Project Thesis

Julie Johanne Uv

Institutt for matematiske fag — NTNU 2018

Contents

Contents

Chapter 1

Introduction

Chapter 2

Financial theory

2.1 Effiecient Market Hypopthesis

Efficient Market Hypothesis - why do we model stocks as a log normal random walk? Financial Market? The financial market is a common term for markets that trade different financial instruments such as stocks, bonds, currencies and different financial derivatives such as futures, forwards, options and swaps. The Efficient Market Hypothesis is stated in several ways, but essentially amounts to the properties that

- the current market price reflects history up until the present, but does not hold any further information about the future
- that the market moves immediately with the arrival of new information.

From the former it follows that stock prices has the Markov Property, i.e. the future price depends only upon the present price.

$$P(X_{t+1} = x_{t+1} | X_t = x_t, \ldots, X_0 = x_0) = P(X_{t+1} = x_{t+1} | X_t = x_t) \tag{2.1}$$

Asinformation essentially moves randomly, it follows from the latter that it is reasonable to model the prices randomly as well. If the market really is efficient, then assets must always be traded at their fair value(reference) and one can never outperform the market. If assets were over- or undervalued, the prices would immediately adjust. The Efficient Market Hypothesis does not mean that one can never profit from the market, but that one must do so at a risk. (Not perfectly efficient in practice.)

2.2 Arbitrage

(A result from EMH - there is no arbitrage.) Arbitrage is a risk-free instantaneous profit which can be made by a exploiting price differences in the same or similar financial instruments or assets (by simultaneous purchase and sale). Arbitrage opportunities will occur as the market is not perfectly efficient, but prices will adjust qiute fast(?). Hence it is reasonable to assume market efficiency and thus no arbitrage in financial pricing models.

2.2.1 Assets

A financial asset is an asset where its value comes from a so called contractual agreement - an agreement that cash flows will be paid to the purchaser at times specified by the contract. Examples of financial assets are cash, equities, indices and commodities. As argued, a financial asset (traded on the market) may be modeled randomly and it is common to value it's price movement by a stochastic differential equation of the following form.

$$dS(t) = u(S,t)dt + w(S,t)dW(t) \tag{2.2}$$

Here $S(t)$ is the value of the asset price at time t, u and w are arbitrary functions of asset price and time. dW is a normally distributed random variable (appendix on r.v?) with the following properties

$$E[dW(t)] = 0 \tag{2.3}$$

$$Var[dW(t)] = dt. \tag{2.4}$$

$dW(t)$ follows what is called a Wiener process(or a Brownian motion)(appendix?) and represents the randomness of the asset (without it we would just have a deterministic differential equation). (u is the drift term and w is the diffusion term.)
Stocks:
By issuing stocks to investors, a company can raise capital. If the company increases it's revenue, the value of the stocks increases. In addition the company can decide to pay out a part of the revenue as dividend per share to its shareholders.
When one values stocks, it is more reasonable to look at the relative change in price rather then the actual level of the price, i.e. $dS(t)/S(t)$ rather then $S(t)$. The contributions to the relative change also known as the return is a drift term containing the average growth of the asset price, μ, and a random term containing the volatility measuring the standard deviation of the returns, σ.

$$dS(t) = \mu S(t)dt + +\sigma S(t)dW(t). \tag{2.5}$$

Again, here $dW(t)$ is the normally distributed random variable as stated above. (Another incentive/reason to evaluate the returns rather then the level is that they tend to increase exponentially by observation.)
μ and σ can both be functions of S and t. The most basic model is taking μ and σ to be constants. The unique solution to (2.5) is (Ito's Lemma appendix?)

$$S(t) = S(0)\exp((\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)) \tag{2.6}$$

from which one can obtain the result

$$\ln\left(\frac{S(t)}{S(0)}\right) \sim N((\mu - \frac{1}{2}\sigma^2)t, \sqrt{t}). \tag{2.7}$$

(Using this one can estimate the parameters using the historical logarithmic returns by the following.

$$r_i = \log\big(\frac{S_{i+1}}{S_i}\big) \tag{2.8}$$

$$m = \frac{1}{ndt}\sum_{i=0}^{n-1} r_i \tag{2.9}$$

$$\hat{\sigma} = \frac{1}{(n-1)dt}\sum_{i=0}^{n-1}(r_i - m)^2 \tag{2.10}$$

$$\hat{\mu} = m + \frac{1}{2}\hat{\sigma}^2) \tag{2.11}$$

(Argue more why lognormal random walk is reasonable? Says something about interest rates, how it also can be modelled as a stochastic variable, maybe not okay to assume its constant, depends on lifespan of option.)

