# Notes on Quantitative Finance

C. Chan

*Essentially all models are wrong, but some are useful.*

George Box (Box & Draper, 1987)

Compiled on June 1, 2019

## Contents

# The Economic Machine

Ref: [Economic Principles]
*Transactions* drives
(1) productivity growth
(2) Short term debt cycle
(3) Long term debt cycle

## Transactions

▷ An economy is simply the sum of the transactions that make it up. Buyers buy with money or credit (sum to total spending); sellers sell the goods, services, or financial assets.

$$\text{Total spendings} = \text{money spent} + \text{credit spent}$$
$$\text{Price} = \frac{\text{total spendings}}{\text{total quantity}} \tag{1.1}$$

Buyers & sellers form a market.

## Participants

▷ *People*, *businesses*, *banks* & *governments* all engage in transactions. The biggest buyer and seller is the government, which consists of 2 important parts: (1) The central government that collects taxes & spends money, and (2) a central bank that controls the amount of money and credit in the economy. The central bank does this by *influencing interest rates* and *printing new money*.

## Credits

▷ Involve a lender & a borrower. Borrower wants to buy something that they cannot afford, or to start a new business. Borrowers promise to repay the amount they borrow, namely the principal, plus an additional amount, called interest.

$$\text{High interest rate} \implies \text{less borrowing}$$
$$\text{Low interest rate} \implies \text{more borrowing}$$

▷ When a *lending* occurs, credit is created out of thin air! and it immediately turns into debt. Debt is an asset to the lender, and a liability to the borrower. In the future, when the principal & the interest is repayed, then the credit & the debt annihilate, and the transaction is settled.
▷ Credit is important. This is because with credit the borrower can increase his spending with a small cost of paying the interest. Recall that spending drives the economy, because one person's spending is another person's income.
▷ Creditworthy borrower has 2 things: (1) the ability to repay (e.g., high income) & (2) collateral. Hence a *self-reinforcing* pattern

$$\uparrow \text{ income } \implies \uparrow \text{ borrowing } \implies \uparrow \text{ spending } \implies \uparrow \text{ income (of another person) } \implies \cdots$$

which leads to economic growth. This leads to *cycles*.

## Cycles

▷ Productivity growth is usually steady and long term. The *fluctuations* of economic growth is driven by the credits or debts. This is because debt allows us to consume more than we produce when we acquire it, and it forces us to consume less than we produce when we pay it back.

▷ There are 2 cycles of debt cycles: The short-term cycle takes ∼5-8 years, and the long-term cycle takes ∼75-100 years.
▷ Credit is **bad** only when it finances over-consumption that cannot be paid back (default). Credit is **good** when it efficiently allocates resources and produces income so you can pay back the debt.

### Short-term debt cycle

▷ Firstly, an expansion of *economy*, then followed by *recession*.
▷ During expansion, price increases [recall eq (1.1)] when the amount of spending & incomes grow faster than the production of goods. This leads to inflation.
▷ The central bank does not want too much inflation. Then it increases interest rate, then fewer people can afford to borrow money, and cost of existing debts rises. Again from the self-reinforcing pattern

$$\uparrow \text{ interest } \implies \downarrow \text{ spending } \implies \downarrow \text{ income (of another person) } \implies \cdots$$

Hence the total spending decreases, and the price also decreases [eq (1.1)]. This is called a deflation. When the economic activity decreases and we have a recession. If the recession is too severe, the central bank will lower interest rates to cause everything to pick up (expand) again.
▷ When credit is (not) easily available, there is an economic expansion (recession).

### Long-term debt cycle

▷ Human nature: they have an inclination to **borrow & spend more** instead of paying back debt! Over long period of time, debts rise *faster* than incomes $\implies$ Long-term debt cycle.
▷ Despite people becoming more indebted, lenders even more freely extend credit. This is because everyone thinks things are going great! **Short-sighted**: People are just focuing on what's been happening *lately*: (1) Income has been rising! (2) Asset values are going up! (3) The stock market roars! It's a Boom! People pay more to buy goods & serivices. $\implies$ Bubble.
▷

$$\text{Creditworthy } \textit{if} \text{ income increase rate} > \text{debt increase rate}$$

This leads us to define

$$\text{Debt burden} = \frac{\text{debt}}{\text{income}}$$
$$\text{Creditworthy } \textit{if} \text{ debt burden} < 1$$

When the demand > supply, lots of people borrow money to buy assets, and causing their prices to rise. Even there is a lot of debts, but rising incomes and asset values help borrowers *remain creditworthy for a long time*. **People feel wealthy**.
But this cannot continue forever. This stops when debt repayments start growing faster than incomes, forcing people to cut back on their spending $\implies$ less income (of another person) $\implies$ less creditworthy $\implies$ less borrowing $\implies$ less spending $\implies \cdots$
This is the long-term debt peak. The economy begins deleveraging. During deleveraging, people cut spending, income fall, credit disappears, asset prices drop, banks get squeezed, the stock market crashes, social tensions rise, and the whole thing starts to feed on itself the other way. As incomes fall and debt repayments rise, borrowers get squeezed. No longer creditworthy, credit dries up and borrower can no longer borrow enough money to make their debt repayments. Scrambling to fill this hole, borrowers are force to sell assets. The rush to sell assets floods the market. This is when the stock market collapses, the real estate market tanks and banks get into trouble. As asset prices drop, the value of collateral borrowers can put up drops. This makes borrowers even less creditworthy. **People feel poor**. The reinforcing pattern of a vicious

cycle: less spending $\implies$ less income $\implies$ less wealth $\implies$ less credit $\implies$ less borrwing $\implies$ less spending (of another person) ...
A deleveraging is different from a recession, because the interest rate cannot be lowered anymore! The debt burden is simply too big. Lenders refuse to lend.

### Cut debt burden

(1) *Spending cuts*, (2) *Debt reduction*, (3) *Wealth redistribution*, (4) *Print money.*
▷ Spending cuts (Austerity 緊縮財政政策): People, businesses, and governments cut their spending $\implies$ falling income $\implies$ not creditworthy borrowers $\implies$ debt burden worse (*deflationary*)
▷ Debt reduction: Debts are reduced via defaults & reconstructurings.
When borrowers cannot repay the debt to the bank, people are nervous that the bank won't be able to repay them, so they rush to withdraw their money from the bank. Banks get squeezed. People, businesses and banks default on their debts. This severe economic contraction is a depression.
Many lenders don't want their assets to disappear and agree to debt restructuring. This means lenders get paid back less or get paid back over a longer time frame or at a lower interest rate that was first agreed. Somehow, a contract is broken in a way that reduces debt. Lenders would still want to have little of something rather than all of nothing.
Even debt disappears, debt restructuring causes income and asset values to disappear *faster*, so the debt burden continues to get worse. Debt reduction is also painful and deflationary.
▷ Wealth redistribution: Wealth is redistributed from the 'haves' to the 'have nots'
Central government collects less taxes, but has to increase spending on financial support to the unemployed. Governments' budget deficits explode in a deleveraging because they spend more than they earn in taxes. To fund their deficits, governments need to (1) raise taxes or (2) borrow money, *from the rich.* This could lead to *social & political instability.*
▷ Print money: The central bank prints new money, which is *inflationary*! The money is used to buy financial assets and government bonds. The central bank and the government cooperate. The central bank buys government bonds, the central bank essentially lends money to the government, allowing it to run a deficit and increase spending on stimulus program on goods & services, and unemployment benefits. This increases people's incomes, as well as the government's debt. Yet it will lower the economy's total debt burden.
▷ Policy makers need to *balance* the deflationary and the inflationary ways in order to maintain stability. If balanced correctly, there can be a *beautiful deleveraging.* In such phase, debts decline relative to income, real economic growth is positive, and inflation isn't a problem.
Would printing money raise inflation? It won't if it offsets the falling credit. The money is printed in such a way that to counter the credit disappearance and maintain a *constant* total spending ( $\implies$ constant price).
▷ *Leveraging* (50+ years) $\to$ *Depression* (2-3 years) $\to$ *Reflation* (7-10 years). *Lost decades* include the depression & the reflation phases.

### Three Rule of Thumb

1. Don't have debt rise faster than income – because your debt burdens will eventually crush you
2. Don't have income rise faster than productivity – because you will eventually become uncompetitive
3. Do all you can to increase your productivity – because, in the long run, that's what matters most

## Gambling Statistically

▷ Kelly Criterion: Given a fixed amount of money $n$, and a random variable $\phi_i$ of outcome for a game and we bet a constant fraction $f$ each time, how to maximize the expected long-term

growth rate?
After $M$ hands I have

$$n\prod_{i=1}^{M}(1+f\phi_i)$$

We can maximize the (logarithmic) growth

$$\frac{1}{M}\sum_{i=1}^{M}\log(1+f\phi_i)$$

**Assume** that the outcome of each hand is independent, then the expected value each term is

$$E[\log(1+f\phi_i)]$$

Suppose (**Big suppose**) we can expand the argument, we have

$$E[f\phi_i - \frac{1}{2}f^2\phi_i^2 + \dots]$$

Also **assume** that $E[\phi_i^2] \gg E[\phi_i]$, then the expected long-term growth rate is approximately

$$f\mu - \frac{1}{2}f^2\sigma^2$$

where $\mu = E[\phi_i]$ and $\sigma^2 = \mathrm{Var}[\phi_i]$.
This is maximized by the choice

$$f^* = \frac{\mu}{\sigma^2}$$

giving an expected growth rate of $\frac{\mu^2}{2\sigma^2}$ per hand.
– Many people believe the Kelly criterion to be a quite aggressive strategy leading to possible large downturns.
– One may also maximize the expected wealth, or minimize the downturn, or maximize some risk-adjusted return (mean-variance analysis).
▷ Gambler's Ruin: Initially a gambler has money $n$, the returns of winning and losing (equal chances) are $+1$ and $-1$ respectively. If the gambler continue the game indefinitely, what is the probability of the wealth becoming zero?
*Answer*: **100%**.

