# Financial Modeling and Analysis
# Course Code: MATH 242
# Module 3: Exploratory Data Anaylsis

Dr. Zahra Lakdawala

July 3, 2020

# Outline

# Day 1: Time series

# Statistical Analysis of financial data

- Randomness means financial risk
- Also, opportunity for profit
- Some time series can show risk-free information
- Forecasting models results in time series with risk
- Use of statistics to understand randomness and predict movement and quantify volatility

# Randomness in data

**Time Series**

Financial data over time comes with some form of random variation

There exist methods for reducing or cancelling the effect due to random variation. 'Smoothing':

1. Simple Moving averages - takes a certain number of past periods and add them together

$$MA_{t+1} = \frac{[D_t + D_{t-1} + \cdots + D_{t-n+1}]}{n}$$

2. reduces irregularities
3. useful for filtering 'white' noise
4. series emphasizes certain informational components in the time series

# Day 2: PDE

# Probability Density

- Relationship between observations and their probability
- Some outcomes have low probability outcomes compared to the other
- Overall shape of the probability density is referred to as a probability distribution
- Calculation of probabilities for specific outcomes of a random variable is done by PDF
- Probability density must be approximated using a probability density estimation

# Content cover

- Histogram plots provide a fast and reliable way to visualize the probability density of a data sample
- Parametric probability density estimation involves selecting a common distribution and estimating the parameters for the density function from a data sample
- Non parametric probability density estimation involves using a technique to fit a model to the arbitrary distribution of the data, e.g. kernel density estimation

# Kernel Density Estimation

- Technique to create a smooth curve given a set of data
- Useful to visualize the 'shape' or distribution of some data
- Continous replacement for the discrete histogram
- Can also be used to generate points that look like they came from a certain dataset
- Used to power simulations - where simulated objects are modelled off of real data
- Inferences about the data 'distribution' is made

# So, how does it work?

- Start with some points sampled from some unknown distribution
- As more points build up, its will start corresponding to a distribution - marginal unconditional distribution
- KDE takes a 'bandwidth' that affects how smooth the resulting curve is.
- If we've seen more points nearby, the estimate is higher, indicating that probability of seeing a point at that location.
- Changing the bandwidth changes the shape of the kernel.
    - lower bandwidth means only points very close to the current position are given any weight

# Kernel Density Estimator

The concept of weighting the distances of our observations from a particular point $x$ is expressed as:

$$\hat{f}(x) = \frac{1}{nb} \sum_{observations} \mathbf{K}(\frac{x - X_i}{b})$$

- $\mathbf{K}$ denotes the Kernel function
- $b$ denotes the bandwidth
- $X_i$ denotes the observations
- Using different kernel and bandwidth produces different estimates

# Choice of bandwidth

- Small/large value of $b$ allows the density estimator to detect/obscure fine features in the true density.
- Small/large value of $b$ permits high/low degree of variation
- Small values of $b$ causes the KDE to have a low bias
- Its a tradeoff between variance and bias
- Overfitting and Underfitting problem
- Are you confused?

# Choice of bandwidth

- Small/large value of $b$ allows the density estimator to detect/obscure fine features in the true density.
- Small/large value of $b$ permits high/low degree of variation
- Small values of $b$ causes the KDE to have a low bias
- Its a tradeoff between variance and bias
- Overfitting and Underfitting problem
- Are you confused?
  - There is no simple answer!

# Choice of K

Kernel density estimate is used to suggest a parametric statistical model

- bell shaped - normal/gaussian distribution
  - normal density mean = sample mean of returns
  - standard deviation = standard deviation of returns
  - There could be more refined ways of chosing the mean and standard deviation using sample median and MAD estimators.
- uniform/tophat, exponential, cosine, and many more

# Summarizing KDE

- KDE suggests a way to model the distribution of the data in the sample
- Parameters must be estimated properly
- Simple to compute, but still comes with a few issues.
- Stay tuned for further improvizations!

# Day 3: ECDF

# Don't we love Gaussian Distributions?!

# Why EDFs?

- Histograms - easy way to visualize a density plot, BUT...
- Bin size problem: wrong bin size = wrong depiction of the data distribution
- Also, what about visualizing multiple variables at the same time?!

**Empirical/Sample distribution functions**

- No binning required
- Visualize many distributions together

# Empirical CDF

$$F_n(y) = \frac{\sum_{i=1}^{n} I\{Y_i \leq y\}}{n}$$

- $I\{.\}$ is the indicator function so that $I\{Y_i \leq y\}$ is 1 is $Y_i \leq y$ and is 0 otherwise.
- Sum in the numerators counts the number of $Y_i$ that are less than or equal to $y$
- True cdf vs sample cdf? Difference comes because of the 'random variation'

# Day 4: Quantiles

# Day5: Transformations

# Dilemma: Data and Statistics

- Statistical methods work best when data is normally distributed (or atleast symmetrically distributed)
  - constant variance
  - less skewness
- Reality does not always conform to the needs of statistics.

# Dilemma: Data and Statistics

- Statistical methods work best when data is normally distributed (or atleast symmetrically distributed)
  - constant variance
  - less skewness
- Reality does not always conform to the needs of statistics.

So what do we do?

# Transformed data

Data Analysts usually don't work with original variables

- Transform data, such that there is constant variance compared to original variables (minimize skewness)
- Commonly used transformations
  - Log transformations
  - Square root transformations
  - Power transformations
- Chose transformation to stabilize variance (removes dependence between conditional variance and conditional mean of a variable.

# Log Transformation

Log Transformation is widely used

## Log strength

- Stabilizes the variance of a variable whose conditional standard deviation is proportional to its conditional mean.
- Changes in log returns have relatively constant variability (compared to changed in returns)
- Super power: log transformations can be embedded into power transformations

# Box Cox Power Transformation

Log Transformation embedded into power transformation

$$y^{\alpha} = \begin{cases} \frac{y^{\alpha}-1}{\alpha} & ,\alpha \neq 0 \\ log(y) & ,\alpha = 0 \end{cases}$$

Since $\lim_{x \to 0} \left( \frac{y^{\alpha}-1}{\alpha} \right) = log(y)$, the transformation is continuous in $\alpha$ at 0.

## Choice of $\alpha$

It is commonly the case that the response is right-skewed and the conditional response variance is an increasing function of the conditional response mean. In such cases, a concave transformation, e.g. a Box–Cox transformation with $\alpha < 1$, will remove skewness and stabilize the variance

- The value of $\alpha$ that is best for symmetrizing the data is not the same value of that is best for stabilizing the variance.