

Delta hedging using constant volatility, GARCH and neural nets to forecast volatility

Dennis Bron

Thesis BSc Informatics & Economics

Erasmus University Rotterdam

Supervisor: dr. ir. Jan van den Berg

Version: November 7, 2005

Abstract

In this thesis I will first explain the concept of delta hedging and the way delta is computed in the traditional way. Next I will zoom in on the volatility parameter within the traditional formula for computing the Black-Scholes option deltas and try to relax the assumption of constant volatility of the price of the underlying asset. I compare the constant volatility model with a GARCH(1,1) model and an Evolutionary Artificial Neural Network. The three models are tested on two artificial and two real world data sets. The conclusion of the thesis is that the amount of heteroskedasticity in real world stock index return is not high enough to give the more flexible models a competitive advantage when delta hedging a European call option.

Index

Index.....	2
1. Introduction.....	3
1.1 General background	3
1.2 Goal.....	3
1.3 Methodology.....	3
2. Derivative pricing and volatility estimation	5
2.1 The Black-Scholes Model	5
2.2 Brownian motion	5
2.3 Assumptions.....	6
2.4 Derivation of the Black-Scholes option pricing formula.....	6
2.5 Volatility estimation.....	9
2.6 Volatility smiles.....	10
2.7 Stochastic volatility.....	11
2.8 Long-term memory in stock prices.....	11
2.9 GARCH.....	12
3. Delta Hedging.....	14
3.1 Hedging.....	14
3.2 The position to hedge.....	14
3.3 Delta hedging.....	15
3.4 Delta hedging with stochastic volatility in discrete time.....	16
4. Evolutionary Artificial Neural Networks.....	18
4.1 Artificial Neural Networks	18
4.2 Evolution Strategy.....	19
4.3 EANNs.....	20
5. Experimental Setup	21
5.1 Choices and assumptions.....	21
5.2 The stock price data.....	21
5.3 Starting value.....	22
5.4 Variance	23
6. Results.....	24
6.1 Hedging costs.....	24
6.2 Risk reduction.....	24
7. Discussions and further research	25
7.1 Discussion of the results	25
7.2 Further research.....	26
References	27

1. Introduction

1.1 General background

This work is written as a concluding thesis in the Artificial Intelligence and Economics major of the bachelor Informatics & Economics at the Erasmus University in Rotterdam.

In the third year of this program the knowledge gathered in the courses of this major is brought into practice in a 14 week seminar in Computational Finance. The final assignment of the seminar was to relax one or more assumptions of the famous Black-Scholes model for pricing options. I have worked on this assignment together with Nees Jan van Eck, Sonja Tomas and Ludo Waltman.

My supervisor for this thesis was dr. ir. Jan van den Berg, who is an Associate Professor at the Department of Computer Science of the Faculty of Economics, also the teacher responsible for the seminar.

1.2 Goal

I try to relax the assumption of constant volatility in the Black-Scholes world while delta-hedging a European call option in order to make the model more flexible and more realistic. The estimation of tomorrow's volatility that is needed when computing the amount of shares Δ , needed for the hedge, is produced by three different models: a model of constant volatility, a GARCH model and an Evolutionary Artificial Neural Network with a structure similar to GARCH(1,1).

In this thesis I will try to find out which of these three methods for forecasting a stock's volatility works best when delta-hedging a short European call option.

1.3 Methodology

The three different methods of volatility estimation are tested on four different sets of time series data: one generated by a process with a constant volatility, one generated by a process that incorporates mean-reversion of the volatility according to the GARCH assumptions, and two series of returns from a major stock index in the real world..

The reason for using the generated data sets is to show that the GARCH model works best when the assumptions underlying this model hold and that the constant volatility models outperforms the rest if the volatility is indeed constant. Given these results and the results on the real world data, conclusions can be drawn about the characteristics of financial time series in the real world.

2. Derivative pricing and volatility estimation

2.1 The Black-Scholes Model

To price derivatives that rely on the value of an underlying asset, it is necessary to describe the process that the price of the underlying asset will follow in the future. The value depends heavily on the nature of the stochastic process followed by the asset price. An example of such a process is a geometric Brownian motion, on which the famous Black-Scholes option-pricing model is based. In this chapter the assumptions and derivation of the Black-Scholes model will be explained.

2.2 Brownian motion

In the Black-Scholes world, the stock prices are generated by a generalized Wiener process called a geometric Brownian motion (Hull (2003)) that describes the price change d_S in terms of a constant drift m_S of the stock, the standard deviation s_S of the stock, a period of time d_t , and a stochastic term g which is a drawing from a standard normal distribution.

$$d_S = m_S d_t + s_S g \sqrt{d_t} \quad (2.1)$$

This kind of process is often referred as a 'random walk' because it has no structure or statistical properties to give traders the opportunities of having an expected return other than m_S , or the risk free rate plus a risk premium that depends on s_S . This property of the Brownian motion is consistent with the Efficient Market Hypothesis (EMH), see Hull (2003).

Because of the assumed Brownian motion, the Black-Scholes option delta incorporates the assumption that the relative change in stock prices are independent drawings from a normal distribution with a constant variance. The first method we use for hedging our position is based on this assumption. We simply take the variance of all past changes to predict tomorrow's variance.

