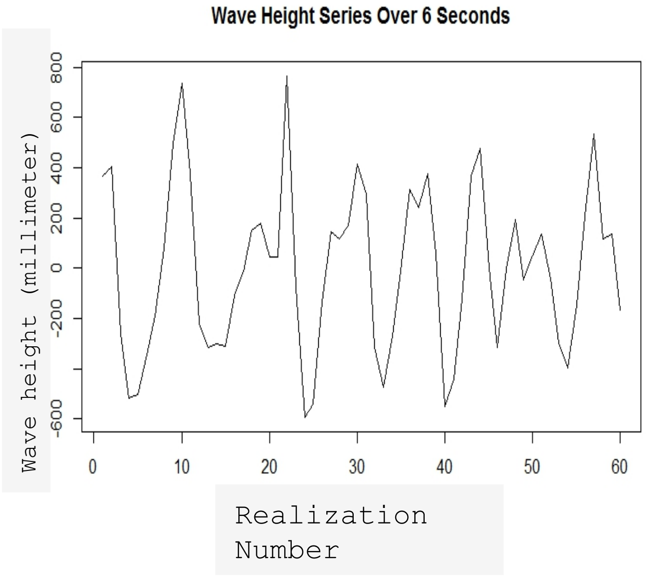




Example 1: Wave Height (3)

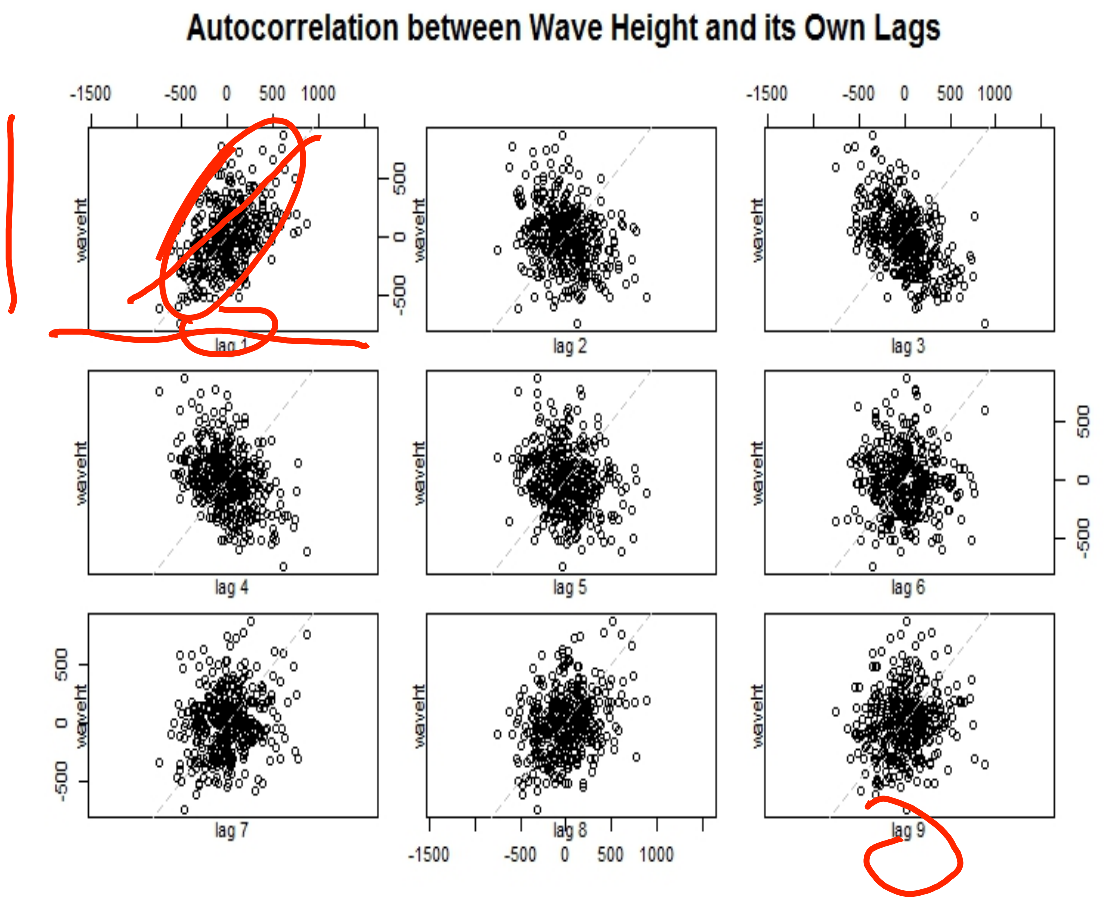
- The graph at the bottom displays a subset of the series, showing only the first 60 recorded wave heights.
- Note that the series appears to fluctuate around a constant mean.
- The series also appears to be generated by an underlying process that has a constant variance.
- Confirming from the graph at the top, the series appears to be a realization from a process with both mean and variance stationary.
- The skip-consecutive values appear to be relatively similar, mimicking that of a rough sea.
- It has quasi-periodicity but no fixed frequency.





