Math

Cox–Ingersoll–Ross (CIR) process

dr_t=$\alpha(b - r_t)$dt+$\sigma\sqrt{r_t}$dW_t
where:
$\sqrt{r_t}$: avoids possibility of negative interest rate.

where:
W_t: Wiener process under the risk neutral framework
b: long term mean level
α: speed of reversion
σ: volatility

In mathematical finance, the Cox–Ingersoll–Ross (CIR) model describes the **evolution of interest rates**. It is a type of "one factor model" (short-rate model) as it describes interest rate movements as driven by only one source of market risk. The model can be used in the valuation of interest rate derivatives.

It ensures **mean reversion** of interest rate towards the **long run value** with **speed of adjustment** governed by the strictly positive parameter.

Valuing MBS:
It is necessary to project its cash flows. The difficult is that the cash flows **are unknown** because of **prepayment.**

The *prepayment rate* is sometimes referred top as *the speed*.

The **factors** that affect **prepayment behavior** are prevailing mortgage rate, characteristics of the underlying mortgage pool, seasonal factors, and economic factors.

The cash flow of MBS is *interest rate path dependent*.

MBS pricing factors:

Spreads: If spreads widen, price of MBS declines. If spreads tighten, price of MBS increases.

Prepayment: If the principal is being returned to the investor at par, this can detract from returns if the MBS was purchased at premium price.

Convexity: MBS price/rate relationship is **negatively convex** due to the likelihood the homeowner will **refinanc**e into a **lower rate** mortgage when **yields falls**, and conversely, the homeowner's tendency to **remain** in the mortgage when **yields rise**.

When rates increase and prepayment slow, the duration extends. This can cause MBS prices to decrease during rate selloffs more than they increase when the rate rallies.

Volatility: Due to the homeowner's embedded options, MBS are exposed to changes in volatility. An **increase in volatility** can result in **MBS underperformance.**

MBS prepayment risk
Spread of current vs original mortgage rates.
Mortgage rate path (refinancing burnout)
Housing turnover
Loan seasoning and property location
Contraction risk occurs as rates fall, prepayments rise, average life falls.
Extension risk occurs as rates rise, prepayments fall (slow), average life rises.

Prepayment risk: contraction risk from faster prepayment; extension risk from slower prepayment.

Estimate the value of Pi using Monte Carlo
The idea is to simulate random (x, y) points in a 2-D plane with domain as a square of side 2r units centered on (0,0). Imagine a circle inside the same domain with same radius r and inscribed into the square. We then calculate the ratio of number points that lied inside the circle and total number of generated points. Refer to the image below:

Take ratio of two areas:

$$\frac{area\ of\ circle}{area\ of\ the\ square} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4}$$

First take ratio of two areas
Second: generate random points for x and y
Third: set boundary condition in which d = x*x + y*y $\leq$ 1
Fourth: we increment the number of points that appears inside the circle.
Fifth: Calculate pi = 4*(circle_points/square_points).

Algorithm:
1. Initialize circle_points, square_points and interval to 0.
2. Generate random point x.
3. Generate random point y.
4. Calculate d = x*x + y*y.
5. If d <= 1, increment circle_points.
6. Increment square_points.
7. Increment interval.
8. If increment < NO_OF_ITERATIONS, repeat from 2.
9. Calculate pi = 4*(circle_points/square_points).
10. Terminate.

Optimal early exercise boundary
https://github.com/cantaro86/Financial-Models-Numerical-Methods/blob/master/2.3%20American%20Options.ipynb

$S_t < S_f$: stopping region
$S_t > S_f$: continuation region
In the stopping region, the value of the options corresponds to its intrinsic value which means $V(t, S) = K - S_t$

In order to find S_f, we have to find the maximum value s such that $V(t, s) - (K - s) = 0$

Covariance

- Covariance is a statistical tool that is used to determine the relationship between the movements of two random variables.
- When two stocks tend to move together, they are seen as having a positive covariance; when they move inversely, the covariance is negative.

Stochastic volatility

y_t = u + $e^{\frac{h_t}{2}} z_t$

z_t~iid. N(0,1)

h_t=h_{t-1}+η_t

This is SSM with states being h_t = lnσ_t^2, but measurement equation is not a linear function of h_t. We wish to estimate h^T=$(h_1, h_2, \ldots, h_T\ )'$.

2.1.2 Stochastic volatility model (SVM)

On the contrary, the SVM, introduced by Taylor (1986), is a nonlinear state space model where the measurement equation is nonlinear. The SVM is especially designed for financial time series that are subject to volatility clustering. One major feature is that the SVM can accommodate time-varying volatility that breaks the assumption of constant variance in the standard volatility model. The SVM can be described as:

$y_t = \mu + \varepsilon_t^y,\ \varepsilon_t^y \sim \mathrm{N}(0, e^{h_t})$ (3)

$h_t = \mu_h + \ \phi_h(h_{t-1} - \ \mu_h) + \varepsilon_t^h,\ \varepsilon_t^h \sim \mathrm{N}(0, w_h^2)$ (4)

Additionally, the **log-volatility** h_t follows a stationary **AR(1) process** with $|\ \phi_h| < 1$ and unconditional mean μ_h.

Implied volatility

In the Black-Scholes model, volatility is the only parameter that can't be directly observed. All other parameters can be determined through market data and this parameter is determined by a **numerical optimization technique** given the Black-Scholes model.

<u>Newton-Raphson method</u>

The Newton-Raphson method uses an **iterative procedure** to solve for a root using information about the derivative of a function.

Newton's method is a method for finding increasingly improved approximations to the roots of a function.

http://www.codeandfinance.com/finding-implied-vol.html

<u>Bisection method (Use this)</u>

The bisection method is considered to be one of the **simplest and robust root finding**

Ways to compute Greek

<u>Finite difference approximations</u>

<u>Pathwise method (Use this)</u>

differentiating the options's payoff and then take the **expectation under the risk neutral measures.**

Where:

Y is the options payoff as Y = $\boldsymbol{e^{-rT}(S_T - K)^+}$

$S_T = S_0 e^{\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z}$

Z~N(0,1).

The **standard normal distribution**, also called the *z*-**distribution**, is a special normal distribution where the mean is 0 and the standard deviation is 1.

Any normal distribution can be standardized by converting its values into *z*-scores. *Z*-scores tell you how many standard deviations from the mean each value lies.

Delta $= \frac{\partial Y}{\partial S_0}$

Vega $= \frac{\partial Y}{\partial \sigma}$

A state space model typically consists of two modelling levels:

The first level: observations are related to latent or unobserved variables called sates according to the observation or measurement equation

The second level: the evolution of the states is modelled via the state or transition equation.

Estimator and BLUE

B: random variable

Let $\tilde{b}$ be an estimator of B

Is b a good estimator?

If E(b) =B, then b is unbiased

Is b the best(minimum variance) estimator?

If Var(B) ≤ Var($\tilde{b}$) for any estimator $\tilde{b}$, then b will be the best estimator

If Gauss Markov are met, then OLS estimators alpha and beta are BLUE-The best linear unbiased estimators.

Best: **Variance of OLS is minimal,** smaller than the variance of any other estimator
Linear: If the relationship is not linear-OLS is not applicable
Unbiased: The **expected value** of estimated beta and alpha **equal** the true values describing the relationship between x and y.

QQ plot
QQ plot displays the standardized residuals on the y-axis and the theoretical quantiles on the x-axis. (Note**: Residual = Observed – Predicted**)

Data that aligns closely to the dotted line indicates a **normal** distribution. If the points **skew drastically from the line**, you could consider adjusting your model by adding or removing other variables in the regression model.

Residual plot
Ideal plots: symmetrically distributed, tending to cluster towards the middle of the plot. There are not any clear patterns.

Problematic plots: not evenly distributed, have outlie, have clear shape.

Forward rate

$$\frac{(1+r_a)^{t_a}}{(1+r_b)^{t_b}} - 1$$

Problems of using regression
Omitted variables, reverse causality, measurement error, sample selection
Solutions: randomized controlled experimental design and instrumental variable

Fixed effect
Fixed effects are variables that are **constant across individuals**; these variables, like age, sex, or ethnicity, don't change or **change at a constant rate over time**. They have fixed effects; in other words, **any change they cause to an individual is the same.** For example, any effects from being a woman, a person of color, or a 17-year-old will not change over time.

Random effect
Random effects regression is suited for longitudinal or panel data. The availability of

repeated observations on the same units allows the researcher to enrich the model by inserting an additional term in the regression**, capturing individual-specific, time-invariant factors** affecting the dependent variable **but unobserved** to the econometrician.

Convex function

A twice-differentiable function of a single variable is convex if and only if its second derivative is nonnegative on its entire domain.

Positive definite matrix (basis of convex optimization)

Positive-definite and positive-semidefinite real matrices are at the basis of convex optimization, since, given a function of several real variables that is twice differentiable, then if its Hessian matrix (matrix of its second partial derivatives) is positive-definite at a point p, then the function is convex near p, and, conversely, if the function is convex near p, then the Hessian matrix is positive-semidefinite at p.

Eigenvector

How the variables of the system fluctuate together

Eigenvalue

Factor by which the eigenvector is scaled

Symmetric matrices

Has only real eigenvalues
Is always DIAGONALIZABLE
HAS ORTHOGONAL EIGENVECTORS.

Orthogonal

an orthogonal matrix in which its **transpose** is **equal** to its **inverse matrix.**

Entropy

Entropy is a way of measuring the **uncertainty/randomness o**f a random variable X. In other words, entropy measures the **amount of information** in a **random variable**. It is normally measured in bits. **KL divergence** is the measure of the **relative difference between two**

probability distributions for a given random variable or set of events. KL divergence is also known as Relative Entropy.

Bayesian

Uncertainty is summarized by the prior. **Information** in the data comes through the **Likelihood.** After observing the data, we update our uncertainty. **Posterior** is the **new** summary of our **uncertainty.**

The central object of interest is the posterior pdf f($\theta|x$) which, by Bayes' theorem is

proportional to the product of the prior and likelihood:

f($\theta|x$) ∝ f((x|θ)f(θ)

CVA

CVA=$\mathbf{-E[}\boldsymbol{I(\tau \leq T) * V(\tau,T)_{+} * (1-\delta)}\mathbf{]}$

$I(\tau \leq T)$: Default time is prior to final maturity

$V(\tau, T)_{+}$: Exposure at default (discounted)

$(1-\delta)$: Loss given default % (LGD)

Increase in credit spread of the counterparty result in increase in the CVA (less negative). However,

the relationship is non-linear. If credit spread increased from 300 to 1200, the CVA **be more negative**. When credit spread increased from 1200 to 9600, the CVA will be less negative.

Increase in recovery increases **the CVA (less negative).**

Math of Logistic regression

Logistic regression **estimates the probability that response variable** belongs to a particular category. It is a process of **modeling the probability of a discrete outcome** given an input variable. The most common logistic regression models a binary outcome; for example, true/false, yes/no, etc.

Logit fits a logit model for a binary response by maximum likelihood; it models the probability of a positive outcome given a set of regressors.

Logistic regression is defined as a supervised machine learning algorithm that accomplishes binary classification tasks by **predicting the probability of an outcome,** event, or observation.

f(x)= $\frac{1}{1+e^{-x}}$

where:
e= base of natural logarithms

The following equation represents logistic regression:

y = $\frac{e^{(b_0+b_1X)}}{1+e^{(b_0+b_1X)}}$

X=input value
y=predicted value

Maximum likelihood estimation

The MLE are those values of the parameters that maximize the likelihood function with respect to the parameter θ.

Step 1: Define the likelihood function L(θ)

Step 2: Take the natural logarithm ln of L(θ)

Step 3: Differentiat ln L(θ) with respect to θ and then equate the derivative to zero

Step 4: Solve of the parameter θ and we will obtain $\hat{\theta}$

Step 5: Check whether it is a maximizer or global maximizer

How to regularize deep NNs:
Shrinkage models (Combat Overfitting)

How to hedge against overfitting: Shrinkage models

Both try to shrink the coefficient by adding a penalty to the sum of squares of the residuals.

L1: Lasso: Reducing variance by reducing weights
L2: Ridge: Reducing cost function and mitigate problem of multicollinearity

Confidence interval: gives range of values the mean value will be between with a given probability.

ANOVA

ANOVA means analysis of variance. Total = regression fit + residual error.

Goodness of fit-coefficient determination

$R^2=\frac{SSE}{SST}=1-\frac{SSR}{SST}$

Supervised machine learning (Both Y and X)

You have an outcome Y and predictor X. You fit the model Y to predict using unseen data X_0.

