Financial Econometrics and Statistical Arbitrage

Master of Science Program in Mathematical Finance
New York University

Lecture 1: Basics on Time Series Analysis

Fall 2005

Administrative Details

- Instructors: Farshid Maghami ASL and Lee Maclin
- Email: fma1@nyu.edu
- Course Web sites:
  - **Blackboard**
  - **http://homepages.nyu.edu/~fma1**

- Teaching Assistant: Junyoep Park
- Email: junyoep@gmail.com
- Office Hours: Mondays 5-7 pm
- Office Location: **WWH 606**

Administrative Details (cont.)

Time: Mondays, 7:10 – 9 pm

First Class: September 12, 2005, Last Class: December 12, 2005. There will be no class on Columbus Day (October 10, 2005). We will make it up on Wednesday 11/23/05.

Homework and Exam:

- There will be six homework sets which will be assigned every other week. Students must write up and turn in their solutions individually within one week.
- Computer assignments can be solved by C/C++/C#, MATLAB, R. For other tools, please coordinate with the TA or the instructor.
- There will be one final exam (no mid-term).
- Final grade will be evaluated based on homework solutions (30%) and the final exam (70%).
- Lecture Notes and Homework will be posted on the course website as they become available

Administrative Details (cont.)

Textbooks: Lectures are drawn from many sources including the following books:

1. Alexander, C. "*Market Models,*" John Wiley and Sons, 2001
2. Brockwell, P.J., Davis, R.A., "*Introduction to Time Series and Forecasting,*" Springer
3. Javaheri, A. "*Inside Volatility Arbitrage : The Secrets of Skewness*" Wiley
4. Tsay, R. S., "*Analysis of Financial Time Series*," Wiley, 2002
5. Wilmott, P. "*Derivatives: The Theory and Practice of Financial Engineering*," Wiley Frontiers in Finance Series
6. Pandit S.M., Wu S.M., "*Time Series and System Analysis with Applications*." Krieger Publishing, Malabar, FL, 2001
7. Hamilton J. D. "*Time Series Analysis*." Princeton University Press, 1994

A number of research articles will be posted on the course webpage.

Expected Background

- Prior knowledge of Linear Algebra, Probability and Statistics is required
- I assume you have taken the following courses:
  - Derivative Securities
  - Continuous Time Finance
  - Scientific Computing / Computing for Finance
- Programming in C/C++ or MATLAB/R is required

What is Statistical Arbitrage?

- Arbitrage is a riskless profit. "Arbitrage Strategy" is a trading strategy that locks in a riskless profit.

Strategies and Implementation Process
(Cointegration based pairs trading, Volatility trading, …)

Financial Econometrics
(Time Series Review and Volatility modeling)

- Statistical Arbitrage covers any trading strategy which uses statistical tools and time series analysis to identify approximate arbitrage opportunities while evaluating the risks inherent in the trades considering the transaction costs and other practical aspects.

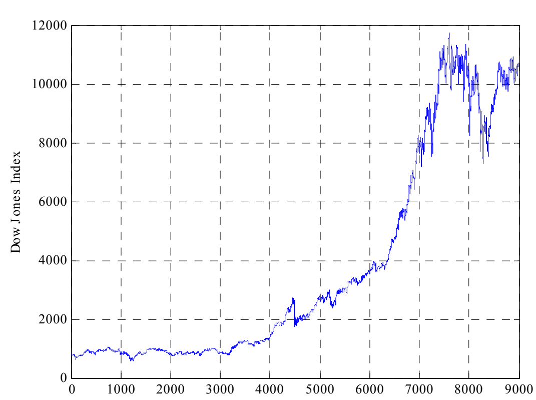
Market Microstructur Theory
(Transaction costs and Optimal Control, Algorithmic Trading,…)

Risk Management
(Practical Risk Measurement and Management Technics)

Course Outline

- Financial Econometrics (8 weeks)
  - Time Series Models Review and Analysis
  - Volatility and Correlation Models in Financial Systems
  - Calibration and Estimation Methods
- Cointegration and Market Microstructure in Practice (3 weeks)
  - Cointegration and Pairs Trading
  - Transaction Costs, and Market Friction
  - Trade Execution Strategies
- Practical Simulation and Risk Management (1 week)
- More on Trading Strategies (1-2 week)

Typical Behavior of Financial Assets



- The unpredictability inherent in asset prices is the main feature of financial modeling.
- Because there is so much randomness, any mathematical model of a financial asset must acknowledge the randomness and have a probabilistic foundation.