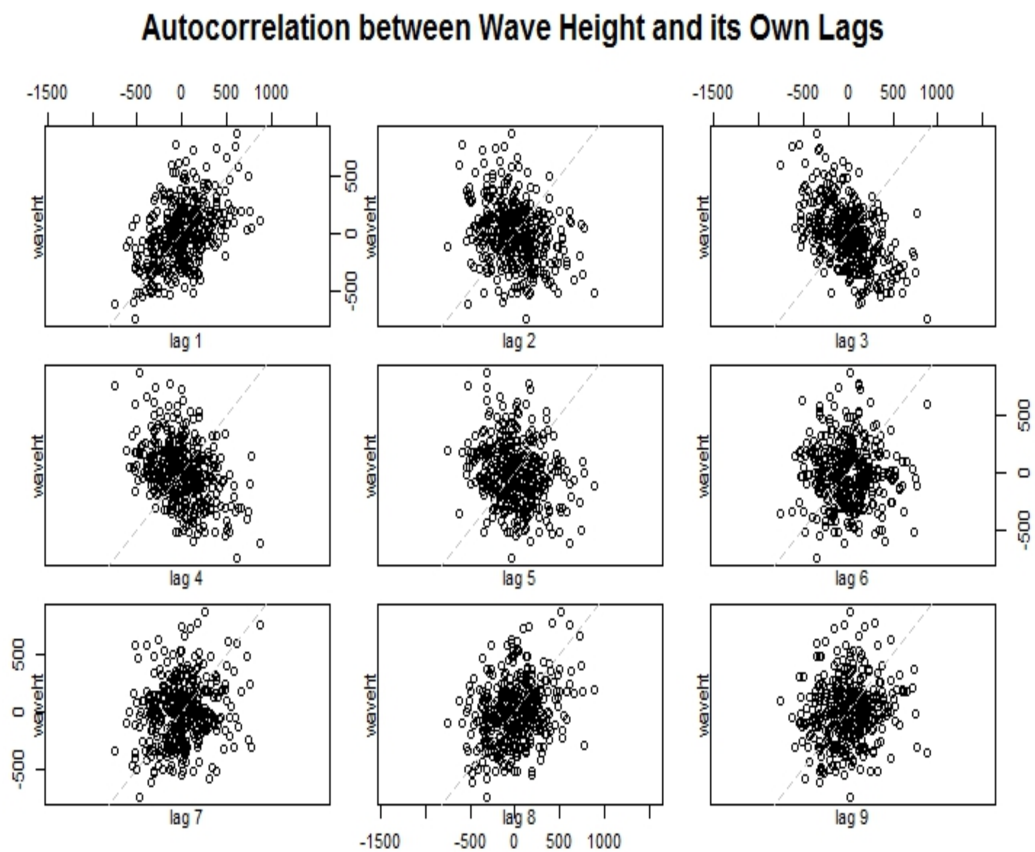
Example 1: Wave Height—Correlation With Lags

- A very useful method to visually inspect the dependency structure of a series is to plot the series against various values of its own lags using a scatter plot matrix.
- Each plot represents the wave height against a particular lag of itself.



Example 1: Wave Height—Correlation With Lags (2)

- Imagine this is a matrix. The scatter plot on the top left (or the (1,1) position of the matrix—first row and first column) can be used to examine the correlation between wave height and its first lag.
- The scatter plot in row 3, column 2 can be used to examine the correlation between wave height and its lag 8 values.

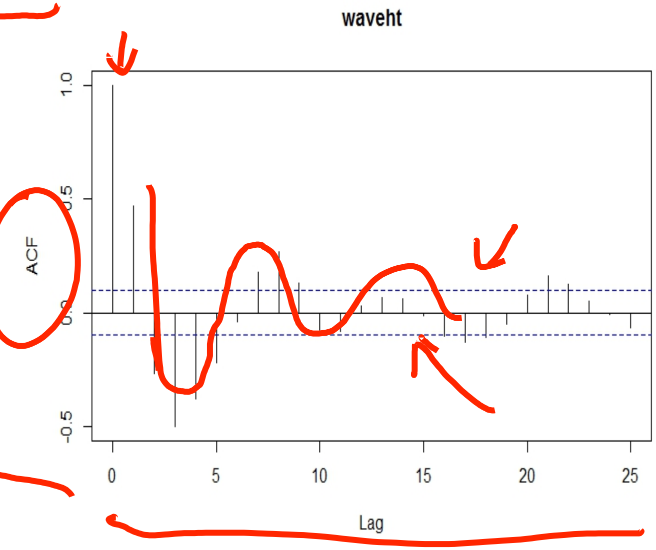


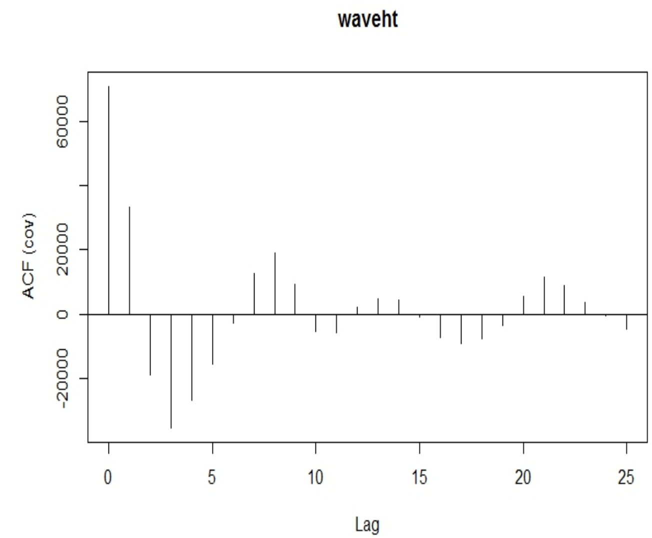
Example 1: Wave Height—Autocorrelation Function (3)

Recall that autocorrelation function takes the form:

$$\frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\frac{1}{T}\sum_{t=1}^{T-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