Unsupervised machine learning (No Y but only X)

No output variable, only inputs.
Dimension reduction: reduce the complexity of your data. Examples: PCA, cluster analysis. Then can be used to generate inputs for supervised learning from the principal component regression.

Dimensionality reduction (Unsupervised)

Dimensionality reduction algorithms **project high-dimensional data to a low-dimensional** space while retaining as much of the variation as possible. There are two main approaches to dimensionality reduction.

- The first one is known as linear projection which involves linearly projecting data from a high-dimensional space to a low-dimensional space. This includes techniques such as principal component analysis (PCA).

- The second approach is known as manifold learning which is also referred to as nonlinear dimensionality reduction. This includes techniques such as Uniform manifold approximation and projection (UMAP).

Dimensionality reduction techniques help to address the curse of dimensionality.

PCA (Unsupervised)

Step 1: Find **highly correlated features** and combine them
Step 2: Represent the data with **smaller** number of linearly uncorrelated features.
Step 3: Algo continues the **correlation reduction**
Step 4: Then **find** the direction of **maximum variance** in the original high dimensional data

Step 5: Project **onto** a **smaller** dimensional space
Step 6: **Newly derived** components are called **principal components**
Step 7: Then we can make prediction based on the important factors

Monte Carlo

Variance reduction
In mathematics, more specifically in the theory of Monte Carlo methods, variance reduction is a procedure used to **increase the precision** of the estimates obtained for a given simulation or computational effort.

Importance sampling
Importance sampling is a variance reduction technique that can be used in the Monte Carlo method. The idea behind importance sampling is that **certain values** of the input random variables in a simulation hav**e more impact** on the parameter being estimated than others. If **these "important" values are emphasized** by **sampling more frequently,** then the **estimator variance can be reduced.** Hence, the basic methodology in importance sampling is to choose a distribution which "encourages" the important values.

The basic idea of importance sampling is to sample the states from a different distribution to lower the variance of the estimation

Antithetic variates
This method attempts to reduce variance by introducing **negative correlation** between **pairs of observations.**

Control variates
The technique of control variates takes advantage of this **additional piece of information** to reduce the variance of a Monte Carlo estimator.

Accept-reject methods
There are many distributions for which the inverse transform method and even general transformations will fail to be able to generate the required random variables. For these cases, we must turn to indirect methods; that is, methods in which we **generate a candidate random variable** and only **accept** it **subject to passing a test.**

3. Description of solution methods

3.1 Solution method for HTSVM

The entire numerical solution for HTSVM requires the Bayesian method that utilizes a branch of Markov Chain Monte Carlo (MCMC) namely the **Metropolis-Hasting algorithm** (MHA) which is estimated in MATLAB. The burn-in (iterations) and loop for the MHA of MCMC are set at 1000 and 5000, respectively, which are arbitrary. Like Monte Carlo integration and Gibbs sampling, the MHA produces **a sequence of draws $\boldsymbol{\theta}^{(r)}$** for r = 1, …, R with the property $\hat{g}$ = $\frac{\sum_{r=1}^{R} g(\theta^{(r)})}{R}$ converges to E[g(θ)|y] as R goes to infinity. It involves drawing from a convenient density related to the importance function, referring to as the candidate generating density. **Candidate draws of $\boldsymbol{\theta}^{(r)}$ are either accepted or rejected w**ith a certain probability referred to as an acceptance probability. If they are rejected, then $\theta^{(r)}$ is set to $\theta^{(r-1)}$.

JCC on MHA:

A transition of the Markov chain from state x is carried out in two phases. Similar to the acceptance-rejection method, first a trial or proposal state Y is drawn from a transition density q(·|x). This state is accepted as the new state, with probability α(x,Y), or rejected otherwise.

The so-called accept-reject methods only require us to know the functional form of the density f of interest up to a multiplicative constant.

We use a simpler density g, called the instrumental or candidate density to generate the random variable for which the simulation is actually done.

Least Square Monte Carlo American options

-Backward induction process: evaluate at each **discrete** time point

-Optimal decision between immediate exercise and payoff of the options and the discounted expected continuation value

-Then calculated as the fitted value of a least squares regression of basis function

-Convergence property: American options value by the LSM converges to the true value as the number of simulations and basis functions increase.

Ordinary least square method

1. Set a difference between dependent variable and its estimation:
2. Square the difference:
3. Take summation for all data.
4. To get the parameters that make the sum of square difference become minimum, take partial derivative for each parameter and equate it with zero,

Swaption (Options on interest rate swap)

There are two types of swaption contracts (analogous to put and call options):[1]

- A payer swaption gives the owner of the swaption the right to enter into a swap where they pay the fixed leg and receive the floating leg.
- A receiver swaption gives the owner of the swaption the right to enter into a swap in which they will receive the fixed leg, and pay the floating leg.

In addition, a "straddle" refers to a combination of a receiver and a payer option on the same underlying swap.

-
- A swaption contract contains terms and conditions of the swaption and the underlying interest rate swap. For example, it specifies two maturities: swaption maturity and underlying swap maturity/tenor.
- The valuation model for pricing a swaption is the Black formula.
- First, one needs to generate the cash flows of the underlying interest rate swap. The generation is based on the start time, end time and payment frequency of each leg, plus calendar (holidays), business convention (e.g., modified following, following, etc.) and whether sticky month end.
- The accrual period is calculated according to the start date and end date of a cash flow plus day count convention
- Then you need to construct interest zero curve by bootstrapping the most liquid interest rate instruments in the market. FinPricing provides useful tools to build various curves, such as swap curve, basis curve, OIS curve, bond curve, treasury curve, etc. Go to the list of the tools
- Any compounded interest yield curves can be used to compute discount factor, of course the formulas will be slightly different. The most common used one is continuously compounded zero rates.

- Another key for accurately pricing an outstanding swaption is to construct an arbitrage-free volatility surface. Unlike a cap implied volatility surface that is 3 dimensional (maturity – strike – volatility), a implied swaption volatility surface is 4 dimensional (swaption maturity – underlying swap tenor – strike – volatility).
- FinPricing is using SABR model to contruct swaption implied volatility. Go to the list of volatility construction tools
-

Exotic options

Asian options

Asian options are securities with payoffs that depend on the average value of an underlying asset over a specific period of time. Underlying assets can be stocks, commodities, or financial indices.

Two types of Asian options are found in the market: average price options and average strike options. Average price options have a fixed strike value and the average used is the asset price. Average strike options have a strike equal to the average value of the underlying asset.

The payoff at maturity of an average price European Asian option is:
max(0,Savg−K) for a call
max(0,K−Savg) for a put

The payoff at maturity of an average strike European Asian option is:
max(0,St−Savg) for a call
max(0,Savg−St) for a put

Lookback options

A *lookback* option is a path-dependent option based on the **maximum** or **minimum** value the underlying asset (e.g. electricity, stock) achieves during the entire life of the option. Basically the holder of the option can ‘look back’ over time to determine the payoff. This type of option provides price protection over a selected period, reduces uncertainties with the timing of market entry, moderates the need for the ongoing management, and therefore, is usually more expensive than vanilla options.

Lookback call options give the holder the right to **buy** the underlying asset at the **lowest** price.

Lookback put options give the right to **sell t**he underlying asset at the **highest price.**

Financial Instruments Toolbox™ software supports two types of lookback options: fixed and floating. The difference is related to how the strike price is set in the contract. Fixed lookback options have a specified strike price and the option pays out the maximum of the difference between the highest (lowest) observed price of the underlying during the life of the option and the strike. Floating lookback options have a strike price determined at maturity, and it is set at the lowest (highest) price of the underlying reached during the life of the option. This means that for a floating strike lookback call (put), the holder has the right to buy (sell) the underlying asset at its lowest (highest) price observed during the life of the option. So, there are a total of four lookback option types, each with its own characteristic payoff formula:

- Fixed call: max(0,Smax−X)
- Fixed put: max(0,X−Smin)
- Floating call: max(0,S−Smin)
- Floating put: max(0,Smax−S)

where:

Smax is the maximum price of underlying asset.

Smin is the minimum price of underlying asset.

S is the price of the underlying asset at maturity.

X is the strike price.

Euclidean norm

On the **n-dimensional Euclidean space** $\mathbb{R}^{\boldsymbol{n}}$, the intuition notion of length of the vector x=$(x_1, x_2, \ldots, x_n)$ is captured by the formula:

$\|x\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$

Similarly:

$\|x\|:=\sqrt{xx}$

The Euclidean norm is also called the L^2 norm

Local volatility model

LVM has one factor: only the stock price is stochastic and so most of the standard BSM schem for perfect replication in terms of a riskless bond and stock still works.

With LVM, we can use **risk neutral valuation methods** to obtain unique arbitrage free options values, just as for BSM.

Evolution of stock price in LVM:

$\frac{dS}{S}$=u(S,t)dt + $\sigma(S,t)$dZ

Note that $\sigma(S,t)$ is a deterministic function of a stochastic variable S.

Advantage of LVM:

Once calibrated, the LVM provides **arbitrage free option values** and hedge ratios for standard and exotic options. It has **closeness to the original BSM** model and its dynamics.

Disadvantage of LVM:

One general objection to LVM is that it needs to be **frequently calibrated**. As time passes and the underlying stock price or index level changes, the implied volatility surface changes, and a new local volatility surface must be extracted from the data.

New hedge ratios and exotic option values must then be calculated from this updated surface. The **parameters** of the model are **not stationary.**

LVM tend to have **difficulty matching the future short term skew.**

Constant Elasticity of Variance (CEV) by Schroder in 1989

CEV is related to **Bessel process**: It is a Brownian motion tiled by a function of the current value. It is a one-dimensional diffusion processes.

The **Bessel process** favor larger values when the α (Brownian motion weighted locally) >0 and favor smaller value when α < 0.

CEV is one of the LVM.

Evolution of stock price in CEV:

$\frac{dS}{S}$=u(S,t)dt + $\sigma S^{\beta-1}$dZ

In this model, the volatility is proportional to $S^{\beta-1}$, where β is a constant to be determined by calibration. If β =1, then the CEVM reduces to standard lognormal geometric Brownian motion. When β =0, returns are normally distributed. In order to account for the observed **skew** in equity markets, $\boldsymbol{\beta}$ **needs to be negative and large in magnitude**.

In Schroder's original CEV paper, the CEV:

$$\sigma(S, t) = \delta S^{(B-2)/2}$$

The **elasticity of return variance** with respect to price equals B−2, and if B<2, **volatility and price** are **inversely related.** The stock price is assumed to be governed by the diffusion process:
dS=uSdt + $\delta S^{B/2}$dZ
where:
dZ is a Weiner process.
If B=2 (The elasticity is zero), prices are lognormally distributed and the variance of returns is constant and is assumed in the Black Scholes model.

The paper presents an efficient algorithm for computing the noncentral chi-square distribution and applies this to the CEV options pricing formula. The Bessel function is also part of the CEV model.

SABR stochastic volatility (For interest rate derivatives)
In mathematical finance, the SABR model is a stochastic volatility model, which attempts to capture the **volatility smile in derivatives markets**. The name stands for "stochastic alpha, beta, rho", referring to the parameters of the model. The SABR model is widely used by practitioners in the financial industry, especially in the **interest rate derivative** markets.

dF_t=$\sigma_t (F_t)^B$dW_t
dσ_t=$\alpha \sigma_t$dZ_t
dW_t dZ_t=ρdt

Where:
F: LIBOR forward rate
σ: volatility of the forward F
W_t and Z_t are two correlated Wiener processes

B: Constant parameter
ρ: instantaneous correlation between the underlying and its volatility
α: height of the ATM implied volatility level

Geometric Brownian Motion
A geometric Brownian motion (GBM) (also known as exponential Brownian motion) is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion (also called a Wiener process) with drift.[1] It is an important example of stochastic processes satisfying a stochastic differential equation (SDE); in particular, it is used in mathematical finance to model stock prices in the **Black–Scholes** model.

