UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2015:14

Pricing American Options using Lévy Processes and Monte Carlo Simulations

Jonas Bergström

Examensarbete i matematik, 30 hp
Handledare och examinator: Maciej Klimek
Juni 2015

Department of Mathematics
Uppsala University

PRICING AMERICAN OPTIONS USING LÉVY PROCESSES AND MONTE CARLO SIMULATIONS

JONAS BERGSTRÖM

Abstract. This paper gives a description of how Lévy processes are utilized in order to establish a more accurate model of the financial market than the standard Black-Scholes model. The Lévy process that is discussed in this paper uses the Normal Inverse Gaussian (NIG) distribution. Stochastic volatility is also addressed with the help of CIR-processes. The end result is a comparison between the standard Black-Scholes model, the model with a NIG distribution and a Black-Scholes model with stochastic volatility. The comparison is done with respect to pricing and measurements of risk in the form of estimating American options and the greek delta. To accomplish this the Black-Scholes model is formulated following a discussion about its inconsistencies with the real financial market. Then Lévy processes and CIR-processes are established followed by a section about how to simulate these processes and a section about estimating American options and greeks.

Acknowledgements. I would first like to thank my supervisor Professor Maciej Klimek for his encouragement and support. He has consistently been available to discuss the subject and scope of the thesis.

1. the black-scholes model

The Black-Scholes model is a mathematical interpretation of a financial market consisting of stocks, bonds and derivatives. The bond is said to be risk-free while the stock has a risk component. This market is assumed to be free of *arbitrage*, meaning a trader can not make an initial investment of zero that later leads to a profit with probability 1 hence with no risk involved.

In the Black-Scholes model the stock has the following stochastic dynamics:

$$dS_t = \mu S_t dt + \sigma S_t dB_t \tag{1}$$

where μ is the expected rate of return of the stock and σ is the volatility i.e standard deviation of returns of the stock.

If a stock is assumed to pay dividend with dividend yield q, then (1) becomes

$$dS_t = (\mu - q) S_t dt + \sigma S_t dB_t \tag{2}$$

The dividend is assumed to be payed continuously to the holder of the stock. From an arbitrage-free perspective, if one were to create two shares on the same company, one paying dividend and the other paying zero dividend, the shares would have had the same total return i.e dividends plus capital gains. This is why, for a stock that pays dividends, the dividend yield is subtracted from μ [1].

However, one can still assume that the stock moves according to (1) if one discounts the present price as $e^{-qt}S_0$. In other words, simulating a stock to time T starting at the price S_0 using (2) yields the same result as using (1) but starting at the price $e^{-qT}S_0$ [1]. In this paper I use the second method.

B_t is a Brownian motion i.e a stochastic process following a normal distribution with zero mean and standard deviation $\sqrt{t}$. Also its increments are stationary and independent.

A Brownian motion can be simulated by first picking a uniformly distributed random variable U and then use the following theorem

Theorem 1.1. *If U is a standard uniformly distributed random variable and F^{-1} is the inverse of the cumulative probability function F, then the random variable $F^{-1}(U)$ has the same distribution as F.*

Proof.

$$\mathbb{P}(F^{-1}(U) \leq t) = \mathbb{P}(F(F^{-1}(U)) \leq F(t)) = \mathbb{P}(U \leq F(t))$$

Since the cumulative probability function for a standard uniformly distributed variable is

$$F_U(x) = x$$

it follows that

$$\mathbb{P}(U \leq F(t)) = F(t)$$

□

With the purpose of finding an explicit solution to (1) one needs to introduce the concept of information, *Itô's lemma* and some theory concerning stochastic integration.

Definition 1.2. *$\mathcal{F}_t^X$ denotes the information generated by the trajectories of the process X_t. If $Y_t \in \mathcal{F}_t^X$ then Y_t is said to be adapted to the process X_t meaning that it is possible to determine the value of Y_t given the information of X_t.*

In this paper $\mathcal{F}_t$ can be interpreted as all market information available to investors at time t.

Itô's lemma can be seen as a stochastic version of the chain rule in real analysis. If one takes a function f on a stochastic process X, with some given stochastic dynamics, then Itô's lemma gives a description of the dynamics of $f(X)$ (see [11])

Theorem 1.3. (Itô's lemma) *Assume that the stochastic process X_t has the dynamics*

$$dX_t = \mu_t dt + \sigma_t dB_t$$

where μ_t, σ_t are two adapted processes.

Define a new stochastic process $f(t, X_t)$, where f is assumed to be smooth. Given the multiplication table

$$\begin{cases} (dt)^2 = 0 \\ dt \cdot dB_t = 0 \\ (dB_t)^2 = dt \end{cases} \tag{3}$$

the stochastic dynamics for f becomes

$$df = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX_t)^2 \tag{4}$$

or equivalently

$$df = (\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{1}{2}\sigma_t^2 \frac{\partial^2 f}{\partial x^2})dt + \sigma_t \frac{\partial f}{\partial x} dB_t$$

Proof. This proof is a simplified version of the one that can be found in [11]. First make a Taylor expansion of f around the point (t, x).

$$\begin{aligned} f(t+h, x+k) &= f(t,x) + h\frac{\partial f}{\partial t}(t,x) + k\frac{\partial f}{\partial x}(t,x) \\ &\quad + \frac{1}{2}(h^2 \frac{\partial^2 f}{\partial t^2} + 2hk\frac{\partial^2 f}{\partial t \partial x} + k^2 \frac{\partial^2 f}{\partial t^2}) + R \end{aligned}$$

where

$$R = \frac{1}{3!}(h\frac{\partial}{\partial t} + k\frac{\partial}{\partial x})^3 f(t+sh, x+sk) \ , \ s \in (0,1)$$

Now let $h \to dt$ and $k \to dX_t$ to get

$$\begin{aligned} f(t+dt, x+dX_t) &= f(t,x) + \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(dt)^2 \\ &\quad + \frac{\partial^2 f}{\partial t \partial x}dt \cdot dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX_t)^2 + R \end{aligned}$$

From (3) it follows that

$$dt \cdot dX_t = dt(\mu_t dt + \sigma_t dB_t) = 0 \text{ and } (\frac{\partial}{\partial t}dt + \frac{\partial}{\partial x}dX_t)^3 = 0$$

therefore

$$df = f(t+dt, x+dX_t) - f(t,x) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX_t)^2$$

Since

$$(dX_t)^2 = \mu_t^2(dt)^2 + 2\mu_t\sigma_t dt \cdot dB_t + \sigma_t^2(dB_t)^2 = \sigma_t^2 dt$$

it follows that df can also be written as

$$df = (\frac{\partial f}{\partial t} + \mu_t\frac{\partial f}{\partial x} + \frac{1}{2}\sigma_t^2\frac{\partial^2 f}{\partial x^2})dt + \sigma_t\frac{\partial f}{\partial x}dB_t$$

□

As discussed in [11], the dynamics

$$dX_t = \mu_t dt + \sigma_t dB_t$$

of the stochastic process X_t is actually just a shorthand for

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s$$

where the first integral is a normal Riemann integral. But how should one interpret the second?

Without going in to details that are beyond the scope of this paper, a stochastic integral can be written as

$$\int_0^t f(s)dB_s = \lim_{n\to\infty}\int_0^t f_n(s)dB_s$$

where $\{f_n\}$ is a sequence of adapted *simple processes* with the condition

$$\int_0^t E[f_n^2(s)dB_s] < \infty, \forall n \in \mathbb{N}$$

The fact that f_n is a simple process, means that there exists a partition of $[0, t]$ s.t.

$$\int_0^t f_n(s)dB_s = \sum_{k=0}^{n-1} f(t_k)(B_{t_{k+1}} - B_{t_k})$$

where the interval $[0, t]$ is portioned as $0 = t_0 < t_1 < ... < t_n = t$. In other words f_n is constant on each portioned subinterval of $[0, t]$.

Theorem 1.4. *Let $f(X_\tau) \in \mathcal{F}_\tau^B$ be a stochastic process s.t $E[f(X_\tau)^2] < \infty$. Then*

$$E[\int_s^t f(X_\tau)dB_\tau|\mathcal{F}_s^B] = 0$$

Proof. Since the complete proof of this theorem is beyond the scope of this paper, the process $f(X_\tau)$ is here assumed to be simple. By the law of total expectation it follows that

$$E[\int_s^t f(X_\tau)dB_\tau|\mathcal{F}_s^B] = E\Big[E[\int_s^t f(X_\tau)dB_\tau|\mathcal{F}_t^B]\big|\mathcal{F}_s^B\Big]$$

where $s < t$.