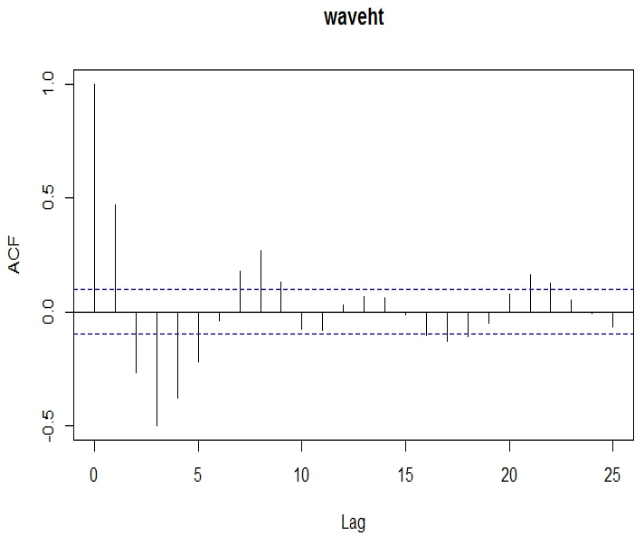
- The plots in the scatter plot matrix show both positive and negative correlations, which is confirmed by the autocorrelation (ACF) function (top graph) and autocovariance function graphs (bottom graph) on the right.
- The blue dotted lines represent the 95% confidence interval (CI) of the ACF.
- Notice that the CI of each of the autocorrelations are the same. It comes from the property of the AR model that both the conditional and unconditional variances are constant.
- The CI takes the form of $-\frac{1}{n} \pm \frac{2}{\sqrt{n}}$, which shrinks as the sample size increases.

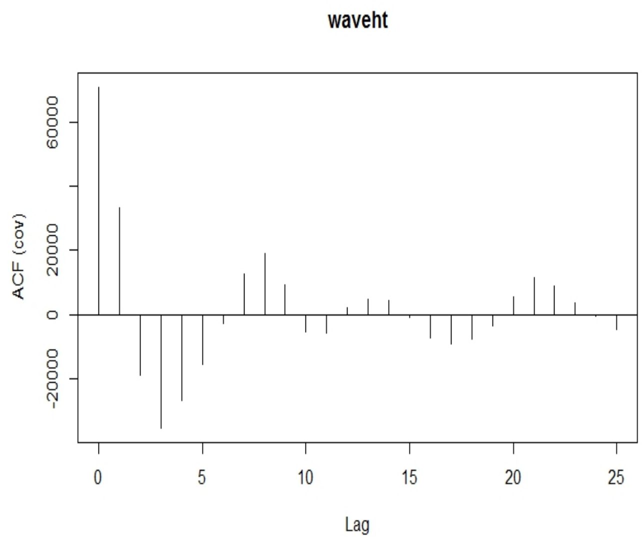




Example 1: Wave Height—Autocorrelation Function (4)

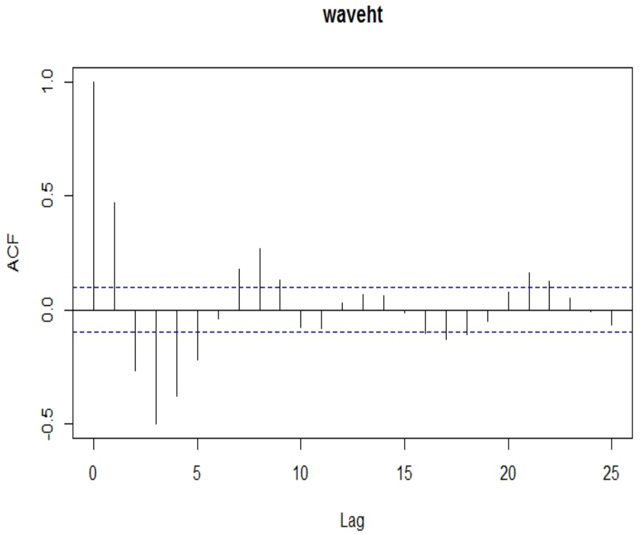
- Although in principle, when the estimated autocorrelation (at any lag k) falling outside the blue dotted line is evidence against the null hypothesis that the correlation at lag k is 0 at the 5% level, we should be very careful about interpreting multiple hypothesis tests. Even if all of the autocorrelations equal to 0 at all lags k, 5% of the estimated autocorrelations could still fall outside the blue lines, by chance. (Recall the hypothesis testing lectures in DATASCI W203.)

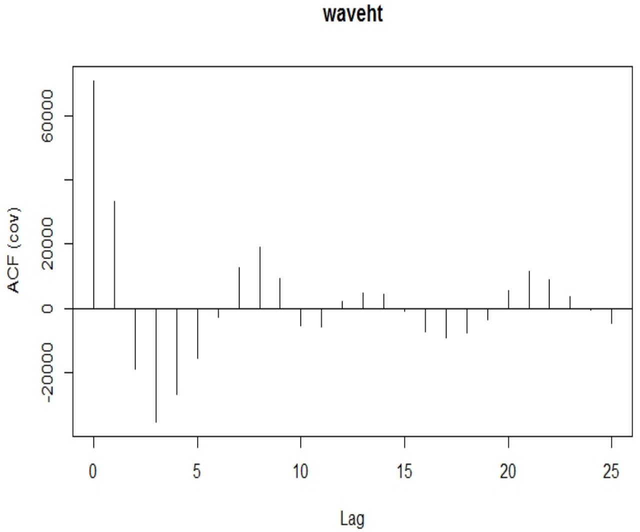




Example 1: Wave Height—Autocorrelation Function (5)

- The correlogram for wave heights has a wavelike shape that resembles that of a shrinking cosine function.
- This is typical of correlograms of time series generated by AR(2) process, as we will see in the next lecture.
- We will explore this aspect in details when we study
- Once we start studying stationary time series models in the next lecture, examining the correlogram (i.e., the graph of ACF function), along with a couple other measures, serves as an important for step to identify the order of an ARIMA model.