2.2.2 Derivatives

Derivatives are instruments where the value is derived from or rely upon an underlying asset (or a group of assets) such as the financial assets stated above or interest rates. (It is a contract between two or more parties.) Examples of derivatives are options, futures, forwards and swaps. This thesis will (focus on) valuing options.
And option is a contract between two parties that gives the holder of the option the right but not the obligation to buy or sell the underlying asset at a specified time and at a specified price called the exercise (or strike) price, in the future. As stated the purchaser of the option is the holder of the option, while the seller is the writer. An option in which the holder has the right to buy the underlying is called a call option, and an option in which the holder has the right to sell the underlying is called a put option. Then there are several types of call and put options. American options for example can be exercised at any time up until expiry, while the one which will be examined(?) further in this thesis is the European option which can only be exercised at expiry. The payoff function for a European call option at expiry T is

$$C(S,T) = max(S(T) - E, 0). \tag{2.12}$$

E is the exercise price. Similarly, the payoff function for a European put option is

$$P(S,T) = max(E - S(T), 0). \tag{2.13}$$

What affects the value of an option? The value of an option is dependent on the underlying asset price and the time to expiry (for obvious reasons) which both will change during the life of the option. In addition, the value of the option will be affected by the exercise price - e.g. the higher the exercise price of a call option, the higher the value of the option - and the interest rate. The interest rate affects the price through

time value as the payoff will be received in the future. Lastly, the volatility will have an effect on the option value(measures the fluctuation, annualised standard dedviation of the returns). Using a Assets, stocks, derivatives, arbitrage (Risk neutrality, ricing derivatives in a risk-neutral world)
Volatility:
Volatility is the standard deviations of the logarithmic returns, so it might be viewed as a measurement of how random the asset price is(?). There are several ways of modelling the volatility. E.g.

- As a stochastic differential equation. This is a very reasonable model as the volatility seems to have a random component just as our asset price model. Then one would have two stochasstic differential equations

$$dS(t) = \mu S(t)dt + \sqrt{\nu(t)}S(t)dW_1(t) \tag{2.14}$$
$$d\nu(t) = \alpha(\nu, t)dt + \beta(\nu, t)dW_2(t) \tag{2.15}$$

  $dW_1(t)$ and $dW_2(t)$ are two Wiener processes(which may be correlated?).

- Historical volatility uses historical values to calculate some parameter estimation for the voltility based on statistics. (This may be the maximum likelihood e.g.). In this paper $\hat{\sigma}$ has been used. A question that arises for the use of the historical volatility is how much historical data one should use in the calculations.

- Implied volatility uses the quoted prices for options today and exploiting the idea that the market "knows" the volatility. When one has today's option price, asset price, life span of option, risk free rate and strike pricec one can solve the Black Scholes formula(which will be discussed in more detail later) backwards to find the implied volatility.

Risk neutrality:
Risk can be interpreted(?) as the variance of the return - how much the price deviates from the expected price(?). There are two types of risk, namely specific and non-specific risk. The former is risk associated with the specific asset/company and can be minimised by diversification(building a portfolio with negatively correlated assets). The latter is risk associated with factors influencing the whole market and can be reduced by hedging(taking opposite psoitions in similar derivatives which gives less return and less non-specific risk). Looking at the risk-neutral case means that investor's risk preferances are irrelevant as all (non-specific) risk can be hedged away. There is no profit to be made above the risk free return. Practically this means that we replace the rate of growth by the risk free rate for all random walks which means that eventhough investors may disagree on the rate of growth, they will value options the same.

Chapter 3

Monte Carlo

Chapter 4

The Black-Scholes Model

When deriving the Black Scholes formula, one makes the following assuptions

- The underlying asset follows a lognormal random walk.
- The risk free rate, r and volatility σ are known functions of time.(This is not a very realistic assumption as both r and σ also are stochastic. There are methods that also model them as stochastic differential equations.)
- There are no transaction costs for trading the underlying assets. (Also not a realistic assumption. Also exists models that take this into account.)
- Delta hedging is possible - underlying can be traded continously and one can buy and sell any number(not nessecarely an integer) of the underlying
- There are no arbitrage oportunities.
- There are no dividends paid out by underlying asset during the life of the option. (This can be adjusted for if one knows when and how much dividend will be paid out in advance.)