Because f is a simple function it follows that the inner expectation equals (see [11])

$$\begin{aligned} E[\int_s^t f(X_\tau)dB_\tau|\mathcal{F}_t^B] &= E[\sum_{k=0}^{n-1} f(X_{\tau_k})(W_{\tau_{k+1}} - W_{\tau_k})|\mathcal{F}_t^B] \\ &= \sum_{k=0}^{n-1} E[f(X_{\tau_k})(W_{\tau_{k+1}} - W_{\tau_k})|\mathcal{F}_t^B] \\ &= \sum_{k=0}^{n-1} E[f(X_{\tau_k})|\mathcal{F}_t^B]E[(W_{\tau_{k+1}} - W_{\tau_k})|\mathcal{F}_t^B] \\ &= 0 \end{aligned}$$

□

By using Itô's lemma one can show that the solution of (1) is

$$S_t = S_0 \exp\big((\mu - \frac{1}{2}\sigma^2)t + \sigma B_t\big) \tag{5}$$

This is done by using Itô's lemma with the function $f(S_t) := \log(S_t)$ where S_t has the dynamics as in (1).

$$f(x) = \log(x) \Rightarrow \frac{\partial f}{\partial t} = 0\ ,\ \frac{\partial f}{\partial x} = \frac{1}{x}\ ,\ \frac{\partial^2 f}{\partial x^2} = -\frac{1}{x^2}$$

Itô's lemma then yields

$$df = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dB_t$$

or equivalently

$$\log(S_t) - \log(S_0) = (\mu - \frac{1}{2}\sigma^2)t + \sigma B_t$$

Taking exponents on both sides yields the result at (5).

An *option* is a contract that gives the holder the right, but not the obligation, to buy or sell an underlying asset e.g a stock for a certain predetermined price called the strike price. The right to buy is called a call option, and the right to sell is called a put option.

A European option can *only* be exercised at a future date called maturity, meanwhile an American option can be exercised at any date between now and maturity.

Assume one has a portfolio consisting of a short position in Δ number of stocks and a long position in an European option with the shorted stock as its underlying asset. Let P_t denote the portfolio price process, V_t the option price process and S_t the stock process. Given that the number Δ is fixed, the portfolio has the dynamics

$$dP_t = dV_t - \Delta dS_t \tag{6}$$

Following from Itô's lemma, dV_t can be written as

$$dV_t = (\frac{\partial V}{\partial t} + \mu S_t \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S^2})dt + \sigma S_t \frac{\partial V}{\partial S} dB_t$$

Inserting this into (6) and changing Δ to $\frac{\partial V}{\partial S}$ yields

$$dP_t = (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S^2})dt$$

Note now that the dynamics of the portfolio has no element of risk or randomness. If instead one had invested $V_t - \frac{\partial V}{\partial S} S_t$ in a bond the dynamics of that position would have been

$$dP_t = r(V_t - \frac{\partial V}{\partial S} S_t)dt$$

Since the Black-Scholes model assumes an arbitrage-free market the two drift terms in the above portfolio positions must be equal. Hence, it follows that

$$\frac{\partial V}{\partial t} + rS_t \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S^2} - rV_t = 0 \tag{7}$$

which is the Black-Scholes PDE.

At maturity, which is denoted by T, one must have that

$$V_T = \Phi(S_T)$$

where Φ describes the payoff structure of the contract. For example the payoff function for a call option is $\Phi(x) = \max\{x - K, 0\}$ where K is the strike price.

The Feynman-Kac theorem uses stochastic processes to find a solution to a boundary valued PDE (see e.g. [11]).

Theorem 1.5. (Feynman-Kac)
Let S_t be the solution to

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

and assume that F solves the boundary value problem

$$\begin{cases} \frac{\partial F}{\partial t} + \mu \frac{\partial F}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 F}{\partial x^2} - rF = 0 \\ F(T, z) = \Phi(z). \end{cases}$$

Also assume that $E\left[|\sigma \frac{\partial F}{\partial x}(t, S_t) e^{-rt}|^2\right] < \infty$ and $\sigma e^{-rt} \frac{\partial F}{\partial x}(t, S_t) \in \mathcal{F}_t^B$.

Then it follows that

$$F(t, z) = e^{-r(T-t)} E[\Phi(S_T) | S_t = z]$$

Proof. Using Itô's lemma on the function $f(t, S_t) := e^{-rt}F(t, S_t)$ yields

$$\begin{aligned} f(T, S_T) &= f(t, S_t) + \int_t^T e^{-rs}\left(\frac{\partial F}{\partial t} + \mu\frac{\partial F}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 F}{\partial x^2} - rF\right)ds + \int_t^T e^{-rs}\sigma\frac{\partial F}{\partial x}dB_s \\ &= f(t, S_t) + \int_t^T e^{-rs}\sigma\frac{\partial F}{\partial x}dB_s \end{aligned}$$

Now, on both sides, take expectations conditioned on the information that $S_t = z$

$$E[e^{-rT}F(T, S_T)|S_t = z] = E[e^{-rt}F(t, S_t) + \int_t^T e^{-rs}\sigma\frac{\partial F}{\partial x}dB_s|S_t = z] = e^{-rt}F(t, z)$$

This is now equivalent to

$$F(t, z) = e^{-r(T-t)}E[\Phi(S_T)|S_t = z]$$

□

Now using the above theorem one can find a solution to (7), namely

$$V = e^{-rT}E[\Phi(S_T)|S_0 = s] \tag{8}$$

which is the discounted expected value of the payoff function at maturity.

Since so called *martingale theory* plays a big role in stochastic processes, it is also important within the field of financial mathematics.

Definition 1.6. (Martingale) *A stochastic process $X_t \in \mathcal{F}_t^B$ is called a martingale if*

- $E[|X_t|] < \infty$, $\forall t \in \mathbb{R}_{\geq 0}$
- $E[X_t|\mathcal{F}_s^B] = E[X_s], \forall s \leq t$

The last point in the above definition basically says that standing at time s, if X is a martingale, the expectation of the future value X_t, given the information $\mathcal{F}_s^B$, is nothing else but the expected value of the stochastic process at time s.

Now follows a short introduction about the connection between probability and measure theory (for more see e.g. [3]).

A *probability space* is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is called the *event space*, $\mathcal{F}$ is a *σ-algebra* on Ω and $\mathbb{P}$ is a *probability measure*. Ω is the set of all scenarios. In a financial context, if $\omega \in \Omega$ then ω is a certain realization on the market e.g. a stock has evolved in a certain way [3].

The σ-algebra $\mathcal{F}$, is a family of subsets of Ω with the following properties

- $\emptyset \in \mathcal{F}$
- If $\{A\}_{n=0}^{\infty}$ is a sequence of disjoint sets then $\bigcup_{n=0}^{\infty} A_n \in \mathcal{F}$
- $\forall A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

Basically $\mathcal{F}$ is the family of sets (called events) from Ω that can be measured.

The probability measure $\mathbb{P}$ assigns, for each event $A \in \mathcal{F}$, a number between 0 and 1. There can exist several probability measures. In this paper $\mathbb{P}$ denotes the so called *original* probability measure on the financial market.

When using the original probability measure $\mathbb{P}$, the stock dynamics are assumed to be

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

The expected rate of return of the stock i.e. μ, plays a vital role when it comes to market assumptions such as completeness and freedom of arbitrage. In order for the model to be free of arbitrage one must find a so called equivalent martingale measure [1].

Definition 1.7. (Equivalent Martingale Measure) *$\mathbb{Q}$ is an equivalent martingale measure of $\mathbb{P}$ if*

- *$\mathbb{Q}$ is equivalent to $\mathbb{P}$ i.e $\mathbb{Q}(A) = 0 \Leftrightarrow \mathbb{P}(A) = 0$ for an event A*
- *$e^{-rt} S_t$ is a martingale under the "new" measure $\mathbb{Q}$*

Theorem 1.8. (The 1st Fundamental Theorem of Asset Pricing) *There exists an equivalent martingale measure iff the market is free from arbitrage.*

Theorem 1.9. (The 2nd Fundamental Theorem of Asset Pricing) *There exists an equivalent martingale measure that is unique iff the market is complete.*

In order to formulate theorem 1.8 and theorem 1.9 precisely, more concepts needs to be introduced. Since this is not necessary for this paper, the theorems are presented as above and a more rigorous explanation can be found in [11].

In the standard Black-Scholes model there exists an unique equivalent martingale measure where the dynamics of the stock becomes

$$dS_t = rS_t dt + \sigma S_t dB_t \tag{9}$$

Here μ has been replaced with the risk-free rate r. This martingale measure is also called the risk-free measure.

2. INCONSISTENCIES WITH THE BLACK-SCHOLES MODEL

Empirical evidence suggests that the standard Black-Scholes model is insufficient as a mathematical depiction of the financial market (see e.g. [1] and [2]).

By setting $\mu = r$ and $\Delta t = t_2 - t_1$ in (5) and rewriting it as

$$\log(\frac{S_{t_2}}{S_{t_1}}) = (r - \frac{1}{2}\sigma^2)\Delta t + \sigma B_{\Delta t}$$

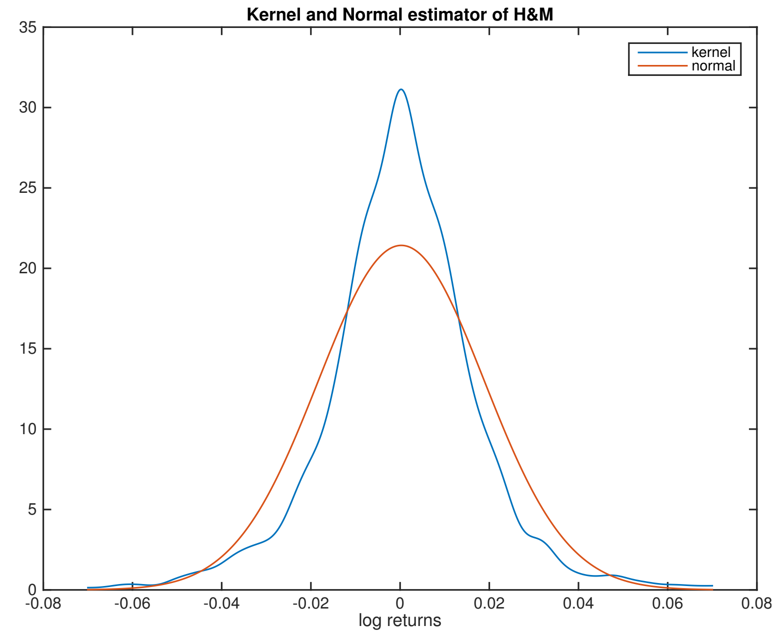
one sees that the *log returns* should be normally distributed. However empirical evidence acknowledge that distributions of log returns from stock markets experiences excess kurtosis and negative skewness [5].