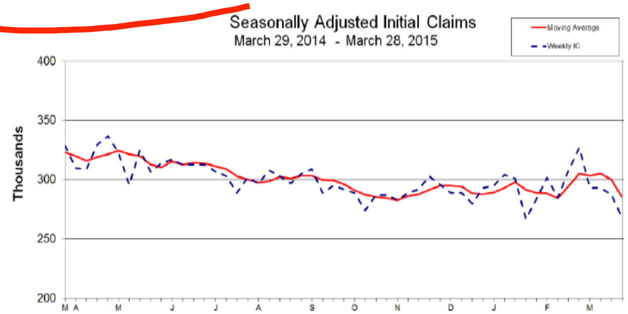




Examining Time Series Correlation—Autocorrelation Function: Example 2

Example: U.S. Initial Jobless Claims

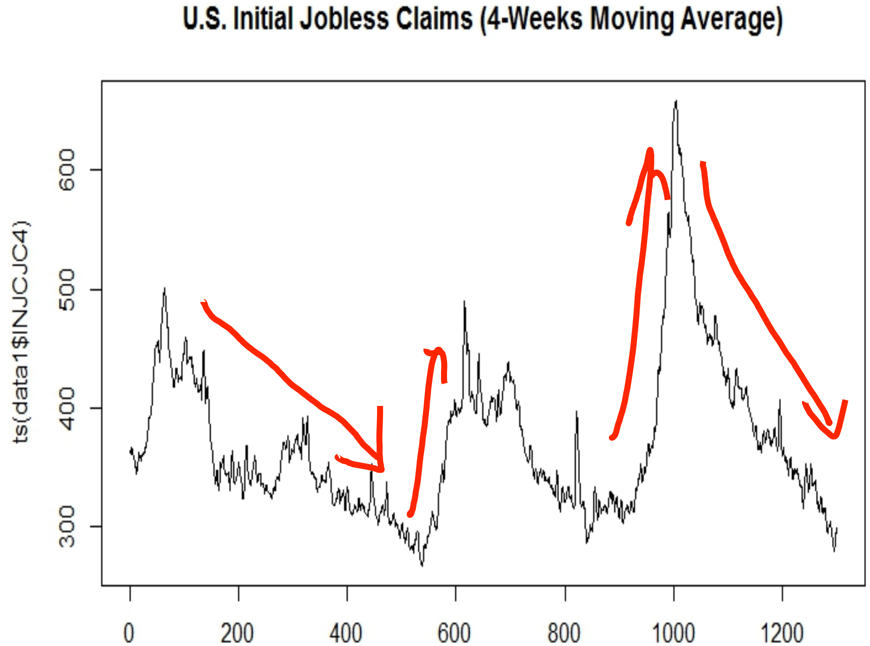
- The U.S. initial jobless claims can be downloaded from various public sources, such as the **Bureau of Labor Statistics**. For instance, the latest report can be downloaded from http://www.dol.gov/ui/data.pdf.
- The following chart is extracted from the report released on April 2, 2015.
- It showed the seasonally adjusted initial unemployment insurance claims per week from March 2014–March 2015 as well as the four-week moving average.



- The data used in this example is downloaded from a Bloomberg terminal.

Example: Initial Jobless Claims (2)

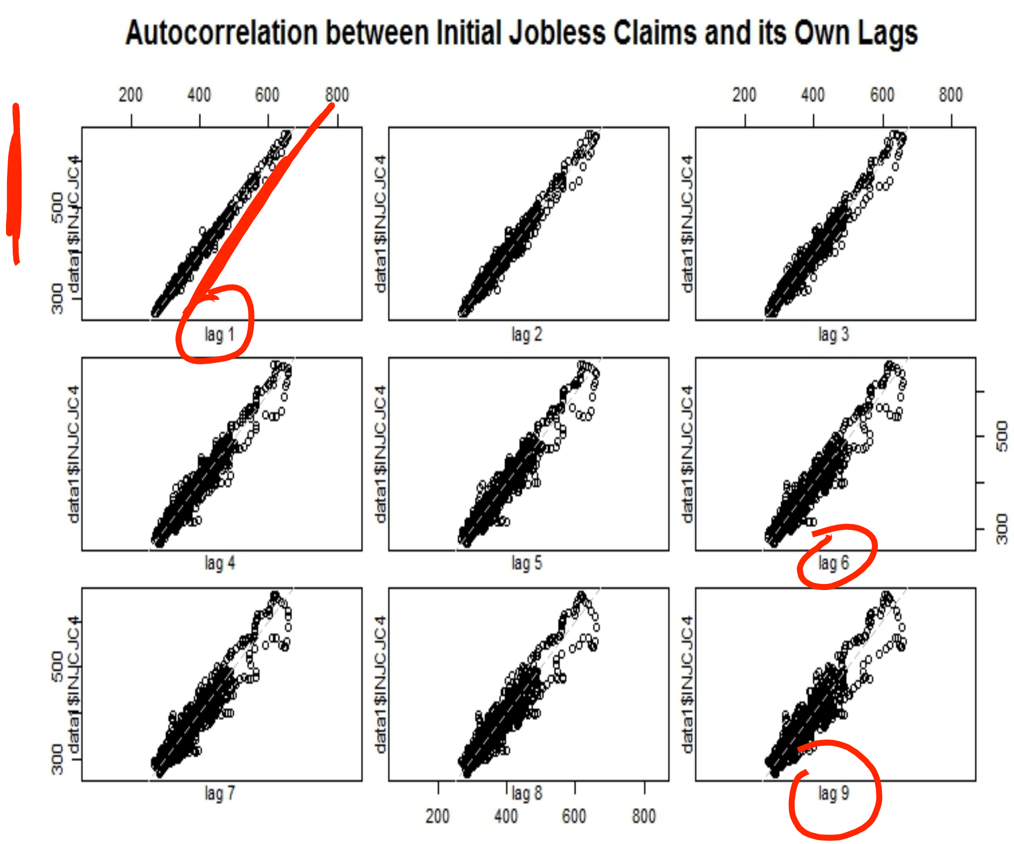
- The series' "ticker name," which is a term used in the Bloomberg Machine, is called **INJCJC4**.
- This is a weekly series from January 5, 1990, to November 28, 2014, with 1,300 observations.
- The series appears very persistence, implying the correlation with its own lags would be high. It also means that stationarity in mean, variance, and autocorrelation may not be satisfied.