2.3 Assumptions

The following assumptions are made when deriving the Black-Scholes-Merton differential equation:

1. the stock price follows a geometric Brownian motion with μ and σ constant;
2. short selling of the asset is allowed;
3. there are no transaction costs or taxes;
4. the asset is perfectly divisible;
5. there are no dividends during the life of the derivative;
6. there are no riskless arbitrage opportunities;
7. asset trading is continuous;
8. the risk-free rate of interest, r_f , is constant and the same for all maturities.

Assumption five means that every riskless asset should earn the same return, namely the risk-free interest rate r_f . Assumption four can be relaxed, as will be discussed later on.

Using these assumptions it is now possible to derive the Black-Scholes model.

2.4 Derivation of the Black-Scholes option pricing formula

This section is based on Hull (2003). Consider the price of a derivative c which depends on the stock price S . Take for example a plain vanilla call option. Because in this case the parameters μ , σ , expiration date T and exercise price K are fixed, the change of c must be a function of S and time t . Using a result known as Ito's lemma (see Hull (2003) for more details) the following equation is obtained from the Brownian motion described in section 2.3.

$$dc = \left(\frac{\partial c}{\partial S} \mu_s + \frac{\partial c}{\partial t} + \frac{1}{2} \frac{\partial^2 c}{\partial S^2} \sigma_s^2 S^2 \right) dt + \frac{\partial c}{\partial S} \sigma_s S g \sqrt{t} \quad (2.2)$$

Because of the fact that the Wiener process underlying S and c in this equation is identical, the Wiener process can be cancelled out by choosing a portfolio of the stock and the derivative. This characteristic will later in this thesis be used for hedging a call option.

Define a portfolio of -1 derivative and $+\partial C/\partial S$ shares. The value Π of this portfolio is:

$$\Pi = -c + \frac{\partial c}{\partial S} S \quad (2.3)$$

The change in the value of this portfolio in the time interval dt is given by

$$d\Pi = -d_c + \frac{\partial c}{\partial S} dS \quad (2.3)$$

Substituting equations (2.1) and (2.2) into equation (2.4) gives:

$$d\Pi = \left(-\frac{\partial c}{\partial t} - \frac{\partial^2 c}{2\partial S^2} \mathbf{s}_S^2 S^2 \right) dt. \quad (2.5)$$

Because this equation does not involve g it is not stochastic anymore. This means that the portfolio must be riskless during the time dt and earn the risk-free interest rate, which follows from the assumption of no arbitrage possibilities. The change in value of the portfolio during dt is therefore

$$d\Pi = r_f \Pi dt, \quad (2.6)$$

where r_f is the risk-free interest rate. After substituting from equations (2.3) and (2.5), this equation becomes

$$\left(\frac{\partial c}{\partial t} + \frac{\partial^2 c}{2\partial S^2} \mathbf{s}_S^2 S^2 \right) dt = r_f \left(c - \frac{\partial c}{\partial S} S \right) dt. \quad (2.7)$$

Rearranging these terms leads to

$$\frac{\partial c}{\partial t} + r_f S \frac{\partial c}{\partial S} + \frac{\partial^2 c}{2\partial S^2} \mathbf{s}_S^2 S^2 = r_f c. \quad (2.8)$$

This is the Black-Scholes-Merton partial differential equation. It has many solutions, based on the type of derivatives and their boundary conditions, which specify the value of a derivative at the boundaries of S and t . In the case of a European call option, this condition is

$$f(S, T) = \max(S_T - K, 0) \quad \text{at time } T \quad (2.9)$$

where $f(S, T)$ is a function of the payoff depending on the stock price at time t . The Black-Scholes-Merton equation can be solved in various ways. Black and Scholes (1973) convert it to the heat-transfer equation for which a solution is given. An alternative method can be applied using Theorem 7.11 from Øksendahl (1985). In this case equation (1.8) can be solved by substituting $e^{-r_f(T-t)} f(S, t)$ for c , where $e^{-r_f(T-t)}$ is the discount factor. After simplification, equation (2.8) becomes

$$\frac{\partial f}{\partial t} + r_f S \frac{\partial f}{\partial S} + \frac{\partial^2 f}{2\partial S^2} \mathbf{s}_s^2 S^2 = 0. \quad (2.10)$$

This equation is equal to Kolmogorov's backward equation, which has the solution $f(S, t) = E[f(S, T)]$ based on the following process:

$$dS = r_f S dt + \mathbf{s}_s S g \sqrt{t}. \quad (2.11)$$

This means that under process (10) the present value of the derivative is the expected value at maturity discounted by the risk-free interest rate:

$$c = e^{-r_f(T-t)} E[f(S_T, T)], \quad (2.12)$$

where $f(S_T, T)$ is the payoff at the predefined boundary condition. This result can be used for all the types of options for which the process of the underlying can be modeled according to process (2.1).

An alternative method to obtain equations (2.9) and (2.11) applies the risk-neutral valuation argument to equation (2.8). The idea behind the risk-neutral valuation argument is based on the fact that none of the variables in this equation depend on the risk preference of investors. The reason is that the only factor involving the risk preference, namely \mathbf{m} is absent. This means that because the risk preference does not affect the solution of (2.8), any type of risk preference can be assumed. It can therefore be assumed that the investors are risk-neutral, which means that the expected return on any security is the risk-free interest rate. Thus, the risk-neutral process for S is equation (2.10). Furthermore, because it can be assumed that the world is risk-neutral, the present value at time t of a derivative can be calculated as the expected payoff at maturity in a risk-neutral world discounted at the risk-free interest rate, which leads to equation (2.11).