Excess kurtosis means that compared to the normal distribution, the probability density function (PDF) have sharp noticeable peaks around its mean and heavy, slowly declining tails [1]. Market log returns are much more clustered around its mean and the probability that there will occur a significant jump in return is not as negligible as it is in the normal distribution.

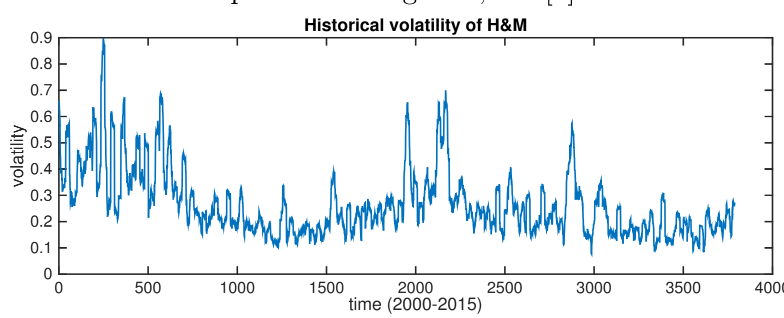
Negative skewness says that the tails are fatter to the left meaning that large declines in stock prices outweigh the equally large increases in stock prices [1]. However, this can not be observed in for example foreign exchange markets, where there is better symmetry [2].

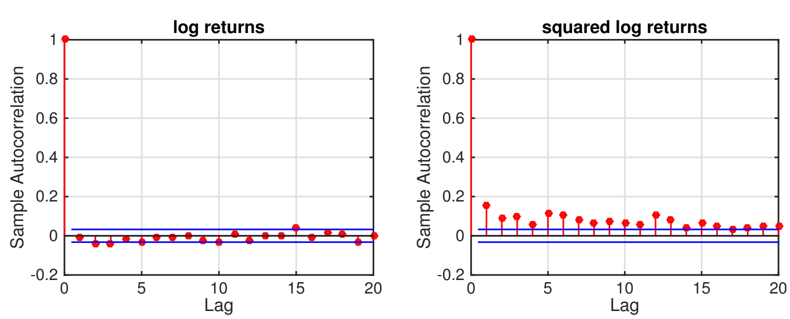
Also traded stock prices are not continuous and are experiencing price jumps. Since Brownian motion, which is built on a normal distribution, has continuous trajectories, jumps can not happen in this mathematical model.

Below is a plot of the normal and the kernel density estimator of the daily log returns of the Swedish retail-clothing company Hennes & Mauritz (H&M) from 2000 to 2015. H&M has the largest market capitalization in Sweden and is one of the most traded shares on the Swedish stock market. The kernel density estimator is used to approximate the empirical density of the stock (see [1]).



Another inconsistency with the Black-Scholes model is how volatility is constructed. Data from financial markets reveal that volatility can also be seen as stochastic. So called *volatility clustering* can be observed meaning that there are instances where an uptick in volatility is likely to be followed by even higher volatility. If one looks at historical volatility for H&M a mean reverting nature can be seen. Moreover while returns themselves are not correlated, squared returns, which are used to measure deviation and not whether returns are positive or negative, are [2].





3. LÉVY PROCESSES

As was seen in the previous section the assumption of normal distributed logarithmic returns of stock prices leave much to be desired. Therefore a more general distribution needs to be found. One that can take into account the issue of excess kurtosis, skewness and jumps. Lévy processes can satisfy all of the above concerns.

Definition 3.1. *Let $X_t \in \mathcal{F}_t$ with $X_0 = 0$. X_t is a Lévy process if*

- *X_t has independent increments i.e $X_t - X_s$ is independent of $\mathcal{F}_s$, $0 \leq s < t < \infty$*
- *X_t has stationary increments i.e $X_t - X_s \stackrel{d}{=} X_{t+\tau} - X_{s+\tau} \forall \tau$, $0 \leq s < t < \infty$*
- *X_t is continuous in probability i.e $\forall \epsilon, \lim_{h \to 0} \mathbb{P}(|X_{t+h} - X_t| > \epsilon) = 0$*

Since Brownian motion satisfies the above definition, it is a Lévy process. While all Lévy processes has right continuous trajectories a.s with left limits, Brownian motion is the *only* Lévy process with continuous trajectories [4].

Note that the third point in the above definition means that at a given fixed t the probability of seeing a jump is zero. Jump times should be random [3] .

Because of independent and stationary increments a Lévy process X_t can be seen as a sum of any number of i.i.d. stochastic variables. This can be seen by fixing t and choosing an $n \in \mathbb{N}$ and writing X_t as

$$X_t = X_{\frac{t}{n}} + (X_{\frac{2t}{n}} - X_{\frac{t}{n}}) + ... + (X_t - X_{\frac{(n-1)t}{n}})$$

where $(X_{\frac{kt}{n}} - X_{\frac{(k-1)t}{n}}) \stackrel{d}{=} X_{\frac{t}{n}}$ [6].

This property is called *infinite divisibility.* It can also be shown that given any infinite divisible distribution there exists a Lévy process with that same distribution (see Corollary 11.6 in [14]).

The second most common Leév́y process is the Poisson process and both it and Brownian motion plays a key role in the important *Lévy-Itô Decomposition* theorem.

A Poisson process is a jump process in the following way. Let $\{\tau_i\}$ be a sequence of exponential distributed jump-times with intensity parameter λ, meaning that λ is the expected number of jumps per unit of time.

Then the Poisson process N_t, with intensity parameter λt, can be defined as

$$N_t = \sum_{k>0} \mathbb{1}_{\{t>T_k\}},\ N_0 = 0$$

where $T_k := \sum_{i=1}^{k} \tau_i$ [7].

The probability mass function for a Poisson process with intensity parameter λt is

$$\mathbb{P}(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

and its characteristic function is

$$\begin{aligned}
\phi(u) = E[e^{iuN_t}] &= \sum_{k=0}^{\infty} e^{iuk} \frac{e^{-\lambda t}(\lambda t)^k}{k!} \\
&= \sum_{k=0}^{\infty} (e^{iu})^k \frac{e^{-\lambda t}(\lambda t)^k}{k!} \\
&= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t e^{iu})^k}{k!} \\
&= \exp(\lambda t(e^{iu} - 1))
\end{aligned}$$

From above it follows that

$$E[N_t] = i^{-1}\phi'(u)|_{u=0} = \lambda t$$

A Poisson process can be simulated as follows [1]. First generate a sequence of $\mathrm{Exp}(\lambda)$ distributed numbers $\{e_n\}$ by letting

$$e_n := -\log(u_n)/\lambda$$

where $\{u_n\}$ is a sequence of standard uniformly distributed random variables.

Now for $n = 1, 2, ...$ let $\tau_n = \tau_{n-1} + e_n$, $\tau_0 = 0$. The sequence $\{\tau_n\}$ is the accumulated jump times. Given theses sequences one can simulate the Poisson process with the observation dates $\{k\Delta t : k = 0, 1, 2, ...\}$ for some small $\Delta t > 0$ by

$$N_n = \max\{k : \tau_k \leq n\Delta\} \text{ for } n \geq 1,\ N_0 = 0$$

This standard version of a Poisson process has it's jumps equal to 1 which has little use from a financial context. However if one let the jump sizes themselves be stochastic one arrives at the so called *compound Poisson process*, namely by introducing i.i.d. random variables Y_i, where Y_i is the ith jump size with some chosen distribution independent from N_t. The compound Poisson process is then defined as

$$X_t = \sum_{k>0} \mathbb{1}_{\{t>T_k\}} Y_k = \sum_{k=1}^{N_t} Y_k \tag{10}$$

(see e.g. [3] for a more rigorous description)

In order to describe the nature of the jumps of the compound Poisson process X_t, one takes a measure on a certain set that counts the number of jump sizes that are in that set. More precisely let $A \subset \mathbb{R}\backslash\{0\}$ be a set contained in some σ-algebra $\mathcal{F}$. Then the measure defined as

$$J_{X_t}(A) = \#\{s \in (0,t], \Delta X_s \in A\} \text{ where } \Delta X_s = X_s - X_{s-}$$

measures the number of times between 0 and t that jumps, with sizes in A, occur. The above definition makes $J_{X_t}(\cdot)$ a measure on sets in $\mathcal{F}$. This measure will depend on some realization $\omega \in \Omega$ i.e a trajectory of X_t, where Ω is the set of all trajectories. Because of this, the measure J_{X_t} is called a *random Poisson measure* [3].