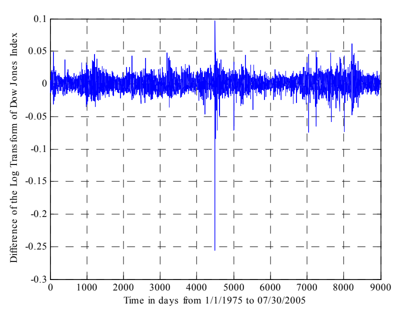
- There are three general types of analysis used in finance and trading
  1. Fundamental Analysis
  2. Technical Analysis
  3. **Quantitative Analysis**

- **Return in financial assets**
  By **return** we mean the percentage growth in the value of an asset, together with accumulated dividends, over some period:

$$\textbf{Return} = \frac{\text{Change in value of the asset} + \text{accumulated cashflows}}{\text{Original value of the asset}}$$

- Denoting the asset value on the i-th day by Si, the return from day i to day $i+1$ is given by

$$R_i = \frac{S_{i+1} - S_i}{S_i}$$



- Supposing that we believe that the empirical returns are close enough to Normal for this to be a good approximation.
- For start, we write the returns as a random variable drawn from a Normal distribution with a known, constant, non-zero mean and a known, constant, non-zero standard deviation:

$$N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\phi^2}$$

$$R_i = \frac{S_{i+1} - S_i}{S_i} = \text{mean} + \text{standard deviation} \times \phi$$

Introduction to Financial Modeling

$$R_i = \frac{S_{i+1} - S_i}{S_i} = \underline{\text{mean}} + \underline{\text{standard deviation}} \times \phi$$

- Time scale δt

Mean return over period δt is $\mu \delta t$	Standard deviation over period δt is $\sigma \delta t^{1/2}$

$$R_i = \frac{S_{i+1} - S_i}{S_i} = \mu \delta t + \sigma \delta t^{1/2} \times \phi$$

$\sigma \delta t^{1/2} \times \phi \rightarrow \delta X$

And in the limit $\delta t \rightarrow 0$

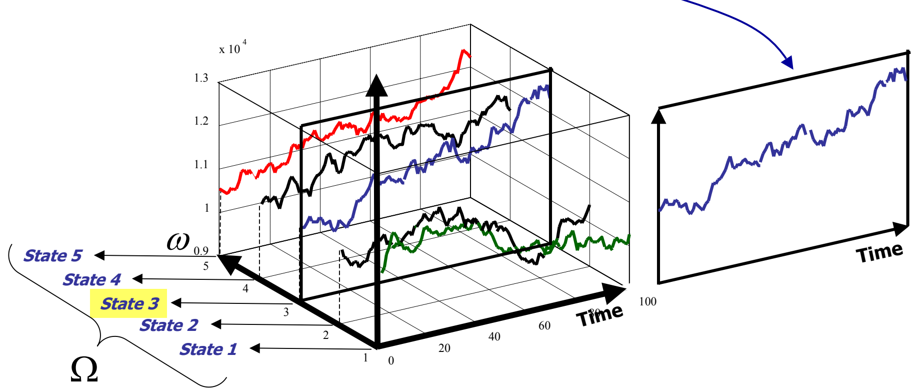
$$R = \frac{dS_t}{S_t} = \mu dt + \sigma dX_t$$

Basic Review

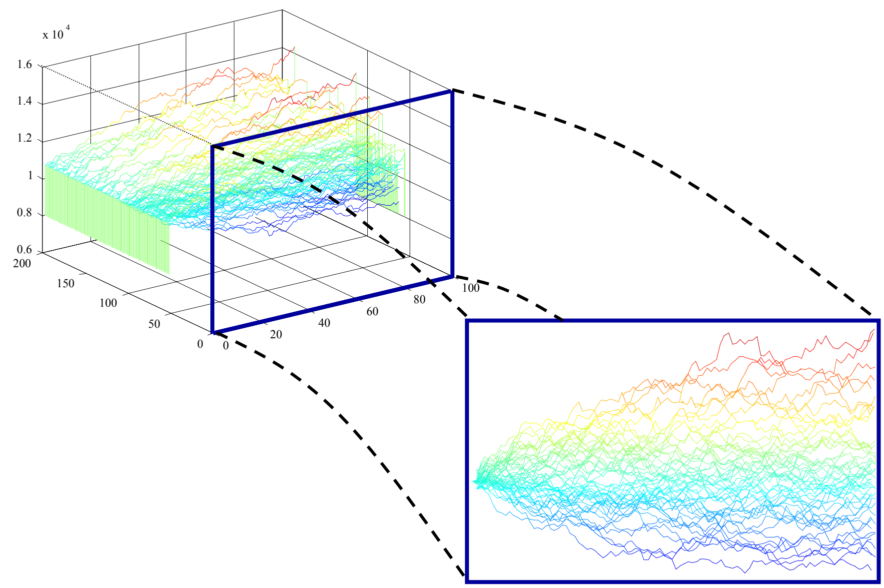
STOCHASTIC PROCESS:

A stochastic process is a collection of random variables $\{X_t(\omega), \ t \in \tau\}$ defined on a probability space $(\Omega \ , F \ , P \ )$.

For a fixed ω , a realization of stochastic process is a function of time (t).



Simulation of a Stochastic Process

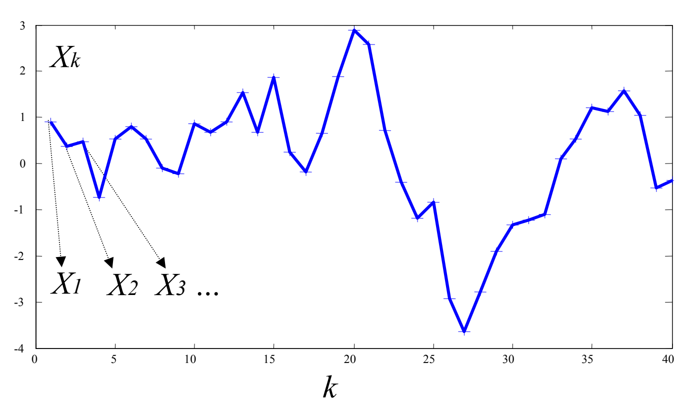


Definition

Time Series:

A time series is a stochastic process where τ is a set of discrete points in time. In other words, it is a discrete time, continuous state process.