Using this result the famous Black-Scholes formula can be derived, see Hull (2003). For a European call option the formula is

$$S_0 N(d_1) - K e^{-r(T-t)} N(d_2), \quad (2.12)$$

with S_0 the initial stock price, $N(x)$ the cumulative standard normal density function and where

$$\begin{aligned} d_1 &= \frac{\ln(S_0 / K) + (r_f + \mathbf{s}_s^2 / 2)(T - t)}{\mathbf{s}_s \sqrt{(T - t)}} \\ d_2 &= \frac{\ln(S_0 / K) + (r_f - \mathbf{s}_s^2 / 2)(T - t)}{\mathbf{s}_s \sqrt{(T - t)}} = d_1 - \mathbf{s}_s \sqrt{(T - t)} \end{aligned} \quad (2.13)$$

This formula can be extended by including a continuous dividend yield q . In that case, the growth rate in equation (2.10) is set to $r_f - q$ instead of r_f .

2.5 Volatility estimation

As mentioned earlier, the nature of the process generating the asset returns is crucial in determining the Black-Scholes option value. As the volatility increases, the chances of the stock price rising and falling both increase equally, because of the symmetric nature of the standard deviation measure for volatility. This results in both call and put option prices rising. Similarly, a decreasing volatility results in falling option prices. Within the equation for the Brownian motion in equation (2.1), the asset's standard deviation \mathbf{s}_s is one of the parameters. In the Black-Scholes world \mathbf{s}_s is constant (homoskedastic) and can be computed by determining the standard deviation of a series of historic relative changes in the stock price. In the real world this assumption hardly holds. It is very likely that the volatility of a stock varies over time with period of high volatility and periods of relative calmth (heteroskedasticity). If today's volatility is correlated with historic price changes it is called autocorrelation. If we succeed in creating a model that makes a better estimation of tomorrow's volatility than the variance of a historic time series, we can improve the approximation of delta, and thus improve our hedge.

$$\text{Autocorrelation } \mathbf{r}(u_{t-1}, u_t) = \frac{\sum u_{t-1} u_t}{T - 1}$$

In recent years, the popularity of using intraday data has increased dramatically. If we look at five minute returns for instance, we can compute the daily, or 8 hour volatility using $\sigma_{daily} = \sigma_{5min} \sqrt{96}$, because there are 96 five minute returns in one day. Of course it is of great importance not to forget the overnight return and volatility. Using intraday data has shown to lead to very good volatility forecasts (Andersen and Bollerslev (1998)). If intraday data is available, volatility turns into an observable variable because it is possible to calculate realized volatilities. Without intraday data, volatilities can never be actually observed but only estimated using the average squared daily returns. Unfortunately historic intraday data is still scarce and not available in this case.

2.6 Volatility smiles

One of the reasons for investigating the impact of non-constant volatilities is the existence of volatility smiles. When Derman and Kani (1994) looked at implied volatilities, or the volatilities that are implied by the Black-Scholes formula They found that the implied volatilities vary with the strike price. This fact points out a major shortcoming of the Black-Scholes framework, namely that it's assumptions clearly don't hold in the real world.

The implied volatility also varies with the time to expiration, creating a three dimensional non-linear volatility surface. Incorporating these surfaces into the Black-Scholes framework gives more accurate option prices, see Derman et al. (1996). But using this, we have to estimate a volatility function before being able to compute the option price. Of course the data needed for estimating this function is not always available.

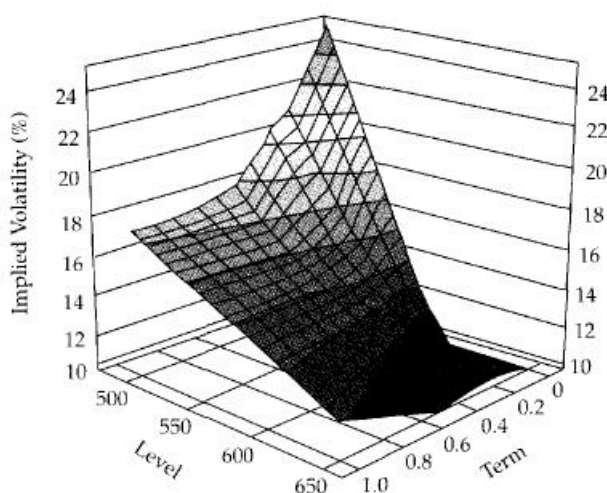


Figure 1 - A non-linear implied volatility surface (Derman and Kani)

In Vahamaa (2003), the author succeeds in significantly improving his delta hedging performance, in terms of Mean Absolute Hedging Error – see chapter 5, by using volatility smiles in the computation of delta. In a similar way, I hope to improve delta hedging performance of the Black-Scholes model by allowing non constant volatility forecasts.

2.7 Stochastic volatility

Hull and White (1987) argue that stock returns are generated by processes with stochastic volatilities, i.e. the volatilities themselves are random drawings from a distribution. If the volatility is stochastic and not correlated with the asset price, it is correct to use the average value of the volatility. If the volatility is stochastic and correlated with the asset price, it gets more complicated.

2.8 Long-term memory in stock prices

Among the first to suggest that there is a persistence of long-term memory of shocks in the volatility of asset was Mandelbrot (1971), who argues that this would lead to arbitrage opportunities. Since then many empirical studies have supported this claim. This had a lot of impact on the existing theories for common economic problems such as determining the optimum between consumption and saving, portfolio allocation and derivative pricing. Because the solutions for these problems had become extremely dependent of the time-horizon. In the long term assuming a random walk or martingale process without autocorrelation between asset returns is still a save assumption, but for short periods the existence of autocorrelation can have a significant impact.