▶ *Proof*: Denote the probability from money $n$ to 0 as $P(n)$ for continuous playing the game, then

$$P(n) = \frac{1}{2}\left[P(n+1) + P(n-1)\right]$$

for $n > 0$ and also $P(0) = 1$. Then we have the recursion relation

$$P(n+1) = 2P(n) - P(n-1)$$

Let $P(1) = a$, then obviously $0 < a \leq 1$. [If $a = 0$, then $P(2) = 2a - P(0) = -1$!] Using the recursion relation, we have

$$P(n) = n(a-1) + 1$$

Since $P(n) \geq 0$ for all $n$. Let $a = 1-\epsilon$, then $P(n) = 1 - n\epsilon \geq 0$ for all $n$. Hence,

$$\begin{aligned} \epsilon \leq \frac{1}{n}, \quad \forall n \\ \implies a = 1 \\ \implies P(n) = 1, \quad \forall n \end{aligned}$$

◀

# Statistics & Probability Theory

## Auto-correlation

– Ref: [Schmidt] *Quantitative Finance for Physicists*
▷ *Covariance*: $\sigma(x,y) = E[(x-\mu_x)(y-\mu_y)]$
▷ *Sample covariance*: $\text{Cov}(x,y) = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$
▷ *Correlation*: $\rho(x,y) = \frac{\sigma(x,y)}{\sigma_x \sigma_y}$
▷ *Sample correlation*: $\text{Corr}(x,y) = \frac{\text{Cov}(x,y)}{\sqrt{\text{Var}(x)\text{Var}(y)}}$
▷ *Mean of a time series* $x_t$: $\mu(t) = E(x_t)$, which is the *population mean* of multiple time series $\{x_t\}$
▷ *Stationary in the mean* if $\mu(t) = \mu$
▷ *Variance of a time series*: $\sigma^2(t) = E[(x_t - \mu)^2]$, **assume** that the time series is **stationary**
▷ *Stationary in the variance* if $\sigma^2(t) = \sigma^2$. **Assumed** for calculating the *sample variance* for a time series

$$\text{Var}(x) = \frac{\sum(x_t - \bar{x})^2}{n-1}$$

▷ *Second order stationary*: see Text
▷ *Auto-covariance*: If a time series is second order stationary, then the (population) *auto-covariance*, of lag $k$ is

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

▷ *Auto-correlation*: $\rho_k = C_k/\sigma^2$. Note that $\rho_0 = 1$
▷ *Sample auto-covariance function*: $c_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})$
▷ *Sample auto-correlation function*: $r_k = c_k/c_0$
▷ *Covariance matrix*: In the general case with $N$ variates (random variables) $X_i$, $i = 1, \ldots, N$, correlations among variates are $\text{Cov}(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)]$.

## Statistics

▷ *Moments*: $m_n \equiv E[X^n] = \int x^n P(x) dx$
▷ *Skewness*: $S = E[(x-\mu)^3]/\sigma^3$. $S = 0$ implies the distribution is symmetrical around its mean value. For $S > 0$ ($S < 0$) indicate long positive (negative) tails.
▷ *Kurtosis*: $K = E[(x-\mu)^4]/\sigma^4$ characterizes the distribution peakedness. Normal distribution has $K = 3$.
▷ *Excess kurtosis*: $K_e = K - 3$. Positive excess kurtosis (or *leptokurtosis*) indicates more frequent medium and large deviations from the mean value than is typical for the normal distribution. Leptokurtosis leads to a flatter central part and to so-called fat tails in the distribution. Negative excess kurtosis indicates frequent small deviations from the mean value. In this case, the distribution sharpens around its mean while the distribution tails decay faster than that of the normal distribution.

## Important Distributions

▷ Uniform distribution
▷ Binomial distribution.
For large $N$ and large $N - n$, binomial distribution $\to$ normal distribution.
For $p \ll 1$, binomial distribution $\to$ Poisson distribution.
▷ Poison distribution
▷ Normal distribution

*Central limit theorem*: the probability density distribution for a **sum** $S_n = \sum_{k=1}^{n} X_k$ of $N$ independent and identically distributed random variables $X_i$ with finite variances and finite means approaches the normal distribution $S_n \sim \mathcal{N}(\mu, \sigma^2)$, as $N$ grows to infinity.
▷ Cauchy distribution (*Lorentzian*)
▷ Extreme value distribution

## *Notes*: Statistical Thinking

Ref: [Poldrack] statistical thinking for the 21st Century, draft 2018-12-07
▷ [5.8] Using simulations to understand statistics
▷ [5.9] $Z$-scores: $Z(x) = \frac{x-\mu}{\sigma}$. For Gaussian: $P(|Z| \le 1) = 68.27\%$, $P(|Z| \le 2) = 95.45\%$, $P(|Z| \le 3) = 99.73\%$. *Note*: $Z$-scores are directly comparable.
▷ [7.4] Central Limit Theorem: As sample sizes get larger, the sampling distribution of the mean will become normally distributed, even if the data within each sample are not normally distributed.
▷ [8] Resampling & Simulations (Need revisit!):
▷ [9.1] NHST (null hypothesis statistical testing): Given the null hypothesis$H_0$ (model), compute conditional probability $P(\text{data}|H_0)$, namely the $p$-value. If the probability is too low, say $P < 0.05$, then we reject $H_0$ and accept the alternative hypothesis $H_1$.
▷ [9.3.4] $t$-test for mean difference = 0? Compute

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

where $\bar{X}_i$, $S_i^2$ and $n_i$ are the estimated means, variances and sizes.
▷ [9.3.5] Randomization method (Need revisit!): Assume 2 distributions are the same, then the data can be randomized into one set $X \ni x_i, x_j$. Then consider $y_{ij} = x_i - x_j$. Finally perform $t$-test on $y_{ij}$. Advantage: It does not requre the assumption of normal distributed data, nor any theoretical distribution. But we need to assume the *exchangability* of the data.
▷ [9.3.6.1] Neyman-Pearson Approach: $\alpha = P(\text{type I error})$, $\beta = P(\text{type II error})$.
▷ [9.4] NHST with Multiple Testing (Need revisit!): Also need to minimize the *familywise error*.
▷ [10.2] Size Effect or Signal-to-Noise ratio: (i) Cohen's $D$, (ii) Pearson's $r$
▷ [10.3] Statistical power: $1-\beta$. If an experiment is *futile*, meaning that it is almost guaranteed to find nothing even if a true effect of that size exists, then it would be a waste of time to perform the experiment!
▷ [11.4.7] Credible intervals for Bayesian statistics
▷ [11.6] Bayesian hypothesis testing with *Bayes factors*:

$$\text{BF} = \frac{p(\text{data}|H_1)}{p(\text{data}|H_2)}$$

characterizes the relative likelihood of the data under 2 different hypotheses
▷ [12.2] Pearson's $\chi^2$ test for categorical relationships:

$$\chi^2 = \sum_i \frac{(\text{observed}_i - \text{expected}_i)^2}{\text{expected}_i}$$

▷ [12.3] Contingency Tables: Test the *independence* between variables $X$ and $Y$. For example, searching of the driver against driver's color.
▷ [15] Comparing Means: (i) Student's $t$-test, (ii) Bayes factor for mean differences, (iii) Paired $t$-tests / Sign test, (iv) comparing more than 2 means.

# Financial Markets

Ref: [Schmidt] *Quantitative Finance for Physicists*

## Market Price Formation

▷ Market price, bid (buy), ask (sell), trader orders (quotes), spread between best (highest) bid & best (lowest) ask prices
▷ Market orders, limit orders, stop orders
▷ Long position, short selling
▷ Market microstructure, liquidity (e.g., low liquidity means that the number of securities available at the best price is smaller than a typical market order), tick (any event that affects the market microstructure)

## Returns & Dividends

▷ Price $S$: its logarithm $p = \log S$ is often used in quantitative anlysis, since $S \geq 0$ while random price innovation can move price into the negative region.
▷ (Simple) Return $R(t) = \frac{S(t)}{S(t-1)} - 1$
▷ (Compounded) Return $R(t,k) = \frac{S(t)}{S(t-k)} - 1$
▷ Portfolio return: Just a weighed sum of returns

$$R_p(t) = \sum_i w_i^p R_i^p(t), \quad \text{with } \sum_i w_i^p = 1$$

▷ Dividend effect: If dividends $D(t+1)$ are paid within the period $[t, t+1]$, the simple return is modified to

$$R(t+1) = \frac{S(t+1) + D(t+1)}{S(t)} - 1$$

▷ Present Value (PV) & Future value (FV): For investment $K$ in a risk-free asset (*zero-coupon bond*) with the interest rate $r$ evey time interval, after $n$ intervals,

$$\begin{aligned} \text{FV} &= K(1+r)^n \\ \text{PV} &= \frac{\text{FV}}{(1+r)^n} \end{aligned}$$

Calculating the present value via the future value is called *discounting*.
*Note*: No coupon (dividend) withdraw before the maturity date for such zero-coupon bond.
*Continuous version*:

$$\begin{aligned} \text{FV} &= Ke^{rt} \\ \text{PV} &= \text{FV}e^{-rt} \end{aligned}$$

▷ Discounted-cashflow model: Determines the stock price via its future cash flow. Assuming a constant return $R$, then the simple return can be written as

$$S(t) = \mathbb{E}\left[\frac{S(t+1) + D(t+1)}{1+R}\right]$$

Recursion for $K$ times, we have

$$S(t) = \mathbb{E}\left[\sum_{i=1}^{K} \frac{D(t+i)}{(1+R)^i}\right] + \mathbb{E}\left[\frac{S(t+K)}{(1+R)^K}\right]$$

In the limit $K \to \infty$, the second term can be neglected if $\lim_{K\to\infty} \mathbb{E}\left[S(t+K)/(1+R)^K\right] = 0$. Finally, the model yields

$$\begin{aligned} S_D(t) &= \mathbb{E}\left[\sum_{i=1}^{\infty} \frac{D(t+i)}{(1+R)^i}\right] \\ &= \frac{1+G}{R-G} D(t), \quad \text{assume that } \mathbb{E}\left[D(t+i)\right] = (1+G)^i D(t) \end{aligned}$$

## Time Scaling of Return's Probability Distribution

▷ Estimation Formulæ: The drift rate $\mu$ and the volatility $\sigma$ of the asset's return can be estimated by

$$\begin{aligned} \mu &= \frac{1}{M\delta t} \sum_{i=1}^{M} R_i \\ \sigma^2 &= \frac{1}{(M-1)\delta t} \sum_{i=1}^{M} (R_i - \mu)^2 \\ &\overset{\delta t \to 0}{\approx} \frac{1}{(M-1)\delta t} \sum_{i=1}^{M} \left[\log S(t_i) - \log S(t_{i-1})\right]^2 \end{aligned}$$

where $T = M\delta t$ is the total time. The approximated expression for variance is due to eq (4.1) below. For $\delta t \to 0$, $S_{i+1} \approx S_i(1 + \sigma\phi\sqrt{\delta t})$, then $\log S_{i+1}/S_i \approx \log(1 + \sigma\phi\sqrt{\delta t}) \approx \sigma\phi\sqrt{\delta t}$.
▷ Time Scaling of Returns: For a (normal) random walk of the asset's returns, the asset price is generated by

$$S_{i+1} = S_i(1 + \mu\delta t + \sigma\phi\sqrt{\delta t}) \tag{4.1}$$

where random variable $\phi \sim \mathcal{N}(0, 1)$ or

$$R \sim \mathcal{N}(\mu\delta t, \sigma^2\delta t)$$

The drift part is easily seen: Ignore the diffusion part for the moment, our return random walk is simply

$$\frac{S_{i+1}}{S_i} - 1 = \mu\delta t$$

After total time $T$, $S_M = S_0(1 + \mu\delta t)^M \approx S_0 e^{\mu T}$. For the variance, in order to obtain a finite $\sigma^2$, each term $(R_i - \mu)^2 \sim \mathcal{O}(\delta t)$, hence variance $\sim \sigma^2\delta t$ or standard deviation $\sim \sigma\sqrt{\delta t}$.
▷ Wiener Process: In the continuous time limit $\delta t \to 0$, we can simply replace $\delta t \to dt$. But for the stochastic term, we write

$$\phi\sqrt{dt} \to dW$$

One can regard $dW$ as a random variable drawn from a normal distribution with mean zero and variance $dt$, namely

$$E[dW] = 0, \quad E[dW^2] = dt$$

## Forward contracts / Futures

– Forward contract is not standard and signed between two parties.
– Futures are standardized and traded through an exchange
– Margins: The present value of futures is zero, since changes in the value of futures are settled each day (known as *marking to market*). To reduce the likelihood of one party defaulting, being unable or unwilling to pay up, the exchanges insist on traders depositing a sum of money, namely *margin*, to cover changes in the value of their positions. As the position is marked to

market daily, money is deposited or withdrawn from this margin account. See [Wilmott] Ch 26 for how to model margin and margin hedge.