Local Volatility model
-Current options price determine a single consistent set of local volatilities that in theory should remain unchanged as time passes and the stock prices moves.
-The surface is a deterministic function of t and S, over time that surface moves in an unpredictable manner.
-Underlying follow**s geometric Brownian motion** with **deterministic forward volatility** under the risk neutral measure.
-Direction computation of local volatility function using finite difference method is problematic.
-The local volatility surface can be very irregular and sensitive to the interpolation method

-**Market prices** based: It used the option prices to lock in the forward volatilities of the underlying, which are a function of the underlying asset price and time.
-Calibration has two approaches: Derman and Kani (1994) and Dupire (1994)

Calibration of market local volatility surface
-**Dupire's equation** to calibrate forward volatility (local volatility surface) to prices of standard European options

Disadvantage of local volatility
Fail to capture the proper dynamics of IV
Hedge ratio incorrect

Parametric Local volatility

-Examples: CEV, shift-lognormal, quadratic
-Calibrate by bootstrap and least squares

Non-parametric local volatility
-Fit the local volatility to Dupire local volatility
-Good for indices with liquid option market.

Stochastic local volatility
-It is equivalent to a multivariate stochastic volatility diffusion, called the "market model" of implied volatility.
-Let v(t) to evolve stochastically

CEV model
-Good for hedging but difficult for pricing due to **parameter instability**

Carol Alexander's calibration method:
Each model was calibrated daily by **minimizing** the **root mean square error (RMSE)** between the **model implied volatilities** and the **market implied volatilities** of the options used in the calibration set.

Yet, for the BS model the deltas and gammas are obtained directly from the market data and there is no need for model calibration.

Heston (1993) model: We used the closed form price based on Fourier transform and chose a volatility risk premium of zero and set the long term volatility at 12%. The calculation of the CEV options price is based on the non-centered chi-square distribution result of Schroder (1989).

RMSE
The RMSE of an estimator $\hat{\theta}$ with respect to an estimated parameter θ is defined as the square root of the mean square error:

$$\mathrm{RMSE}(\hat{\theta}) = \sqrt{MSE(\hat{\theta})} = \sqrt{E\left((\hat{\theta}-\theta)^2\right)}.$$

For unbiased estimator, the RMSE is just the square root of the variance, known as the standard deviation.

Normal Mixture options pricing model

The two options prices will be very close for near to ATM options but, as explained in the text, the normal mixture prices will tend to be well above the Black-Scholes prices for OTM and ITM options. For very OTM or ITM the difference between the two pricing models again becomes very small.

The normal mixture prices are also used as 'market' prices to 'back-out' implied volatilities (see section 2.1). These are shown in the Implied Volatility chart. For most options these normal mixture implied volatilities will have the familiar 'smile' shape. Thus normal mixture models can explain the volatility smile. Sujit's dissertation will be showing that normal mixture prices are closer to market prices for currency options.

As usual, the user must input the option characteristics. But now, instead of a single volatility to price the option under the assumption of returns normality, the user needs to input the parameters of a zero mean normal mixture. These parameters could be based on historical returns of the underlying, and found using the method of moments (use the worksheet normal mixture estimation.xls).

Heston model (One of the most popular Stochastic volatility model)

-Volatility is driven by a mean reverting square root process

-Generates closed form expression for call and put prices

Assume that the asset price S and variance V follow the model by Heston (1993):

$$dS_t = u(V_t)\, S_t\, dt + \sqrt{V_t} S_t \left(\rho dW_t^1 + \sqrt{1-\rho^2} dW_t^2\right)$$
$$dV_t = k(\theta - V_t)\, dt + \sigma \sqrt{V_t}\, dW_t^1$$

The variance process follows a square toot process and it is bounded below by zero. The relative rate of return u is taken to be a deterministic function of the variance.

Sticky strike

An option with a fixed strike will always have the same IV, that a particular IV value sticks to each strike. Under this rule, options with different strikes can still have different IV.

Sticky Delta

Sticky moneyness means that an option's volatility depends only on its moneyness K/S. This is an attempt to shift the skew as the stock price moves by adjusting for the option's moneyness. No matter what the stock price is, the volatility of the most liquid options should be the same for ITM and OTM.

Dupire

-Forward equation of option pricing

-allows one to price calls and puts with different strikes and maturities at once.

-no arbitrage

Calibration

-First approach: construct a continuous IV surface **matching the market quotes** by using either some parametric or non-parametric regression and then generates the corresponding surface via the Dupire formula.

-Second approach: Direct solution of the Dupire equation using either analytical or numerical methods.

Bond pricing

Duration

- **Macaulay duration** estimates **how many years** it will take for an investor to **be repaid** the bond's price by its total cash flows.
- **Modified duration** measures the price change in a bond given a 1% change in interest rates.

V=$\sum_{i=1}^{n} PV_i$

MacD = $\frac{\sum_{i=1}^{n} t_i PV_i}{\sum_{i=1}^{n} PV_i}$

PV_i: PV of the ith cash payment from an asset

t_i: time in years unitl the ith payment will be received

V is the present value of all future cash payments from the assets

ModD = $\frac{\text{MacD}}{(1+y_k/k)}$

k: compounding frequency per year (1 for annual, 2 for semi-annual, 12 for monthly, 52 for weekly, etc)

y_k: yield to maturity for an asset, periodically compounded.

Duration more review

The longer the maturity, the longer the duration

The lower the coupon, the longer the duration

Convexity

Convexity is the second derivative of the price function with respect to yield.

Convexity = $\frac{1}{P}\frac{\partial^2 P}{\partial y^2}$

Skew and kurtosis

Negative skew: Frequent small gain and few large loss

Positive skew: Frequent small loss and few large gain

<u>Entry order</u>

Kurt < 0 and if skew >0 is buy and if skew <0 is sell

Distribution

https://vitalflux.com/types-probability-distributions-defined-examples/

Tracking error

TE: $\sqrt{Var(Return_{trackign\ portfolio} - Return_{TW50})}$

Current situation

Dow's high was December 2021

10 year treasury yield's low was July 2020

Business Cycle

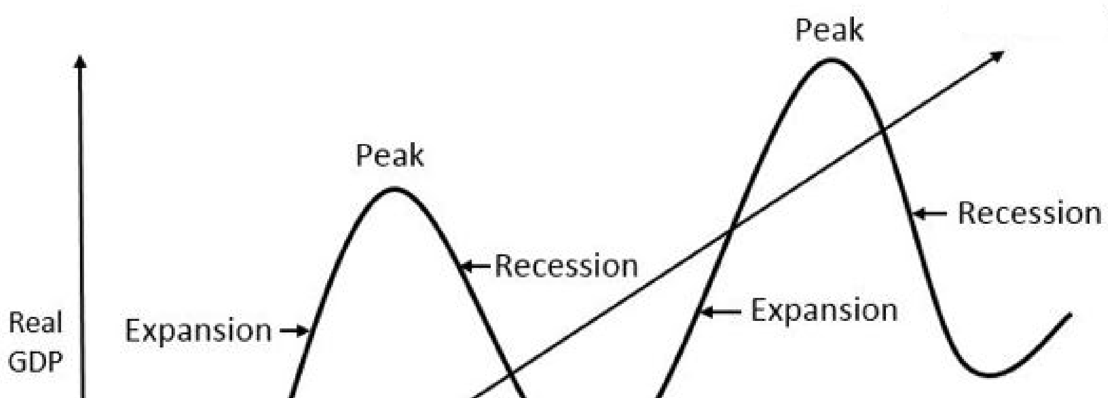
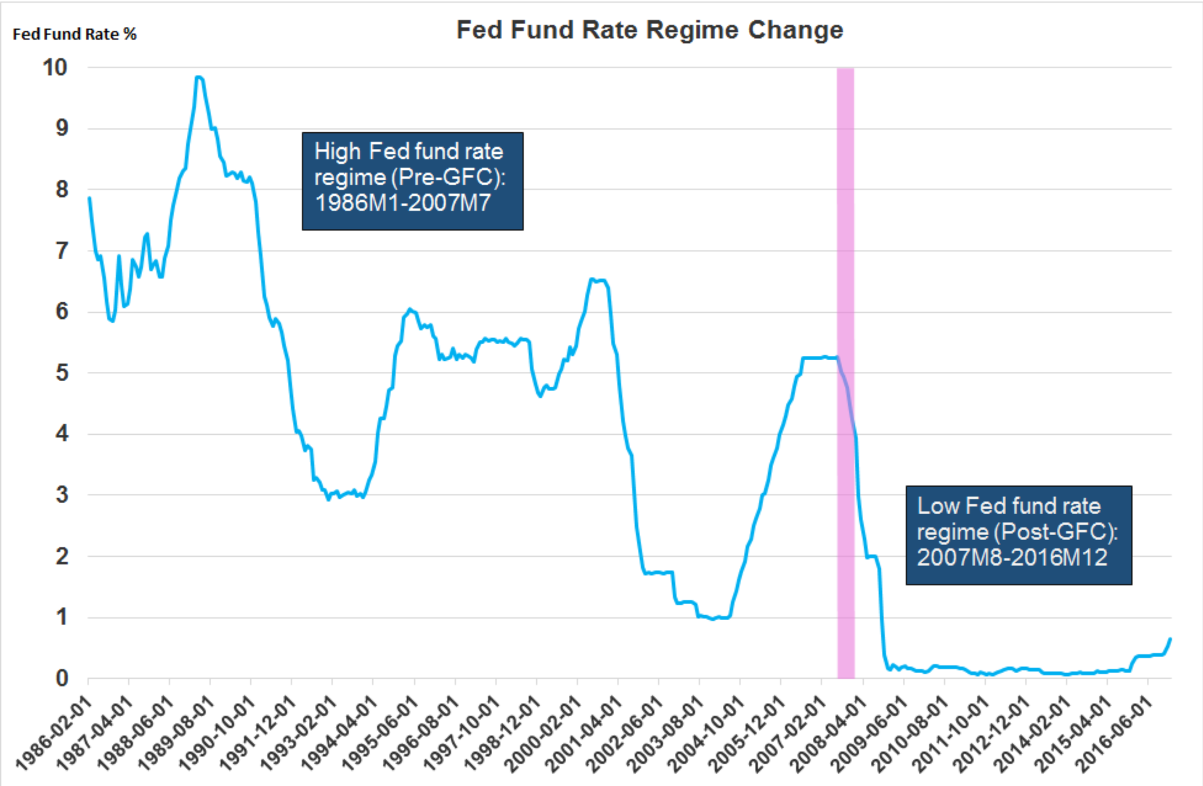


Figure 1: Trend and regime change in Fed fund rate from 1986 to 2016



Before the global financial crisis which began in August 2007, the Fed fund rate was also in increasing trend from 2004. The Fed Fund rate increased from 1% to 5%.

Iraq war: 2003 to 2011

Fed Fund rate now is at: 3.25

GARCH
Alpha: how volatility reacts to new information
Beta: The persistence of volatility
Alpha + Beta: overall persistence of volatility

GARCH (p, q) where AR(p) models the variance of the residuals and MA(q) models the variance of the process.

The **persistence of** a garch model has to do with how fast large volatilities decay after a shock. For the garch(1,1) model the key statistic is the sum of the two main parameters (alpha1 and beta1, in the notation we are using here).

The sum of alpha1 and beta1 should be less than 1. If the sum is greater than 1, then the predictions of volatility are explosive — we're unlikely to believe that. If the sum is equal to 1, then we have an exponential decay model.
It is possible to express the persistence as a half-life. The **half-life** is log(0.5)/log(alpha1 + beta1), where the units will be the frequency of the returns. When alpha1+beta1 hits 1, the half-life becomes infinite.

GARCH is a **weighted average** of **past squared residuals** but it has declining weights which never go completely to zero. It is good at **predicting conditional variance**. It asserts the **best** predictor of the **variance in the next period** is a **weighted average of the long run average variance**, the variance predicted for this period and the new information this period which is the **most recent squared residual**. Both EWMA and GARCH place more weight on recent information.

Note: A **residual** is the difference between an **observed** value and a **predicted** value in regression analysis.

Unconditional SD: Long run average
Conditional SD: Just like volatility which can be up and down

EWMA
Gives the **heaviest weight** on the last data. **Exponentially** weighted model give immediate reaction to the market crashes or huge changes.

EWMA vs GARCH

The exponentially weighted moving average (EWMA) volatility model is the recommended model for forecasting volatility by the Riskmetrics group. For monthly data, the lambda parameter of the EWMA model is recommended to be set to 0.97.

*EWMA:The EWMA does not consider the long run variance.
*GARCH: It believes that the variance will return to an average level in the long run.

EGARCH

In this paper, I select EGARCH for volatility estimation. Nelson (1991) invented EGARCH model so that good news and bad news will have different effect on the volatility (p. 351) or technically speaking, EGARCH is able to uncover the **asymmetric relationship between asset return and its volatility.**

Impulse response

impulse response functions are used to describe how the economy reacts over time to exogenous impulses, which economists usually call shocks, and are often modeled in the context of a vector autoregression.