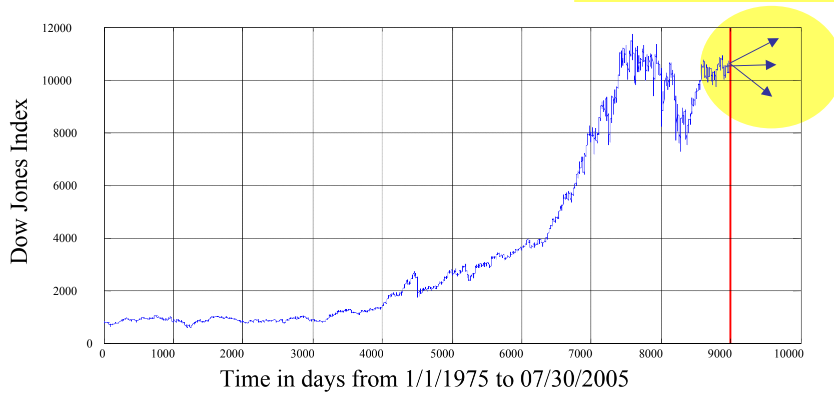
In this course we consider $\tau = \{ \text{all integers} \}$



Goals of Studying Time Series

1- Forecasting

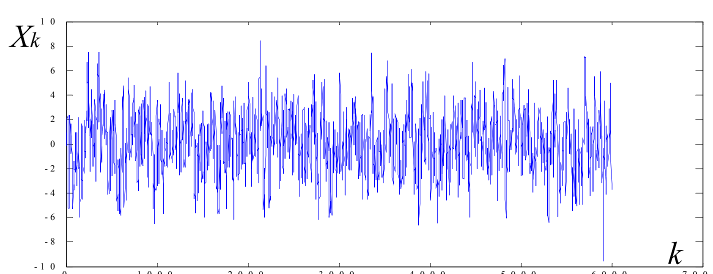
We want to forecast distributions

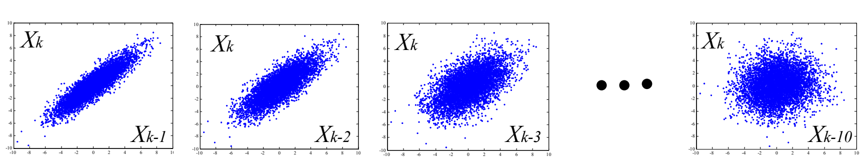


2- Understanding the statistical characteristics and building trading strategies based on them

Basic Review

An Example of a Time Series:





$$X_k = \varphi_1 X_{k-1} + \varphi_2 X_{k-2} + \ldots + e_k$$

Auto-Regression as a Dynamic System? We will get back to this

Definition

Autocovariance Function:

Let {*Xt*} be a time series. The autocovariance function of process {*Xt*} for all integers *r* and *s* is:

$$\gamma_X(r,s) = \mathrm{cov}(X_r, X_s)$$

$$\gamma_X(r,s) = E[(X_r - E(X_r))(X_s - E(X_s))]$$

$$\gamma_X(r,s) = E[X_r X_s - X_r E(X_s) - X_s E(X_r) + E(X_r)E(X_s)]$$

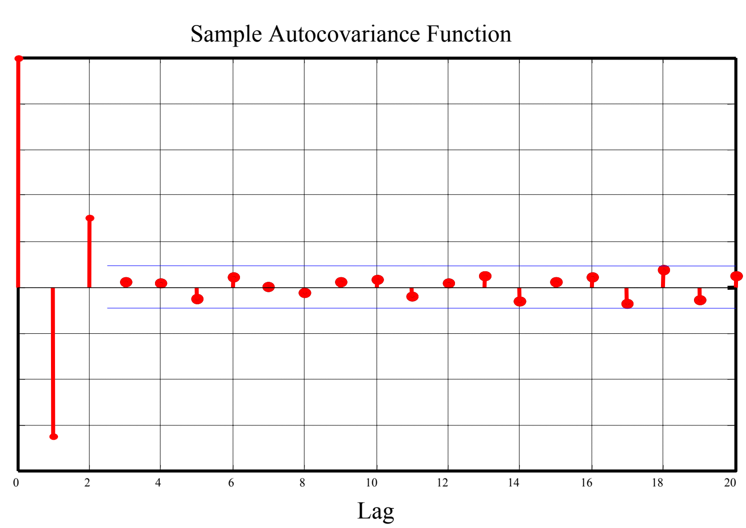
$$\gamma_X(r,s) = E(X_r X_s) - E(X_r)E(X_s) \boxed{- E(X_s)E(X_r) + E(X_r)E(X_s)}$$

$= 0$

$$\gamma_X(r,s) = E(X_r X_s) - E(X_r)E(X_s)$$

Note that $\gamma_X(r,r) = E(X_r^{\,2}) - E(X_r)^2 = \mathrm{var}(X_r) \geq 0$

Autocovariance Function



Definition

Stationary Process:

A time series $\{X_t\}$ is stationary (weakly) if:

1. $E(X_t^2) < \infty$
2. $E(X_t) =$ Some constant m for all t
3. $\gamma_X(r,s) = \gamma_X(r+t, s+t)$

i.e. $Cov(X_r, X_s)$ only depends on r and s and not on t.