| Holding | Worth today $(t)$ | Worth at maturity $(T)$ |
|---|---|---|
| Forward | 0 | $S(T) - F$ |
| Short stock | $-S(t)$ | $-S(T)$ |
| Cash | $S(t)$ | $S(t)e^{r(T-t)}$ |
| Total | 0 | $S(t)e^{r(T-t)} - F$ |

We show above the cashflows in this hedged portfolio. Since we began with a portfolio worth zero, and we end up with a predictable amount, that predictable amount should also be zero. (If we start with a portfolio worths $M$ at $t$, then it should worth $Me^{-r(T-t)}$.) Hence $F = S(t)e^{r(T-t)}$, which relates the spot price $S(t)$ and forward price $F$.

– FX futures: Suppose that the interest rate for the foreign currency (or similarly dividend) is $r_f$. We can construct a similar cashflow as above. The difference is that the cash at $T$ worth $S(t)e^{(r-r_f)(T-t)}$ since we gain a rate $r$ by holding cash, but we also lose a rate $r_f$ by shorting the FX, hence the cash at $T$ is

$$S(t)\left(1 + \frac{r}{m} - \frac{r_f}{m}\right)^{m(T-t)} \to S(t)e^{(r-r_f)(T-t)}$$

Finally, the forward price is $F = S(t)e^{(r-r_f)(T-t)}$.

## Stochastic (random) Processes

Ref: [Gardiner] Handbook of Stochastic Methods 3ed

▷ General: Assumed that the joint probability density $f(x_1, t_1; x_2, t_2; \dots)$ exists and defines the system completely. The conditional probability density function is defined as

$$f(x_1, t_1; x_2, t_2; \dots; x_k, t_k \mid x_{k+1}, t_{k+1}; x_{k+2}, t_{k+2}; \dots) = \frac{f(x_1, t_1; x_2, t_2; \dots; x_k, t_k; x_{k+1}, t_{k+1}; \dots)}{f(x_{k+1}, t_{k+1}; x_{k+2}, t_{k+2}; \dots)}$$

where the time orders as

$$t_1 > t_2 > \dots > t_k > t_{k+1} > \dots$$

▷ Simple stochastic process: No dependence on the past,

$$f(x_1, t_1; x_2, t_2; \dots) = \prod_i f(x_i, t_i)$$

### Markov Processes

✡ Markov process: The future depends *only* on the present (or initial condition) $t_{k+1}$ of value $x_{k+1}$ but not on the past,

$$f(x_1, t_1; x_2, t_2; \dots; x_k, t_k \mid x_{k+1}, t_{k+1}; x_{k+2}, t_{k+2}; \dots) = f(x_1, t_1; x_2, t_2; \dots; x_k, t_k \mid x_{k+1}, t_{k+1})$$

One can easily show that

$$f(x_1, t_1; x_2, t_2; \dots; x_k, t_k) = \left[\prod_{i=1}^{k-1} f(x_i, t_i \mid x_{i+1}, t_{i+1})\right] f(x_k, t_k)$$

▷ Chapman-Kolmogorov (semigroup) equation:

$$f(x_1, t_1 | x_3, t_3) = \int dx_2 f(x_1, t_1 | x_2, t_2) f(x_2, t_2 | x_3, t_3)$$

▷ Fokker-Planck equation: **Assume** to neglect the *discontinuous jumps* of the random variable given by

$$R(x|z,t) = \lim_{\delta t\to 0}\frac{1}{\delta t} f(x,t+\delta t|z,t)$$

which is uniform in $x, z$ and $t$ for $|x-z| \geq \epsilon$. Then the Fokker-Planck eq reads

$$\partial_t f(x,t|x_0,t_0) = -\partial_x[\mu(x,t)f(x,t|x_0,t_0)] + \frac{1}{2}\partial_x^2[D(x,t)f(x,t|x_0,t_0)]$$

where $\mu(x,t)$ and $D(x,t)$ are the drift and diffusion coefficients given by [up to $\mathcal{O}(\epsilon)$]

$$\mu(x,t) = \lim_{\delta t\to 0}\frac{1}{\delta t}\int_{|x-z|<\epsilon} dz(z-x)f(x,t+\delta t|z,t)$$
$$D(x,t) = \lim_{\delta t\to 0}\frac{1}{\delta t}\int_{|x-z|<\epsilon} dz(z-x)^2 f(x,t+\delta t|z,t)$$

▶ *Proof* (brief): [Gardiner] Consider time evolution of the expectation of a twice differentiable function $\psi(z)$,

$$\begin{aligned}
&\partial_t \int_x \psi(x)f(x,t|x_0,t_0)\\
&= \lim_{\delta t\to 0}\frac{1}{\delta t}\int_x \psi(x)\left[f(x,t+\delta t|x_0,t_0) - f(x,t|x_0,t_0)\right]\\
&= \lim_{\delta t\to 0}\frac{1}{\delta t}\left[\int_{x,z}\psi(x)f(x,t+\delta t|z,t)f(z,t|x_0,t_0) - \int_z \psi(z)f(z,t|x_0,t_0)\right]
\end{aligned}$$

Then expand

$$\psi(x) = \psi(z) + f'(z)(x-z) + \frac{1}{2}f''(z)(x-z)^2 + (x-z)^2R(x,z)$$

where $R(x,z)$ is the residue with $|R(x,z)| \to 0$ as $x \to z$, since $\psi(x)$ is twice differentiable. ◀

▶ *Aside*: The Fokker-Planck equation, also known as *forward (Kolmogorov) equation*, is a forward parabolic PDE, requiring initial conditions at time $t$ and to be solved for $t' > t$. On the other hand, we have the **backward equation** (or backward Kolmogorov eq)

$$\frac{\partial f}{\partial t} + \frac{1}{2}B(x,t)^2\frac{\partial^2 f}{\partial x^2} + A(x,t)\frac{\partial f}{\partial x} = 0,$$

for which a final conditions at time $t$ is given and to be solved for $t' < t$. See [Wilmott2] eq (10.4). Note that the BS equation is essentially a backward equation. ◀

▷ Brownian motion (Wiener process): **Assume** drift $\mu(x,t) = 0$ and diffusion $D(x,t) = D =$ const, we have the Brownian motion eq,

$$\partial_t f(x,t|x_0,t_0) = \frac{D}{2}\partial_x^2 f(x,t|x_0,t_0)$$

with the analytic solution of a Gaussian form

$$f(x,t|x_0,t_0) = \frac{1}{\sqrt{2\pi D(t-t_0)}}\exp\left[-\frac{(x-x_0)^2}{2D(t-t_0)}\right]$$

Mean and variance are

$$\mathbb{E}[x(t)] = x_0$$
$$\mathrm{Var}[x(t)] = \sigma^2 = D(t-t_0)$$

The *standard* Wiener process $W_t$ is defined with $D = 1$.
– *Summary*:
(1) $W_{t=0} = 0$
(2) independent increment: $\forall t, u > 0$, $W_{t+u} - W_t$ are independent of the past values $W_s$ for $s < t$
(3) Gaussian increment: $W_{t+u} - W_t \sim \mathcal{N}(0, \sigma_t^2)$ is normally distributed
(4) with probability 1, $W_t$ is continuous with $t$
– The Brownian motion can also be derived as the continuous limit for the discrete random walk. See [Schmidt] eq (4.2.7) for details.
▷ Generalization: (i) Markov chain with memory, (ii) Hidden Markov chain

## Stochastic Integral

The stochastic integral for a random process $X_t$ and the induced process $W_t$ is

$$\begin{aligned} W_t &= \int_0^t f(\tau) dX(\tau) \\ &= \lim_{n\to\infty} \sum_{j=1}^{n} f(t_{j-1})[X(t_j) - X(t_{j-1})] \end{aligned} \tag{5.1}$$

with $t_j = jt/n$. **Note** (non-anticipatory property): the function $f(t)$ is evaluated in the summation at the *left-hand point* $t_{j-1}$. It is crucial that each function evaluation does not know about the random increment $dX(t) \approx X(t_j) - X(t_{j-1})$. This is just *causality* that $X \implies f$. We define the **shorthand** from differentiating (5.1) as

$$dW_t = f(t) dX_t$$

where $dX_t \sim_{\text{i.i.d.}} \mathcal{N}(0, dt)$. Similarly,

$$dW = g(t)dt + f(t)dX \iff W_t = \int_0^t g(\tau)d\tau + \int_0^t f(\tau)dX(\tau)$$

▷ Mean Square Limit: Consider

$$\int_0^t (dX)^2 = \mathbb{E}\left[\left(\sum_{j=1}^n (X(t_j) - X(t_{j-1}))^2 - t\right)\right]$$

where $t_j = jt/n$. One can show that [Wilmott] in the limit $n \to \infty$

$$\int_0^t (dX)^2 = t \quad \text{or} \quad (dX)^2 = dt$$

## Stochastic Differential Equation

▷ Brownian motion in differential form:

$$dS_t = \mu dt + \sigma dW_t$$

where the term $dt$ is *deterministic*, and the stochastic term $dW_t \equiv W_{t+dt} - W_t$ has the following properties

$$\mathbb{E}[dW] = 0,\ \mathbb{E}[dW dW] = dt,\ \mathbb{E}[dW dt] = 0$$

Then up to order $\mathcal{O}(dt)$, the *quadratic covariance*

$$\mathbb{E}[(dS)^2] \approx \sigma^2 dt + \cdots$$

is deterministic.
▷ Itô Lemma: Consider a function $F(S_t, t)$ that depends on both deterministic variable $t$ and stochastic variable $S_t$. We expand the differential for $F(S, t)$ into the Taylor series retaining linear terms,

$$\begin{aligned} dF(S,t) &= \frac{\partial F}{\partial S}dS + \frac{\partial F}{\partial t}dt + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}(dS)^2 \\ &= \frac{\partial F}{\partial S}dS + \left[\frac{\partial F}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 F}{\partial S^2}\right]dt \end{aligned}$$

Note that the term $\frac{\sigma^2}{2}\frac{\partial^2 F}{\partial S^2}dt$ has stochastic nature.