Using this random Poisson measure it follows that the compound Poisson process X_t can be written as

$$X_t = \sum_{0\leq s\leq t} \Delta X_s \mathbb{1}_{\mathbb{R}\backslash\{0\}}(\Delta X_s) = \int_{\mathbb{R}\backslash\{0\}} x J_{X_t}(dx)$$

The same definition of J_{X_t} also works for other Lévy processes and not just the compound Poisson process.

All Lévy processes has a so called *intensity measure* that corresponds to the expected number of jumps with jump sizes in a certain set per unit of time. This is called the Lévy measure.

Definition 3.2. *Let X_t be a Lévy process. A measure ν on $\mathbb{R}\backslash\{0\}$, defined as*

$$\nu(A) = E\big[\#\{s \in [0,t]; \Delta X_s \in A\}\big]$$

is the Lévy measure of X_t

From the definition of a Lévy measure, it can be shown (see e.g. [3]) that ν must satisfy

$$\int_{\mathbb{R}} \min\{1, x^2\}\nu(dx) < \infty$$

The Lévy measure gives a description of the nature of the jump sizes in (10) because

$$\mathbb{P}(Y_k \in [a,b]) = \frac{1}{\lambda}\int_a^b \nu(dx) = \frac{\nu([a,b])}{\lambda} \tag{11}$$

In this context λ can be written as $\lambda = \int_{\mathbb{R}\setminus\{0\}} \nu(dx)$ [12].

There are some processes, e.g. the NIG distribution, where $\nu(dx)$ can be written as $\nu(dx) = f(x)dx$. Here $f(x)$ is called a Lévy density.

The following theorem gives a unique representation of any given Lévy process.

Theorem 3.3. (Lévy-Itô Decomposition) *Let X_t be a Lévy process. Then there exists a Brownian motion, with drift and diffusion components $\gamma \in \mathbb{R}$ respectively $\sigma \in \mathbb{R}_{>0}$, and an independent Poisson random measure J_{X_t} on $\mathbb{R}\setminus\{0\}$ s.t*

$$X_t = \gamma t + \sigma B_t + \int_{|x|<1} x\tilde{J}_{X_t}(dx) + \int_{|x|\geq 1} xJ_{X_t}(dx) \tag{12}$$

where

$$\tilde{J}_{X_t}(dx) = J_{X_t}(dx) - t\nu(dx)$$

A measure $\mathcal{M}$ is said to be independent if for some sequence of disjoint measurable sets $A_1, A_2, ..., A_n \in \mathcal{F}$, where $\mathcal{F}$ is a σ-algebra, then $\mathcal{M}(A_1), \mathcal{M}(A_2), ..., \mathcal{M}(A_n)$ are independent of each other [3].

So according to the Lévy-Itô Decomposition any Lévy process can be identified by its triplet (γ, σ, ν). The Lévy process can be written as the sum of a Brownian component and a jump component. Let us look closer at the last line in theorem 3.3 by integrating it over $A := \{|x| < 1\}$.

$$\int_A \tilde{J}_{X_t}(dx) = \int_A (J_{X_t}(dx) - t\nu(dx)) = J_{X_t}(A) - t\nu(A)$$

Here $J_{X_t}(A)$ can be viewed as a Poisson process with jumps occurring in A with intensity $E[J_{X_t}(A)] = t\nu(A)$.

Proposition 3.4. *Let $X_t \in \mathcal{F}_t$ be a Lévy process with $E[|X_t|] < \infty$. Then $X_t - E[X_t]$ is a martingale.*

Proof. Let $s < t$

$$E[X_t|\mathcal{F}_s] = E[X_t - X_s|\mathcal{F}_s] + E[X_s|\mathcal{F}_s] = E[X_t] - E[X_s] + X_s$$

$$\Rightarrow E[X_t - E[X_t]|\mathcal{F}_s] = E[X_t|\mathcal{F}_s] - E[X_t] = X_s - E[X_s]$$

□

Proposition 3.4 makes $\tilde{J}_{X_t}(dx) = J_{X_t}(dx) - t\nu(dx)$ a martingale.

Another way of identify a Lévy process is by looking at its characterstic function.

Theorem 3.5. (Lévy Khintchine formula) *Let X_t be a Lévy process with the Lévy triplet (γ, σ, ν). Then its characteristic function, denoted by ϕ, equals*

$$\phi(u) = E[e^{iuX_t}] = e^{t\psi(u)}$$

where

$$\psi(u) = i\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} (e^{iux} - 1 - iux\mathbb{1}_{\{|x|<1\}})\nu(dx)$$

Proof. From the Lévy-Itô decomposition theorem it follows that a Lévy process can be written as

$$X_t = \gamma t + \sigma B_t + \sum_{n=0}^{N_t} Y_k - E\big[\sum_{n=0}^{N_t} Y_k \mathbb{1}_{\{|Y_k|<1\}}\big]$$

The characteristic function of the Brownian part with the constant drift is

$$E[\exp(iu(\gamma t + \sigma B_t))] = \exp(iu\gamma t - \frac{1}{2}\sigma^2 u^2 t)$$

The characteristic function for the compound Poisson process equals

$$\begin{aligned} E[\exp(iu\sum_{n=0}^{N_t} Y_n)] &= \sum_{k=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!} E[\exp(iu\sum_{n=0}^{k} Y_n)] \\ &= \sum_{k=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!} (E[\exp(izY_1)])^k \end{aligned}$$

From (11) it follows that $E[\exp(iuY_1)] = \int_{\mathbb{R}} e^{iux}\frac{\nu(dx)}{\lambda}$ which yields [12]

$$\begin{aligned} E[\exp(iu\sum_{n=0}^{N_t} Y_n)] &= e^{-\lambda t}\sum_{k=0}^{\infty} \frac{(\lambda t \int_{\mathbb{R}} e^{iux}\frac{\nu(dx)}{\lambda})^k}{k!} \\ &= \exp(\lambda t(\int_{\mathbb{R}} e^{iux}\frac{\nu(dx)}{\lambda} - 1)) \\ &= \exp(t\int_{\mathbb{R}} (e^{iux} - 1)\nu(dx)) \end{aligned}$$

From Wald's equation it follows that

$$E\big[\sum_{n=0}^{N_t} Y_k \mathbb{1}_{\{|Y_k|<1\}}\big] = E[N_t]E\big[Y_1\mathbb{1}_{\{|Y_1|<1\}}\big] = \lambda t \int_{\mathbb{R}} x\mathbb{1}_{\{|x|<1\}}\frac{\nu(dx)}{\lambda}$$

From the above expressions it follows that

$$E[e^{iuX_t}] = e^{t\psi(u)}$$

where

$$\psi(u) = iu\gamma - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} (e^{iux} - 1 - x\mathbb{1}_{\{|x|<1\}})\nu(dx))$$

□

If one wants to use expected values on financial assets e.g. when setting up a simulation, it is crucial to cut out arbitrage opportunities. According to theorem 1.8 this is done by finding an equivalent martingale measure (EMM).

However, with a few exeptions, utilizing Lévy processes leads to incomplete markets [1], i.e theorem 1.9 is not satisfied. This means that not all contracts can be perfectly hedged. If this is a correct view of the real market is of course a subject for discussion. One argument for the belief in market incompleteness is that if all financial derivatives could be *perfectly* hedged or replicated using stocks and bonds, why do they even exist.

Theorem 1.8 is satisfied by adding an additional drift term to the original process X_t. Under this new measure $\mathbb{Q}$, called the mean correcting martingale measure, the stock follows the process

$$S_t = S_0 e^{Y_t} \text{ where } Y_t = X_t + mt$$

It is the choice of m that makes the measure $\mathbb{Q}$ an EMM [8].

Theorem 3.6. *Let $S_t = S_0 e^{X_t}$ be the stock process under the original measure $\mathbb{P}$ where X_t is a Lévy process. Let $\mathbb{Q}$ be an equivalent measure to $\mathbb{P}$ with the property that the stock price follows $S_t = S_0 e^{Y_t}$ where $Y_t = X_t + mt$.*

Then $\mathbb{Q}$ is an EMM if $m = r - \log(\phi(-i))$

Proof. For $\mathbb{Q}$ to be an EMM we need $E^{\mathbb{Q}}[e^{-rt}S_t|\mathcal{F}_u] = e^{-ru}S_u$ for $u < t$.

$$\begin{aligned} E^{\mathbb{Q}}[e^{-rt}S_t|\mathcal{F}_u] &= e^{-rt}E^{\mathbb{Q}}[S_u e^{Y_{t-u}}|\mathcal{F}_u] = e^{-rt}S_u E^{\mathbb{Q}}[e^{X_{t-u}+(t-u)m}|\mathcal{F}_u] \\ &= e^{-rt+(t-u)m}S_u E^{\mathbb{Q}}[e^{X_{t-u}}] = e^{-ru}S_u e^{(t-u)(m-r+\log(\phi(-i)))} \end{aligned}$$

where the second to last equality follows from independent increments and the last equality follows from theorem 3.5. So it follows that

$$m - r + \log(\phi(-i)) = 0$$

in order for $\mathbb{Q}$ to be an EMM [8]. Note that $\phi(-i)$ is here the characteristic function of X_1.