Note:

A strict (strong) stationary time series $\{X_t,\ t = 1, 2, \ldots, n\}$ is defined by the condition that realizations $(X_1, X_2, \ldots, X_n)$ and $(X_{1+h}, X_{2+h}, \ldots, X_{n+h})$ have the same joint distributions for all integers h and $n > 0$.

Note: If $\{X_t\}$ is stationary, then

$$\gamma_X(r,s) = \gamma_X(r-s, s-s) = \gamma_X(r-s, 0) = \gamma_X(r-s)$$

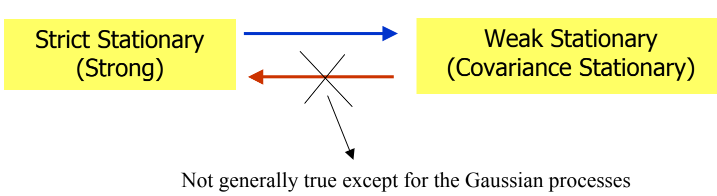
Define $h = r - s$

$$\gamma_X(r-s) = \gamma_X(h) = \operatorname{cov}(X_t, X_{t+h})$$

Does not depend on t

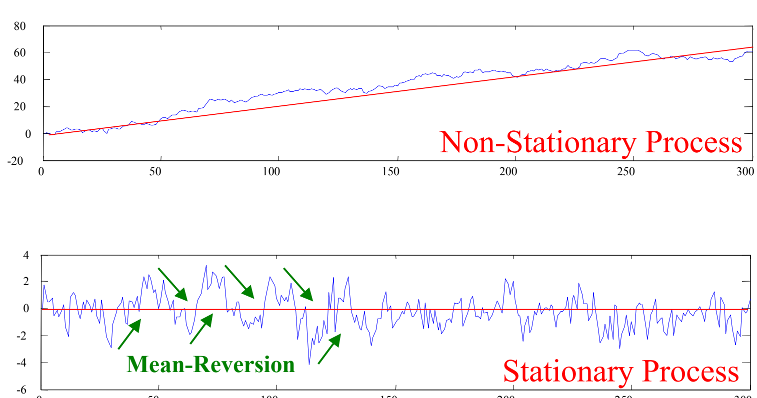
Definition

Note:



Stationary Process and Mean Reversion

- We are interested in stationary time series because many models and tools are developed for stationary processes.
- A stationary process can never drift too far from its mean because of the finite variance. The speed of mean-reversion is determined by the autocovariance function: Mean-reversion is quick when autocovariances are small and slow when autocovariances are large.
- Trends and periodic components make a time series non-stationary.



General Approach to Time Series

Time Series Analysis

1. Plot time series and check for trends or sharp changes in behavior (most of the time non-stationary)
2. Transform into a stationary time series
3. Fit a model ← If bad
4. Perform diagnostic tests (residual analysis,…)
5. Generate forecasts (find predictive distributions) and invert the transformations performed in 2.

   Note for option pricing:

6. Find a risk neutral version of the model
7. Obtain predictive distributions under the risk neutral model

Building Blocks of Financial Models

White Noise Process

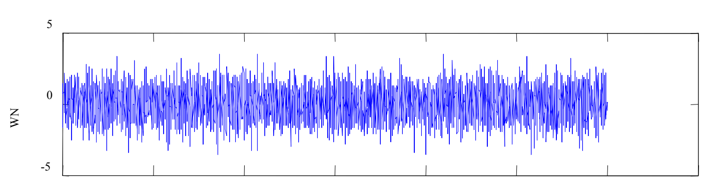
If $\{X_t\}$ is a sequence of random variables with $E(X_t)=0$, $E(X_t^2)=\sigma^2$ and

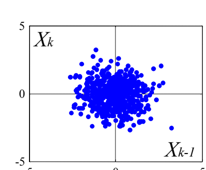
$$\gamma_X(r,s)=\begin{cases}\sigma^2 & r=s \quad (\sigma^2<\infty)\\ 0 & otherwise\end{cases}$$

$\{Xt\}$ is called White Noise and it is written as $WN(0,\sigma^2)$

Note that $E[X_t X_s]=0$ for $t \neq s$ → Uncorrelated r.v.'s

If X_t and X_s independent for $t \neq s$ → $IID(0,\sigma^2)$





White Noise Process (Is it Stationary?)

$$E(X_t) = 0$$

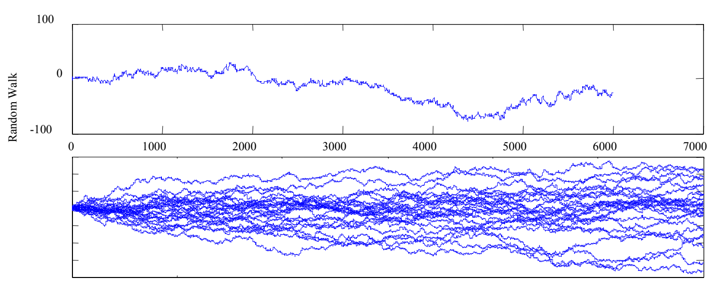
$$\gamma_X(r,s) = \begin{cases} \sigma^2 & r = s \\ 0 & otherwise \end{cases} \qquad (\sigma^2 < \infty)$$