VaR

VaR is non-convex and discontinuous wrt confidence level. **Difficult to control for non-normal distributions.**

It is the **maximum loss** of a firm's position due to market movement over a given holding period with a **given level of confidence.**

It offers a probability statement about the potential change in the value of portfolio resulting from a change n market factors over a specified period of time.

VaR is about how likely the threshold, eg 99^{th} percentile, will be exceeded.

Delta-Normal approach VaR

This approach is based on the assumption that the underlying market factors have a multivariate normal distribution. Using this assumption, one can determine the distribution of mark-to-market portfolio profits and losses has been obtained, standard mathematical properties of the normal distribution are used to determine the loss that will equaled or exceeded x percent of the time (ie, the VaR).

Value at risk = $-[(\text{Expected change in portfolio value}) - 1.65(\text{Standard deviaton of change in portfolio value})]$

VaR for one asset (parametric method)
VaR = Vol * Sqrt [T] *confidence level * notional amount
Where confidence level can be 2.33 if 1-day VaR 99%

Z:
90%: 1.29
95%: 1.65
99%: 2.33

Advantage of Delta Gamma VaR Monte Carlo
-does not require distributional assumptions on the risk factors. In particular, risk factors can be drawn from **skewed and leptokurtic distributions or historical samples.**
-does not require the computation of the moments of the distribution of DPV (unlike most parametric methodologies).
-does not require exact revaluation of every position in a given portfolio (unlike full Monte Carlo).

Expected Shortfall (ES) or CVaR

ES or CVaR can be interpreted as the expected loss that is incurred **when VaR is exceeded**. It is the **conditional expectation of loss** given that **VaR is exceeded.**
CVaR equals the **average** of some percentage of the **worse case loss scenarios.**
CVaR asks "If we are in the tail, what is the expected loss"?

Duration
Duration approximates the percent changes in price for a 100 basis point change in rates.

Modified duration
It is the approximate change in a bond's price given a 1% change in its YTM.

Macaulay duration

It estimates **how many years** it will take for an investor to **be repaid** the bond's price by its total cash flows.

Convexity

Curvature; relationship between **bond prices and bond yields.**

It is the second order derivative of the price wrt conventional yield.

Factors affecting duration:

<u>Time to maturity</u>: Shorter maturity, lower duration and less risk

<u>Coupon rate</u>: The higher the coupon rate, the lower the duration, and the lower the interest rate risk.

Risk neutral measure, arbitrage, and hedging

We have constructed a risk neutral measure $\tilde{P}$ a measure which is equivalent to P, and which makes the **discounted stock price a martingale.**

The mean rate of return of any asset under the risk neutral measure is equal to the risk free rate r.

Implied Volatility

With historical volatility, the initial price of an option can be established. However, after its introduction, an option is traded through an exchange and the price is determined by the market.

While there is no closed form for inverting Black-Scholes to solve for volatility including the option's price, the formula can easily be **solved numerically**. A simple and adequate method is **bisection.**

The method **starts with an estimate of volatility** σ_1 that is too low, and an estimate σ_2 that is too high. The computed options price for σ_1 is less than the actual price but for σ_2 it is greater. In this way the root we are seeking σ_0 is bracketed. Next, the option price is calculated for the **mid-volatility** $\sigma_m = \frac{1}{2}(\sigma_1 + \sigma_2)$

Historical volatility

HV = $\sqrt{\frac{\Sigma \ (R_i - R_{avg})^2}{n-1}}$

Degree of freedom

- Degrees of freedom refers to the **maximum number** of **logically independent values,** which are values that have the freedom to vary, in the data sample.
- Degrees of freedom is calculated by subtracting one from the number of items within the data sample.

Annualized volatility

AV = HV * $\sqrt{252}$

Mixed Integer Linear Programming (MILP)

MIP models with a **quadratic objective** but **without quadratic constraints** are called Mixed Integer Quadratic Programming (MIQP) problems.

Share a time when facing an unexpected challenge

Last semester we had to replicate a very difficult econometrics paper based on the semiparametric bootstrap for copula.
Moreover, the authors did not provide any code to the public, making this project even much more difficult.

At first, we went through all the similar academic review and then went through Github for coding ideas. At the end, my partner and I eventually successfully figure out how to structure the coding for the semiparametric bootstrap step by step.

In this project, we summarize the contribution of Chen and Fan (2006) by discussing their estimation methodology on the ***copula-based semiparametric*** *time series and making balanced critique on their strengths and weaknesses. Further, we update the literature review in order to know the latest development ever since their publication in 2006. Most importantly, we replicate the semiparametric* ***bootstrap method*** *suggested by Chen and Fan (2006) and finally, we conduct copula simulation and extend the research with empirical studies on interest spread with exceedence correlation and tail dependence.*

OU process

The **Ornstein–Uhlenbeck** process is a **stationary Gauss–Markov process**, which means that it is a *Gaussian process, a Markov process,* and is temporally homogeneous. In fact, it is the only nontrivial process that satisfies these three conditions, up to allowing linear transformations of the space and time variables.[1] Over time, the process tends to drift towards its mean function: such a process is called **mean-reverting.**

Half Life

Half-life indicates how long the spread typically takes to revert back to the mean. A half-life of 10 days for example indicates that this pair typically takes 10 days to revert.

$\boldsymbol{\alpha}$: a positive number called the **decay constant of the decaying quantity.**

Given OU process:
https://mathtopics.wordpress.com/2013/01/07/ornstein-uhlenbeck-process/

$\mathbf{d}\boldsymbol{x_t}\mathbf{=}\boldsymbol{\alpha}(\boldsymbol{\mu}-\boldsymbol{x_t})\mathbf{dt+}\ \boldsymbol{\sigma}\mathbf{d}\boldsymbol{W_t}$

$x_t{=}e^{-\alpha t}x_0+\ \mu(1-e^{-\alpha t})+\int_0^t\ \sigma\, e^{\alpha(s-t)}\ \mathrm{d}W_t$

Where:
x_t: OU process
α: speed of mean reversion
W_t: Wiener process

Since $\mathrm{x}\left(t_{1/2}\right)-\ \mu{=}\frac{x_0-\mu}{2}$

$\boldsymbol{t_{1/2}}\mathbf{=}\frac{\mathbf{ln}\ (\mathbf{2})}{\boldsymbol{\alpha}}$

Semidefinite Positive matrices
The symmetric matrix A is said positive semidefinite if all its **eigenvalues** are **non-negative.**

Symmetric matrices
Has only real eigenvalues
Is always DIAGONALIZABLE
HAS ORTHOGONAL EIGENVECTORS.

Orthogonal
an orthogonal matrix in which its transpose is equal to its inverse matrix.

Higher order Greeks
Vanna: DdeltaDvol
Volga: DvegaDvol
Ultima: DvolgaDvol

What is logistic regression?
Logistic regression models the **probabilities for classification problems** with **two possible outcomes.** It's an extension of the linear regression model for classification problems.

Logistic regression is a process of **modeling the probability of a discrete outcome** given an input variable. The most common logistic regression models a binary outcome; for example, true/false, yes/no, etc.

What is GMM? (Moments of Random variables instead of entire distribution)
The generalized method of moments (GMM) is a method for constructing estimators, analogous to maximum likelihood (ML). GMM uses assumptions about specific **moments of the random variables** instead of assumptions about the entire distribution, which makes **GMM more robust than ML**, at the cost of some efficiency. The assumptions are called moment conditions.

GMM generalizes the method of moments (MM) by allowing the number of moment conditions to be **greater than the** number of parameters. Using these extra moment conditions makes GMM more efficient than MM. When there are **more moment conditions than parameters,** the estimator is said to be overidentified. GMM can efficiently combine the moment conditions **when the estimator is overidentified.**

What is robust regression?

Robust regression is an alternative to least squares regression when data are contaminated with outliers or influential observations, and it can also be used for the purpose of **detecting**

influential observations.

What is endogeneity problem in economics?
Wrong causal claims

What is threshold regression?

Threshold models are often applied to time-series data. The **threshold can be a time.** For example, if you think investment strategies changed as of some **unknown date**, you can fit a model to obtain **an estimate of the date** and obtain estimates of the different coefficients **before** and **after** it.

Or the threshold can be in terms of another variable. For example, **beyond a certain level of inflation, central banks increase interest rates.** You can fit a model to obtain an estimate of the threshold and the coefficients on either side of it.

What is DID? (treatment and control group)

DID is a quasi-experimental design that makes use of longitudinal data from **treatment** and **control** groups to obtain an appropriate counterfactual to estimate a **causal effect**. DID is typically used to estimate the effect of a specific intervention or treatment (such as a passage of law, enactment of policy, or large-scale program implementation) by **comparing the changes in outcomes over time between** a population that is enrolled in a program (the intervention group) and a population that is not (the control group).

What is m-estimation?
M-estimators are especially useful when your data **has outliers o**r is contaminated because **one outlier (or heavy tailed errors) can render the normal-distribution** based OLS useless; In that case, you have two options: remove the badly-behaving outliers, or use the robust M-estimator.

M-estimation typically involves minimizing or maximizing a specified objective function defined in terms of data and unknown population parameters.

New Stochastic calculus (Heston SVM)

Textbook:

Heston stochastic volatility model

dS(t)=rS(t)dt+$\sqrt{V(t)}$S(t)d$\widetilde{W_1}$(t),

where interest rate r is constant and the volatility $\sqrt{V(t)}$ is a SP govern by:

dV(t)=$(a - b$V(t))dt+$\sigma\sqrt{V(t)}$ d$\widetilde{W_2}$(t).

Online:

$\mathbf{d}\boldsymbol{S_t}$=$\mathbf{r}\boldsymbol{S_t}$+$\sqrt{\boldsymbol{V_t}}\boldsymbol{S_t}\mathbf{d}\boldsymbol{W_t^S}$

$\mathbf{d}\boldsymbol{V_t}$=$\mathbf{k}(\boldsymbol{\theta} - \boldsymbol{V_t})$**dt+**$\zeta\sqrt{\boldsymbol{V_t}}$ $\mathbf{d}\boldsymbol{W_t^V}$

where:

ζ is the volatility of volatility

k is the rate at which V_t returns to 0

θ is the long-run price variance

W_t^S is the Brownian motion corresponding to the underlying asset's price

W_t^V is the Brownian motion corresponding to the variance

Both BMs are negatively correlated. So for example, a large **sharp drop** in equity prices **increases the volatility.**

Heston model is a type of **state-space stochastic volatility model** that describes th**e evolution of volatility** of an asset.

The variance is the instantaneous variance given by the **Cox-Ingersoll-Ross (CIR)** process which is used for modelling interest rate in the market. Its assumption is that the rates are **mean-reverting** and **never negative.** The "square root" over the variance is used to keep only positive values.

Vanilla options

American options

American options **do not have a closed-form (analytical)** solution.

The finite difference method is one of the most widely method approximation to solve the

PDE equation for American options. The Black-Scholes PDE can also be used to price American options. The main difference between European options and American options is that the latter can be executed any time prior to the expiry date.

The **European option**s under BS assumption has an **analytical solution for the fair pri**ce. In contrast, the **American option** does not have an analytical solution so the PDE has to be solved using technique like the **finite difference numerical methods.**

https://finpricing.com/lib/EqAmerican.html

Finite difference method

FDM are similar to binomial trees. However, instead of discretizing asset prices and passage of time in a tree structure, it discretizes in a grid-with time and price steps by calculating the value at every possible grid points.

Step 1: Generate the grid by specifying grid points.
Step 2: Specify the **final or initial** conditions.
Step 3: Use **boundary conditions** to calculate option values and step back down the grid to fill it.

Local volatility surface

The local volatility surface is the **surface of all forward volatilities** for different S and t that is locked in by the prices of standard **European options.**

If we use market prices of options, we get the market local volatility surface. And if we use model prices, we get the model local volatility surface. We obtain the local volatility surface of **forward instantaneous volatilities** using finite differences to approximate the options price derivatives in Dupire’s equation. Market implied volatility surface float with S, but market local volatilities are static.

Black Scholes Model

*The BSM assumes that the **distribution of asset return** is **normal** which implies that the **distribution of prices** is **lognormal**.