Time Period in Week (5 Jan 1990-28 Nov 2014)

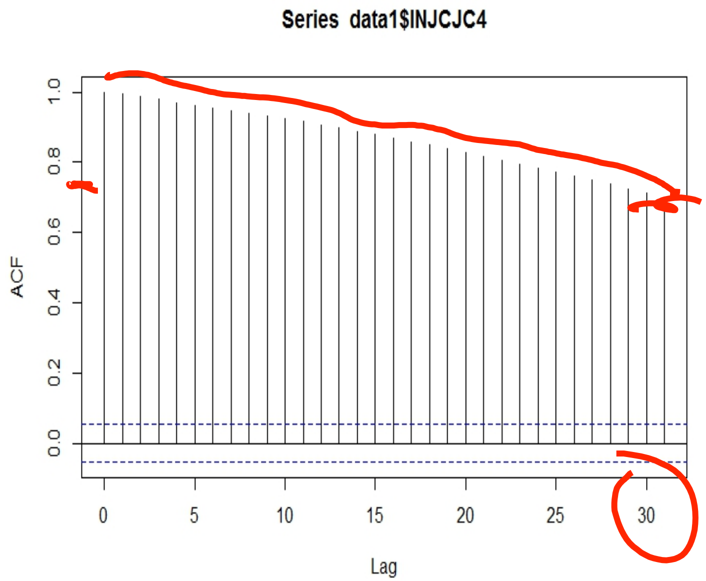
Example 3: Initial Jobless Claims (3)

- As hinted above, the persistent initial jobless claims should have very high correlation with its own lags, as shown in the following scatter plot matrix.
- The scatter plot shows that the variable and its lag almost fall on a straight line/
- Even after nine weeks, the correlation remains very high.



Example 3: Initial Jobless Claims (4)

- The persistence of the series as shown in the time series plot and the high correlation with its own lags as shown in the scatter plot matrix are evidenced in the correlogram, the graph of autocorrelation function (ACF).
- If we were to plot the ACF (ignoring the strong trend), the correlogram shows that even after 20 lags, the correlation remains above 0.8.
- Remember that the definition of ACF requires stationarity in mean and variance.
- To analyze the series using a stationary model, the trend (and seasonal effects) need to be first removed.



Notion of Stationarity: Definitions

Strict and Weak Stationarity

A time series x_t is *strictly stationary* if the joint distributions $F(x_{t_1}, \ldots, x_{t_n})$ and $F(x_{t_1+m}, \ldots, x_{t_n+m})$ are the same, $\forall\ t_1, \ldots t_n$ and m. This is a very strong condition; it implies that the distribution is unchanged for any time shift!

- A weaker and more practical stationarity condition is *weakly stationary* (or *second order stationary*).
- A time series x_t is *weakly stationary* if it is mean and variance stationary and its autocovariance $Cov(x_t, x_{t+k})$ depends only the time displacement *k* and can be written as $\gamma(k)$.
- Second-order stationarity plays an important role in many of the time series models we will discuss in this course: If a time series is second-order stationary, once a distribution assumption, such as normality, is imposed, the series can be completely characterized by its mean and covariance structure.
- Stationarity is a very important property and plays a very crucial role in time series analysis. The most popular models that are used in practice, namely autoregressive models, are stationary models, and oftentimes if a series is not stationary, it is transformed into a stationary series and modeled it using as stationary model.

Introduction to This Section: Stationarity

- In this section, we will illustrate the concepts of stationarity using four foundational processes, in their simplest form.
- Each of these processes are either building blocks of other stochastic processes and time series models or are useful by themselves very often in practice.
- We will
    1. Define the process in a mathematical form.
    2. Study the mean, variance, autocovariance, and autocorrelation of the process.
    3. Simulate the process using R.
    4. Examine the empirical behavior of the simulated realizations.

Stationarity: Example 1

1. White Noise

White Noise Recall that a *white noise* process, w_t is a sequence of random variables indexed by t, that are *independently* and *identically distributed* with mean zero and variance σ_w^2. Therefore, the process's first two moments can be written as

$$\begin{aligned} E(w_t) &= \mu_w \\ &= 0 \end{aligned}$$

$$\gamma_k = Cov(w_t, w_{t+k}) = \begin{cases} \sigma_w^2 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

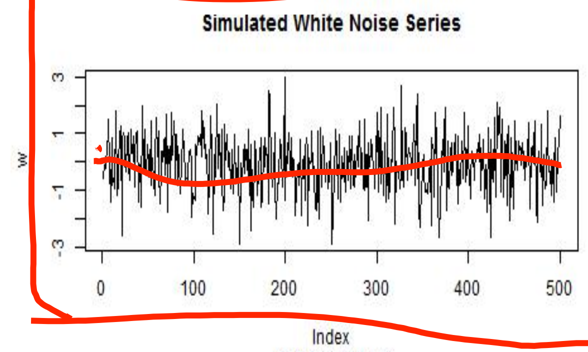
and the corresponding autocorrelation function is

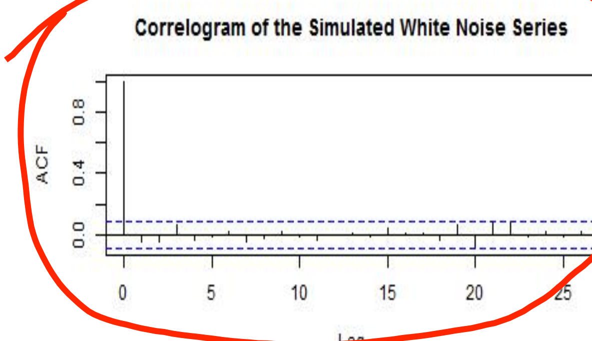
$$\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

1. White Noise: Simulations

A time series $\{w_t : t = 1, 2, \ldots, n\}$ is *discrete white noise* (DWN) if the variables $w_1, w_2, \ldots, w_n$ are *independent* and *identically* distributed with a mean of zero. This implies that the variables all have the same variance σ^2 and $\text{Cor}(w_i, w_j) = 0$ for all $i \neq j$. If, in addition, the variables also follow a normal distribution (i.e., $w_t \sim \text{N}(0, \sigma^2)$) the series is called *Gaussian* white noise.