▶ *Formally*, let $S_t$ be an Itô process given by $dS_t = \mu dt + \sigma dW_t$. Let $g(t,s) \in C^2([0,\infty) \times \mathbb{R})$, meaning that $g$ is twice continuously differentiable on $[0,\infty) \times \mathbb{R}$. Then

$$F_t = g(t, S_t)$$

is again an Itô process, and

$$dF_t = \frac{\partial g}{\partial t}(t, S_t)dt + \frac{\partial g}{\partial s}(t, S_t)dS_t + \frac{1}{2}\frac{\partial^2 g}{\partial s^2}(t, S_t) \cdot (dS_t)^2$$

where $(dS_t)^2 = (dS_t) \cdot (dS_t)$ is computed according to the rules

$$dt \cdot dt = dt \cdot dW_t = dW_t \cdot dt = 0, \quad dW_t \cdot dW_t = dt.$$

◀

## Options

▷ Stock price $S$, strike price $K$, time to maturity $T$, stock price volatility $\sigma$, risk-free interest rate $r$, dividends $D$ paid during the life of the option
▷ **Intrinsic value**: $I_t = S_t - K$ for call, $I_t = K - S_t$ for put. **Time value**: $= P - I_t$, the value that the option has above its intrinsic value. **In the money**: if $I_t > 0$. **Out of the money**: if $I_t < 0$. **At the money**: $I_t \approx 0$.
▷ **Factors** affecting option price (other factors kept fixed):
– Longer $T$ increases the value of an American option (than European one), since its holders have more opportunity to exercise it with profit.
– Growing $\sigma$ increases the value of both call and put options, since it yields better chances to exercise them with higher payoff.
– Effect of $r$ is NOT straightforward:
(i) At a fixed $S$, the rising of $r$ increases the value of the call option, since the defer payment to buy the stock can have return $r$
(ii) At a fixed $S$, the rising of $r$ decreases the value of the put option, since the defer receive of the payment from selling the stock loses the chance of return $r$
(iii) The above result is based on the assumption of fixed $S$. However, rising $r$ often lead to falling stock prices.
– Dividends $D$ effectively reduce the stock prices, hence it decreases (increases) value of call (put) options.
▷ **Payoffs** $P$ (pr is the premium, priced by the option writing institution)
– Long call option: $P = \max(S - K, 0) - \text{pr}$
– Short call option: $P = \min(K - S, 0) + \text{pr}$
– Long put option: $P = \max(K - S, 0) - \text{pr}$

– Short put option: $P = \min(S - K, 0) + \text{pr}$

▷ Put-call parity (*European* option): Assuming the share does not pay dividends, and no arbitrage opportunity, the following portfolios have the same value at maturity: (i) one call option at price $c$ and zero-coupon bond of present value $Ke^{-r(T-t)}$, (ii) one put option at price $p$ and one share at price $S$. Both call and put option has the same strike price $K$ and maturity time $T$. Both portfolios at maturity have the same value (see [Hull-9ed] Table 11.2)

$$\max(S_T, K)$$

Hence they must have the same present value,

$$c + Ke^{-r(T-t)} = p + S$$

Dividends affect the put-call parity. Namely, $D$ being paid during the option lifetime have the same effect as the cash future value. Thus, we have

$$c + D + Ke^{-r(T-t)} = p + S$$

▷ Risk-free arbitrage: Suppose there are 2 portfolios $\phi_{1,2}$ of the same underlying asset, and their values differ. WLOG, we assume $\phi_1 < \phi_2$. Then we can contruct a portfolio $\phi = \phi_1 - \phi_2$ to *immediately* gain a profit. The operation is to buy $\phi_1$ and then immediately sells it at $\phi_2$ with profit $\phi_2 - \phi_1 > 0$.

## Binomial options pricing model

*Assumption*: The current stock price $S$ can change at the next moment to higher value $Su$ ($u > 1$) or the lower value $Sd$ ($d < 1$).
Consider a portfolio $\Pi$ that consists of $\Delta$ long shares and one short option of price $F$. The rationale behind is that the portfolio is *risk-free* (or *risk-neutral*), meaning $\Pi$ in the next moment worths the same irrespective to the stock's direction (up or down).
After a time step $\tau$: This portfolio is risk-free if its value does not depend on whether the stock price moves up or down, namely

$$\begin{aligned} Su\Delta - F_u &= Sd\Delta - F_d \\ \implies \Delta &= \frac{F_u - F_d}{Su - Sd} \end{aligned}$$

Using *risk-neutral valuation* (that is no arbitrage opportunity), the portfolio's PV related to the FV by

$$(S\Delta - F)e^{r\tau} = Su\Delta - F_u = Sd\Delta - F_d \tag{6.1}$$

where $r$ is the risk-free interest rate, and $\tau$ is the time interval. Then we get the present price of the option

$$\begin{aligned} F &= S\Delta - e^{-r\tau}(Su\Delta - F_u) \\ &= e^{-r\tau}[pF_u + (1-p)F_d] \\ p &= \frac{e^{r\tau} - d}{u - d} \end{aligned}$$

The factor $p$ and $(1-p)$ have the sense of (*risk-neutral*) probabilities for the stock price to move up or down, respectively. Then the expectation of the stock price at time $\tau$ is

$$\mathbb{E}[S_\tau] = \mathbb{E}[pSu + (1-p)Sd] = Se^{r\tau}$$

meaning that $S$ grows on average with the risk-free rate $r$.

After $2\tau$:

$$F = e^{-2r\tau}[p^2 F_{uu} + 2p(1-p)F_{ud} + (1-p)^2 F_{dd}]$$

**Note**: The *risk-neutral* probability $p$ above is a mathematical construct, not necessarily equal to the *real* probability $p'$ of stock fluctuations. [Wilmott]
**Note**: The option price $F$ is independent of $p'$ (related to the drift $\mu$ in geometric Brownian motion for stock price), and is completely determined by $r, d, u$.
After $n\tau$ [wiki]: This procedure can be generalized for a tree with an arbitrary number of steps. Specifically, the stock prices at every node are calculated by going forward from the first node to the final nodes. When the stock prices at the final node are known, we can determine the option prices at the final nodes by using the relevant payoff relation. Then we calculate the option prices at all other nodes by going backward recursively.
Continuum Limit: In this limit, the model reduces to the Black-Scholes model. [Wilmott]
Early Exercise: [Wilmott] For American options, we use the same binomial tree with the same $u, d, p$, but we must ensure that there is no arbitrage opportunities at any of the nodes.
Estimate the factor $u$ and $d$: From historical stock price volatility. For a geometric Brownian motion $dS = S(\mu dt + \sigma dW)$, the price changes within the time interval $[0, t]$ are described by the lognormal distribution

$$\begin{aligned}
\ln S_t &= \mathcal{N}(\ln S_0 + (\mu - \sigma^2/2)t, \sigma\sqrt{t}) \\
\mathbb{E}[S_t] &= S_0 e^{\mu t} \\
\mathrm{Var}[S_t] &= S_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1)
\end{aligned}$$

With a further assumption of $u = 1/d$, we can derive [Schmidt]

$$\begin{aligned}
p &= \frac{e^{r\delta t} - d}{u - d} \\
u &= 1/d = e^{\sigma\sqrt{\delta t}}
\end{aligned}$$

where $\delta t$ is a small time interval. The assumption ensures that $(Su)d = (Sd)u = S$.
Generalizations:
(i) dividends and varying $r$ can be included
(ii) trinomial tree model (with up, down, unchanged states) can be considered

### Other pricing methods

To see how other methods for pricing options, see [Josphi] section 2.1, [Wilmott2] Ch 59.

## Black-Scholes Theory

▷ Geometric Brownian motion (lognormal random walk): Given Itô SDE $\frac{dS}{S} = \mu dt + \sigma dW$ and $W$ is a Wiener process, consider a function $F(S, t)$ with differential

$$dF = \frac{\partial F}{\partial S} dS + \frac{\partial F}{\partial t} dt + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}(dS)^2$$

Note that $(dS)^2 \approx \sigma^2 S^2 dt$. Then for $F(S, t) = \ln S$,

$$dF = \frac{dS}{S} + \frac{1}{2}\left(-\frac{1}{S^2}\right)\sigma^2 S^2 dt$$

► *Aside*: $100(1-0.2) = 80$, $80(1+0.2) = 96 < 100$, hence the minus sign for $F''(S) = -\frac{1}{S^2}$. ◄

Hence

$$d(\ln S) = \sigma dW + \left(\mu - \frac{\sigma^2}{2}\right) dt$$
$$\Longrightarrow \quad S_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t\right]$$

▷ Black-Scholes equation:
– Ref: [Andreasen] *Eight valuation methods for Black-Scholes formula*, Mathematical Scientist [link]
*Assumptions*:
(i) The option price $F(S_t, t)$ is a continuous function of time $t$ and its underlying asset's price $S_t$, and the price $S_t$ follows the geometric Brownian motion.
(ii) No market imperfections (price discreteness, transaction costs, taxes, trading restrictions...)
(iii) Unlimited risk-free borrowing at a constant interest rate $r$
(iv) No arbitrage opportunities
(v) No dividend payments during the life of the option
*Derivation*: Starting with Itô lemma,

$$dF = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial S}dS + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}(dS)^2 = \left[\frac{\partial F}{\partial t} + \frac{\sigma^2}{2}S^2\frac{\partial^2 F}{\partial S^2}\right]dt + \frac{\partial F}{\partial S}dS$$

Then we build a portfolio $\Pi$ with eliminated random contribution $dW$ as

$$\Pi = -F + \Delta S$$

The change of portfolio value within the time interval $dt$ is

$$\begin{aligned} d\Pi &= -dF + \Delta dS \\ &= -\left[\frac{\partial F}{\partial t} + \frac{\sigma^2}{2}S^2\frac{\partial^2 F}{\partial S^2}\right]dt + \left(\Delta - \frac{\partial F}{\partial S}\right)dS \end{aligned}$$

Since we want a *risk-neutral* portfolio, then we should make the coefficient of the stochastic process $dS$ vanish, namely $\Delta = \frac{\partial F}{\partial S}$ (*Delta hedging*). Since there are no arbitrage opportunities, this change is related to interest rate $r$ as

$$d\Pi = r\Pi dt$$

The BS eq follows

$$\begin{aligned} &\frac{\partial F}{\partial t} + rS\frac{\partial F}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 F}{\partial S^2} - rF = 0 \\ \text{or} \qquad &\Theta + rS\Delta + \frac{1}{2}\sigma^2 S^2 \Gamma - rF = 0 \end{aligned}$$

**Note**: this equation does NOT depend on the stock price drift $\mu$.
**Note**: The portfolio $\Pi$ is built on the assumption that the option price $F$ depends *linearly* on the stock price $S$ for small time increment $dt$. Then if we require the portfolio $\Pi = -F + \Delta S$ at the moment $t + dt$ is a constant irrespective of the value of $S$, then we have $\Delta = \frac{\partial F}{\partial S}$.
**Note**: Option is a *non-linear* instrument due to the convexity term $\frac{\partial^2 F}{\partial S^2}$. Recall the Jensen's inequality: $E[F(S)] \geq F(E[S])$ for a convex function $F(S)$. [For instance, consider $F(S) = \max(S - K, 0)$ at maturity is a convex function.] Now separate the stock price into the mean and fluctuations $S = \bar{S} + \epsilon$ where $E[\epsilon] = 0$. Then

$$E[F(S)] \approx F(\bar{S}) + \frac{1}{2}F''(\bar{S})E[\epsilon^2] = F(E[S]) + \frac{1}{2}F''(E[S])E[\epsilon^2]$$

*Interpretation*: If the stock price at maturity $T$ of the option is $E[S]$, then the option price is $F(E[S])$. Since the option price has finite probability of being positive, the option price $E[F(S)]$ before $T$ should be higher than the price $F(E[S])$ at $T$ by the amount $\approx \frac{1}{2}F''(E[S])E[\epsilon^2]$, which related to the convexity of an option, and the variance of the underlying. [Wilmott] Section 4.3. **Note**: [Wilmott2] Section 51.6. In the BS world, or the stochastic volatility world, the **option value** can be interpreted as the **present value of the expected payoff under a risk-neutral random walk**.