The autocorrelation for time lag k is computed as follows:

$$r_k = \frac{\sum_{i=0}^{N-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=0}^N (y_i - \bar{y})^2}.$$

Where y_0, y_1, \dots, y_N is a time series of stock returns.

The following figure shows that the effects of a ‘shock’ in the volatility remain present in the time series of asset prices for a long time. (Remember section 2.8 Long term memory in stock prices)

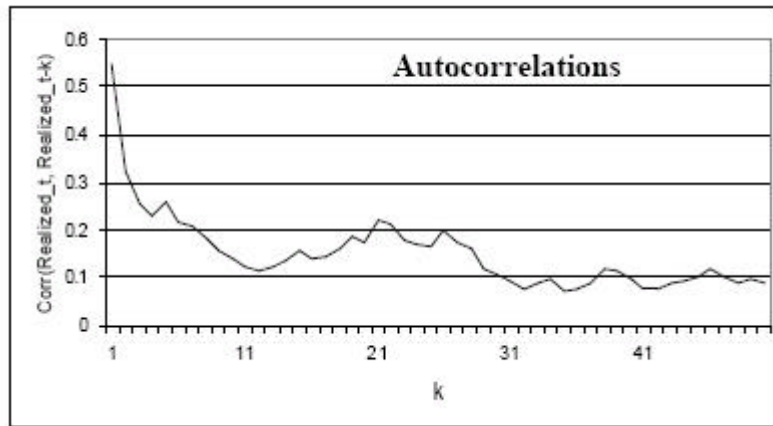


Figure 2 - Figure 2 - Autocorrelations for different time lags - Andersen and Bollerslev (1998)

2.9 GARCH

In 1986 Bollerslev introduced a model that deals with heteroskedasticity: Generalized AutoRegressive Conditional Heteroskedasticity. In this model the estimation of tomorrow's volatility is based on historic relative changes, with the most recent change having the most influence and the influence declining exponentially over time. The model is based on the assumption that a time series, that is in a state of unusually large or small volatility, will gradually return to its long-term variance (mean reversion).

Every estimation is made using the following formula:

$$s_t^2 = w + a u_{t-1}^2 + b s_{t-1}^2. \quad (2.14)$$

Where a , b and w are determined by maximizing the Maximum Likelihood Score (MLS).

$$MLS = \max P(D | h) = \Sigma(-\ln(v_i) - u_i^2/v_i) \quad (2.15)$$

The long-term variance is then given by:

$$V_L = \frac{w}{1-a-b}. \quad (2.16)$$

Maximizing the MLS is normally done by a non-linear solver program.

Since its introduction, GARCH models have become very popular and many empirical studies have showed the relevance of especially the GARCH(1,1) model, where tomorrow's volatility is forecasted using one historic return and one historic volatility forecast. Because each estimation of the volatility is based on b times the previous one, the influence of the historic returns declines exponentially, even using GARCH(1,1).

3. Delta Hedging

3.1 Hedging

Hedging is a way to control the risk in a company. If a company, like an airline, is vulnerable to changes in financial value, e.g. oil prices, it might want to control this risk by buying financial derivatives, e.g. crude oil futures. One of the reasons why companies attempt to hedge these risks is that they try to ensure their future cashflows in order to be sure that they can commit to their financial obligations. The shareholders might appreciate this effort, because it reduces the risk of bankruptcy.

Another reason for companies to hedge certain risks is that they might have shareholders that are not able to diversify their portfolio. In general shareholders are able to reduce the risk of investments by having multiple, less than perfectly correlated investments. See Markowitz' Modern Portfolio Theory (1952).

Companies also have a strong incentive not to hedge their risks, namely the cost of hedging. In the case of relatively small business-related risks and diversified shareholders, it will usually not be in the shareholder's interest to hedge the risk. In short: if the cost of hedging for the company is lower than for the shareholders, hedging increases the shareholders' value.

3.2 The position to hedge

In this thesis, the position I try to hedge is a short position in a European style call option contract. A European call option is the right to buy a stock at a given price: the strike price of the option. This right can only be exercised on its expiration date, this in contradiction to an American stock option, where the right to buy can be exercised at any given time before or on the expiration day.

The short position means that I 'write' the option contract and sell it to someone else. This means I sell the right to buy. If the buyer decides to exercise the option, I have to supply the share against the strike price. In practice, the option is usually closed with a cash-settlement. This means that I have to pay the difference between stock price and the strike price if the stock price is higher than the strike price.

3.3 Delta hedging

A common way to hedge a short position in a European call option contract is by keeping exactly Δ stocks, where $\Delta = \partial c / \partial S$, with c the price of the option, S the price of the stock and $\partial c / \partial S$ the partial derivative of c to S . This results in a portfolio $-c + \Delta S$. For this portfolio,

$$\frac{\partial \Pi}{\partial S} = -\frac{\partial c}{\partial S} + \Delta \frac{\partial S}{\partial S} = -\frac{\partial c}{\partial S} + \frac{\partial c}{\partial S} \frac{\partial S}{\partial S} = 0.$$

Because the price of the portfolio is insensitive to change of the stock price, as long as the amount of shares is adjusted to Δ , the portfolio is riskless. This is called delta hedging.