- Not surprisingly, the simulated series appears random, and its autocorrelation function (acf) shows no statistical significant correlation with any lags.
- The blue dotted lines represent the 95% confidence interval of the autocorrelation.
- Keep these patterns in mind because we will be comparing estimated residual series (in later lectures) and examine if they resemble the dynamics of white noise.





Stationarity: Example 2

2. A Stochastic Model With a Deterministic Linear Trend

A Stochastic Model with a Linear Trend

Consider a model with a deterministic linear trend:

$$x_t = \boxed{a + bt} + w_t$$

w_t

where ~~x_t~~ is a white noise with mean 0 and variance σ_w^2

The expected value of x_t is

$$\begin{aligned} E(x_t) &= E(a + bt + w_t) \\ &= a + bt + E(w_t) \\ &= a + bt \end{aligned}$$

2. A Stochastic Model With a Deterministic Linear Trend (2)

As such, a stochastic model with a linear trend is not mean stationary, as the mean changes with time. If $b > 0$, then the mean is an increasing function of the time index t. On the other hand, if $b < 0$, then the mean is an decreasing function of the time index t.

$$
\begin{aligned}
Var(x_t) &= Var(a + bt + w_t) \\
&= Var(w_t) \\
&= \sigma_w^2
\end{aligned}
$$

which is a constant. So, while the model is not mean stationary, it is variance stationary.

2. A Stochastic Model With a Deterministic Linear Trend (3)

$$
\begin{aligned}
Cov(x_t, x_{t-1}) &= Cov\left(a + bt + w_t, a + b(t-1) + w_{t-1}\right) \\
&= Cov\left(w_t, w_{t-1}\right) \\
&= 0
\end{aligned}
$$

since w_t is a white noise series.

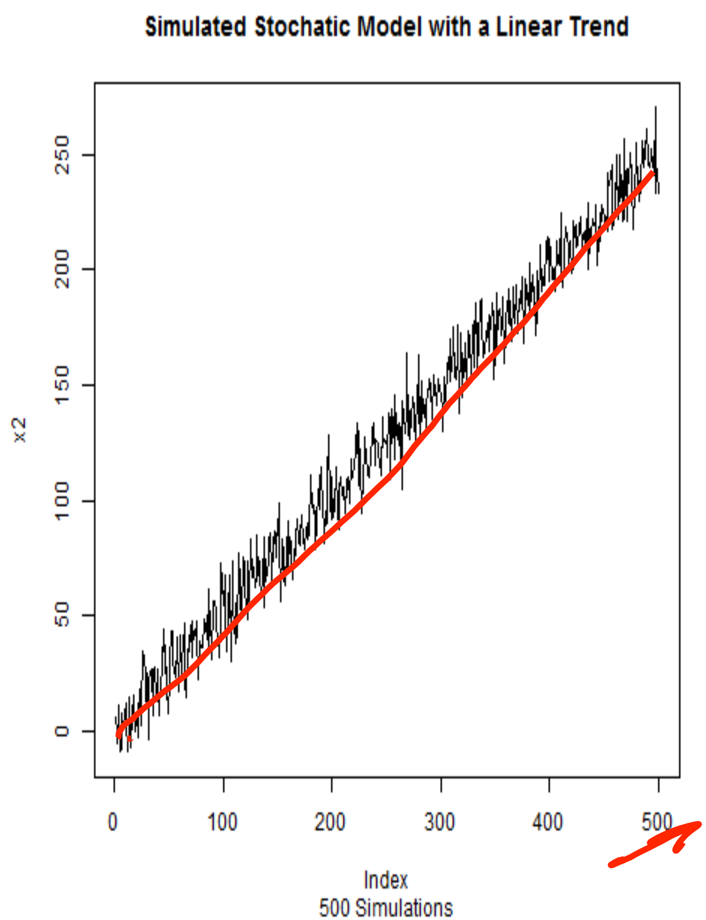
Therefore, the stochastic model with a deterministic linear trend is not (strictly or weakly) stationary, although as we will see in a few lectures, it can be easily transformed into a stationary model.

2. A Stochastic Model With a Deterministic Linear Trend: Simulations

$x_t = 1 + 0.5t + w_t$ where w_t is a series of independent Gaussian white noise with mean 0 and variance 10

```
set.seed(898)
sigma_w = 10
beta0 = 1
beta1 = 0.5
t = seq(1,500)
w <- rnorm(500,0,sigma_w)
x2 <- beta0 + beta1*t + w
cbind(t, x2, w)
summary(x2)
mean(x2)
sd(x2)
```

- The simulated series appears as a linear trend, as if the white noise does not affect the trend.
- The acf shows that the series is very persistence, meaning that is highly correlated with its lags.



Stationarity: Example 3

3. Moving Average Model of Order 1:

MA(1) An textitMA(1) model takes the form

$$x_t = w_t + \beta w_{t-1}$$

where w_t is a white noise series with mean zero and variance σ_w^2.