It is reasonable to assume the optionvalue which will be denoted by V is a funciton of time, t, (even more specific, time to expiry, $T-t$) and the underlying asset S. Consider the following portfolio, Π

$$\Pi = V(S,t) - \Delta S \tag{4.1}$$

where Δ is some quantity of the underlying asset, called the delta. Next we want to consider a infinitessimal change of the portfolio for some time interval $(t, t+dt)$. We assume that S follows a lognormal random walk (eqref). By Itô's Lemma

$$dV = \frac{\partial V}{\partial S} dS + \frac{\partial V}{\partial t} dt + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} dt. \tag{4.2}$$

Hence we obtain the following expression for $d\Pi$

$$d\Pi = \Big(\mu S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} - \mu \Delta S\Big) dt + \sigma S \Big(\frac{\partial V}{\partial S} - \Delta\Big) dW. \tag{4.3}$$

By chosing $\Delta = \frac{\partial V}{\partial S}$ we eliminate the random/stochastic term and thereby eliminate the risk connected to the randomness(?). This is called delta hedging and is a dynamic hedging strategy. Our remaining epression for $d\Pi$ is

$$d\Pi = \Big(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\Big)dt \tag{4.4}$$

and thus completely deterministic. As mentioned, this is a riskless portfolio and the change in value of the portfolio should therefore correspond to the amount we would get on return from a risk-free account (by the assumption of no arbitrage). I.e.

$$d\Pi = r\Pi dt \tag{4.5}$$

$$\Big(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 + \frac{\partial^2 V}{\partial S^2}\Big)dt = r\Big(V - \Delta S\Big)dt \tag{4.6}$$

From the latter we obtain the Black Scholes partial differential equation

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\ rS\frac{\partial V}{\partial S} - rV = 0. \tag{4.7}$$

In order to solve (4.7) we need some boundary and final conditions.(opposite of inital conditions). For a european call otpion, this would be

$$V(S,T) = max(S(T) - E, 0) \tag{4.8}$$

$$V(0,t) = 0 \tag{4.9}$$

$$V(S,t) = S \text{ as S } \to \infty \tag{4.10}$$

This leads to the following closed form solution for European call options.

$$V(S,t) = SN(d_1) - Ee^{-r(T-t)}N(d_2) \tag{4.11}$$

$N(\cdot)$ is the cumulative normal distribution and

$$d_1 = \frac{\log(S/E) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \tag{4.12}$$

$$d_2 = \frac{\log(S/E) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \tag{4.13}$$

Chapter 5

Artificial Neural Networks

5.1 Deep Feedforward Networks

Multilayer perceptrons (MLPs) are a type of deep feedforward networks. It is called feed-forward because the input $\mathbf{x}$ floats forward through the network to finally produce an output $\mathbf{y}$. The goal of the network is to approximate some function f so that $\mathbf{y} = f(\mathbf{x})$. It does so by defining a mapping $\mathbf{y} = f^*(\mathbf{x}; \theta)$ for which it optimizes the parameters θ by learning/training.

(Difference between linear perceptron and mulitlayer non-linear activation function. Why can it approximate non-linear functions?)

The network is deep if it consists of two or more hidden layers. Hidden layers are classified by being between the input and the output layer. Each layer in the network defines an activation function w and initializes weights W_{ij} for each node and a bias $b^{(i)}$. If the number of hidden layers is n, this is called the depth of the network, the number of nodes in each hidden layer is m_i for $i = 1, ..., n$, this is called the width of the network and $\mathbf{x} \in \mathbb{R}^d$, $W^(1) \in \mathbb{R}^{m_1 \times d}$, $W^(i) \in \mathbb{R}^{m_i \times m_{i-1}}$ for $i = 2, \ldots n$, $b^{(i)} \in \mathbb{R}$ for $i = 1, \ldots n$ then the process in vector notation would look like

$$
\begin{aligned}
\mathbf{h}^{(1)} &= w(b^{(1)} + W^{(1)}\mathbf{x}) \\
\mathbf{h}^{(2)} &= w(b^{(2)} + W^{(2)}\mathbf{h}^{(1)}) \\
&\vdots \\
\mathbf{h}^{(n)} &= w(b^{(n)} + W^{(n)}\mathbf{h}^{(n-1)}) \\
\mathbf{y} &= w^{(out)}(b^{(out)} + \mathbf{w}^{\mathrm{T}}\mathbf{h}^{(n)})
\end{aligned}
$$