Δ is defined by a part of the Black-Scholes formula explained in chapter 2.

$$d_1 = \frac{\ln(S_0/K) + (r + \mathbf{s}^2/2)T}{\mathbf{s}\sqrt{T}}, \quad (3.1)$$

$$\Delta = N(d_1), \quad (3.2)$$

where S_0 is the stock price at $t = 0$, K is the strike price of the option, r is the risk-free interest rate, \mathbf{s} is the volatility of the stock and T the time to expiration. $N(x)$ is the cumulative normal distribution function.

If it would be possible to adjust the amount of stock continuously to the changes in the option delta, the hedge would be perfect, i.e. the risk would be reduced to zero without any costs. Because of the practical impossibility of continuous hedging and the presence of transaction costs, the hedge has to be adjusted with a certain time interval, e.g. daily.

Adjusting the portfolio with a certain delay always results in making hedging costs, with a bigger delay leading to more costs. This is because if the price of the stock has dropped, the more time goes by before one sells shares to adjust the portfolio, the less money one receives. If the price goes up, the opposite effect occurs and one will buy too expensive. If we take fixed transaction costs into account, an optimal adjustment interval can be determined, see Clewlow and Hodges (1997).

Because the price change during an interval is uncertain, the cost of adjusting the portfolio is uncertain too. This is a highly undesired effect because we started hedging

to loose the uncertainty in the first place. The uncertainty in the hedging costs, or hedging risk is defined as the standard deviation of the hedging costs. In this thesis, I will measure hedge performance by the costs of the hedge and the uncertainty – the variance - in the costs. To clarify the differences in performance between the different models, I compute the level of risk reduction, or the hedging risk as a percentage of the asset risk.

The fact that discontinuous delta hedging leads to a remaining hedging risk can also be explained in a more mathematical way. Keeping Δ stocks and a short derivative of this stock results in a portfolio that is delta-neutral. Because delta is the first derivative of the option value to the stock price, keeping delta stocks only results in a riskless portfolio when the first-order derivative is linear. Or in other words, if the second derivative is constant. Because in reality this is never the case, delta-hedging in discrete time never results in a complete disappearance of risk.

3.4 Delta hedging with stochastic volatility in discrete time

The only paper on delta hedging with the constant volatility assumption relaxed I have been able to find is Geyer and Schwaiger (). Strangely enough, they don't use the stochastic volatility model proposed by Hull and White (1987), but they try to incorporate GARCH volatility forecasts into delta hedging.

The paper starts with explaining that the Black-Scholes model assumes a geometric Brownian motion with constant diffusion – or variance – and states that it's purpose is to investigate the consequence of relaxing the assumption of constant diffusion.

The paper describes three ways of adopting the computed GARCH volatilities into the calculation of the option delta:

1. assuming a GARCH process leading to an estimation of tomorrow's volatility and assuming that constant for the rest of the life of the option and compute the delta using the normal Black-Scholes formula;
2. compute an aggregate volatility based on the GARCH process using it's last prediction of σ_t^2 as σ_{t-1}^2 and the expected value of u_{t-1}^2 , namely 0, and using this in the Black-Scholes formula;
3. assuming the GARCH process to continue for the rest of the option's life and find a numerical way to approximate the option delta;

In the second case, the aggregate variance for the rest of the option's life is computed using:

$$\sum_{k=t+1}^T E(\mathbf{s}_k^2), \quad (3.3)$$

where \mathbf{s}_k^2 is the volatility estimation for day k . This method is used in this thesis, where the aggregate volatility will be something between the forecast for tomorrow and the long-term variance V_L .

The third case leads to a different option delta under the GARCH assumptions:

$$\Delta_t^G = e^{-(T-t)r_f} E\left(\frac{S_T^*}{S_t^*} f(S_T^*, K)\right), \quad (3.4)$$

where S_T^* is the value of the stock at expiration time T , S_t^* is the current value, $f(S_T^*, K)$ is a function that is 1 when the option is exercised and 0 if not, and K is the strike price of the option. For a European call option, $f(S_T^*, K)$ is:

$$f(S_T^*, K) = \begin{cases} 1 & , S_T^* \geq K \\ 0 & , S_T^* < K \end{cases}. \quad (3.5)$$

Because the GARCH option delta Δ_t^G depends on the stochastic evolution of the transformed price process S^* , it is impossible to derive a closed form solution. In Geyer and Schwaiger () a numerical way to approximate the delta is explained. Unfortunately the authors only test the three different methods on a generated GARCH data set and not on real world data.

4. Evolutionary Artificial Neural Networks

4.1 Artificial Neural Networks

Artificial Neural Networks (ANN's) or simply neural networks are networks existing of a number of layers of interconnected simple logic units or nodes, see Jang et al. (1997). These networks have been invented in the 1950s and were inspired by the way scientists believed the human brain worked. The use of ANN's however, was limited strongly by the lack of suitable training methods. This changed in the mid-1980s with the reformulation of the back propagation algorithm by Rumelhart et al. (1986) The logical units in feedforward neural networks - as opposed to recurrent ones - are called perceptrons. These perceptrons model a human brain's neuron that 'fires' on the output side when a certain threshold is reached. In perceptrons the input x is a weighted linear combination of the outputs of perceptrons in the previous layer and a so called 'bias' (always equal to 1). The output is computed by using a nonlinear, differentiable activation function called a 'transfer function' or the identity function $f(x) = x$. The following activation functions are most commonly used.