□

4. NORMAL INVERSE GAUSSIAN DISTRIBUTION

The Lévy process I have chosen to work with in this paper follows a NIG distribution. For more details about the NIG distribution see e.g. [1]. The probability density function of the NIG distribution is

$$f(x) = \frac{\alpha\delta}{\pi} \exp(\delta\sqrt{\alpha^2 - \beta^2} + \beta x) \frac{K_1(\alpha\sqrt{\delta^2 + x^2})}{\sqrt{\delta^2 + x^2}} \tag{13}$$

where $\alpha > 0$, $-\alpha < \beta < \alpha$, $\delta > 0$. $K_v(z)$ is the modified Bessel function of the third kind with index v and is written as

$$K_v(z) = \frac{\pi}{2} \frac{\Psi_v(z) - \Psi_{-v}(z)}{\sin(v\pi)}$$

where

$$\Psi_v(z) = \left(\frac{z}{2}\right)^v \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{\Gamma(v+k+1)k!}$$

and

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx$$

The NIG process is identified by the Lévy triplet $(\gamma, 0, \nu)$ where

$$\gamma = \frac{2\alpha\delta}{\pi} \int_0^1 \sinh(\beta x) K_1(\alpha x) dx$$

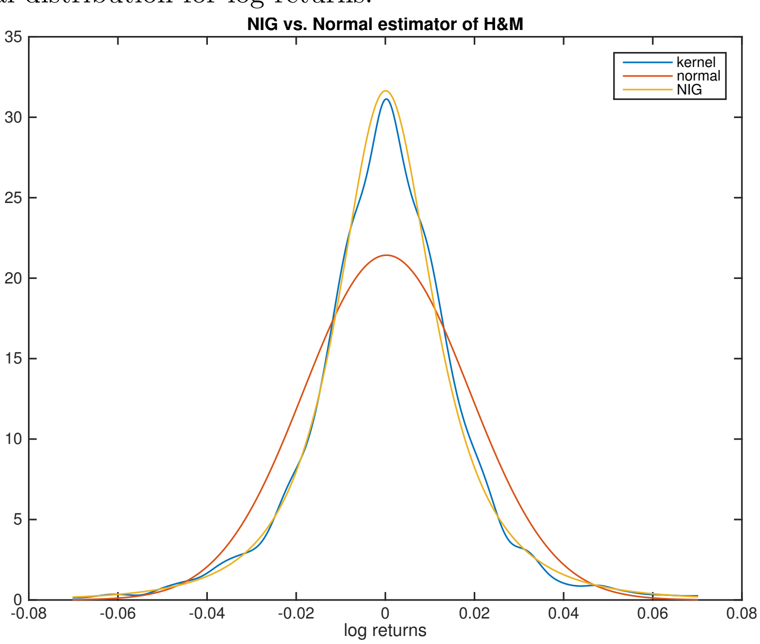
and

$$\nu(dx) = \frac{\alpha\delta}{\pi} \frac{\exp(\beta x) K_1(\alpha |x|)}{|x|}$$

The drift term m, that is added to the Lévy process in order to create a mean correcting martingale measure, equals

$$m = r + \delta(\sqrt{\alpha^2 - (\beta+1)^2} - \sqrt{\alpha^2 - \beta^2})$$

Since this distribution has more parameters than the normal distribution, the issues about kurtosis and skewness can be addressed. Below is a plot of the kernel density estimator together with the normal and NIG distributions. As the plot shows the NIG distribution is much more accurate than assuming a normal distribution for log returns.



The parameters of the NIG distribution are adapted to the stock H&M using the maximum likelihood method.

Given a number of observations $x_1, x_2, ..., x_n$, in this case daily closing prices from H&M between 2000 and 2015, one creates a *Likelihood function* $L(\theta)$, where θ is a family of parameters. In the case of the NIG distribution these parameters are α, β and δ from the PDF of the NIG distribution.

The likelihood function is defined as

$$L(\theta) = \Pi_{i=1}^{n} f(x_i; \theta)$$

where f is the PDF of the NIG distribution defined as in (13). In order to find the parameters that best describes the distribution of the observations $x_1, x_2, ..., x_n$ one maximizes $L(\theta)$ w.r.t θ. In other words find a θ where $L'(\theta) = 0$.

This is done more easier by instead of maximizing $L(\theta)$ one maximizes

$$l(\theta) := \log(L(\theta))$$

Since the derivative of $l(\theta)$ equals

$$\frac{d}{d\theta} \log(L(\theta)) = \frac{L'(\theta)}{L(\theta)}$$

finding the maximum of $L(\theta)$ is equivalent to finding it for $l(\theta)$ [10].

5. SIMULATION OF NORMAL INVERSE GAUSSIAN PROCESSES

Here is a description of an algorithm for simulating NIG processes. The algorithm is mostly taken from [1]. The method uses the fact that a NIG process can be written as a *time-changed* Brownian motion.

Let X_t be a NIG process with the parameters α, β and δ where $\alpha > 0$, $-\alpha < \beta < \alpha$ and $\delta > 0$. It can be shown that X_t can be written as

$$X_t = \beta\delta^2 I_t + \delta B_{I_t}$$

where I_t is an *inverse gamma* (IG) process. An IG process is a Lévy process with the probability density function

$$f(x) = \frac{a}{\sqrt{2\pi}} \exp(ab) x^{-3/2} \exp(-\frac{1}{2}(a^2 x^{-1} + b^2 x))$$

An IG process with parameters $a > 0$ and $b > 0$ can be simulated as follows.

First generate a normally distributed random variable z, with mean 0 and variance 1. Then define two new numbers y and x as $y = z^2$ and

$$x = \frac{a}{b} + \frac{y}{2b^2} - \frac{\sqrt{4aby + y^2}}{2b^2}$$

Now generate a standard uniform number u. If it it is the case that

$$u \leq \frac{a}{a+bx}$$

then appoint x to be the generated number from the IG distribution. Otherwise choose $\frac{a^2}{b^2 x}$ to be the generated number from the IG distribution.

When simulating a NIG process as a time-changed Brownian motion, one uses an IG process with the parameters $a = 1$ and $b = \delta\sqrt{\alpha^2 - \beta^2}$. In this way one can generate a NIG process.

6. Stochastic Volatility

Instead of changing to a more adjustable distribution like the NIG distribution, one can try to improve upon the Black Scholes model by adding stochastic volatility.

As discussed above, one of the inconsistencies in the standard Black Scholes model is how volatility acts in real financial markets compared to in the model. Volatility is not constant but moves in a stochastic nature over time.

One of the characteristics of observed historical volatility is the so called clustering effect where a large move in the stock price is likely to be followed by further large movements. The same goes for small movements i.e one can observe, from historical volatility, periods that are experiencing low volatility and periods with high volatility [2].

With the purpose of mathematical modeling volatility can also be assumed to be mean-reverting i.e there are periods when volatility fluctuates over and under a certain mean. This can be seen by looking for example at historical volatility of H&M. So in order to model this mathematically one needs a positive stochastic process that is mean-reverting. One stochastic process that has these characteristics is the Cox-Ingersoll-Ross (CIR) process.

The CIR process, y_t has the following dynamics [3]

$$dy_t = a(b - y_t)dt + c\sqrt{y_t}dZ_t$$

The parameter b is the long-run mean of the process y_t. This is the value that y_t will fluctuate around. a is the rate of mean-reversion and it describes how fast y_t returns to its long-run mean b. The last parameter c is the volatility of y_t and Z_t is a Brownian motion. The Brownian motion Z_t is usually defined by

$$Z_t = \rho B_t^{(1)} + \sqrt{1-\rho^2}B_t^{(2)} \text{ , where } \rho \in [-1, 1]$$

Here $B_t^{(1)}$ and $B_t^{(2)}$ are two independent Brownian motions where $B_t^{(1)}$ is the Brownian part of the underlying asset. In this way volatility is correlated with the process that drives the asset price [3].

This model can also be used to simulate interest rates. There are also models where volatility and the returns of the underlying assets possesses jumps. In this model volatility and returns share the same jump component (see e.g. [3]).

In order to simulate the above CIR process one can use an Euler scheme by setting $y_0 = b$ and then discretizes the process as

$$y_{t+\Delta t} = y_t + a(b - y_t)\Delta t + c\sqrt{y_t}Z_t$$

where Z_t is defined as above.

The problem with this scheme is that the process can reach zero. Since this is not consistent with historical volatility one must find a way around it. There are a number of solutions to this problem, for example instead using an implicit scheme while invoking a relationship between a, b and c.
However the simplest way to get around the issue is to use the modified Euler scheme

$$y_{t+\Delta t} = y_t + a(b - y_t^+)\Delta t + c\sqrt{y_t^+}Z_t$$

where $y_t^+ = \max\{y_t, 0\}$ [13].