Methods for deriving the BSM:

Risk neutral pricing

BS PDE solution

Binomial tree model

Black-Scholes Assumptions

The Black-Scholes model makes certain assumptions:

- No dividends are paid out during the life of the option.
- Markets are random (i.e., market movements **cannot be predicted**).
- There are no transaction costs in buying the option.
- The risk-free rate and volatility of the underlying asset are **known and constant**.
- The returns on the underlying asset are **log-normally** distributed.
- The option is European and can only be exercised at expiration.

https://en.wikipedia.org/wiki/Greeks_(finance)

Delta = $\frac{\partial V}{\partial S} = \Delta$

Vega=$\frac{\partial V}{\partial \sigma}$

Theta=$-\frac{\partial V}{\partial T}$

Rho=$\frac{\partial V}{\partial r}$

Gamma=$\frac{\partial \Delta}{\partial S}$

Ways to compute Greek

Finite difference approximations

Pathwise method (Use this)

differentiating the options's payoff and then take the **expectation under the risk neutral measures.**

Where:

Y is the options payoff as Y = $e^{-rT}(S_T - K)^+$

S_T=$S_0 e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}Z}$

Z~N(0,1).

The **standard normal distribution**, also called the *z*-**distribution**, is a special normal distribution where the mean is 0 and the standard deviation is 1.

Any normal distribution can be standardized by converting its values into *z*-scores. *Z*-scores tell you how many standard deviations from the mean each value lies.

Delta =$\frac{\partial Y}{\partial S_0}$

Vega = $\frac{\partial Y}{\partial \sigma}$

<u>Likelihood ratio method</u>

Similar to the pathwise method, but **differentiate the density.**

The LL ratio method differentiates a probability density with respect to the parameter of interest, θ. It provides a good potential alternative to the pathwise method when Y is not continuous in θ.

What is the law of large numbers?

Strong law: average of a large number of independent identically distributed integrable random variable variables **converges almost surely** to their **common mean**

Averaging removes randomness

$\bar{y}$=$\frac{1}{N}\sum_{i=1}^{N} y_i$ →u as N→ ∞ or plim($\bar{y}$)=u.

Weak law: Convergence is only in probability

What is the central limit theorem?

The CLT states that the limiting distribution of the centered and scaled sum of an IID sequence of RV is a normal distribution if the common distribution of the RV has finite variance.

What is importance of Martingale in options pricing?

If a market model is arbitrage free, then there exists a risk neutral probability so that each **discounted asset price process** under the risk neutral probability is a martingale. By applying Ito's formula, the discounted price process of the derivative is a martingale under risk neutral probability.

What is Girsanov's theorem?
It is providing a way to change the drift of Wiener process by defining a new probability measure via Radon-Nikodym derivative.

The collection of the theorems that tell us **how to make drift disappear** is called the Girsanov theory.

In probability theory, the Girsanov theorem tells how stochastic processes change under **changes in measure**. The theorem is especially important in the theory of financial mathematics as it tells how to convert from the physical measure which describes the **probability that an underlying instrument** (such as a share price or interest rate) will take a particular value or values to the risk-neutral measure which is a very useful tool for evaluating the value of derivatives on the underlying.

Radon–Nikodym
In mathematics, the **Radon–Nikodym** theorem is a result in measure theory that expresses the relationship between two measures defined on the same measurable space. A **measure** is a set function that assigns **a consistent magnitude to the measurable subsets** of a **measurable space**. Examples of a measure include area and volume, where the subsets are sets of points; or the probability of an event, which is a subset of possible outcomes within a wider probability space.

Brownian Bridge
A Brownian bridge is a **continuous-time stochastic process** B(t) whose probability distribution is the **conditional probability distribution of a standard Wiener process** W(t) (a mathematical model of Brownian motion) subject to the condition (when standardized) that W(T) = 0, so that the process is pinned to the same value at both t = 0 and t = T. In a Brownian bridge process on the other hand, not only is **B(0) = 0** but we also require that **B(T) = 0**, that is the process is "tied down" at t = T as well. Just as a literal bridge is supported by pylons at both ends, **a Brownian Bridge** is required to satisfy conditions at both ends of the interval [0,T].

Martingale

An $\mathcal{F}_t$ adapted, value valued process M is called a martingale (with respect to the filtration $\mathcal{F}_t)\ if$:

E$|M_t|$<∞ for all t∈T

E$(M_t|\mathcal{F}_s)$ = M_s for al s≤t

Binomial options pricing

Given:

S_0

K

U

D=1/U

R: risk free rate

Calculation:

p=$(1 + R - D)$/U-D

q=1-p

Valuation:

$$\frac{(p * value\ of\ up\ t + 1\ node + q(value\ of\ down\ t + 1\ node))}{1 + risk\ free\ rate}$$

N(d_2) is the probability that the call will be exercised

N(d_1) does not lend itself to a simple probability interpretation. Stock price *N(d_1)is the present value using the risk free interest rate of the expected asset price at expiration.

N(d_1) is the factor by which the **present value of contingent receipt of the stock** exceeds the current stock price.

What's risk neutral

The risk neutral condition of BS assumes the **ability to delta hedge continuously** and there is **no bid/ask**, **no transaction cost**, **full liquidity.**

5.5.4 Uniqueness of the risk neutral measure

Definition 5.4.8. A market model is complete if every derivative security can be hedge.

Theorem 5.4.7 First fundamental theorem of asset pricing.
If a market model has a risk neutral probability measure, then it does not admit arbitrage.

In the **Black Scholes model,** the option value under the **risk neutral measure** can be expressed as the expectation:
V(t, S_t)=$e^{-r(T-t)}E^Q[G(S_T)|\mathcal{F}_t]$ where S_t follows the dynamics:
dS_t=r$S_t$dt+σS_tdX^Q(t) in which X^Q(t) is a Brownian motion under Q.

Put call parity of American options
However, **put-call parity does NOT hold for American options due to early exercise.**
C: not optimal to early exercise: C=c
P:maybe optimal to early exercise: P≥p
p+S≤P+S

C+Ke^{-rT} ≤P+S
Due to the fact that **C not optimal to exercise** and **P optimal to exercise.**

Combine: S-K≤C-P≤S- Ke^{-rT}

Is it ever optimal to early exercise an American call options? What about a put options?
For call options, an option will always be **worth more than its value if exercised immediately.**

For put options, there is not an equivalent argument for put options.
There are times when the **optionality value is worth less than receiving the payoff today** and hence it is optimal to early exercise.

Why is there a Skew?
1. Risk aversion
2. The leverage effect: Equity volatility should increase as the equity value decreases.

Why never optimal to early exercise American call?
The reason is that exercise requires payment of the strike price X. By holding onto X until the expiration time, the **option holder saves the interest on X.**

Binomial options pricing

Given:

S_0

K

U

D=1/U

R: risk free rate

Calculation:

p=$(1 + R - D)$/U-D

q=1-p

Valuation:

$$\frac{(p(value\ of\ up\ t+1\ node) + q(value\ of\ down\ t+1\ node))}{1 + risk\ free\ rate}$$

Put-call parity

C+Xe^{-rt}=P+S

- Put-call parity implies that call and put prices should **change by the *same amount* when volatility changes**

European options without dividends

c+Ke^{-rT} = p+S

Implied volatility

In the Black-Scholes model, volatility is the only parameter that can't be directly observed. All other parameters can be determined through market data and this parameter is determined by a **numerical optimization technique** given the Black-Scholes model.

<u>Newton-Raphson method</u>

The Newton-Raphson method uses an **iterative procedure** to solve for a root using information about the derivative of a function.

Newton's method is a method for finding increasingly improved approximations to the roots of a function.

http://www.codeandfinance.com/finding-implied-vol.html

Bisection method
The bisection method is considered to be one of the simplest and robust root finding algorithm

Implied Volatility
With historical volatility, the initial price of an option can be established. However, after its introduction, an option is traded through an exchange and the price is determined by the market.

While there is no closed form for inverting Black-Scholes to solve for volatility including the option's price, the formula can easily be **solved numerically**. A simple and adequate method is **bisection.**

The method starts with an estimate of volatility σ_1 that is too low, and an estimate σ_2 that is too high. The computed options price for σ_1 is less than the actual price but for σ_2 it is greater. In this way the root we are seeking σ_0 is bracketed. Next, the option price is calculated for the mid-volatility $\sigma_m = \frac{1}{2}(\sigma_1 + \sigma_2)$

Exotic options

A path dependent option is an exotic option that's value depends not only on the price of the underlying asset but the **path that asse**t took during all or part of the life of the option. There are many types of path-dependent options including **Asian, chooser, lookback, and barrier options.**

Heston stochastic volatility model
dS(t)=rS(t)dt+$\sqrt{V(t)}$S(t)d$\widetilde{W}_1$(t),

where interest rate r is constant and the volatility $\sqrt{V(t)}$ is a SP govern by:
dV(t)=$(a - b$V(t))dt+$\sigma\sqrt{V(t)}$ d$\widetilde{W_2}$(t).

Online:
d$\boldsymbol{S_t}$=r$\boldsymbol{S_t}$+$\boldsymbol{\sqrt{V_t}S_t}d\boldsymbol{W_t^S}$
d$\boldsymbol{V_t}$=k($\boldsymbol{\theta - V_t}$)dt+ζ$\boldsymbol{\sqrt{V_t}}$ d$\boldsymbol{W_t^V}$

where:
ζ is the volatility of volatility
k is the rate at which V_t returns to 0
$\boldsymbol{\theta}$ is the long-run price variance
W_t^S is the Brownian motion corresponding to the underlying asset's price
W_t^V is the Brownian motion corresponding to the variance
Both BMs are negatively correlated. So for example, a large sharp drop in equity prices increases the volatility.

Heston model is a type of **state-space stochastic volatility** model that describes th**e evolution of volatility** of an asset.

The variance is the instantaneous variance given by the **Cox-Ingersoll-Ross (CIR)** process which is used for modelling interest rate in the market. Its assumption is that the rates are **mean-reverting** and **never negative.** The "square root" over the variance is used to keep only positive values.

Cliquet options
Cliquet options are designed to protect against **downside risk y**et with **significant upside** potential. Capping the maximum, as in this **globally floored, locally capped,** ensures that the payoff is **never too extreme** and therefore that the value of the contract is **not too outrageous.**

Cliquet Option: A series of call or put options with rules determining how the strike price is determined. For example, a cliquet might consist of 20 at-the-money three-month options. The total life would then be five years When one option expires a new similar at-the-money is comes into existence

A cliquet option is an exotic option consisting of a series of **consecutive forward start**

options, each struck at-the-money on the date it becomes active. Both compound cliquet (CC) and compounding compound cliquet (CCC) are options on a cliquet option, whose payoff is based on a single underlying index.

The key model assumptions are:

- The asset values can be accurately expressed using the volatility skew model.
- The volatilities of the underlying assets are stochastic and capable of rising without a movement in spot prices.
- The stochastic volatility model parameters derived from SPX index-based cliquets can be applied to other indices based cliquets.
- Interest rates are deterministic.
- The projected dividend yields are known and fixed.

Barrier options
Barrier options have a payoff that is contingent on the underlying asset reaching some specified level before expiry.

Asian options
Have a payoff that depends on the **average value of the underlying asset** over some period before expiry. They are strongly path dependent.

Lookback options
Lookback option payoffs are calculated using the optimal value (maximum for a call, minimum for a put).

Binary (Digital) options
A binary option is a type of options contract in which the payout will depend entirely on the **outcome of a yes/no proposition** which related to whether the price of a particular asset that underlies the binary option will rise above or fall below a specified amount. Unlike other types of options, a binary option does **not give the holder the right to purchase or sell t**he underlying asset. When the binary option expires, the option holder wil**l receive either a pre-determined amount of cash or nothing at all.** Given the all-or-nothing payout structure, binary options are called all-or-nothing options or fixed-return options.