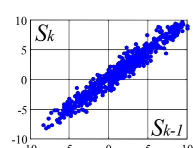
Random Walk Process

If $\{X_t\}$ be a sequence of $IID(0, \sigma^2)$ random variables , a sequence $\{S_t\}$ with $S_0=0$ and

$$S_t = \sum_{j=1}^{t} X_j \qquad \text{(Integrated Process)}$$

Is called a Random Walk.





Random Walk Process (Is it Stationary?)

$$S_t = \sum_{j=1}^{t} X_j \qquad \{Xt\} \sim IID(0, \sigma^2)$$

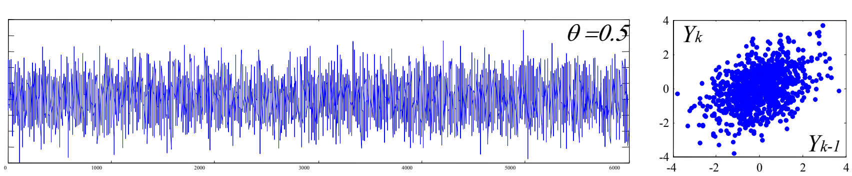
Moving Average Process

Let $\{Xt\}$ be $WN(0, \sigma^2)$, and consider the process

$$Y_t = X_t + \theta X_{t-1}$$

Where θ could be any constant. This time series model is called a first-order moving average process, denoted *MA(1).*

The term "Moving Average" comes from the fact that Yt is constructed from a weighted sum of the two most recent values of Xt.



Moving Average Process (Is it Stationary?)

$\{X_t\}$ is $WN(0, \sigma^2)$ $\qquad Y_t = X_t + \theta X_{t-1}$

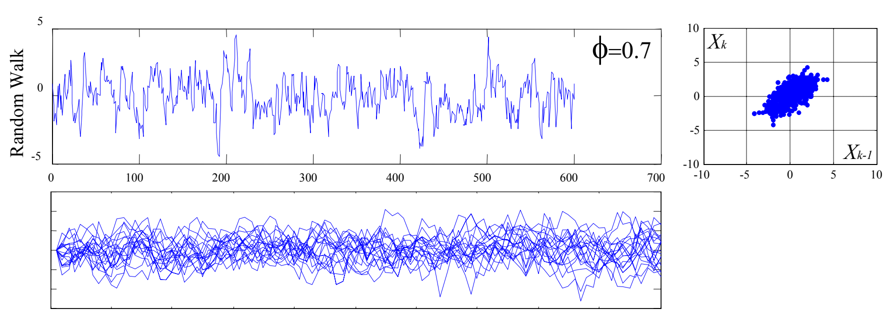
Autoregressive Process

Let $\{Z_t\}$ be $WN(0, \sigma^2)$, and consider the process

$$X_t = \phi X_{t-1} + Z_t$$

Where $|\phi| < 1$ and Z_t is uncorrelated with X_s for each $s<t$. This time series model is called a first-order Autoregressive process, denoted *AR(1)*.

It is easy to show that $E(X_t)=0$



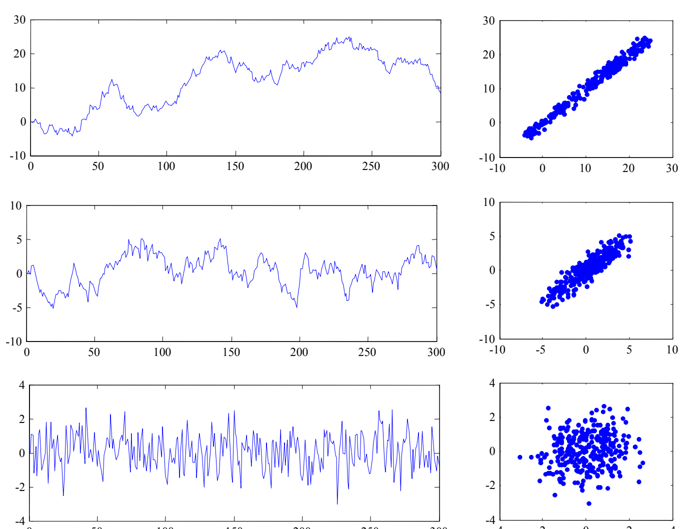
Autoregressive Process (Is it Stationary?)

$\{Z_t\}$ is $WN(0, \sigma^2)$, and $X_t = \phi X_{t-1} + Z_t$

Where $|\phi| < 1$ and Z_t is uncorrelated with X_s for each $s < t$.