Logistic function:

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

Hyperbolic tangent function:

$$f(x) = \tanh\left(\frac{x}{2}\right) = \frac{1 - e^{-x}}{1 + e^{-x}}$$

An ANN is normally trained on a dataset with an algorithm such as back propagation to adjust the weights of the perceptrons. In the back propagation algorithm the error, or the desired value minus the model output, is propagated through the nodes towards the input side while adjusting the weights so that the error decreases. ANNs are frequently regarded as black-box models or non-parametric models.

Because the real variance on a specific day remains unknown, there is no way to determine the networks error when tomorrow's variance is forecasted. The absence of direct error feedback is the reason that traditional training algorithms for neural networks cannot be used.

We can however choose an neural net that maximizes the probability that the data available is generated by the given neural network, like in 3.3. This is done by using the Maximum Likelihood Criterion. Because there is no simple mathematical dependency between the Maximum Likelihood Score (MLS) and the network's weights, the search for the neural net with Maximum Likelihood has to be done randomly. A structured way to do a random search is by using an Evolution Strategy.

4.2 Evolution Strategy

Evolution Strategies are based on the biological concept of evolution. A population of a certain size is simulated during a certain number of generations. Parents create offspring that is subject to mutations. If the offspring is generated by multiple parents it is called crossover. The concept of mutation is vital for an Evolution Strategy to prevent it from getting trapped in a local minimum or maximum.

To decrease the tendency to get trapped in a local minimum or maximum further, we use games to determine which members of the population survive. Every member plays against a given number of opponents and receives a point for every game that is won.

To measure how the EANN performs, I use the same Maximum Likelihood Score as with the GARCH model.

An Evolution Strategy can also be useful for maximizing the MLS when fitting a GARCH model. In an experiment fitting a GARCH model with an Evolution Strategy easily outperformed fitting it with Microsoft Excel's Solver. In order to prevent the EANN having a positive bias caused by the better optimizer, I also use an Evolution Strategy to fit the GARCH model.

One of the problems when using evolutionary methods for fitting models is choosing the appropriate standard deviation for mutation. Mutation works as follows:

$$x = x + N(0, \mathbf{s})$$

with $N(\mathbf{m}, \mathbf{s})$ a drawing from the standard normal distribution with mean \mathbf{m} and standard deviation \mathbf{s} . When \mathbf{s} is too small, it takes too long before the optimum is reached, when \mathbf{s} is too large the mutations will be too large to find the optimum. Schwefel (1974) has solved this problem by introducing self-adaptation, where \mathbf{s} is

also subject to mutation. See also Beyer (). In this way, the s will gradually decrease as the population reaches the optimum.

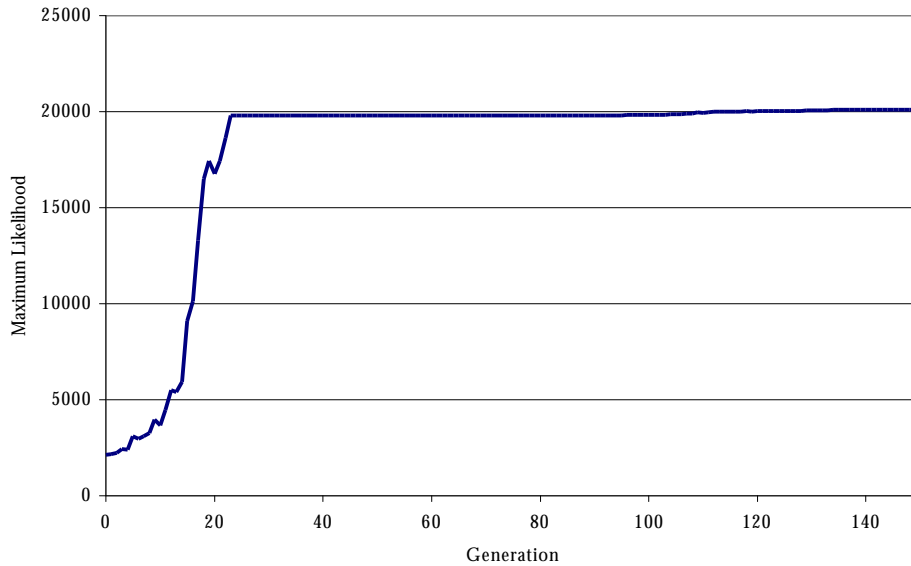


Figure 3 - The evolution of an EANN in 150 generations

4.3 EANNs

As there are three design issues in creating an ANN - the weights, the structure, and the training rules - it is possible to apply evolution to to ANNs on three different levels, according to Yao and Liu (1998).

Because I have a good estimation of the complexity of the model needed to produce a reasonable estimation of tomorrow's volatility based on the input parameters, I determine the structure of the ANN upfront and only let the weights evolve.

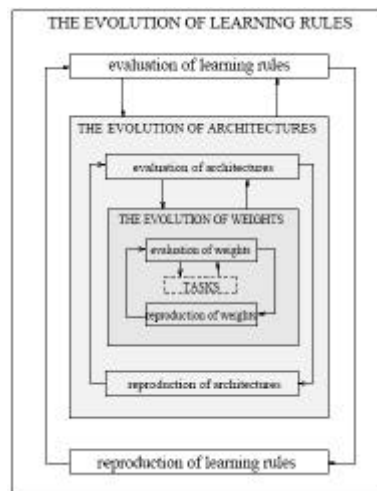


Figure 1-Three level of evolution in creating an EANN - Yao and Liu (1998)

5. Experimental Setup

5.1 Choices and assumptions

The European call option I short sell has one underlying stock. The option is at-the-money at the start of the experiment, meaning the strike price equals the current stock price. The option expires at the end of the experiment.