Features of a Digital Option:

- A digital option is a currency pair-based short-term trade.
- The maximum losses for a digital option can be close to 100%, whereas profits can be as high as 900%.
- The traders can close the trade of a digital option any time before the expiry. If the trader believes that he/she will receive more profits in case of an early exit, the trader can exit before the expiry of the trade.
- A digital option comes with a life of five minutes. Hence, traders are allowed to enter a trade up until the last 30 seconds of expiry.
- Trading in digital options closes only from 7:30 p.m. to 10:30 p.m. (UTC).
- A digital option enables a trader to open a lot of trades until the last 30 seconds of the trade's life. However, only experienced traders are capable of making money by entering multiple trades. Traders can lose a huge amount of money if they lose control.
- The amount of potential profits varies with the strike price. The potential profits increase if there is a wide gap between the strike price and the actual price, and vice-versa. However, traders need to remember that if the strike price is pushed away from the actual price, risk will also increase.
-

(Digital Option – PDE Approach)

Let $\{W_t: t \geq 0\}$ be a $\mathbb{P}$-standard Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let the stock price S_t follow a GBM with the following SDE:

$$dS_t = S_t(u - D)d_t + S_t\sigma dW_t$$

where u is the drift parameter, D is the continuous dividend yield, σ is the volatility parameter, and r denotes the risk-free rate. A **digital (cash-or-nothing) call** option is a contract that pays **\$1 at expiry time T is the spot price $S_T \geq K$ and nothing if $S_T \leq K$.** By denoting $C_d(S_t, t; K, T)$ as the price of a European digital call satisfying the following PDE with boundary conditions:

$$\frac{\partial C_d}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C_d}{\partial S_t^2} + (r - D)S_t \frac{\partial C_d}{\partial S_t} = rC_d(S_t, t; K, T)$$

$C_d(0, t; K, T) = 0$ and $C_d(S_T, T; K, T) = 1_{(S_T > K)}$ and letting the solution of the SDE be in the form:

$C_d(S_t, t; K, T) = e^{\alpha x+\beta\tau}u(x, \tau)$ where $x = \log(S_t/K)$ and $\tau = \frac{1}{2}\sigma^2(T-t)$, show that by setting: $\alpha = -\frac{1}{2}K_1$ and $\beta = -\frac{1}{4}K_1^2 - K_0$ where $K_1 = \frac{r-D}{\frac{1}{2}\sigma^2} - 1$ and $K_0 = \frac{r}{\frac{1}{2}\sigma^2}$, the Black-Scholes equation for $C_d(S_T, T; K, T)$ is: $\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2}$, $u(x, 0) = f(x)$, $x \in (-\infty, \infty)$, $\tau > 0$ where $f(x) = e^{\frac{1}{2}K_1 x}1_{(x>0)}$.

(Asset-or-Nothing Option – Probabilistic Approach)

Let $\{W_t: t \geq 0\}$ be a $\mathbb{P}$-standard Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let the stock price S_t follow a GBM with the following SDE:

$dS_t = S_t(u - D)d_t + S_t\sigma dW_t$

where u is the drift parameter, D is the continuous dividend yield, σ is the volatility parameter, and r denotes the risk-free rate.

An **asset-or-nothing call option** is a contract that pays S_T at expiry time T if the spot price $S_T > K$ and nothing if $S_T \leq K$. In contrast, an asset-or-nothing put pays $S_T < 0$ at expiry time T if the spot price $S_T < K$ and nothing if $S_T \geq K$.

By denoting $C_a(S_t, t; K, T) = S_t e^{-D(T-t)}\Phi(d_+)andP_a(S_t, t; K, T) = S_t e^{-D(T-t)}\Phi(-d_+)$, where: $d_+ = \left(\frac{\log\left(\frac{S_t}{K}\right)+\left(r-D-\frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right)$ and $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}u^2}du$.

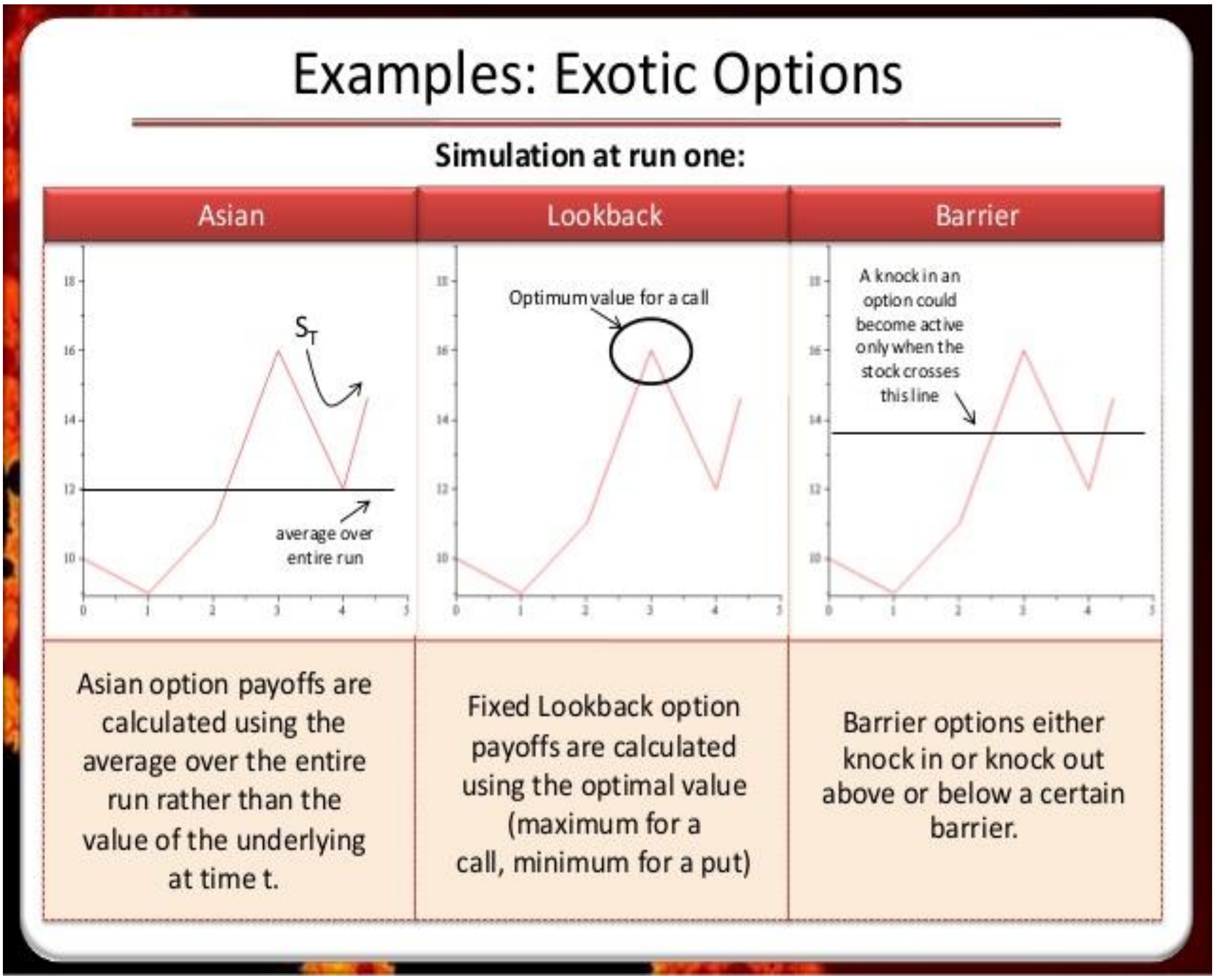
Monte Carlo Simulation

We take the population and then we sample it by **drawing a proper subset**. And then we make an inference about the population based upon some set of statistics we do on the sample.

And, the key fact that makes them work, that if we choose the sample at random, the sample will tend to exhibit the same properties as the population from which it is drawn.

It requires **generation** of **quasi random numbers** and relies on their low latency.

A Monte Carlo simulation is a model used to **predict the probability of different outcomes** when the intervention of random variables is present.
Monte Carlo simulations help to explain the **impact of risk and uncertainty in prediction** and forecasting models.



Interest rate, swap, and credit

Duration

- **Macaulay duration** estimates **how many years** it will take for an investor to **be repaid** the bond's price by its total cash flows.
- **Modified duration** measures the price change in a bond given a 1% change in interest rates.

Convexity

Curvature; relationship between **bond prices and bond yields.**

LIBOR

LIBOR is the benchmark for floating short-term interest rates and is set daily.

SWAP rate

The "swap rate" is the fixed interest rate that the receiver demands in exchange for the uncertainty of having to pay the short-term LIBOR (floating) rate over time.

Yield curve fitting

Ho & Lee spot interest rate model

The yield curve

The yield curve is given the distillation of many different market-observable prices into one condensed representation. Those can include interbank cash deposit rates, government bond repo rates, interest rate futures, and swap rates.

Inverted yield curve (Short term rate higher than long term rate)

When the yield curve is inverted, it indicates a view among investors that there is **greater risk to the economy in the short run,** encouraging central banks to eventually lower interest rates to combat recession.

The HJM model

The HJM models the **yield curve as s whole**. With the HJM, we evolve the entire forward curve. Calibration means **linear factorization of forward rate volatility.** PCA reveals the data structure: the key factors of curve moment are level, steepness/flatness and curvature. Pricing under the HJM is done using the Monte-Carlo.

Variance swap

VS is an OTC derivative contract in which two parties agree to buy or sell the realized volatility of an index or single stock on a future date.

The payoff for an investor who buys a variance using a swap is equal to the difference between the **realized and strike variance**, multiplied by the national amount of the swap.

The profitability of **a variance swap** has **a quadratic relationship** to realized volatility. Variance swaps are often used for **portfolio protection.**

Volatility swap
The payoff of a volatility swap is directly proportional to realized volatility.

The **difference between variance and volatility is called convexity.

Interest rate swap
corporations could lock into paying the **prevailing fixed rate** and **receive payments** that matched their **floating-rate debt.**

Credit Default Swap
A product to buy or sell **protection against a credit event** such as a **default**

CDS spread
Reflect the cost of credit risk. Credit spread is composed of **risk premium** and real-world **default loss.**

Credit derivatives
Merton
Extension to Merton: Black and Cox, KMV, Capital structure arbitrage.

CDO
CDO is a type of asset backed security, it is an investment on a **pool of diversified assets** in the form of trenched securities.

Default correlation
Modelling the joint probability of two events, knowing the individual probabilities.

High yield bonds
HYB are more affected by spread change and investment grade by general market interest rate changes.
Liquidity risk is significant for both investment grade (IG) and HYB, but more so for HYB.

Credit analysis:
Investment grade: BBB or above

Default risk
Probability of default

Loss severity
Percent of value lost if borrower default

Expected loss
Default risk X loss severity

Recovery rate = 1 - Expected loss percentage

Time Series

Energy paper/VAR at ANU
This research examines the dynamic relationship among impacts of prices and volatilities of real macro activities, interest rate, and two major energy markets: crude oil and natural gas, with main concerns on both price and volatility shocks to energy markets and monetary policy.
This research examines the dynamic relationship among impacts of prices and volatilities of real macro activities, interest rate, and two major energy markets: crude oil and natural gas, with main concerns on both price and volatility shocks to energy markets and monetary policy.

Technically, VAR models a system of endogenous variables that **depend only on their lagged values and helps capturing the dynamics of interrelationship.**

Shock in crude oil induces three basis point increase in T-bill yield by the third month
By contrast, **shock in natural gas** causes seven basis points increase in T-bill yield

Altogether, ever **since the breakout of GFC,** variations of most variables, whether in either form of price or volatility, have became much more sensitive to exogenous shocks than before the crisis.

This research examines the dynamic relationship among impacts of prices and volatilities of real macro activities, interest rate, and two major energy markets: crude oil and natural gas, with main concerns on both price and volatility shocks to energy markets and monetary policy. I estimate conditional volatilities of all variables with EGARCH model and estimate the dynamic relationship from January 1986 to December 2016 with **non-recursive SVAR**. Parameter estimation from EGARCH suggests that crude oil volatility has negative correlation with crude oil return and natural gas volatility has positive correlation with natural gas return. Further, the inference from EGARCH also explains why implied volatility smile of options are skewed. Moreover, I conduct subsample analysis that splits estimation between pre-global financial crisis (GFC) which is characterized with high discount rate regime and post-GFC which is characterized with the so-called zero lower bound on discount rate regime. During the post-GFC when the Fed fund rate approaches zero lower bound, a positive monetary policy shock becomes effective since prize puzzle in either crude oil or natural gas no longer exists.

Additionally, S&P 500 and industrial production showed that they can be resilient to either crude oil or natural gas price shocks. Interestingly, T**-bill yield exhibits positive response to crude oil price shock but negative response to natural gas price shock**.

In the dimension of volatility transmission, **all volatilities of macro activities will be heightened after crude oil volatility shock.**

In contrast, volatilities of crude oil, T-bill, and S&P 500 will be cooled down by natural gas volatility shock. Altogether, ever since the breakout of GFC, variations of most variables, whether in either form of price or volatility, have became much more sensitive to exogenous shocks than before the crisis. Policy makers should especially avoid huge surprise to the market and adopt phase-in approach to smooth out the impacts on real macro activities during the era of zero lower bound on interest rate.

AR model

Lagged values of y_t are used **as predictor variables**. Lags are where results from one time period affect following periods. The order p can be determined by looking at **the PACF.** It gives the partial correlation of a stationary time series with its **own lagged values**.