▶ *Aside* (BS eq from parameter fixing): Consider the general linear diffusion equation of the form

$$\frac{\partial F}{\partial t} + a\frac{\partial F}{\partial S} + b\frac{\partial^2 F}{\partial S^2} + cF = 0$$

The solution of cash $F(S,t) = e^{-r(T-t)}$ gives $c = -r$. The solution of the stock itself $F(S,t) = S$ gives $a = rS$. The remaining unknown parameter $b$ is of course related to volatility $\sigma$, that cannot be easily inferred from simple solutions. ◀

▷ Solution: With expiry $T$, exercise price $E$, risk-free interest rate $r$, dividend rate $D$, the BS equation reads

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

which is a *backward diffusion/heat equation*. The solution is

$$V(S,t) = \frac{e^{-r(T-t)}}{\sigma\sqrt{2\pi(T-t)}} \int_0^\infty \frac{d\varsigma}{\varsigma} \text{Payoff}(\varsigma) \exp\left\{ -\frac{\left[\log\frac{S}{\varsigma} + \left(r - \frac{\sigma^2}{2}(T-t)\right)\right]^2}{2\sigma^2(T-t)} \right\} \tag{7.1}$$

See appendix [Black-Scholes solution] for derivation.
– Call option value:

$$\begin{aligned} V_{\text{call}}(S,t) &= Se^{-D(T-t)}N(d_1) - Ee^{-r(T-t)}N(d_2) \\ d_1 &= \frac{\log\frac{S}{E} + (r - D + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &= d_1 - \sigma\sqrt{T-t} \end{aligned}$$

– Put opton value:

$$V_{\text{put}}(S,t) = -Se^{-D(T-t)}N(-d_1) + Ee^{-r(T-t)}N(-d_2)$$

where $N(x)$ is the CDF (cumulative distribution function) for the standardized normal distribution

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\phi^2/2} d\phi$$

# Δ-Hedging

## Discrete Hedging

Choose the underlying price as

$$\begin{aligned} S &= e^x \\ \delta x &= \left(\mu - \frac{\sigma^2}{2}\right)\delta t + \sigma\phi\sqrt{\delta t} \end{aligned}$$

where $\phi \sim \mathcal{N}(0,1)$.
We construct a "hedged" BS portfolio

$$\Pi = V - \Delta S$$

Note that we no longer have Itô lemma in discrete time, we shall use the Taylor expansion, and we get

$$\begin{aligned}
\delta\Pi &= \sqrt{\delta t}A_1 + \delta t A_2 + \mathcal{O}(\delta t^{3/2}) \\
A_1 &= \sigma\phi S\left(\frac{\partial V}{\partial S} - \Delta\right) \\
A_2 &= \frac{\partial V}{\partial t} + S\left(\frac{\partial V}{\partial S} - \Delta\right)\left(\mu + \frac{1}{2}\sigma^2(\phi^2 - 1)\right) + \frac{1}{2}\sigma^2\phi^2 S^2\frac{\partial^2 V}{\partial S^2}
\end{aligned}$$

▶ *Derivation*: We have $\delta x^2 \sim \delta t\phi^2\sigma^2$, $\partial^n S/\partial x^n = S$,

$$\begin{aligned}
\delta\Pi &= \frac{\partial V}{\partial t}\delta t + \left(\frac{\partial V}{\partial S} - \Delta\right)S\delta x + \frac{1}{2}\left(\frac{\partial^2 V}{\partial x^2} - \Delta S\right)\delta x^2 \\
\frac{\partial^2 V}{\partial x^2} &= S^2\frac{\partial^2 V}{\partial S^2} + S\frac{\partial V}{\partial S}
\end{aligned}$$

and the result follows. ◀

The hedging strategy is to choose $\Delta$ that minimizes the variance of $\delta\Pi$, and value the option by setting the *expected* return on $\Pi$ equal to the risk-free rate. The hedge is NOT risk-free (deterministic) anymore. [incomplete...]

### Perfect $\Delta$-Hedging?

| Possible | Not Possible |
|---|---|
| Binomial | Trinomial etc |
| Black-Scholes | Discrete time heding + lognormal |
| | Stochastic volatility |
| | Jump diffusion |
| | Fat-tailed returns ($\infty$ std dev) |

## Stochastic Volatility

**Suppose** the volatility $\sigma(t)$ is stochastic, meaning that the time scales for the randomness of both the asset and the volatility is comparable.
Consider the two-parameter SDE

$$\begin{aligned}
\frac{dS}{S} &= \mu dt + \sigma dW_1 \\
d\sigma &= p(S,\sigma,t)dt + q(S,\sigma,t)dW_2 \\
dW_1 dW_2 &= \rho dt
\end{aligned}$$

### Pricing Equation

Suppose we have two options $V$ and $V_1$ of the same underlying $S$, consider the following portfolio

$$\Pi = V - \Delta S - \Delta_1 V_1$$

where both stochastic variables $V$ and $V_1$ depend on $S, \sigma, t$. The change of this portfolio in time $dt$ is

$$d\Pi = \mathscr{F}[V]dt - \Delta_1 \mathscr{F}[V_1]dt + \left(\frac{\partial V}{\partial S} - \Delta_1 \frac{\partial V_1}{\partial S} - \Delta\right) dS + \left(\frac{\partial V}{\partial \sigma} - \Delta_1 \frac{\partial V_1}{\partial \sigma}\right) d\sigma$$
$$\mathscr{F}[V] \equiv \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma q S \frac{\partial^2 V}{\partial S \partial \sigma} + \frac{1}{2} q^2 \frac{\partial^2 V}{\partial \sigma^2}$$

To eliminate randomness we must choose

$$\Delta_1 = \frac{\partial V / \partial \sigma}{\partial V_1 / \partial \sigma}$$
$$\Delta = \frac{\partial V}{\partial S} + \Delta_1 \frac{\partial V_1}{\partial S}$$

The no-arbitrage argument set $d\Pi = r\Pi dt$, or

$$\left(\frac{\partial V}{\partial \sigma}\right)^{-1} \left(\mathscr{F}[V] + rS\frac{\partial V}{\partial S} - rV\right) = \left(\frac{\partial V_1}{\partial \sigma}\right)^{-1} \left(\mathscr{F}[V_1] - rS\frac{\partial V_1}{\partial S} - rV_1\right)$$

▶ **Aside**: The sign of the $rS\partial V_1/\partial S$ term is wrong in [Wilmott2] p.856. ◀

Since the two options will typically have different payoffs, strikes, or expiries, the only way for an equation with a functional of $V$ or $V_1$ on either side is that both sides are independent of the contract type. In other words, both sides must be a **universal** function of all of the **independent** variables common to **all** options, namely $S, \sigma, t$. We shall choose such universal function to be $-(p - \lambda q)$, of which the reason will be manifest later. The resultant equation for $V$ is

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma q S \frac{\partial^2 V}{\partial S \partial \sigma} + \frac{1}{2} q^2 \frac{\partial^2 V}{\partial \sigma^2} + (p - \lambda q)\frac{\partial V}{\partial \sigma} - rV = 0 \tag{9.1}$$

where $\lambda(S, \sigma, t)$ is the market price of volatility risk.

### Market Price of Volatility Risk

If we $\Delta$-hedge the option $V$, namely $\Pi = V - \Delta S$. Then

$$d\Pi = \mathscr{F}[V]dt + \left(\frac{\partial V}{\partial S} - \Delta\right) dS + \frac{\partial V}{\partial \sigma} d\sigma$$

With $\Delta$-hedgin the coefficient of $dS$ vanishes. Finally, we have

$$d\Pi - r\Pi dt = q\frac{\partial V}{\partial \sigma}(\lambda dt + dW_2)$$

*Interpretation*: For every unit of volatility risk, represented by $dX_2$, there are $\lambda$ units of extra return, represented by $dt$. Hence the name "market price of volatility risk." The quantity $p - \lambda q$ is the *risk-neutron drift rate* of the volatility; while the real volatility drift rate is $p$.

## Portfolio Management

### MPT (modern portfolio theory) / Mean-variance analysis

*Assumptions*: (i) Hold the portfolio for a single period, (ii) asset returns $R_i \sim \mathcal{N}(\mu_i T, \sigma_i^2 T)$, (iii) correlations between assets $\rho_{i \neq j}$
*Disadvantage*: Too many (to be determined) parameters!

▷ Risk-free + Risky portfolio: For a portfolio with weight $\alpha$ on the risky asset $r$ and $1-\alpha$ on risk-free asset $f$, the expected return is

$$\begin{aligned}\mathbb{E}[R] &= R_f + s\sigma \\ s &= \frac{\mathbb{E}[R_r]-R_f}{\sigma_r} \\ \sigma &= \alpha\sigma_r\end{aligned}$$

where $s$ is the slope of the *risk-return trade-off line*, $\sigma$ is the s.d. of the portfolio.
▷ Multiple risky assets: Assuming the asset weight $W_i$, then the expected return is

$$\mathbb{E}[R] = \frac{1}{T}\mathbb{E}\left[\frac{\delta\Pi}{\Pi}\right] = \sum_{i=1}^{N} W_i\mathbb{E}[R_i]$$

and the risk of the portfolio $\Pi$ is

$$\sigma = \frac{1}{\sqrt{T}}\sqrt{\mathrm{Var}\left[\frac{\delta\Pi}{\Pi}\right]} = \sqrt{\sum_{i,j=1}^{N} W_iW_j\rho_{ij}\sigma_i\sigma_j}$$

These defines the *efficient frontier*. The straight line tangent to the efficient frontier and has intercept $R_f$ is called the *capital market line*.
– Diversification over two risky assets: Assume that $\rho_{12}=0$, and let one weight be $\gamma$, then

$$\begin{aligned}\mathbb{E}[R] &= \gamma\mathbb{E}[R_1] + (1-\gamma)\mathbb{E}[R_2] \\ \sigma^2 &= \gamma^2\sigma_1^2 + (1-\gamma)^2\sigma_2^2\end{aligned}$$

These defines the *mean-variance efficient portfolio* in the graph of $\mathbb{E}[R]$ against $\sigma$. One can attain a minimum risk at

$$\begin{aligned}\gamma_{\min} &= \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} \\ \sigma_{\min}^2 &= \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} \le \sigma_{1,2}^2\end{aligned}$$

## Performance measure

▷ Sharpe ratio: Maximize the expected return, while minimize the risk

$$\frac{\mathbb{E}[R]-r}{\sigma}$$

where $r$ is the risk-free interest rate.