Furthermore I assume there is a constant risk-free interest rate of 4%. This is a realistic number and will not influence the results. The stock I use for hedging is perfectly divisible. Because I use the same updating frequency (daily) for each of the models, transaction costs will be roughly the same and are not taken into account.

Because neural networks results in a complex, non-linear, continuous function, there is no way to exclude the possibility of forecasting negative variances. Of course this is a mathematical impossibility. After replacing the logistic function in the neural net's perceptrons with the tanh function the problems we encountered due to negative variances disappeared.

I used an Evolutionary Artificial Neural Network (EANN) with two input nodes u_{t-1}^2 and s_{t-1}^2 and one hidden layer with two nodes and one output node: s_t^2 . The main reason I chose for this network architecture lies in the limited availability of training data. After a few experiments with more hidden units and more lagged inputs, this architecture lead to the best results.

The strategy used for fitting the GARCH model involves a population size of 50, 350 generations and 10 games per individual. The strategy for the EANN also involves a population size of 50 and 10 games per individual, but needs 500 generations.

5.2 The stock price data

I use four datasets for my experiments:

1. Random walk: a generated 10-year series of stock prices with a constant variance;
2. GARCH data: a generated 10-year series with mean reversion of the volatility;
3. AEX: a 10-year series of the Dutch AEX-index;
4. S&P 500: a 5-year series of the Standard and Poor's 500 index.

The reasons I chose stock index data in stead of single stock data are:

1. most similar research is done using stock index data and therefore the result will be better comparable;
2. it is more realistic because most investors will hold a diversified portfolio, similar to a major index;
3. the index pays no cash of stock dividends, has no stock splits or secondary offerings.

Clearly, I expect the constant variance method to perform best on the first dataset and the GARCH model to perform best on the GARCH dataset. If the assumption of constant volatility does not hold in the real world, the GARCH model and EANN will outperform the constant variance model on real world data. Because the EANN is a more flexible model than GARCH, it is possible that it will perform better. However, the risk of overfitting is larger for the EANN than for the GARCH model.

I repeatedly use four months of data for fitting the model and then one month for hedging. The first four months of data are solely used for training and every following month is used for hedging, giving us 116 experiments in ten years (120 months).

5.3 Starting value

Because the GARCH model and the EANN use today's prediction for predicting tomorrow's variance, they need a value to start with.

$$\mathbf{s}_i^2 = f(\mathbf{s}_{i-1}^2)$$

For \mathbf{s}_{-1} the last instance of the training set is used: u_{-1}

$$u_{-1} = \frac{(S_{-1} - S_{-2})}{S_{-2}}$$

When S_{-1} equals S_{-2} this leads to problems because u_{-1} then equals zero. The exact value of u_{-1} has very little influence, but the total absence of variance leads to strange results. Because the constant volatility model doesn't need a starting value, it does not have this problem and outperforms the other models.

The solution for this problem is simple: use the average of the last three squared relative changes of the training set.

$$s_{-1}^2 = \frac{(u_{-1}^2 + u_{-2}^2 + u_{-3}^2)}{3}$$

5.4 Variance

Black and Scholes' d_1 use the variance for the rest of the life of the option, while the models give a daily variance. There is a simple way to compute the variance for the remaining option life (Hull (2003)):

$$s_{life} = s_{daily} \sqrt{T}$$

with T the rest of the life of the option in days.

5.5 Performance

In literature, there are a few ways to compute hedge performance. The different measures can roughly be divided into two categories:

1. measures based on the mean hedging error for each interval, or total resulting hedging costs;
2. measures based on the remaining risk, or variance in the total resulting hedging costs.

I would argue that, since hedging is always done for reasons of risk aversion, it is not fair to measure hedge performance solely in terms of average hedging costs for a certain period, as for instance in Vahamaa (2003), and thus completely ignoring a new source of risk that lies in the variance of the hedging costs. Therefore I will show both the average hedging costs and its standard deviation in the following chapter.

Because a simple standard deviation is only meaningful when comparing the results of different methods, I have added a measure called risk reduction. This measure shows the size of the standard deviation relative to the risk of the underlying asset.

$$risk\ reduction = \left(1 - \frac{s_{hedging\ costs}}{s_{underlying\ asset}} \right) \cdot 100\%$$

6. Results

6.1 Hedging costs

The three volatility models are compared on the way they fit on the set of training data (MLS) and on out-of-sample hedge performance. Hedge performance is measured in the average costs of the hedge and in the standard deviation of the hedging costs.

	Generated data sets		Real world	
	Constant volatility	GARCH	S&P 500	AEX
Constant volatility model				
Average MLS	655.44	533.62	228.23	672.17
Average hedging costs	1.25	8.73	7.00	9.56
Standard deviation of the costs	0.42	7.32	4.37	6.75
GARCH(1,1)				
Average MLS	656.42	554.93	229.60	6.76.65
Average hedging costs	1.26	8.17	7.00	9.45
Standard deviation of the costs	0.65	5.32	4.68	6.87
EANN				
Average MLS	658.57	555.99	232.71	677.15
Average hedging costs	1.27	8.15	6.96	9.29
Standard deviation of the costs	0.71	5.15	4.55	6.85

Table 1 - Comparing the hedging costs

6.2 Risk reduction

A way to clarify the relative results of the models is by computing the level of risk reduction, that is the percentage difference between the risk of the underlying asset and the resulting standard deviation of the hedging costs.