We will see this later

$$X_t = \phi X_{t-1} + Z_t$$



$\phi = 1$
Random Walk

$\phi = 0.9$
AR(1)

$\phi = 0.1$
AR(1)

Classical Decomposition

$$X_t = m_t + S_t + Y_t$$

- X_t: Original Time series (Nonstationary)
- m_t: Trend
- S_t: Seasonal component
- Y_t: Stationary Time series (zero-mean)

Seasonal component S_t satisfies

$$S_{t+d} = S_t \quad \text{where } d = \text{period of seasonality}$$

Also for mathematical convenience assume $\sum_{j=1}^{d} S_j = 0$

Most observed time series are non-stationary but they can be transformed to stationary processes.

Classical Decomposition

$$X_t = m_t + S_t + Y_t$$

Idea of transformation is to estimate m_t and S_t by $\hat{m}_t$ and $\hat{S}_t$, then work with the stationary process:

$$X_t^* = X_t - \hat{m}_t + \hat{S}_t$$

Assume there is no seasonal component ($S_t=0$)

$$X_t = m_t + Y_t$$

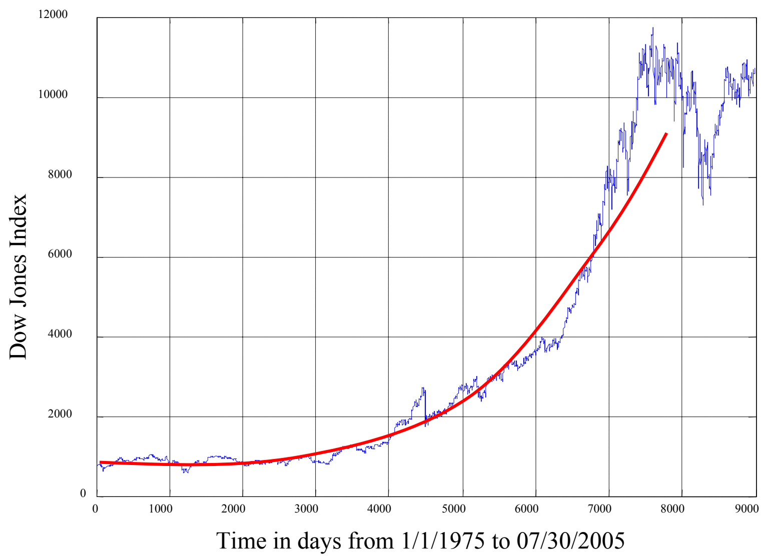
Consider a parametric form for $\hat{m}_t$ *e.g.*

$$\hat{m}_t = a_0 + a_1 t + a_2 t^2$$

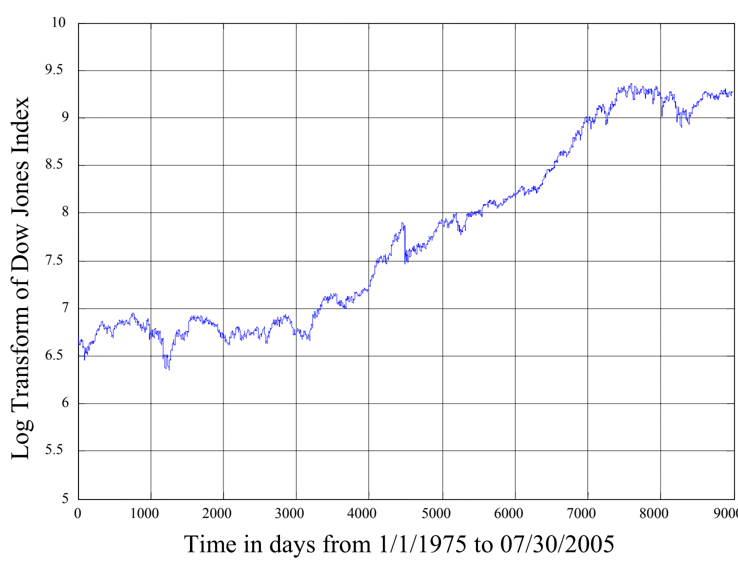
Using observed data $X_1, X_2, \ldots X_n$, choose $\alpha_0, \alpha_1, \alpha_2$ to minimize

$$\sum_{t=1}^{n} (X_t - \hat{m}_t)^2$$

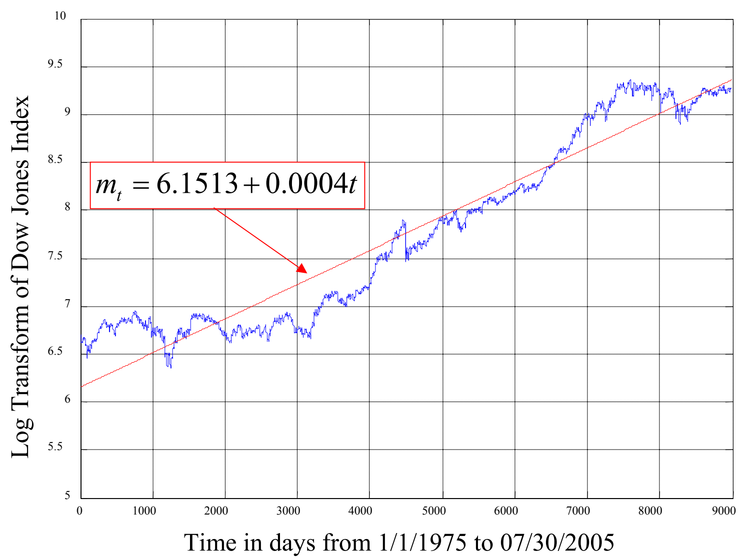
Transforming a Non-Stationary Process to a Stationary Process