MA model

It accounts for **very short-run autocorrelation.** It basically states that the **next observation is the mean of every past observation.** MA(q) can be estimated by looking at the ACF plot. If ACF plot take a very long time to converge, this means the MA of the time series would not fix the problem of not having a stationary time series. Thus, the MA model will not be a good model to forecast with.

ARMA model

A combination of AR and MA models.

ARIMA(p, d, q)

Now it a d which represents the number **of nonseasonal differences** needed for stationarity. Simply, **d makes nonstationary data stationary by removing trends.**

Stationary

If a time series **has trend or seasonality**, then it is **not stationary**.

AIC and BIC

The AIC and BIC provide measures of model performance that account for model complexity. The AIC is a measure of **the goodness of fit** of any estimated statistical model. The BIC is a model selection among a class of parametric models with different numbers of parameters.

The test of Ljung-Box:

Ho: The variable is independently distributed; so exhibits white noise process.
Ha: The variable is not independently distributed; so exhibits autocorrelation.

Autocorrelation

- Autocorrelation, also known as serial correlation, refers to the degree of correlation of the same variables between two successive time intervals.
- The value of autocorrelation ranges from -1 to 1. A value between -1 and 0 represents negative autocorrelation. A value between 0 and 1 represents positive autocorrelation.

Optimal lag selection and consequence of autocorrelation.

3.4 Optimal lag selection

To eliminate autocorrelation that results in **inaccurate computed variance and standard error of forecast,** we use Likelihood Ratio (LR) test, Hamilton and Herrera (2004) argue that LR test works best for adjusting for seasonal factors of oil price in a monthly VAR (p. 277).

The three equations of ADF test are:

Equation (1): the pure random walk model.

Equation (2): adds a drift term to the previous model, Equation (1).

Equation (3): adds a drift and a linear time trend to the random walk model.

The hypothesis test in ADF is described below:

Ho: The variable contains a unit root

Ha: The variable is stationary

Unit root process

a unit root is a feature of some stochastic processes (such as random walks) that can cause problems in statistical inference involving time series models. A **linear stochastic process** has a unit root if 1 is a root of the process's characteristic equation. Such a process is non-stationary **but** does not always have a trend.

Engle-Granger test

Ho: No cointegration among the variables

Ha: Cointegration exist among the variables

The forecast error variance decomposition (FEVD):

1. Most of the forecast error variances in the return of rx are not explained by changes in itself but by changes in ry. Thus, rx is not exogenous.

2. The entire forecast error of variance in the return of ry is due to change in changes in itself. Thus, ry is exogenous.

3. Most of the forecast error of variances in the return of rz are not due to changes in itself but by changes in ry followed by changes in rx. Hence, rz is not exogenous because it can be explained by other two variables.

Definition.

The lag operator L applied to a stochastic process $\{y_t\}$ transforms a realisation at time t into a realisation at time $t-1$, i.e.

$$y_{t-1} = Ly_t.$$

This helps us write polynomials in the lag operator:

- $a(L) = 1 - a_1 L - \cdots - a_p L^p$,
- $\beta(L) = 1 + \beta_1 L + \cdots + \beta_q L^q$.

Then, the ARMA(p, q) can be concisely expressed as $a(L)y_t = \beta(L)\varepsilon_t$.

AR and MA representations of the ARMA(p, q)

If the ARMA(p, q) is stationary, then $a(L)$ can be *inverted* to produce the pure MA representation $y_t = \theta(L)\varepsilon_t$, where

$$\theta(L) = \frac{\beta(L)}{a(L)}.$$

We will also refer to this as the Wold representation.

If the ARMA(p, q) is invertible, then $\beta(L)$ can be *inverted* to produce the pure AR representation $\phi(L)y_t = \varepsilon_t$, where

$$\phi(L) = \frac{a(L)}{\beta(L)}.$$

What is stationary time series?

A stationary TS has no trend or seasonal effects. In addition, it has good quality of mean reversion.

What is stationary time series?

A stationary TS has no trend or seasonal effects. In addition, it has good quality of mean reversion. Otherwise, spurious regression:

Spurious regression is a regression that provides **misleading statistical evidence** of a linear relationship between independent non-stationary variable.

Non-stationary has unit root:
A linear stochastic process has a **unit root if 1** is a root of the process's characteristic equation. Such a process is **non-stationary** but does not always have a trend. If a root of the process's characteristic equation **is larger than 1**, then it is called an **explosive process**, even though such processes are sometimes inaccurately called unit roots processes.

ARMA
ARMA(p,q) refers to the model with **p autoregressive terms** and **q moving-average terms.** We can use the ARMA to **predict future** values. We can analyze response of variables to unforeseeable shocks.

If the ARMA(p, q) is **stationary,** then a(L) can be **inverted** to produce the pure MA representation y_t=θ(L)ε_t, where θ(L)=$\frac{\beta(L)}{a(L)}$. We also refer to this as the **Wold representation.**

If the ARMA(p, q) is **invertible, t**hen B(L) can be inverted to produce the AR representation $\emptyset$(L) y_t=ε_t, where $\emptyset$(L) = $\frac{a(L)}{B(L)}$

ARCH
It assumes that the variance of tomorrow's return is an **equally weighted average** of the **squared residuals** from the last 22 days. ARCH shows that periods of high volatility are followed by more high volatility and periods of low volatility are followed by more low volatility. ARCH measures volatility and forecast it into the future.

GARCH
GARCH is a weighted average of **past squared residuals** but it has declining weights which never go completely to zero. It is good at **predicting conditional variance**. It asserts the **best** predictor of the **variance in the next period** is a **weighted average of the long run average variance**, the variance predicted for this period and the new information this period which is the **most recent squared residual**. Both EWMA and GARCH place more
weight on recent information.

Unconditional SD: Long run average

Conditional SD: Just like volatility which can be up and down

AMS 517 Time Series

Skewness: $\sum_{i=1}^{n} \frac{(r_i-\bar{r})^3}{\widehat{\sigma^3}}$

Skewness is **a measure of the asymmetry of a distribution**.

Kurtosis: $\sum_{i=1}^{n} \frac{(r_i-\bar{r})^4}{\widehat{\sigma^4}}$

Kurtosis is a measure of the tailedness of a distribution

Sample Skewness: $\frac{1}{n}\sum_{i=1}^{n} \frac{(r_i-\bar{r})^3}{\widehat{\sigma^3}}$

Sample Kurtosis: $\frac{1}{n}\sum_{i=1}^{n} \frac{(r_i-\bar{r})^4}{\widehat{\sigma^4}}$

Estimation of trend with seasonal effect

Y_t=m_t+S_t+X_t

Remove the trend (No seasonal effect)

Δy_{t+1}=$y_{t+1} - y_t$ = $(m_{t+1} - m_t)$ + $X_{t+1} - X_t$

Backward shift operator

Δy_t=$y_t - y_{t-1}$ = $(1 - B)y_t$

AR(1) process: y_t=α_0+$a_1 y_{t-1}$+⋯+$a_p y_{t-p}$+μ_t

MA(1) process: μ_t=ε_t+$\beta_1\varepsilon_{t-1}$+⋯+$\beta_q\varepsilon_{t-q}$

ARMA(1,1): y_t=α_0+$a_1 y_{t-1}$+$\beta_1\varepsilon_{t-1}$+ε_t

Stationary (UQ note):

E(y_t) = μ

Var(y_t)=σ_y^2=γ_0

Cov(y_t, y_{t-k})=γ_k

Constant u (mean) for all t

Constant variance for all t
The autocovariance function between y_t and y_{t-k} only depends on the interval t_1 and $t_{2.}$

Weakly stationary (covariance stationary)
EX_t=Constant
Cov(X_t, X_s)=r$(t-s)$= r$(s-t)$

Stationary
EX_t=0

The AR(1) model with $|a_1|$ ≥1 is called non-stationary. $|a_1|$< 1 is stationary.
ε_t is the uncorrelated innovation in the process.

Stationary:
E(y_t) = μ
Var(y_t)=σ_y^2=γ_0
Cov(y_t, y_{t-k})=γ_k

The AR(1) model with $|a_1|$ ≥1 is called non-stationary.
ε_t is the uncorrelated innovation in the process.

Putting AR(1) and MA(1) together, we get:
y_t=α_0+$a_1 y_{t-1}$+$\beta_1 \varepsilon_{t-1}$+$\varepsilon_t$
Where ε_t are uncorrelated.
This is called the ARMA(1,1).

β_1=0 means an AR(1) process for y_t.
a_1=0 means an MA(1) process for y_t.
If $|a_1|$<1, then there exists an equivalent pure MA process with infinitely many lags.
If $|\beta_1|$<1, then there exists an equivalent pure AR process with infinitely many lags.

The **lag operator L** applied to a stochastic process y_t transform a realization at time t into realization at time t−1, for example: y_{t-1}=Ly_t

Polynomials in the lag operator:
a(L) = 1$-a_1$L$- \cdots - a_p L^p$
β(L)= 1+β_1L$+ \cdots + a_q L^q$

Then the ARMA(p, q) can be expressed as a(L) y_t= β(L) ε_t

If the ARMA(p, q) is **stationary,** then a(L) can be **inverted** to produce the pure MA representation y_t=θ(L)ε_t, where θ(L)=$\frac{\beta(L)}{a(L)}$. We also refer to this as the **Wold representation.**

If the ARMA(p, q) is invertible, then B(L) can be inverted to produce the AR representation $\emptyset$(L) y_t=ε_t, where $\emptyset$(L) = $\frac{a(L)}{B(L)}$

ARMA(1,1)
y_t=α_0+$a_1 y_{t-1}$+$\beta_1 \varepsilon_{t-1}$+$\varepsilon_t$

One-step Forecasting:
$\widehat{y_{T+1}}$ = a_0+$a_1 y_T$+⋯+$a_p y_{T-p}$ + $\beta_1 \widehat{\varepsilon_T}$ + ⋯+$a_p \beta_1 \varepsilon_{\widehat{T-q+1}}$

VAR
AR models a system of endogenous variables that depend only on their lagged values and helps capturing the dynamics of interrelationship.

The reduced form VAR gives us reduced form errors, ε_t, which are linear combinations of structural errors, v_t.

Unit root test
To avoid spurious regression: misleading statistical relation and false causality.

Autocorrelation and optimal lag selection
To eliminate autocorrelation that results in **inaccurate computed variance and standard error of forecast**

Cointegration

Consequently, we found no evidence of cointegration among them, then the five-variable system (Δlog(industrial production), Δ(T-bill yield), Δlog(real crude oil), Δlog(real natural gas), and Δlog(real stock price of S&P 500)) may be modelled as a vector autoregression.

State-Space Model

We should discuss the SSM first since the SVM is part of the state space model (SSM) family pioneered by Kalman (1960). The most conventional one is the linear SSM which has been commonly used for times series estimation and forecasting. The SSM can be simply expressed as:

$y_t = \alpha_t + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma^2)$ (1)

$\alpha_t = \alpha_{t-1} + \mu_t$, $\mu_t \sim N(0, \sigma^2 q^2)$ (2)

Equation (1) is the first level of the SSM and is called the **measurement equation** which contains the observation and where the **observable y_t is the sum of unobserved α** and the error term. The subscript t is a time series notation; for example, t and t – 1 represent the current period and previous one period, respectively. **Equation (2)** is the second level of SSM where **α_t is the sum of its previous one period and the error term**. Equation (2) can be either called the **state or transition equation that captures the evolution of the state.**

Kalman filter: (Forecast from a linear function of previous observatons)

The Kalman filter is a **forecast of the state vector** as a linear function of **previous observations.** The **estimates** are **updated with weighted average**, the estimate with **greater certainty** carries **more weight**. It is **a recursive algorithm** for sequentially **updating a linear projection** for the system. It can calculate exact finite-sample forecast and the exact likelihood function for Gaussian process and can estimate vector autoregression coefficient that change over time.

Stochastic Calculus

Fundamental theorem of asset pricing (FTAP):

FTAP states that the following two conditions are equivalent.

(a) There is no arb

(b) There is an equivalent risk neutral measure

Radon Nikodym Derivative (RND)

ξ_t is a positive risk neutral probability. (ξ_t>0).

$\frac{dQ}{dP}|_{\mathcal{F}_t}=\xi_t$ = RND

Filtered space

$(\Omega, \mathcal{F}, P; F)$ where F={$\mathcal{F}_t$}

P is probability measured defined

on $(\Omega, \mathcal{F})$ and tells the probability of

each event in $\mathcal{F}$.