## CAPM (capital asset pricing model)

– Single-index model

$$R_i = \alpha_i + \beta_i R_M + \epsilon_i, \quad \text{with } \beta_i = \frac{\mathrm{Cov}[R_iR_M]}{\mathrm{Var}[R_M]}$$

where $M$ represents the market or index value, and $\epsilon_i \sim \mathcal{N}(0, e_i^2)$. Here $\beta_i$ is the correlation of the asset to the market (**beta**). A **beta-neutral** strategy of a portfolio $\Pi$ means $\beta_\Pi = 0$.

Then we have

$$\mathbb{E}[R_i] = \alpha_i + \beta_i \mathbb{E}[R_M]$$
$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + e_i^2$$
$$\mathbb{E}[R_\Pi] = \alpha_\Pi + \beta_\Pi \mathbb{E}[R_M]$$
$$\sigma_\Pi^2 = \sum\nolimits_{i,j} W_i W_j \beta_i \beta_j \sigma_M^2 + \sum\nolimits_i W_i^2 e_i^2$$

where $\alpha_\Pi = \sum_i W_i \alpha_i$, $\beta_\Pi = \sum_i W_i \beta_i$, and the sum is $\sum_{i=1}^N$. If the weights are all $\sim N^{-1}$ (*well-diversified portfolio*), then term $\sum W_i^2 e_i^2 \sim \mathcal{O}(N^{-1})$. Thus, as $N \to \infty$,

$$\sigma_\Pi = |\beta_\Pi| \sigma_M$$

The term $\sum W_i^2 e_i^2$ vanishes as $N$ increases is called the *diversifiable risk*. The remaining risk correlated with the market is called the *systematic risk*.
The advantage of CAPM over mean-variance analysis is that CAPM considers fewer parameters. With 100 assets, for example, there are $N(N-1)/2 = 4950$ different pairwise covariances. In contrast, the factor structure in CPM involves only 100 parameters, namely the $\beta_i$'s, plus the variance of the market. This considerably simplifies the analysis.

### APT (arbitrage pricing theory)

*Assume*: $K$ risk factors $f_j(t)$ and volatility $\epsilon_i(t)$ with

$$\mathbb{E}[f_j(t)] = 0$$
$$\mathrm{Cov}[f_j(t), f_j(t')] = \mathrm{Cov}[\epsilon_i(t), \epsilon_i(t')] = 0,\ t \neq t'$$
$$\mathrm{Cov}[f_j(t), \epsilon_i(t)] = 0$$

but no constraints on $\mathrm{Cov}[f_j(t), f_k(t)]$ and $\mathrm{Cov}[\epsilon_i(t), \epsilon_j(t)]$. The return is

$$R_i(t) = a_i + \sum\nolimits_{i,j} \beta_{ij} f_j + \epsilon_i(t)$$

and the APT theorem states that there exist $K+1$ constants $\lambda_j$, $\vec{\lambda} \neq \vec{0}$ that

$$\mathbb{E}[R_i(t)] = \lambda_0 + \sum\nolimits_j \beta_{ij} \lambda_j$$

where $\lambda_0$ is related to the risk-free asset return (no arbitrage opportunity), and the remaining $\lambda_i$'s are risk premiums for the corresponding risk factors.

## Risk Measurement

▷ VaR (value at risk): Assume normal distribution of returns $R \sim \mathcal{N}(\mu, \sigma^2)$. For the chosen confidence level $\alpha$,

$$\mathrm{VaR}(\alpha) = -\sigma z_\alpha - \mu \begin{cases} > 0 & \text{Loss} \\ < 0 & \text{Profit} \end{cases}$$

where $z_\alpha$ is determined from the CDF (cumulative distribution function)

$$\Pr(Z < z_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\alpha} dz e^{-z^2/2} = 1 - \alpha$$

– See also *coherent risk measures*.
▷ ETL (expected tail loss)

$$\mathrm{ETL} = \mathbb{E}[L | L > \mathrm{VaR}]$$

# Time Series Analysis

Ref: [Halls-Moore1] Successful AlgoTrading Ch 10
Ref: [Halls-Moore2] Advanced AlgoTrading Ch 7-14
Ref: [Hamilton] Time Series Analysis
✡ Mean Reversion: A process of which a time series displays a tendency to revert to a historical mean value. One can ascertain whether a time series is mean reverting by a statistical test to see if it differs from the behavior of a *random walk*.
$\Longrightarrow$ Trading strategies (used in *statistical arbitrage* quant hedge funds). It can be applied to *both* long or short time scales.
▷ Mathematically, such a (continuous) time series is an Ornstein-Uhlenbeck process. The corresponding SDE is

$$dx_t = \theta(\mu - x_t)dt + \sigma dW_t$$

where $\theta$ is the rate of reversion to the mean $\mu$, and $\sigma^2$ is the variance of the Wiener process.
▷ ADF test (Augmented Dickey-Fuller test): To test for the presence of a **unit root** in an auto-regressive time series sample.

▶ *Aside* (unit root): [Hamilton] Consider a discrete-time stochastic process $y_t$, and suppose that it can be written as an AR process of order $p$:

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \epsilon_t$$

where $\epsilon_t$ is a serially uncorrelated, zero mean stochastic process with constant variance $\sigma^2$. For convenience, we assume $y_0 = 0$. If $m = 1$ is a root of the characteristic equation

$$m^p - \sum_{i=1}^{p} m^{p-i} a_i = 0$$

then the stochastic process has a unit root, or is integrated of order 1, denoted as $I(1)$. If $m = 1$ is a root of multiplicity $r$, then the stochastic process is integrated of order $r$, denoted as $I(r)$. See also *ARIMA model* below. ◀

Consider a *linear lag model of order* $p$: Change in the value of the time series $\propto$ a constant, the time $t$ itself, and the previous $p$ values of the time series, along with an error term:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{j=1}^{p-1} \delta_j \Delta y_{t-j} + \epsilon_t$$

where $\Delta y_t = y(t) - y(t-1)$.
ADF hypothesis test is to ascertain statistically the *null hypothesis* $\gamma = 0$ (no mean reverting).
(1) Calculate the test statistics

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

then (2) use the *distribution* and the *critical values* to decide whether to reject the null hypothesis.
▷ Hurst Exponent for Stationary Time Series:
*Strongly Stationary*: Joint probability distribution of a time series is invariant under translations in time or space.
For any time lag $\tau$, the variance of $\tau$ is given by [**check!**]

$$\mathrm{Var}(\tau) = \left\langle |\log P(t+\tau) - \log P(t)|^2 \right\rangle \sim \tau^{2H}$$

where $0 \leq H \leq 1$.
Compare the rate of diffusion to that of a GBM (geometric Brownian motion):
– $H < 0.5$: Mean reverting
– $H = 0.5$: GBM
– $H > 0.5$: Trending
✡ Cointegration:
A linear model between the two stock prices

$$y(t) = \beta x(t) + \epsilon(t)$$

$\Longrightarrow$ Trading strategies – namely *Pairs Trade* [wiki].
▷ Cointegrated ADF test: (1) Find the optimal $\beta_0$ by linear regression, (2) construct another time series $\epsilon(t) = y(t) - \beta_0 x(t)$ and use the ADF test check for stationary.

## Time Series Models

▷ *Backward Shift operator* or *lag operator*: $\boldsymbol{B}x_t = x_{t-1}$
▷ *Difference operator*: $\nabla x_t = x_t - x_{t-1} = (1 - \boldsymbol{B})x_t$
✡ Discrete White Noise: Consider a time series $\{w_t : t = 1, \ldots, n\}$. If the elements of the series $w_t$ are i.i.d (independent and identically distributed), with a mean of zero, variance $\sigma^2$ and no auto correlation, that is $\mathrm{Corr}(w_i, w_j) = 0, \forall i \neq j$ then the time series is DWN.
Properties:

$$\begin{aligned}
\mu_w &= E(w_t) = 0 \\
\gamma_{k=0}(t) &= \sigma^2 \\
\rho_k &= \mathrm{Corr}(w_t, w_{t+k}) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}
\end{aligned}$$

✡ Random Walk: is a time series model $x_t$ such that $x_t = x_{t-1} + w_t$, where $w_t$ is a DWN. One can easily show that

$$x_t = w_t + w_{t-1} + w_{t-2} + \cdots$$

Hence the random walk is simply the sum of the elements from a DWN series.
Properties: A random walk is *non-stationary*

$$\begin{aligned}
\mu_x &= 0 \\
\gamma_k(t) &= \mathrm{Cov}(x_t, x_{t+k}) = t\sigma^2 \\
\rho_k(t) &= \frac{\mathrm{Cov}(x_t, x_{t+k})}{\sqrt{\mathrm{Var}(x_t)\mathrm{Var}(x_{t+k})}} = \frac{1}{\sqrt{1 + k/t}}
\end{aligned}$$

Note that for long time series (large $t$) with short term lags (small $k$), then $\rho_k(t) \lesssim 1$.
To fit a random walk, first find the difference series $w_t = x_t - x_{t-1}$, and plot the ACF (auto-correlation function) of lag $k$.

## Stationary Models

Includes AR($p$), MA($q$), ARMA($p, q$) mdoels. Stationary (or covariance-stationary) means that the ensemble mean $\mu_t$ and ensemble variance $\sigma^2$ are both independent of time $t$. Moreover, ARIMA($p, d, q$) model is non-stationary, but is related to ARMA by considering the difference time series $\nabla^d x_t$.