The expected value of x_t is

$$\begin{aligned} E(x_t) &= E(w_t + \beta w_{t-1}) \\ &= \alpha E(w_t) + E(w_{t-1}) \\ &= 0 \end{aligned}$$

since $E(w_t) = 0 \forall t$

It is straight-forward (left as an exercise) to derive the autocorrelation function for the MA(1) model:

$$\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} \beta_i^2} & \text{if } k = 1 \dots q \\ 0 & \text{if } k > q \end{cases}$$

3. Moving Average Process: Simulations (1)

- As seen in various places in the course, a useful way to learn about the empirical patterns of a econometric model is to use simulate to simulate realizations from a theoretical model.
- In this example, we simulated from a MA(1) model 500 realization:

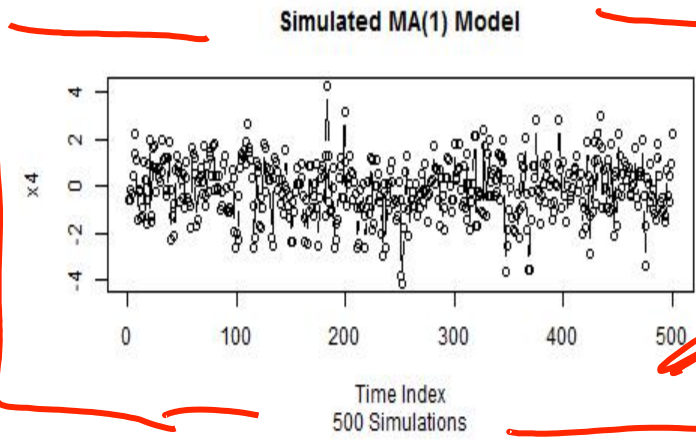
$$
\begin{aligned}
x_t &= \omega_t + \beta\omega_{t-1} \\
&= \omega_t + 0.8\omega_{t-1}
\end{aligned}
$$

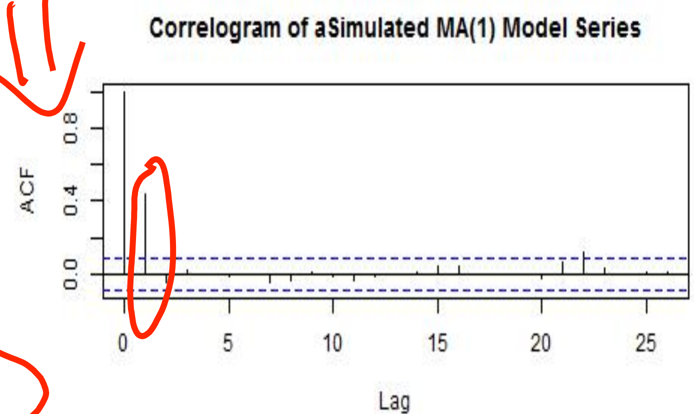
```
## 4. MA(1) Model
# Simulation
set.seed(898)
beta1 = 0.8
x4 <- w <- rnorm(500,0,sigma_w)

for (t in 2:500) x4[t] <- w[t] + beta1*w[t-1]
```

```
> summary(x4)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-41.5400  -8.9360  -0.8613  -0.9833   7.5360  43.0900
```

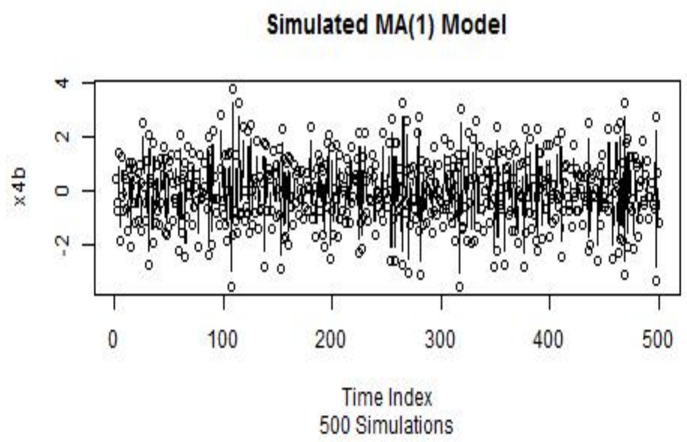
```
> sd(x4)
[1] 12.28474
```

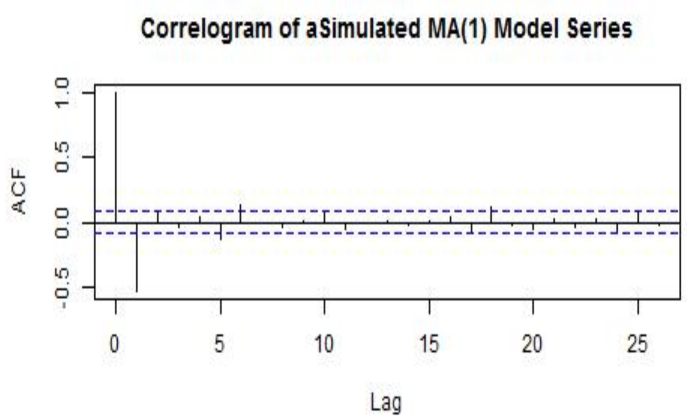




3. Moving Average Process: Simulations (2)

- A distinguish property of a MA(q) model is that its ACF drops off abruptly at q lags.
- In the case of MA(1) model, the ACF drops off to almost zero after the first lag, as shown in the correlogram.





Stationarity: Example 4

4. Autoregressive Model of Order 1

AR(1) An AR(1) model takes the form

$$x_t = \alpha x_{t-1} + w_t$$

where w_t is a white noise series with mean zero and variance σ_w^2.