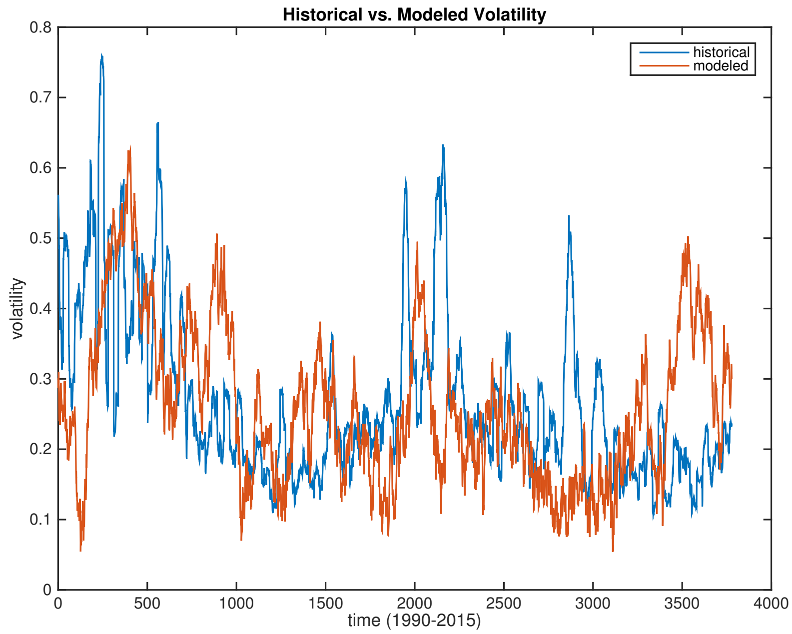
Volatility is thereafter defined as

$$\sigma_t = \sqrt{y_t^+}$$

and the underlying stock is discretized as

$$S_{t+\Delta t} = S_t \exp\left((\mu - \tfrac{1}{2}\sigma_t^2)\Delta t + \sigma_t B_{\Delta t}^{(1)}\right)$$

Below is a plot of the historical 30-day period volatility of H&M and the square root of a CIR process with $a = 2$, $b = 0.3^2$ and $c = 0.45$. These are the parameters I use when simulating the stochastic volatility in the Black Scholes model.



7. AMERICAN OPTIONS

An American option gives the holder the right but not the obligation to buy or sell an underlying asset for a predetermined price K, called the strike price. The option can be exercised *at any time* between its creation and a future maturity date T.

In contrast to its European counterpart there is no analytic formula for American options with one exeption. An American call option on a non-dividend paying stock has an analytic formula since its price is equal to its European version, i.e it is not profitable to exercise the option before T [11]. However, for American option pricing in general one needs to turn to numerical methods. The numerical method I have chosen to work with in this paper is based on Monte Carlo simulations.

As seen in (8), pricing option contracts are mostly about calculating an expected value given some information and then discounting it to the present. For this purpose Monte Carlo simulations are much suitable.

Assume X to be a continuous random variable with a known probability density function f_X. The expected value of a function V written on X is then

$$E[V(X)] = \int_{\mathbb{R}} V(x) f_X(x) dx$$

The introduction about Monte Carlo methods below, is based on [16].

Presume that one has simulated X in some way n times with the attained values $x_1, x_2, ..., x_n$. By evaluating V at all these observations and calculating its mean, one derives the so called Monte Carlo estimate

$$\bar{V}_n = \frac{1}{n}\sum_{i=1}^{n} V(x_i)$$

By looking at each observation as a version of the random variable X one can construct the *random* variable

$$\bar{V}_n(X) = \frac{1}{n}\sum_{i=1}^{n} V(X_i)$$

Now, from the weak law of large numbers it follows that $\bar{V}_n(X)$ will converge in probability to its expected value $E[V(X)]$ as $n \to \infty$ that is for any $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|\bar{V}_n(X) - E[V(X)]| > \epsilon) = 0$$

To establish a coherent mathematical formulation of American options it is good to introduce the concept of *stopping time.*

Definition 7.1. *A random variable $\tau : \Omega \to [0,\infty)$, where Ω is a sample space, is called a stopping time if*

$$\{\tau \leq t\} \in \mathcal{F}_t \ , \forall t \geq 0$$

for some set of information $\mathcal{F}_t$

The above definition can be illustrated by the following example. Let's say that one is observing a continuous stochastic process $X_t \in \mathcal{F}_t$ in the set $\mathbb{R}_{>0}$. Let $A \subset \mathbb{R}_{>0}$ be defined by

$$A = \{x \in \mathbb{R}_{t>0} : a < x < b\}$$

Then define a new process τ as

$$\tau = \inf\{t > 0 : X_t \in A\}$$

Here τ will quantify the first time that the process X_t enters into A and it is also a stopping time since one has that

$$\{\tau \leq t\} = \{X_s \in A, \text{for some } s \in [0,t]\} = \bigcup_{s=0}^{t}\{X_s \in A\}$$

Because X_t is adapted to the information $\mathcal{F}_t$ it is true that

$$\forall s \in [0,t] \ , \{X_s \in A\} \in \mathcal{F}_s \subset \mathcal{F}_t \Rightarrow \{\tau \leq t\} = \bigcup_{s=0}^{t}\{X_s \in A\} \in \mathcal{F}_t$$

When holding an American option one needs make a choice of either exercising the option or hold it to a future date. This is where stopping times come in.

Assume that a trader is holding an American option. Standing at time $t < T$ the trader must decide if t is the *optimal* stopping time i.e whether or not exercising the option at time t will gain the trader the greatest payoff of all potential exercising dates between t and T.

If the option is exercised at time t the trader will receive $\Phi(S_t)$ where Φ is the payoff function of the option. The optimal value the trader, standing at time t, can receive between t and T is equal to

$$\sup_{t \leq \tau \leq T} E[e^{-r(\tau - t)}\Phi(S_\tau)|\mathcal{F}_t]$$

i.e the supremum of all discounted expected payoffs between t and T [11]. The optimal stopping time $\hat{\tau}$ is then defined as

$$\hat{\tau} = \arg \sup_{t \leq \tau \leq T} E[e^{-r(\tau - t)}\Phi(S_\tau)|\mathcal{F}_t]$$

Let V_t denote the value of the option at time t with payoff function Φ. If instead of immediate exercise of the option, the trader choses to wait until a future date $t_1 > t$, the value of that choice, at time t, will be

$$e^{-r(t_1 - t)}E[V_{t_1}|\mathcal{F}_t]$$

Thus from an arbitrage-free argument the option value at time t_1 must be equal to [11]

$$V_t = \max\{\Phi(S_t), e^{-r(t_1 - t)}E[V_{t_1}|\mathcal{F}_t]\}$$

So in order to price American options using numerical methods, computing conditional expectations such as $E[V_{t_1}|\mathcal{F}_t]$ is of great importance.

The subsequent lemma and theorem gives the link between approximations of conditional expectations and the *least square method* that is used in this paper to price American options. The proof of theorem 7.3 is taken from [15] where also a proof of lemma 7.2 can be found.

Lemma 7.2. *Let Y and X be random variables, then $E[Y|X] = h(X)$ for some Borel function h.*

Theorem 7.3. *Let Y and X be random variables with finite variance. If $\mathcal{H} = \{h; h$ is a Borel function with $Var[h(X)] < \infty\}$ then*

$$E[Y|X] = \mathbb{h}(X), \textit{ where } \mathbb{h} = \arg\min_{h \in \mathcal{H}} E[(Y - h(X))^2]$$

Proof. **(Theorem 7.3)** Expand $E[(Y - h(X))^2]$ as

$$\begin{aligned} E[(Y - h(X))^2] &= E[(Y - E[Y|X] + E[Y|X] - h(X))^2] \\ &= E[(Y - E[Y|X])^2] - 2E[(Y - E[Y|X])(h(X) - E[Y|X])] \\ &\quad + E[(h(X) - E[Y|X])^2] \end{aligned}$$

In the last equality the third expected value is minimized by taking $h(X) = E[Y|X]$ and the second expectation can be computed as follows.

From lemma 7.2 it follows that $E[Y|X]$ can be seen as a Borel function of X. Therefore

$$E[(Y - E[Y|X])(h(X) - E[Y|X])] = E[(Y - E[Y|X])g(X)]$$

for some Borel function g. The law of total expectation yields that

$$\begin{aligned} E[g(X)(Y - E[Y|X])] &= E[E[g(X)(Y - E[Y|X])|X]] \\ &= E[g(X)E[(Y - E[Y|X])|X]] \\ &= E[g(X)(E[Y|X] - E[Y|X])] = 0 \end{aligned}$$

Therefore $E[(Y - h(X))^2]$ is minimized by $h(X) = E[Y|X]$. □

According to lemma 7.2 a conditional expectation of a random variable can be seen as a function h, of the given information. Theorem 7.3 tells us that this function should be the one that minimizes the square distance between the random variable and the function of the information.

Hence $E[Y|X]$ can be viewed as the orthogonal projection of the space of $\{Y; Var(Y) < \infty\}$ onto the space $\{h(X); h \in \mathcal{H}\}$ since this will minimize $(Y - h(X))^2$ [15].

Below follows a description of the least least square regression method for pricing American options. This method was independently introduced by [17] and [18]. The method was further developed by [19].