► *Aside* (Bounded time series $y_t$): If there exists a finite number $\bar{y}$ such that $|y_t| < \bar{y}$ for all $t$. ◄

► *Comment*: This is just a mathematical requirement for time series analysis. In the real financial market, time series can be *unbounded* (financial bubble). ◄

### AR($p$) model (Auto-regression of order $p$)

$$x_t = \sum_{i=1}^{p} \alpha_i x_{t-i} + w_t$$

where $\{w_t\}$ is white noise with $E(w_t) = 0$ and variance $\sigma^2$, and $\alpha_i \in \mathbb{R}$. The model can be written as

$$\theta_p(\boldsymbol{B})x_t = w_t$$
$$\theta_p(\boldsymbol{B}) = 1 - \sum_{i=1}^{p} \alpha_i \boldsymbol{B}^i$$

Second order properties:

$$\mu_x = E(x_t) = 0$$
$$\gamma_k = \sum_{i=1}^{p} \alpha_i \gamma_{k-i}, \quad k > 0$$
$$\rho_k = \sum_{i=1}^{p} \alpha_i \rho_{k-i}, \quad k > 0$$

### MA($q$) model (moving average of order $q$)

$$x_t = w_t + \sum_{i=1}^{q} \beta_i w_{t-i}$$
$$\equiv \sum_{i=0}^{q} \beta_i w_{t-i}, \quad \text{with } \beta_0 = 1$$
$$\text{OR } x_t = \phi_q(\boldsymbol{B})w_t$$
$$\phi_q(\boldsymbol{B}) = 1 + \sum_{i=1}^{q} \beta_i \boldsymbol{B}^i$$

Second order properties:

$$\mu_x = E(x_t) = \sum_{i=0}^{q} E(w_i) = 0$$
$$\text{Var } \gamma_0 = \sigma_w^2(1 + \beta_1^2 + \cdots + \beta_q^2)$$
$$\text{ACF } \rho_k = \begin{cases} 1 & k = 0 \\ \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} \beta_i^2} & k = 1, \dots, q \\ 0 & k > q \end{cases}$$

### ARIMA($p, d, q$) model (Auto-regression Integrated Moving Average)

$$\theta_p(\boldsymbol{B})(1 - \boldsymbol{B})^d x_t = \phi_q(\boldsymbol{B})w_t$$

▷ A series is called *integrated* of order $d$ denoted as $I(d)$ if $\nabla^d x_t = w_t$. Hence the name for ARIMA model.
▷ When $d = 0$, it is the **ARMA($p, q$) model**. Obviously, if $p \geq 2$ or $q \geq 2$, then ARMA model is *non-Markovian*.

▷ The application of the difference operator $\nabla = 1 - \boldsymbol{B}$ is to reduce a non-stationary $I(d > 0)$ series to a stationary $I(d = 0)$ one.
▷ AR is modeled for processes like *momentum* and *mean-reversion*.
For example: mean-reversion $dx_t = (\mu - x_t)dt + \sigma dW_t$ or in the discretized time

$$x_t - \frac{1}{2}x_{t-1} - \frac{1}{2}\mu = \frac{\sigma}{2}w_t$$

▷ MA is modeled for the "shock" or "source". Notice the similarity of the model to a differential equation $\hat{L}\psi(t) = f(t)$, where $\hat{L}$ is some linear differential operator, which governs the equation of motion of $\psi(t)$ over time $t$, and $f(t)$ is the source term.

## Model Selection

Prevent overfitting of too many parameters
▷ AIC (Akaike Information Criterion)

$$\text{AIC} = -2\log L + 2k$$

where $L$ maximizes the likelihood of $k$ parameters.
▷ BIC (Bayesian Information Criterion)

$$\text{BIC} = -2\log L + k\log n$$

where $n$ is the number of data points in the time series.
▷ Ljung-Box Test: Hypothesis of correlations between time series.
See [Halls-Moore2] section 10.6.2.

## Fitting parameters

▷ Apply ACF to the residuals series $\implies$ If $\sim$ white noise, then good model.
▷ **MLE** (maximum likelihood estimation)
We shall discuss an example for the AR(1) process. See [Hamilton] for more examples and general methods.
Consider a Gaussian AR(1) process:

$$Y_t = c + \alpha Y_{t-1} + w_t$$

with $w_t \sim_{\text{i.i.d}} \mathcal{N}(0, \sigma^2)$. The vector of population parameters to be estimated is $\boldsymbol{\theta} = (c, \alpha, \sigma^2)$. For the first observation $t = 1$, $E(Y_1) = \mu = c/(1-\alpha)$ and variance $E[(Y_1 - \mu)^2] = \sigma^2/(1-\alpha^2)$. Since $w_t$ is Gaussian, so does $Y_1$. Hence the likelihood probability density of the first observation takes the form

$$f_{Y_1}(y_1; \boldsymbol{\theta}) = \sqrt{\frac{1-\alpha^2}{2\pi\sigma^2}}\exp\left[-\frac{(1-\alpha^2)\left(y_1 - \frac{c}{1-\alpha}\right)^2}{2\sigma^2}\right]$$

For $t = 2$, notice that it is a conditional probability $P(Y_2|Y_1 = y_1)$ or

$$f_{Y_2|Y_1}(y_2|y_1; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(y_2 - c - \alpha y_1)^2}{2\sigma^2}\right]$$

and the *joint probability density* of observations 1 and 2 is just the product

$$f_{Y_2,Y_1}(y_2, y_1; \boldsymbol{\theta}) = f_{Y_2|Y_1}(y_2|y_1; \boldsymbol{\theta})f_{Y_1}(y_1; \boldsymbol{\theta})$$

This process continues and the likelihood of the complete sample is

$$f_{Y_T,Y_{T-1},\ldots,Y_1}(y_T,\ldots,y_1;\boldsymbol{\theta}) = f_{Y_1}(y_1;\boldsymbol{\theta})\prod_{t=2}^{T} f_{Y_t|Y_{t-1},\ldots}(y_t|y_{t-1},\ldots;\boldsymbol{\theta})$$

The log likelihood function $\mathcal{L}(\boldsymbol{\theta})$ is

$$\mathcal{L}(\boldsymbol{\theta}) = \log f_{Y_1}(y_1;\boldsymbol{\theta}) + \sum_{t=2}^{T} \log f_{Y_t|Y_{t-1},\ldots}(y_t|y_{t-1},\ldots;\boldsymbol{\theta})$$

Clearly the value of $\boldsymbol{\theta}$ that maximizes log likelihood $\mathcal{L}(\boldsymbol{\theta})$ also maximizes the original likelihood. Now our objective becomes

$$\arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$$

▹ Conditional MLE
An alternative is to regard $y_1$ as deterministic and maximize the likelihood conditioned on the first observation

$$f_{Y_T,Y_{T-1},\ldots,Y_2|Y_1}(y_T,\ldots,y_2|y_1;\boldsymbol{\theta}) = \prod_{t=2}^{T} f_{Y_t|Y_{t-1},\ldots}(y_t|y_{t-1},\ldots;\boldsymbol{\theta})$$

The objective is then to maximize

$$\log f_{Y_T,Y_{T-1},\ldots,Y_2|Y_1}(y_T,\ldots,y_2|y_1;\boldsymbol{\theta}) = -\frac{T-1}{2}\log(2\pi\sigma^2) - \sum_{t=2}^{T}\frac{(y_t - c - \alpha y_{t-1})^2}{2\sigma^2} \tag{12.1}$$

Maximize eq (12.1) with respect to $c$ and $\alpha$ is equivalent to minimization of

$$\sum_{t=2}^{T}(y_t - c - \alpha y_{t-1})^2$$

which is achieved by an ordinary least squares (OLS) regression of $y_t$ on a constant and its own lagged value. See [Hamilton] for how to find $\sigma^2$.

### ✡ GARCH$(p,q)$ models (generalized auto-regressive conditional heteroskedastic)

A time series $\{\epsilon_t\}$ of mean zero for all $t$ and

$$\begin{aligned}
\epsilon_t &= \sigma_t w_t \\
\sigma_t^2 &= \beta_0 + \sum_{i=1}^{q}\beta_i\epsilon_{t-i}^2 + \sum_{i=1}^{p}\alpha_j\sigma_{t-j}^2 \\
\text{or simply}\quad \epsilon_t &= w_t\sqrt{\beta_0 + \sum_{i=1}^{q}\beta_i\epsilon_{t-i}^2 + \sum_{i=1}^{p}\alpha_j\sigma_{t-j}^2}
\end{aligned}$$

where $w_t$ is white noise with mean zero and variance 1.
▹ The model describes *autoregressive (and moving average) process for the variance* itself.
*Example*: Consider ARCH(1)= GARCH($p=0, q=1$) model, The (time-dependent) variance is

$$\begin{aligned}
\mathrm{Var}(\epsilon_t) &= \mathrm{E}(\epsilon_t^2) - \cancelto{0}{\mathrm{E}(\epsilon_t)^2} \\
&= \cancelto{1}{\mathrm{E}(w_t^2)}\mathrm{E}(\beta_0 + \beta_1\epsilon_{t-1}^2) \\
&= \beta_0 + \beta_1\mathrm{Var}(\epsilon_{t-1})
\end{aligned}$$

where in the 3rd equality, we utilize the following:

► *Aside*: [link] If two random variables $X, Y$ have a joint distribution, then they are independent iff the corresponding CDF (cumulative distribution function) satisfy

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

Then we have

$$\mathrm{E}(XY) = \int xydF_{X,Y}(x,y) = \int xydF_X(x)dF_Y(y) = \int xdF_X(x)\int ydF_Y(y) = \mathrm{E}(X)\mathrm{E}(Y)$$

◄

▷ Procedure: (1) Find optimal parameters for ARIMA$(p,d,q)$ model, (2) Fit the residuals with GARCH$(p,q)$.

### Fractionally Differentiated Features

Ref: [de Prado]
Most econometric analyses follow one of the following:
(1) Box-Jenkins: Returns are stationary, however memoryless
(2) Engle-Granger: Log-prices have memory, however, they are non-stationary. Cointegration is the trick that makes regression work on non-stationary series, so that memory is preserved.
▷ The FFD Method: Recall the binomial series $(1+x)^d = \sum_{k=0}^{\infty}\left(\begin{array}{c} d \\ k \end{array}\right)x^k$. Similarly, we have

$$\begin{aligned}\nabla^d = (1-B)^d = \sum_{k=0}^{\infty}\left(\begin{array}{c} d \\ k \end{array}\right)(-B)^k &= \sum_{k=0}^{\infty}\frac{\prod_{i=0}^{k-1}(d-i)}{k!}(-B)^k \\ &= \sum_{k=0}^{\infty}(-B)^k\prod_{i=0}^{k-1}\frac{d-i}{k-i}\end{aligned}$$

where $B$ is the backward operator. Hence,

$$\nabla^d X_t = \tilde{X}_t = \sum_{k=0}^{\infty}\omega_k X_{t-k}$$

together with the recursion relation

$$\begin{aligned}\omega_k &= -\omega_{k-1}\frac{d-k+1}{k} \\ \omega_0 &= 1\end{aligned}$$

Applying the FFD (fixed-width window fracdiff) method, we have compute the minimum coefficient $d^*$ such that the resulting fractionally differentiated series $\tilde{X}_t$ is stationary (at certain confidence level, say 95%). This coefficient $d^*$ quantifies the amount of memory that needs to be removed to achieve stationarity

**Black noise process** [url] Long Memory & Regime Shifts in Asset Volatility
– Denote the Hurst Exponent as $H$: fractal random walk for black noise ($H > 0.5$); white noise ($H = 0.5$), and mean-reverting, pink noise ($H < 0.5$) process.
– Long memory models such as ARFIMA & FIGARCH
— Long term serial corrletaion can be also be resulted from structural breaks $\Longrightarrow$ spurious (欺騙性的) autocorrelations
– Useful for detecting regime shift (in backtest)

## Financial ML

Ref: [de Prado] Advances in Financial ML
**Comments**: [de Prado] argues that many old assumptions stated above are wrong!
(1) non-IID processes – Could be corrected by using dollar bars (instead of time bars)
(2) memoryless Markov processes – Preserve more memory by fractionally differentiation, and modeling using NN
▷ Meta-strategy paradigm: Research manual for teams as research factory, not individuals. The teamworks yields discoveries at a predictable rate.
Production chain:
(1) *Data curators*
(2) *Feature analysts*, transforming raw data into informative signals
(3) *Strategists*, make sense of all observations & formulate a general theory that explains them
(4) *Backtesters*, find out the weaknesses & strengths of a proposed strategy. Prevent overfitting!
(5) *Deployment team*
(6) *Portfolio oversight*

### Financial Data Structures

We want to work with *unstructured raw* financial data, and from that to derive a structured dataset amenable to ML algorithms and then to obtain informative features.
▷ Time Bars: Should be avoided for 2 reasons: (1) today's markets are operated by algorithms that trade with loose human supervision, for which CPU cycles are much more relevant than chronological intervals. This means that time bars oversample information during low-activity periods and undersample information during high-activity periods. (2) time-sampled series often exhibit poor statistical properties, like series correlation, heteroscedasticity, and non-normality of returns.
▷ Dollar Bars: Formed by sampling an observation every time a pre-defined market value is exchanged. The resultant variable is more Gaussian than time bars, tick bars, volume bars.