The expected value of x_t is

$$\begin{aligned} E(x_t) &= E(\alpha x_{t-1} + w_t) \\ &= \alpha E(x_{t-1}) + E(w_t) \\ &= \alpha E(x_{t-1}) \\ &= 0 \end{aligned}$$

because using recursive substitution, an *AR(1)* model can be written as a linear process in terms of the sum of infinite white noises:

$$\boxed{x_t = \sum_{i=0}^{\infty} \alpha^i w_{t-i}}$$

4. Autoregressive Process of Order 1 (3)

With the linear process (of white noise) representation, the autocovariance can be derived as follow:

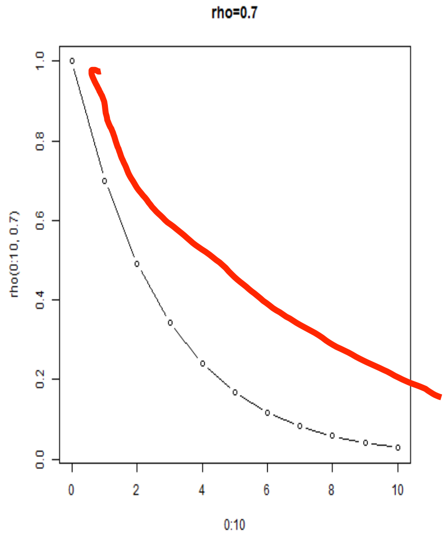
$$
\begin{aligned}
\gamma_k &= Cov(x_t, x_{t+k}) \\
&= Cov\left(\sum_{i=0}^{\infty} \alpha^i w_{t-i}, \sum_{j=0}^{\infty} \alpha^j w_{t+k-j}\right) \\
&= \sum_{j=k+i} \alpha^i \alpha^j Cov\left(w_{t-i}, w_{t+k-j}\right) \\
&= \alpha^k \sigma^2 \boxed{\sum_{i=0}^{\infty} \alpha^{2i}} \\
&= \frac{\alpha^k \sigma^2}{(1-\alpha^2)}
\end{aligned}
$$

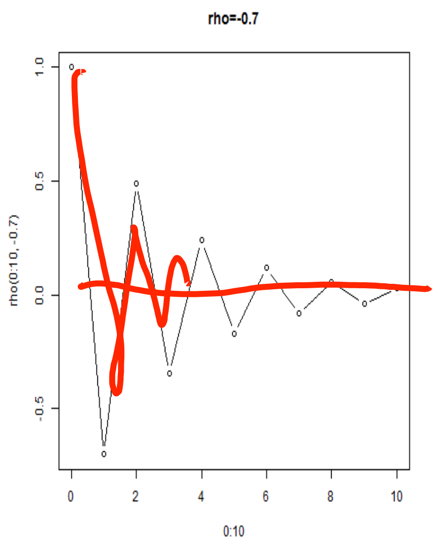
4. Autoregressive Process of Order 1: Simulations (1)

```
#--------------------------------------------------------------
# Examine the exponential decay behavior of AR(1) correlogram
rho <- function(k, alpha) alpha^k
plot(0:10, rho(0:10, 0.7), type="b", main="rho=0.7")
plot(0:10, rho(0:10, -0.7), type="b", main="rho=-0.7")
#--------------------------------------------------------------
```

- Top graph: The pattern of a theoretical AR(1) model with a positive correlation decays exponentially.

- Bottom graph: The patterns of a theoretical AR(1) model with a negative correlation oscillate between positive and negative correlation.





4. Autoregressive Process of Order 1: Simulations (1)

Recall that an AR(1) model takes the form:

$$\tilde{x}_t = \alpha \tilde{x}_{t-1} + w_t$$

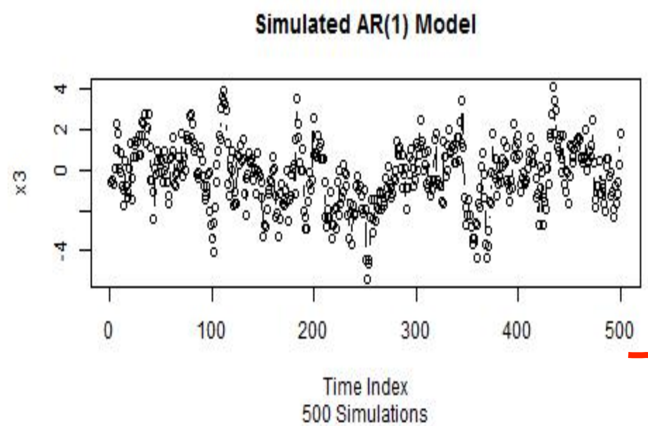
where w_t is a white noise series with mean zero and variance σ_w^2

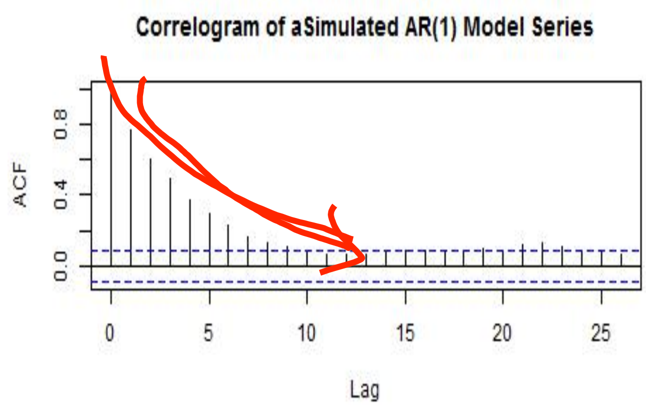
Consider $\tilde{x}_t$ as series of deviations from μ, which is a parameter that determines the "level" of the process.

```
set.seed(898)
sigma_w = 1
alpha0 = 0
alpha1 = 0.8
x3 <- w <- rnorm(500,0,sigma_w)

for (t in 2:500) x3[t] <- alpha0 + alpha1*x3[t-1] + w[t]
```

- Note the exponential decay of the autocorrelation of the simulated AR(1) series.





Berkeley SCHOOL OF INFORMATION