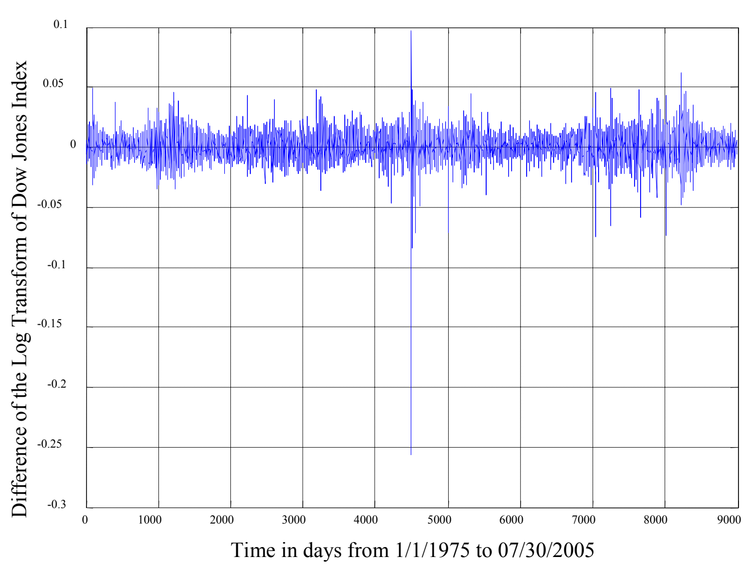
Transforming a Non-Stationary Process to a Stationary Process

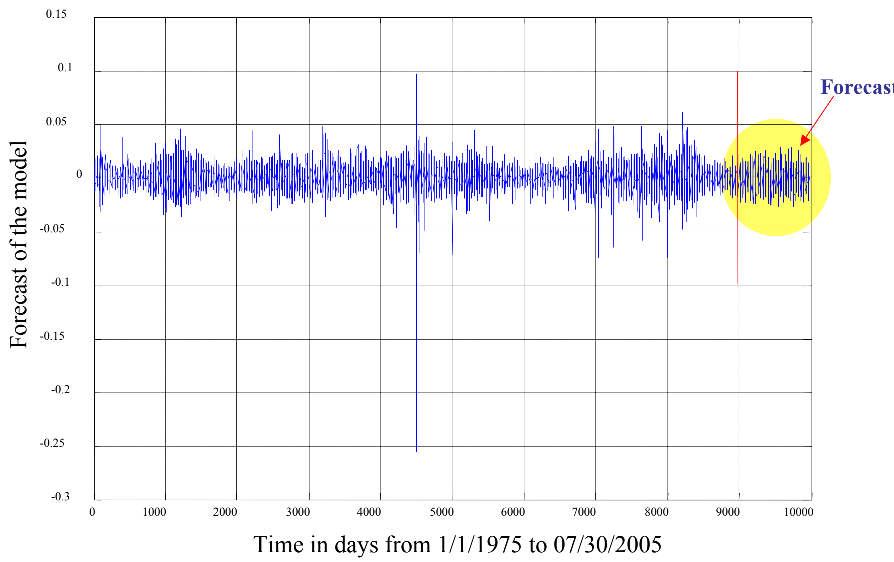


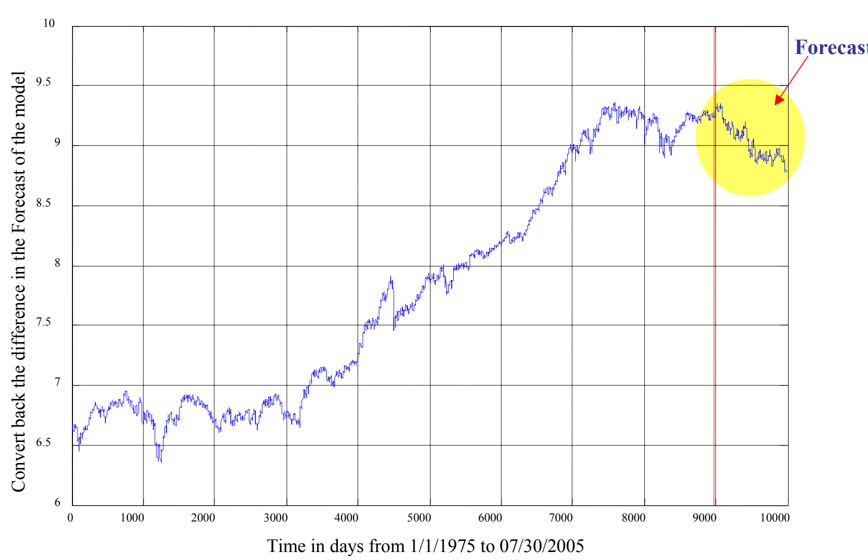
Transforming a Non-Stationary Process to a Stationary Process

