

Valuation and Hedging of Correlation Swaps

Mats Draijer
4298829

September 27, 2017

Abstract

The aim of this thesis is to provide a formula for the value of a correlation swap. To get to this formula, a model from an article by Bossu is inspected and its resulting expression for fair the fair value of a correlation swap is simulated. The Jacobi process will be defined and two discretization schemes will be compared for it. Methods are discussed to make simulations of the Jacobi process as accurate as possible, making sure it crosses its boundaries -1 and 1 as little as possible. It will be shown that a correlation swap can be hedged by dynamically trading variance dispersion trades. The main result of the thesis is a partial differential equation for the fair value of a correlation swap. It will be shown that the expression for the value of a correlation swap obtained by Bossu's model satisfies this partial differential equation.

Contents

1	Introduction	3
2	Financial background	4
2.1	What are correlation swaps?	4
2.1.1	Basic financial concepts	4
2.1.2	Volatility	5
2.1.3	Correlation	6
2.1.4	Swaps	7
2.2	Why correlation swaps?	8
2.2.1	Hedging	8
3	Bossu's toy model	10
3.1	Statistical definitions used	10
3.1.1	Standard Brownian motion/Wiener process	10
3.1.2	Adapted process to a filtration	10
3.2	Definitions in Bossu (2007)	11
3.2.1	Realised volatility and correlation	11
3.2.2	Implied volatility and correlation	13
3.2.3	Proxy formulas	13
3.3	Comparison of the definitions in [1] to the regular definitions	14
3.3.1	Using the definitions of variance and correlation	15
3.3.2	Justifying the formulas in [1]	15
3.3.3	Implicit assumptions	17
3.3.4	The case $\rho = 0$	18
3.4	The toy model	19
3.4.1	One factor: variance swaps	19
3.4.2	Two factors: correlation swaps	21
3.5	Without the assumption $\sigma^I \leq \bar{\sigma}^S$	26
4	The Jacobi Process	30
4.1	What is the Jacobi process?	30
4.2	Discretization schemes	31
4.2.1	Euler discretization	32
4.2.2	Milstein discretization	32

4.2.3	Comparison of the discretization schemes	33
4.3	Exceeding the boundaries	35
4.4	Correlation swaps	38
5	Hedging	40
5.1	Introduction	40
5.2	The method of variance dispersion trades	40
5.2.1	What are variance dispersion trades?	40
5.2.2	The hedging property	41
5.3	Fair strike of a correlation swap	44
5.3.1	The vega-neutral portfolio	45
5.3.2	Comparison to the solution in [1]	47
5.3.3	From no-arbitrage price to fair strike	48
6	Conclusion	50
6.1	Further research	51

Chapter 1

Introduction

In recent years, volatility has become an important factor on the financial market. It is a traded asset with great liquidity and further derivatives on it are being widely traded as well. Volatility swaps and variance swaps are popular financial products, but they bring with them an exposure to correlation. To hedge that exposure, correlation swaps can be used.

In this thesis, a model by Bossu [1] for correlation swaps will be inspected. All its definitions and assumptions are being explored, resulting in a formula for the fair value of a correlation swap. This result will be simulated and these simulations analyzed. Its drawbacks will be pointed out and explained.

In addition, the Jacobi process will be defined. This is a bounded process with boundaries -1 and 1 and it is easy to simulate numerically. This makes it a fitting process for simulating correlation paths. The Euler discretization scheme and the Milstein discretization scheme will be compared for the Jacobi process. Next, we will look at why the simulations cross the boundaries -1 and 1. We will look at several possible options to reduce the amount of paths within a simulation that exceed the boundaries. The effect of these options will be observed and explained.

Furthermore, the hedging of the correlation swaps is looked into. We define variance dispersion trades and look at how they can be used to hedge correlation swaps. A hedged portfolio consisting of correlation swaps and variance dispersion trades is then inspected to arrive at a partial differential equation for the fair value of a correlation swap. Finally, this partial differential equation is used to confirm that the formula given by [1] is the no-arbitrage price of a correlation swap.

Chapter 2

Financial background

2.1 What are correlation swaps?

2.1.1 Basic financial concepts

In finance, the an important aspect we deal with is a financial **market**. Mathematically, a market is represented as a probability space. In this probability space, we can use random processes to represent **stocks**. Stocks are basically small pieces of ownership of a company, but they are often bought and sold only for trading, rather than the buyer actually being interested in the company's policy. Sometimes, stocks are put together in a **stock index**. If you invest in a stock index, you immediately invest in a lot of different stocks, and usually their share in the index is weighted. The constituent stocks and the weights are chosen and frequently changed by the index managers. A stock index can be modeled by taking the appropriate linear combination of the stocks in the index, the **constituent stocks**.