For the purpose of numerical computations one must restrict the set $\{h(X); h \in \mathcal{H}\}$ to a smaller set of vector functions. One way to do that is to use linear combinations of monomials which are defined as

$$l_k(x) = x^{k-1}, \text{ for } k = 1, 2, ...$$

In order to price American options one first chooses the number of observation dates between today and maturity by partition $[0, T]$ as

$$0 = t_0 < t_1 < ... < t_M = T$$

These are the dates that the option can be exercised. Then simulate N independent trajectories of a stock that takes on values at each of the predetermined observation dates. This creates a stock matrix S that looks like

$$S = \{S_{t_i}^{(n)}\} \text{ where } 1 \le n \le N \ , \ 0 \le i \le M$$

Next define

$$Z = e^{-rt}\Phi(S_t) \ , \text{ where } \Phi(x) = \max\{K - x, 0\}$$

Z is the intrinsic value of the stock discounted to time 0.

Using monomials one sets up the following vector space

$$H_t^{(m,N)} = \text{Lin}\left\{ \begin{bmatrix} l_k(S_t^{(1)}) \\ ... \\ l_k(S_t^{(N)}) \end{bmatrix} \in \mathbb{R}^{N\times 1} \, : \, k = 1, ..., m \right\} \subset \mathbb{R}^{N\times 1}$$

This vector space is composed of all linear combinations of the vectors

$$\begin{bmatrix} l_k(S_t^{(1)}) \\ ... \\ l_k(S_t^{(N)}) \end{bmatrix} \in \mathbb{R}^{N\times 1} \text{ , where } k = 1, ..., m$$

One can make computations easier by defining

$$e(x) := [l_1(x), l_2(x), ..., l_m(x)] \in \mathbb{R}^{1\times m}$$

and construct the matrix

$$E_t = \begin{bmatrix} e(S_t^{(1)}) \\ ... \\ e(S_t^{(N)}) \end{bmatrix} \in \mathbb{R}^{N\times m}$$

and rewriting $H_t^{(m,N)}$ as

$$H_t^{(m,N)} = \{E_t\alpha : \alpha \in \mathbb{R}^{m\times 1}\} \subset \mathbb{R}^{N\times 1}$$

By projecting the payoff of the option at time T, which is known to be $V_T = Z_T$, onto the space $H_{t_{M-1}}^{(m,N)}$, one can derive the value of $V_{t_{M-1}}$ i.e the payoff at the observation date before T. This is done by using the standard inner product and the projection

$$P_t : \mathbb{R}^{N\times 1} \to \mathbb{R}^{N\times 1}$$

$$P_t(\nu) = E_t\beta$$

where $\nu \in \mathbb{R}^{N\times 1}$ and $\beta = (E_t^T E_t)^{-1} E_t^T \nu \in \mathbb{R}^{m\times 1}$.

For each i, where $1 \leq i \leq M$, starting from M, V_{t_M} is projected onto the space $H_{t_{M-1}}^{(m,N)}$. After that the projected value, $P_{t_{M-1}}(V_{t_M}) = E_{t_{M-1}}\beta$, must be compared with the discounted value of exercising the option at time t_{M-1} i.e $Z_{t_{M-1}}$. In other words, the value of the option at time t_{M-1} is estimated as

$$V_{t_{M-1}} = \max\{Z_{t_{M-1}}, E_{t_{M-1}}\beta\}$$

Next project this $V_{t_{M-1}}$ onto the space $H_{t_{M-2}}^{(m,N)}$ and so on until one reaches time t_0. To get the option value at time t_0 one takes the average of the option values at the first observation time after t_0 and compare it to the value of early exercising. Hence the option value will be the same for all trajectories.

Therefore, for $1 \leq n \leq N$

$$V_{t_0}^{(n)} = \max\{Z_{t_0}, \frac{1}{N}\sum_{n=1}^{N} V_{t_1}^{(n)}\}$$

8. Greeks

When analyzing options it is of great importance to comprehend how the price might be influenced by changes in the underlying market parameters. The parameters that an option depends on are the current stock price S_0, strike price K, volatility σ, time to maturity T and the risk-free rate r. Movements of these parameters will have various effects on the option price.

With the purpose of quantify these potential effects on the option V standing at time t where $0 \leq t \leq T$, a family of risk measures called *greeks* are introduced. The measures are called greeks because they are identified by greek letters. The most common of the greeks are

$$\Delta = \frac{\partial V}{\partial S}\ , \Gamma = \frac{\partial^2 V}{\partial S^2}\ , \mathcal{V} = \frac{\partial V}{\partial \sigma}(\text{vega})$$
$$\Theta = \frac{\partial V}{\partial t}\ , \rho = \frac{\partial V}{\partial r}$$

The impact from small changes in r is mostly negligible, therefore ρ is not that important to follow. However one can note that ρ is positive for call options and negative for puts.

This is because if there is a rate hike, all else being equal, a trader would be more inclined to have less equity in the market and more money in the bank. Since a call option gives the holder the benefits of going long a stock but using less of an investment, the trader can put more money in the bank. Hence ρ is positive for call options.

On the other hand, if a trader is long a put option he or she will receive a future cash-flow (potentially zero) buy selling the underlying asset for the price of K. After a rate hike, all else being equal, it is less beneficial to receive a payoff in the future since its net present value will be discounted even more.

The greek Θ is negative for both call and put options. When the time to maturity decreases, the probability that the option will be so called in the money (ITM) i.e $\Phi(S_t) > 0$, also decreases. The underlying stock needs time to move in a desirable range. Therefore after each passing day, all else being equal, the value of the option will decrease. This is called "time decay" and it is important to be aware of since it can erode the value of a position fairly quickly.

In this paper I will estimate Δ using finite difference quotients together with Monte Carlo methods as seen in [16].

Let's assume that one has a stochastic process V that depends on some parameter θ. Then let

$$\zeta(\theta) := E[V(\theta)]$$

The question now is how to simulate $\zeta'(\theta)$.

This can be done by simulating, for some small $h > 0$, $V_1(\theta - h), V_2(\theta - h), ..., V_n(\theta - h)$ and then simulate $V_1(\theta + h), V_2(\theta + h), ..., V_n(\theta + h)$. Then calculate

$$\bar{V}_n(\theta \pm h) := \frac{1}{n} \sum_{i=k}^{n} V_i(\theta \pm h)$$

and construct the so called central difference estimator

$$\zeta'(\theta)^* = \frac{\bar{V}_n(\theta + h) - \bar{V}_n(\theta - h)}{2h}$$

If $\zeta(\theta) \in \mathcal{C}^2$ i.e ζ is a continuous function with continuous first and second derivatives, in some neighborhood of θ, then by Taylor expansion around θ

$$\zeta(\theta \pm h) = \zeta(\theta) \pm \zeta'(\theta)h + \zeta''(\theta)\frac{h^2}{2} + \mathcal{O}(h^3)$$

Since obviously $E[\zeta'(\theta)^*] = \frac{\zeta(\theta+h)-\zeta(\theta-h)}{2h}$ the bias of the estimator $\zeta'(\theta)^*$ is

$$E[\zeta'(\theta)^* - \zeta'(\theta)] = \frac{\zeta(\theta + h) - \zeta(\theta - h)}{2h} - \zeta'(\theta) = \mathcal{O}(h^2).$$

For the greek Δ the parameter θ is the underlying stock price. For call options one has that $\Delta \in [0, 1]$ and for puts $\Delta \in [-1, 0]$.

9. SIMULATION

In this section I simulate H&M's share price using the standard Black Scholes model (with and without stochastic volatility) and the NIG process. The parameters of the NIG process has been adapted, using the maximum likelihood method, to H&M's historical price movements between 2000 and 2015. They are estimated to be

$$\alpha = 42.8475,\ \beta = 0.8055,\ \delta = 0.0144$$

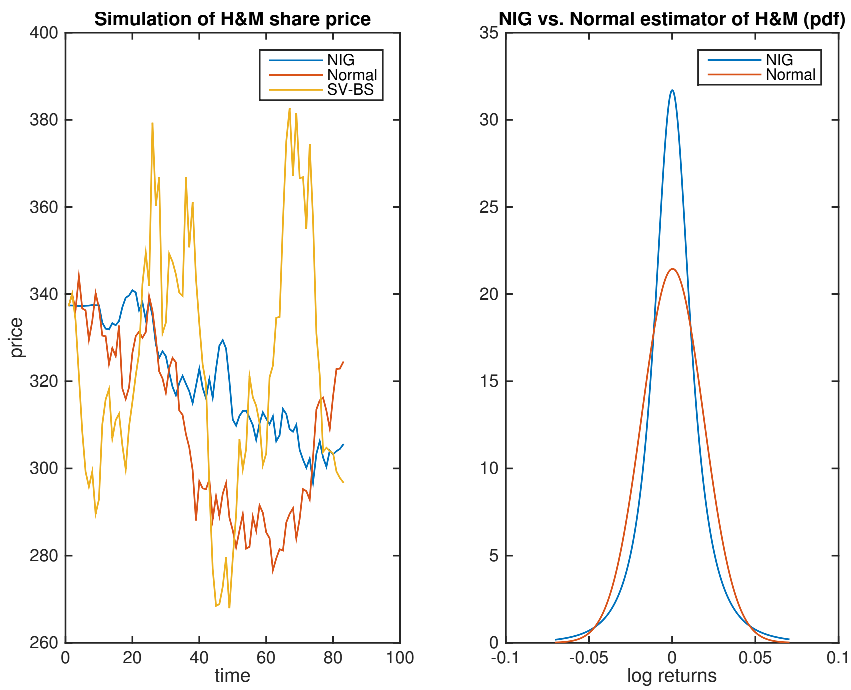
I simulated stochastic volatility as a CIR process. The parameters in the CIR process was chosen by me to be

$$a = 2,\ b = 0.3^2,\ c = 0.45$$

These values were assigned to the parameters of the CIR process because the resulting trajectory of the process looked fairly similar to the historical volatility. One needs to keep in mind that this model, with these parameters, does not use an EMM. Finding an EMM, when one has incorporated stochastic volatility, is beyond the scope of this paper. The correlation coefficient denoted by ρ is chosen to be 0.8.