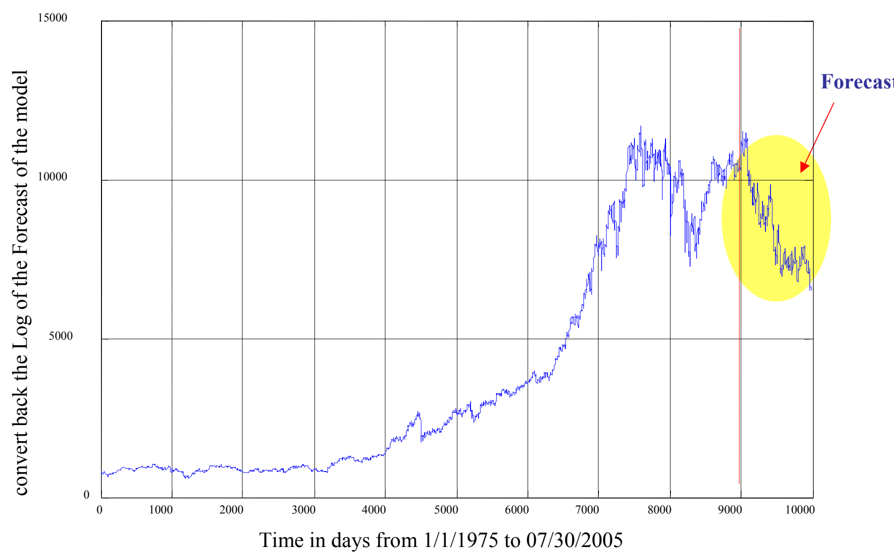


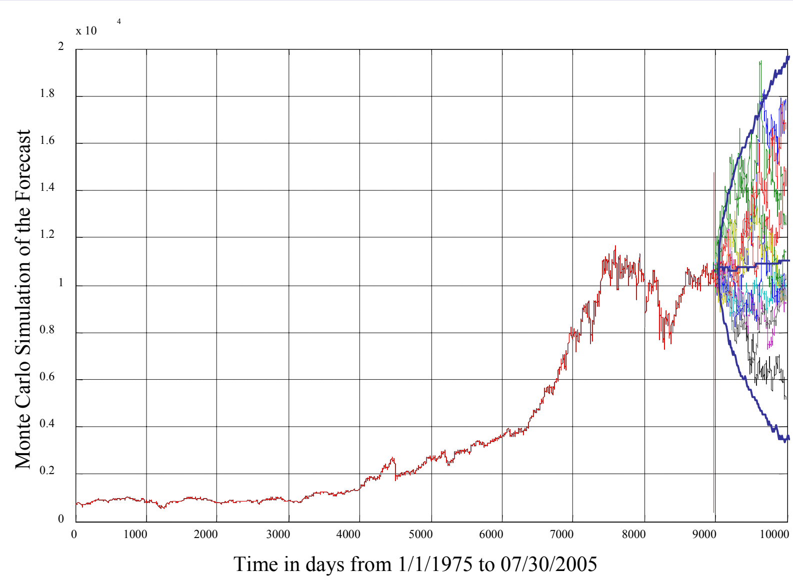
Transforming a Non-Stationary Process to a Stationary Process



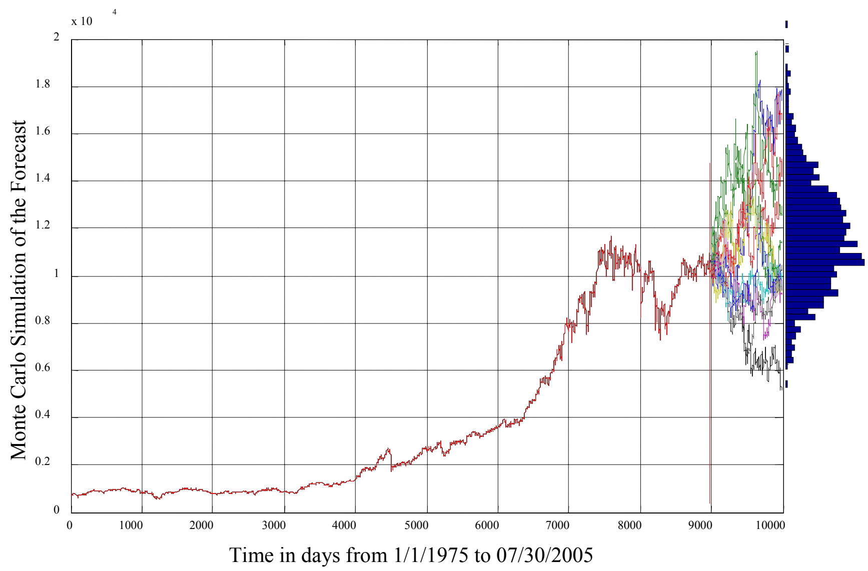

Forecast of the model
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Time in days from 1/1/1975 to 07/30/2005
Forecast


Convert back the difference in the Forecast of the model
10
9.5
9
8.5
8
7.5
7
6.5
6
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Time in days from 1/1/1975 to 07/30/2005
Forecast


15000
10000
5000
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
convert back the Log of the Forecast of the model
Time in days from 1/1/1975 to 07/30/2005
Forecast


x 10
4
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Monte Carlo Simulation of the Forecast
Time in days from 1/1/1975 to 07/30/2005

Transforming a Non-Stationary Process to a Stationary Process



Transforming a Non-Stationary Process to a Stationary Process