	Generated data sets		Real world	
	Constant volatility	GARCH	S&P 500	AEX
Constant volatility model	86%	69%	83%	55%
GARCH(1,1)	78%	77%	81%	54%
EANN	76%	78%	82%	54%

Table 2 - Risk reduction levels

7. Discussions and further research

7.1 Discussion of the results

When we look at the hedging costs, it is obvious that there are no significant differences between the models on any of the data sets using a student t-test, because the largest difference is still less than one-eighth of the concerning standard deviation and the number of observations is 116.

Looking at the remaining or hedging risk, the expectations stated in section 1.3 are clearly matched by the results above. On the random walk data set, the performance of the constant volatility model is better than the performance of the other models. The difference in performance is significant at a 99% confidence level, testing for the equality of standard deviations. The other models have a better fit on the training data – they have a higher MLS – but every outlier suggesting non-constant volatility they react on is nothing more than coincidence and thus can by definition never lead to a better result on the test set. This is a clear example of the impact overfitting a model on a training data set. The EANN is a more flexible model than GARCH and overfits more, resulting in a high standard deviation of the hedging costs

On the GARCH data set, the constant volatility model leads to a much higher hedging risk (standard deviation of the costs), significant at a 95% confidence level. The sample average hedging costs are also higher than for GARCH and EANN, but this difference is by far not as significant as the difference in hedging risk. Strangely enough the EANN outperforms the GARCH model. However, this is clearly not statistically significant.

Considering external shocks and their impact on short-term volatility as mentioned in section 2.8, it might be that the period the impact of the shock is present in the stock returns is too long or that the impact of the shocks is too small for the non-constant volatility models too have a significant advantage over the random walk model.

In the end, using non-constant volatilities for delta-hedging does not lead to a better hedge performance. The amount of heteroskedasticity in real world stock index data is not high enough to give the GARCH or EANN model a competitive advantage over the simple constant volatility model.

7.2 Further research

In reaction to my conclusions about the amount of heteroskedasticity, it can be argued that this could be higher when looking at a single stock in stead of index data. If this is the case, it could be that hedging a single stock with non-constant volatilities proves to be better. Furthermore it would be interesting to test the numerical method for determining the option delta in a GARCH world referred to in section 3.4 on real world data.

References

- Andersen, T.G., Bollerslev, T. and Lange, S. (1998), Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*
- Beyer, H.-G. (1995). Towards a Theory of Evolution: Self-Adaptation. *Evolutionary Computation*, (<http://ls11-www.informatik.uni-dortmund.de>).
- Black, F. and Scholes, M.S. (1973). The Pricing of options and corporate liabilities. *Journal of Political Economy*, 7, 637-654.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, Vol. 31, 1986, pp. 307-327.
- Chou, R.Y. (1988). Volatility persistence and stock valuations: some empirical evidence using GARCH. *Journal of Applied Econometrics*, Vol. 3 No. 4 (Oct.-Dec. 1988) 279-294.
- Clewlow, L. and Hodges, S. (1997). Optimal Delta-Hedging under Transaction Costs. *Journal of Economic Dynamics and Control*, (<http://www.ingentaconnect.com>).
- Derman, E. and Kani, I. (1994). Riding on a smile. *Risk*, vol. 7 no. 2 (February):32-39.
- Derman, E., Kani, I. and Zhou, J.Z. (1996). The local volatility surface: unlocking the information in index option prices. *Financial Analysts Journal*. July/August 1996
- Franses, Ph. H. (1998). *Time Series Models for Business and Economic Forecasting*. Cambridge University Press.
- Geyer, A.L.J. and Schwaiger, W.S.A., Delta hedging with stochastic volatility in discrete time. (<http://www.wu-wien.ac.at/wwwu/institute/or/geyer/pdf/deltaFMPFM.pdf>).
- Hull, J.C. (2003) *Options, Futures and other Derivatives*. Prentice Hall.
- Hull, J.C. and White, A. (1987). The pricing of options on asset with stochastic volatilities. *Journal of Finance*, 42 (June 1987), 281-300

Jang, J.-S.R, Sun, C.-T., and E. Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing*. Prentice Hall.

Low, A.E. (1989). Long-term memory in stock market prices.

Mandelbrot, B. (1971). When can price be arbitrated efficiently? A limit to the validity of random walk and Martingale models. *Review of Economics and Statistics*, 53, 225-236.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, .

Øksendal, B. (1985). *Stochastic Differential Equations*, Springer, Berlin.

Schwefel, H.-P. (1974). Adaptive Mechanismen in der Biologischen Evolution und ihr Einfluss auf der Evolutionsgeschwindigkeit. *Technical Report. Technical University of Berlin. Abschlussbericht zum DFG-Vorhaben Re 215/2.*

Vahamaa, S. (2003) Delta Hedging with the Smile. *Proceedings of the 2003 Multinational Finance Society*. (<http://mfs.rutgers.edu/conferences/10/mfcindex/files/MFC-099%20Vahamaa.pdf>).

Yao, X. and Liu, Y. (1998). Towards designing artificial neural networks by evolution. *Applied Mathematics and Computation*. Elsevier Science Inc.