$\boldsymbol{F}$**:** sequence of sub σ-algebras of $\mathcal{F}$

$\boldsymbol{\mathcal{F}}$**:** a σ-algebra

$\boldsymbol{\Omega}$**:** set of events

Black Scholes PDE

As a consequence of the previous argument, we know that the hedging strategy is

$\Delta_t = \frac{\partial c}{\partial x}(t, S_t)$

Δ_t is also the number of shares needed in a replicating portfolio.

$c(t, x)$ satisfies the PDE: $\frac{\partial c}{\partial t}$+r$S_t \frac{\partial c}{\partial x} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 c}{\partial x^2}$=r $c(t, x)$

With boundary conditions: c(T, x)=$(x-K)^+$, c(t,0)=0, $\lim_{x\to\infty} \frac{c(t,x)}{x}$=1

Wiener process

It is a continuous time stochastic process. W(t) with W(0)=0 and such that increment W(t)-W(s) is Gaussian with **mean 0** and **variance t-s** and increments for nonoverlapping time intervals are independent. **Brownian motion** is an example of a Wiener process.

Think **dx** as being a **random variable**, drawn from a **normal distribution** with **mean zero** and **variance dt:**

E[dX]=0 and E[dX^2]=dt.

SDE

dS_t=$(r_t - d_t)S_t d_t$+$\sigma_t S_t$dW_t

where:

r_t: Instantaneous risk free rate

W_t: Wiener process, representing the inflow of randomness

σ_t: The amplitude of randomness is measured by this instant volatility

A SDE is an equation of the form:

dX(u)=$\beta(u, X(u))$du+$\gamma(u, X(u))$dW(u).

Here β(u,x) and γ(u, x) are functions, called the **drift** and **diffusion**, respectively. In addition, an initial condition of the form X(t)=x where t≥0 and x∈ℝ, is specified. The problem is then to find a stochastic process X(T), defined for T≤t, such that X(t)=x,

SDE is composed of two main components: **deterministic** and **random.**

Mathematically, ds=dt+dx The randomness is captured in dx. In other words, the change in stock price has a predictable outcome and a random component.

SDE equation from CQF

dG=a(G, t)dt+ b(G, t)dX

Where:

a$(\boldsymbol{G}, \boldsymbol{t})$ is **deterministic** and the coefficient of dt is known as **drift or growth**.

b$(\boldsymbol{G}, \boldsymbol{t})$ is **random** and the coefficient of dX is known as the **diffusion or volatility**.

What is drift?

The drift describes the **long-run behavior** of the process; it tells us whether it tends to some steady-state, explodes, or oscillation.

What is diffusion?

A diffusion process is a solution to a SDE. It is a **continuous time Markov process** with almost surely continuous sample paths. **Brownian motion and OU process** are examples of **diffusion process.**

SDE for interest rate that goes by OU process or the Vasicek model

dr = γ$(\bar{r} - r)$dt+σdX.

Risk neutral measure, arbitrage, and hedging

We have constructed a risk neutral measure $\tilde{P}$ a measure which is equivalent to P, and which makes the discounted stock price a **martingale**.

The **mean rate of return** of any asset under the risk neutral measure is **equal to the risk free rate r.**

Now what? There are two implications:

(i) The market is arbitrage free

(ii) We can **perfectly hedge**! The main tool is Martingale representation theorem

To talk about arbitrage and hedging, we need to first specify what trading strategies are allowed.

As before, we require portfolios to be self financing:

dX_t=r($X_t - \Delta_t S_t$)dt+Δ_tdS_t

Brownian motion

Are **independent** and each of these **increments** is **normally distributed** with:

E[W$(t_{i+1}) -$ W(t_i)]=0

Var[W$(t_{i+1}) -$ W(t_i)]=t_{i+1}-t_i. The increment has the **Normal (0, t) distribution.**

Let (Ω, F,P) be a probability space. For each w∈Ω, suppose there is a **continuous function** W(t) of t ≥0 that satisfies W(0) =0 and that depends on w. Then W(t), t ≥0 , is a Brownian motion if for all 0=t_0<t_1<…<t_m the **increments**

W(t_1)=W(t_1)-W(t_0), W(t_2)-W(t_1),…, W(t_m)-W(t_{m-1})

Brownian motion is **both** a **martingale** and a **Markov process**.

BM as a stochastic process W_t satisfying:

1. W_0 = 0
2. The sample path t→W(t) are continuous
3. Independent increments: $W_{t_4} - W_{t_3}$, $W_{t_2} - W_{t_1}$ are independent
4. **Normally distributed** increments: $W_t - W_s$ ~N$(0, |t-s|)$

Properties of Brownian motions:

1. Finiteness: Any scaling of the bet size ore increments with time step would have resulted in either a **random walk going infinity in a finite time.**
2. Continuity: **Paths are continuous**

3. Markov: Future probability are determined by its most recent values.
4. Martingale: Expectation is constant over time.
5. Quadratic variation
6. **Normality**

Stochastic process

A SP is a **collection** of **random variable indexed by time.**

Markov process: A random process whose **future** prob. are determined by its **most recent** values.

Martingale

Martingale is a stochastic process that has no dream. It is the idea of a fair **random game.**

A **martingale** is a **sequence of random variables** (i.e., a stochastic process) for which, at a particular time, the **conditional expectation** of the **next value** in the sequence is **equal to the present value**, regardless of all prior values. The **expectation** of a Martingale **is constant over time.**

Girsanov theory

The collection of the theorems that tell us how to **make drift disappear** is called the Girsanov theory.

The Feyman Kac formula

By the **Feynman kac formula**, the value V(t, S_t) of the option solves the **boundary value problem**: $\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}+rS\frac{\partial V}{\partial S}-rV=0$

V(T,s)=G(s) which is the Black Scholes PDE.

Consider the coin example in which head is probability p.

If p = 1/2: martingale (gambler neither wins nor loses)

If p<1/2: supermartingale (gambler loses money over time)

If p>1/2: submartingale (gambler wins money over time)

Reflection principle
The RP of BM states that BM reflected at **some stopping time** is still a BM.

European options: discounted price process is a **martingale**
American options: discounted price process is **supermartingale.**

Stopping time:
A **random time** that makes the **decision to stop (exercise)** without looking ahead. The exercise strategy of the owner of an **American options** should be a stopping time.

Prove that the IV of put and IV of call are the same
Use Put Call parity

Is it ever optimal to early exercise an American call options? What about a put options?
For call options, an option will always be **worth more than its value if exercised immediately.**

For **put options**, there is not an equivalent argument for put options.
There are times when the **optionality value is worth less than receiving the payoff today** and hence it is **optimal to early exercise.**

What happens to price of a vanilla call options as volatility tends to infinity?
The price of a vanilla call option is **monotone increasing in volatility,** so as volatility tends to infinity the option price will tend to its maximum value.

In the pricing of options, why doesn't it matter if the stock price exhibits mean reversion?
Mean reversion is the real word dynamics of the stock process but to price an option we are only interested in the **risk neutral dynamics**. The real world drift be it mean reverting or not, does not change the price of the option.

Explain the Longstaff-Schwartz algorithm for pricing an early exercisable options with Monte Carlo.

1. Run a certain number of Monte Carlo paths and store the exercise value divided by the numeraire at each possible exercise date for each path.
2. Anow look at the second last exercise date and perform a regression of the value at the last exercise data against the basis function.
3. Use these coefficients we can calculate the **continuation value** at the second last exercise time for each path.
4. We now replace the deflated exercise value stored initially with the continuation value if the continuation value is greater.
5. Continue back through the exercise dates until reaching time 0. The biased value of the product will be the average of the time zero values multiplied by the initial numeraire value.

Probability and other math for brainteaser

32^{19}-32

Pattern of unit of power of 2: 2, 4, 8, 6
Since there are 4 number repeat
19/4=4…3
So the third number is 8
8-2=6

A bag has 4 white, 5 black, and 3 red, random pick 3 balls, find

1. probability of one red, white, and black: $\frac{C_1^4 C_1^5 C_1^3}{C_3^{12}} = \frac{3}{11}$

2. probability of three same color: $\frac{C_3^4 + C_3^5 + C_3^3}{C_3^{12}} = \frac{3}{44}$

2. probability of at least one black = $1 - (Probability\ of\ none\ of\ three\ is\ black)$: $1 - \frac{C_3^7}{C_3^{12}}$

Bag A has 6 red and 4 white and Bag B has r red and 6 white, now random pick 2 from bag A and B, find probability of all 4 balls have the same color?

P(A) = $\frac{C_2^6 C_2^4}{C_2^{10} C_2^{10}} = \frac{2}{45}$; P(B)= $\frac{C_2^4 C_2^6}{C_2^{10} C_2^{10}} = \frac{2}{45}$

P(A)+ P(B) = $\frac{4}{45}$

Bag has 3 white, 4 black, and 5 red, random pick 1 without replacement, draw for 3 times.

(1) Probability of white, black, and red in the order

$\frac{3}{12} \text{ x } \frac{4}{11} \text{ x } \frac{5}{10} = \frac{1}{22}$

(2) Probability of black in the second draw

$\frac{4}{12}\frac{3}{11} + \frac{8}{12}\frac{4}{11}$

Repeated experiment: Bag has 2 red and 3 white, random pick 1 and then put it back, draw for 5 times

(1) Probability of exactly 3 red

$C_3^5 \left(\frac{2}{5}\right)^3 \left(\frac{3}{5}\right)^2 = \frac{144}{625}$

(2) Probability of 3 red and the last draw is also red

$C_2^4 \left(\frac{2}{5}\right)^2 \left(\frac{3}{5}\right)^2 \frac{2}{5} = \frac{432}{3125}$

A box has 50 draws, among them 3 draws are the reward draws, now random draw 3.

(1) Probability of only 1 is reward

$\frac{C_1^3 C_2^{47}}{C_3^{50}}$

(2) Probability of at least 1 is reward

$\frac{C_3^{50} - C_3^{47}}{C_3^{50}}$

Two cards chosen at random from a deck of 52 playing cards. Find prob that they

(a)Are both aces?

Total: 52 cards; Aces: 4 cards; Draw: 2 cards

Method: 2 cards chosen randomly

P(two cards are both aces) = $\frac{C_2^4}{C_2^{52}}$ = $\frac{6}{1326}$ = 0.0045

(b) Have the same value?

Since within 52 cards, there are 13 decks numbered from ace (1) to K (13) and each deck has 4 cards.

P(two cards have the same value) = 13$\left(\frac{C_2^4}{C_2^{52}}\right)$ = 0.0585

Given 20 people and if everyone does handshake with each other:

$\binom{20}{2} = \frac{20!}{2!(20-2)!}$ = 190 handshakes

Given 10 lifters and among them: 3 from US, 4 from Russia, 2 from China, 1 from Canada.

Different possible outcomes:

$\frac{10!}{3!4!2!1!}$ = 12600 outcomes

1. Let $\{X_i\}$ be iid with E(X_i)= u and VAR(X_i) = σ^2. Obtain the probability limit and asymptotic distribution of 1/$\bar{X}$

Step one: [Weak Law of Large Numbers]

$\bar{X}$ = 1/n $\sum X_i \xrightarrow{p}$ u

Step two: [Slustky's theorem]

Consider g(x)=1/x

$\bar{X} \xrightarrow{p}$ u → 1/$\bar{X} \xrightarrow{p}$ 1/u

In probability theory, Slutsky's theorem extends some properties of algebraic operations on **convergent sequences** of real numbers to sequences of random variables.

Step three: [Lindeberg-Levy]

$\bar{X} \xrightarrow{d}$ N(u, $\frac{\sigma^2}{n}$)

Finally, [Delta method]

g(σ) $\xrightarrow{d}$ N(g(u), $[g'(u)]^2 \frac{\sigma^2}{n}$) → $\frac{1}{\bar{X}} \xrightarrow{a}$ N($\frac{1}{u}$, $\frac{\sigma^2}{nu^4}$)

The delta method is a general method for **deriving the variance of a function of asymptotically normal random variables with known variance**.

Hilbert space

In mathematics, Hilbert spaces (named after David Hilbert) allow generalizing the methods of linear algebra and calculus from (**finite-dimensional**) Euclidean vector spaces to spaces that may be infinite-dimensional. Hilbert spaces arise naturally and frequently in mathematics and physics, typically as function spaces. Formally, a Hilbert space is a **vector space** equipped with an **inner product** that defines a **distance function** for which the space is a complete metric space.

A Hilbert space is a complete inner product space.

Banach space

B is a complete normed space, that is, a normed space in which every Cauchy sequence **converges in norm.**