### Labeling

– To apply supervised learning, the data needs labels. We can label a *bar-frame* (timeframes, if using time bars) with its sign of P&L, denoted as `side=sgn(P&L)`. By feeding the data, features, and labels, a ML model can be trained to attain a high rate of positive |`side`| (suppose that we can long and short).
▷ Triple-barrier method:
▷ Learning *Bet Side*:
▷ Meta-Labeling – Learning *Bet Size*: As in the human trading, one may pass a trading opportunity for some reason (low return and/or high risk), a ML model should also learn the bet size. The As argued in [de Prado], this should be another layer of model built on top of the model for bet side. The feeds for a supervised ML model are data, features, output of bet side model as input, and the overall return as the output. The bet size for each non-zero `sgn(P&L)` event is trained.

### Sample Weights

– In the labeling methods above, the bar-frames of labeled events might overlap. These overlappings lead to financial series seemingly generated from non-IID processes. While most of ML methods is based on the IID assumption. Thus it is important to "remove" the non-IID labels, by designing sampling and weighting schemes that correct for the undue influence of overlapping outcomes.

### Fractionally Differentiated Features

See section 12.3

– It is known that, as a consequence of arbitrage forces, financial series exhibit low signal-to-noise ratios. To make things worse, standard stationary transformations, like integer differentiation, further reduce that signal by *removing memory*. *Price series have memory*, because every value is dependent upon a long history of previous levels. Our aim is to transform the data to ensure stationary while preserving as much memory as possible.

Relevant: [1604.00105] option pricing under fast-varying long-memory stochastic volatility

## Some Mathematics

### Factor of $n-1$ in sample variance estimation

▷ *(Point) Estimator* or *Statistic*: Let $\{x^{(i)}\}$ be a set of $m$ i.i.d. data points. An estimator $\hat{\theta}_m$ of the parameter $\theta$ is any function of the data:

$$\hat{\theta}_m = g(\{x^{(i)}\})$$

▷ *Bias*: of an estimator is defined as

$$\text{bias}(\hat{\theta}_m) = E(\hat{\theta}_m) - \theta$$

where the expectation is over the data (seen as samples from a random variable) and $\theta$ is the true underlying value of $\theta$ used to define the data generating distribution. An estimator is unbiased if $\text{bias}(\hat{\theta}_m) = 0$, and is asymptotically unbiased if $\lim_{m\to\infty} \text{bias}(\hat{\theta}_m) = 0$.

▷ The *estimate* for variance for a sample $\{y_i\}$ is defined as

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{N}(y_i - \bar{y})^2$$

▷ Let $Y_i$ denote the random variable whose process is "to choose a random sample $y_1, y_2, \ldots, y_n$ of size $n$" from the random variable $Y$, and whose value for that choice is $y_i$.

▷ A useful identity: $\text{Var}(Y) = E(Y^2) - [E(Y)]^2$ or $E(Y^2) = \text{Var}(Y) + [E(Y)]^2$

▷ For random variable $Y_i$, the estimate for variance is

$$\begin{aligned}(n-1)S^2 &\equiv \sum\nolimits_i \left(Y_i - \frac{1}{n}\sum\nolimits_i Y_i\right)^2 \\ &= \sum\nolimits_i Y_i^2 - \frac{1}{n}\left(\sum\nolimits_i Y_i\right)^2\end{aligned}$$

Then

$$\begin{aligned}(n-1)E(S^2) &= \sum\nolimits_i E(Y_i^2) - \frac{1}{n}E\left[\left(\sum\nolimits_i Y_i\right)^2\right] \\ &= \sum\nolimits_i \left\{\text{Var}(Y_i) + [E(Y_i)]^2\right\} - \frac{1}{n}\left\{\text{Var}\left(\sum\nolimits_i Y_i\right) + \left[E\left(\sum\nolimits_i Y_i\right)\right]^2\right\}\end{aligned}$$

By choice, each $Y_i$ has the same distribution (hence the same mean and variance) as $Y$. Thus $E(Y_i) = \mu$ and $\text{Var}(Y_i) = \sigma^2$ for each $i$. Also, $Y_i$'s are independent. Hence,

$$\begin{aligned}(n-1)E(S^2) &= \sum\nolimits_i \left(\sigma^2 + \mu^2\right) - \frac{1}{n}\left\{\text{Var}\left(\sum\nolimits_i Y_i\right) + \left[\sum\nolimits_i E(Y_i)\right]^2\right\} \\ &= n\left(\sigma^2 + \mu^2\right) - \frac{1}{n}\left[\sum\nolimits_i \text{Var}(Y_i) + (n\mu)^2\right] \\ &= (n-1)\sigma^2\end{aligned}$$

where in the second line we utilize the independence of $Y_i$'s that $\text{Var}\left(\sum_i Y_i\right) = \sum_i \text{Var}(Y_i)$. We conclude that $S^2$ is a *unbiased* estimator of $Y_i$ such that $E(S^2) = \sigma^2 = \text{Var}(Y)$.

## Range of correlation

$\rho(x,y) = \frac{\sigma(x,y)}{\sigma(x)\sigma(y)} \in [-1,1]$ by Cauchy-Schwarz inequality

$$\left|\sum_{i=1} u_i \bar{v}_i\right|^2 \leq \sum_{j=1}^{n} |u_j|^2 \sum_{j=1}^{n} |v_j|^2$$

for $u_i, v_i \in \mathbb{C}$ with $i = 1, \dots, n$.

## Black-Scholes solution

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

▷ Step 1 (remove the sink term):

$$V(S,t) = e^{-r(T-t)}U(S,t)$$
$$\Longrightarrow \frac{\partial U}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 U}{\partial S^2} + rS\frac{\partial U}{\partial S} = 0$$

▷ Step 2 (transform into a forward eq):

$$\tau = T - t$$
$$\Longrightarrow \frac{\partial U}{\partial \tau} = \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 U}{\partial S^2} + rS\frac{\partial U}{\partial S}$$

▷ Step 3 (Price $S \to$ return $\xi$):

$$\xi = \log S$$
$$\Longrightarrow \frac{\partial U}{\partial \tau} = \frac{1}{2}\sigma^2 \frac{\partial^2 U}{\partial \xi^2} + \left(r - \frac{1}{2}\sigma^2\right)\frac{\partial U}{\partial \xi}$$

We used $\frac{\partial}{\partial S} = e^{-\xi}\frac{\partial}{\partial \xi}$ and $\frac{\partial^2}{\partial S^2} = e^{-2\xi}\left(\frac{\partial^2}{\partial \xi^2} - \frac{\partial}{\partial \xi}\right)$.
▷ Step 4 (remove the $\partial_\xi U$ term): Let $U = W(x,\tau)$, together with

$$x = \xi + \left(r - \frac{1}{2}\sigma^2\right)\tau$$
$$\Longrightarrow \frac{\partial W}{\partial \tau} = \frac{1}{2}\sigma^2 \frac{\partial^2 W}{\partial x^2}$$

which is a *heat equation* [wiki]. We used $\frac{\partial U}{\partial \xi} = \frac{\partial W}{\partial x}$ and $\frac{\partial U}{\partial \tau} = \left(r - \frac{1}{2}\sigma^2\right)\frac{\partial W}{\partial x} + \frac{\partial W}{\partial \tau}$.
▷ Step 5 (Solve for $W$): Suppose

$$W(x,\tau) = \tau^\alpha f(\eta) \equiv \tau^\alpha f\left(\frac{x - x'}{\tau^\beta}\right)$$

The aim is to find a differential equation with *one variable* $\eta$:

$$\tau^{\alpha-1}\left(\alpha f - \beta\eta\frac{df}{d\eta}\right) = \frac{1}{2}\sigma^2\tau^{\alpha-2\beta}\frac{d^2 f}{d\eta^2}$$

which then requires

$$\alpha - 1 = \alpha - 2\beta \quad \Longleftrightarrow \quad \beta = \frac{1}{2}$$

Another desired property is that the scaling solution is independent of $\tau$ after integral over $\xi \in (-\infty, \infty)$ or

$$\text{const} = \int_{-\infty}^{\infty} dx \tau^{\alpha} f\left(\frac{x - x'}{\tau^{\beta}}\right) = \int_{-\infty}^{\infty} d\eta f(\eta) \tau^{\alpha+\beta}$$

which requires

$$\alpha = -\beta = -\frac{1}{2}$$

Now $f$ satisfies

$$0 = \sigma^2 \frac{d^2 f}{d\eta^2} + \frac{d(\eta f)}{d\eta} = \frac{d}{d\eta}\left(\sigma^2 \frac{df}{d\eta} + \eta f\right)$$

which can then be integrated. We shall **assume** that integration constant vanishes, then

$$f(\eta) = b \exp\left[-\frac{\eta^2}{2\sigma^2}\right]$$

where $b = \frac{1}{\sqrt{2\pi}\sigma}$ is the normalization constant such that $\int_{-\infty}^{\infty} d\eta f(\eta) = 1$. Finally, we have

$$W_f(x, \tau) = \frac{1}{\sqrt{2\pi\tau}\sigma} \exp\left[-\frac{(x - x')^2}{2\sigma^2 \tau}\right]$$

As $\tau \to 0^+$, the kernel $W_f(x, \tau) \to \delta(x - x')$. Hence the general solution for the heat equation is a convolution of the kernel with the initial boundary condition

$$W(x, \tau) = \int_{-\infty}^{\infty} W_f(x, \tau; x') \text{Payoff}(e^{x'}) dx'$$

After some manipulation, we have the final result eq (7.1).