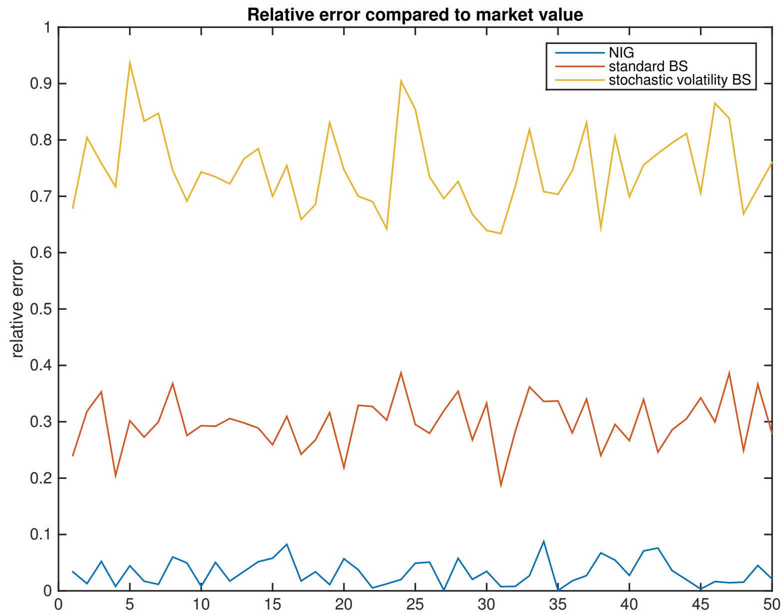
Below is a plot of simulations using the NIG process, standard Black Scholes and Black Scholes with stochastic volatility.

As expected, the trajectory with stochastic volatility is more volatile than the others. When comparing the standard Black Scholes model with the NIG process the NIG process evolves less rapidly. This can be explained by looking at the PDF's for the NIG distribution and the normal distribution. The NIG distribution has fatter tails than the normal distribution but is much more centered around zero. This means that the NIG process will experience smaller returns compared to the Black Scholes model.



In order to compare the different models, I estimate an American call option and its Δ. The option is written on H&M with 82 days to maturity, strike 310, risk-free rate 0.78% (the yield on the Swedish 10-year government bond). The stock was trading at 340.6 with dividend yield 2.9%. When using the standard Black Scholes model the (constant) volatility is estimated to be 29.56%.

This option had a bid price of 33.5 and an ask price of 36.5. The mean of these, which I use as the market price, equals 35. Below is a plot of the absolute relative error compared to the market price.



The absolute relative error, when using the Black Scholes model with stochastic volatility, is the highest and most unstable of the three models. This might be related to the choice of the underlying parameters in the CIR process. Since these were not selected by any rigorous statistical method, the result is a high absolute relative error.

The NIG process is here superior compared to the standard Black Scholes model in terms of stability and rate of absolute relative error. The absolute relative error of the Black Scholes model is between 20-40% while the absolute relative error, when using the NIG process, rarely moves above 10%.

The following table consists of 95% confidence intervals for each of the three models. In terms of confidence intervals, the NIG model surpasses the other models.

Table 1. Confidence intervals for the option price

NIG	Standard BS	BS with stochastic volatility
[34.7142 , 35.4947]	[45.0302 , 45.8941]	[60.4884 , 61.863]

One needs to keep in mind that these results are only based on one sample from the market. If one was to do a more rigorous analysis, one would need more samples of historical option prices from the market. It might also be interesting to compare the models when pricing options that are deep ITM and "out of the money" (OTM) i.e. in the case of a call option, the stock price is far below the strike price. Since OTM options carry more risk than an ITM option, there could be a significant difference when using the NIG model compared to the standard Black Scholes model, since the NIG model is better at capture market movements.

Below is a plot of Δ for each of the three models followed by a table with 95% confidence interval. The h that is used in the finite difference quotients is equal to 3.406.

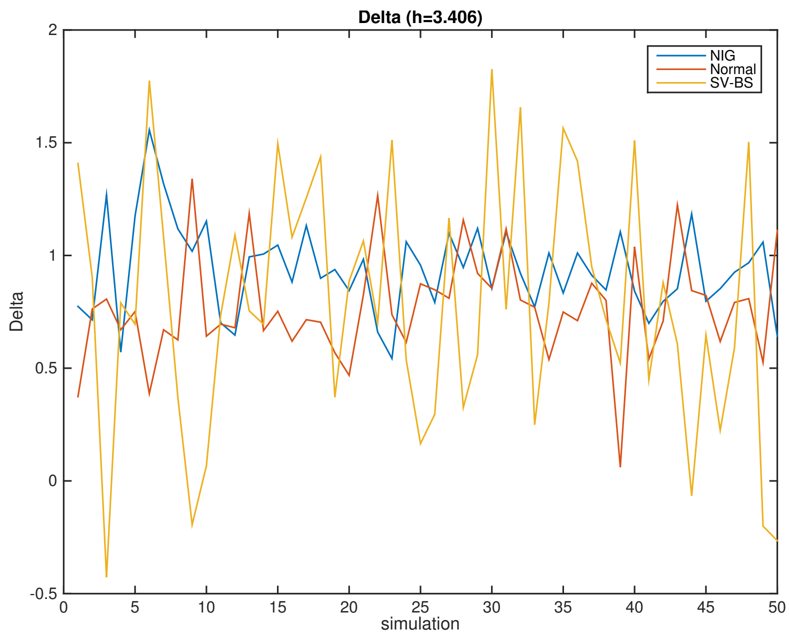


Table 2. Confidence intervals for Δ

NIG	Standard BS	BS with stochastic volatility
[0.88291 , 0.99311]	[0.70334 , 0.83484]	[0.62306 , 0.9341]

As one can observe, the results are highly unstable, even when h is not particularly small. Therefore the result in this comparison can be seen as inconclusive.

This instability probably stem from the fact that the underlying estimations of American options are themselves volatile. Hence, when calculating finite difference quotients, the result is unstable. From this one can conclude that a different and more rigorous discretization of Δ is preferable.

References

[1] SCHOUTENS, W. Lévy Processes in Finance: Pricing Financial Derivatives. Wiley (2003).

[2] CONT, R. Empirical Properties of Asset Returns: stylized facts and statistical issues. Quantitative Finance volume 1 (2001) 223-236.

[3] CONT, R. and PETER, T. Financial Modelling With Jump Processes. Chapman & Hall/CRC (2004).

[4] APPLEBAUM, D. Lévy Processes and Stochastic Calculus. School of Computing and Mathematics The Nottingham Trent University (2003).

[5] KOU, S. G. Lévy Processes in Asset Pricing. Columbia University.

[6] GEMAN, H. Pure Jump Lévy Processes for Asset Price Modelling. University Paris IX Dauphine and ESSEC. Published in the Journal of Banking and Finance, July 2002.

[7] TANKOV, P. Lévy Processes in Finance and Risk Management. World Congress on Computational Finance.

[8] YAO, L. YANG, G. YANG, X. A Note on the Mean Correcting Martingale Measure for Geometric Lévy Processes. Applied Mathematics Letters 24 (2011) 593-597.

[9] ANDERSON, C. E. Monte Carlo Methods and Importance Sampling Lecture Notes for Stat 578C Statistical Genetics.

[10] ALM, E. S. and BRITTON, T. Stokastik: sannolikhetsteori och statistikteori med tillämpningar. Liber (2008).

[11] BJÖRK, T. Arbitrage theory in continuous time. Third edition. Oxford university press. 2009.

[12] CROSBY, J. Introduction to jump and Lévy processes
Lecture notes from 2012 for M.Sc. course in Mathematical Finance at the Mathematical Institute, Oxford University.

[13] ANDERSEN, L. Efficient Simulation of the Heston Stochastic Volatility Model, 2006.

[14] SATO, K. Lévy Processes and Infinitely Divisible Distributions Cambridge University Press 1999.

[15] KLIMEK, M. Lecture notes from Monte Carlo Methods with Applications to Finance, 2014 Uppsala university.

[16] GLASSERMAN, P. Monte Carlo Methods in Financial Engineering Springer 2003.

[17] KLIMEK, M. AND PITERA, M. The least squares method for option pricing revisited, 2014.

[18] TSITSIKLIS, J. AND VAN ROY, B. Regression methods for pricing complex American style options, IEE Trans. on Neural Networks, 2001, 694-703.

[19] CLÉMENT, E., LAMBERTON, D. AND PROTTER, P.
An analysis of a least square regression method for American option pricing, Finance and Stochastics (2002), 449-471.