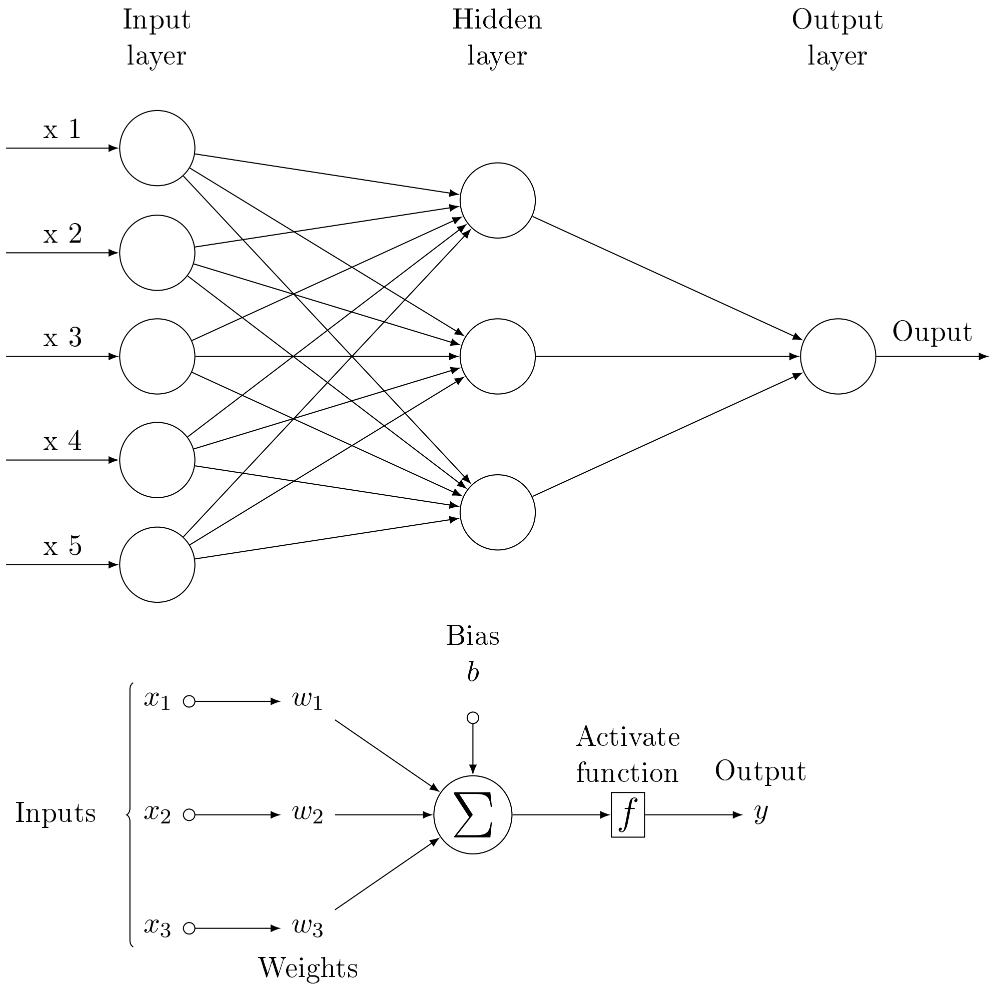

Figure 5.1: Feedforward neural network

5.1.1 Activation function

Different choices for activation functions are the sigmoid function

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

hyperbolic tangent, tanh,

$$f(z) = \frac{1 + e^{-2z}}{1 - e^{-2z}},$$

rectified linear units, ReLU,

$$f(z) = max(0, z),$$

softplus,

$$f(z) = \ln(1 + e^{z})$$

and several others. (ReLU have become especially popular as of lately and is used for all convolutional networks.)

5.2 Gradient-Based Training

5.2.1 Backpropagation

In order to train the model, one needs to specify a cost function to minimise. This cost function often involves a cross-entropy between the training data and the predicted data(?) and a regularising term (penalising complex models to avoid overfitting). (write more about cost function?).
In order to minimise the cost function, the gradient of the model has to be calculated. This is done using the backpropagation algorithm.

5.2.2 Stochastic Gradient Descent

Write about the different loss functions, optimisers, activation functions, dropout rate and other regularising techniques. Argument about choice made.

(Define input and output of the neural network, why are these reasonable?)