On the financial market, not only stocks are traded, but also many other products that have to do with stocks. Traded products that are derived from stocks are called **(financial) derivatives** or simply **financial products**. In this thesis, the term 'financial products' will be used to prevent confusion with mathematical derivatives of a function. An example of a financial product is a **future contract**. This is a deal between two parties saying that one party will buy a certain amount of a certain stock from the other party at a predetermined date (**expiry date**) and against a predetermined price (**strike price**). If the market stock price at the expiry date is higher than the strike price, the buying side of the future contract makes a profit, because he can get the stock cheaper than usual. If this is more likely to happen than the stock price being lower than the strike price, the seller is likely to lose. In this case, the seller may demand a payment from the buyer when entering the contract, as compensation for the expected loss. This payment is called the **price** of the contract. These definitions extend to many other types of financial derivatives, including swaps,

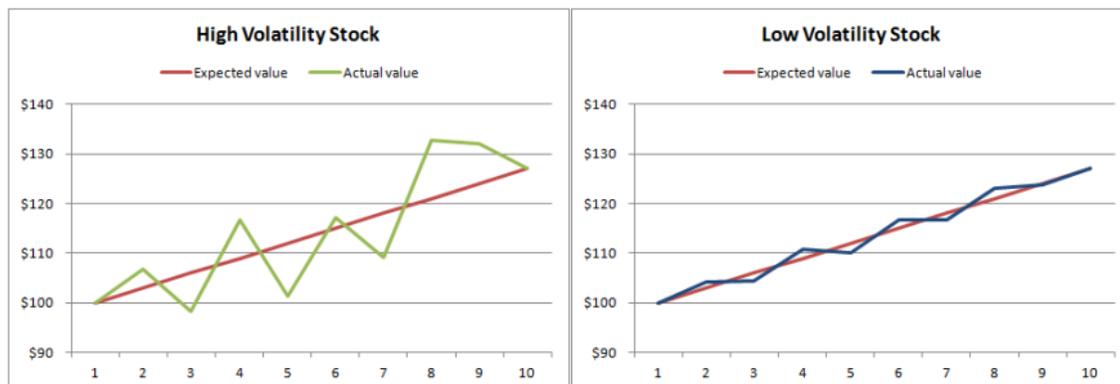


Figure 2.1: High volatility on the left versus low volatility on the right. The stock with high volatility fluctuates around its expected value much more than the stock with low volatility. [5]

which will be explained in section 2.1.4.

Correlation swaps have several things in common with future contracts, but there are some large differences as well. To see this, we need several more definitions.

2.1.2 Volatility

The **volatility** of a stock price is, simply put, the square root of its variance, so it is mathematically equivalent to the standard deviation. But since we are dealing with stock prices, and not just any random variable, there is a bit more to it than that. As can be seen in figure 2.1, the volatility describes how much the stock price wiggles around its expected value, not to be confused with its total difference from the previous value or previous average. This is in line with our understanding of variance and standard deviation: the higher the variance, the more observations will differ from expectations. This is the same with volatility.

In mathematical terms, the volatility of a random variable X is defined as the square root of the variance of X ;

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{\mathbb{E}(X^2) - (\mathbb{E}(X))^2}.$$

This is the exact definition of volatility, and here we view a stock price as ‘just’ a random variable. In reality, stock prices are modeled in various ways, and the volatility may be approximated with different definitions to go along with that to make calculations simpler. This will be evident when the toy model from [1] is discussed.

It is important to note the difference between **implied** and **realized volatility**. Volatility is a variable that will change over time. This is partly because within the company selling the stocks, there are periods of stability (low volatility, stock price moves very little), and periods of uncertainty (high volatility). This means future volatility is uncertain, and needs to be predicted. If a model yields such a prediction for the future volatility, this is called **implied volatility**. The **realized volatility**, however, represents the exact value the volatility will have in the future. Implied volatility is often taken as the expected value of the realized volatility, using a distribution or path and a probability space given by the model. The use of the two different terms implied and realized volatility helps to distinguish when predictions for future volatility are discussed (implied volatility), or when there is talk of a future situation in which the volatility at that point in time will be known (realized volatility).

2.1.3 Correlation

Volatility is an essential variable when regarding stock prices, but it is not enough when several stock prices or a stock index are involved. Different stock prices may not move completely independently from each other, e.g. when one bank collapses, other banks may start doing bad as well. This means the stock prices of those banks are **correlated**/there is a **correlation**. Two random variables may be represented by a **correlation coefficient**, a number to describe how much influence one random variable has on another one. It is similar to the statistical idea of covariance, but the correlation coefficient is scaled to be between -1 and 1. The correlation coefficient for two random variables is defined if and only if their covariance is defined.

Since the economic situation is very important for the stock market, it is logical that every stock is influenced by a good or bad economy. This gives every stock a basic (but very low) level of correlation with the other stocks, which is further increased if the companies are in the same sector, e.g. technology or banking. Many stock indices pick their constituent stocks from the same sector, and in addition, there is a logical correlation between the index and the constituent stocks. All of this means that correlation is an important concept in the world of finance.

The mathematical definition of the correlation coefficient $\rho_{X,Y}$ between two stocks X and Y is:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sqrt{[\mathbb{E}(X^2) - (\mathbb{E}(X))^2][\mathbb{E}(Y^2) - (\mathbb{E}(Y))^2]}}$$

where $\text{cov}(X,Y)$ is the covariance between X and Y, and σ_X and σ_Y are the volatilities of X and Y respectively. As with volatility, different definitions of correlation or correlation coefficients may be used to simplify models. Several different examples of correlation are shown in figure 2.2.

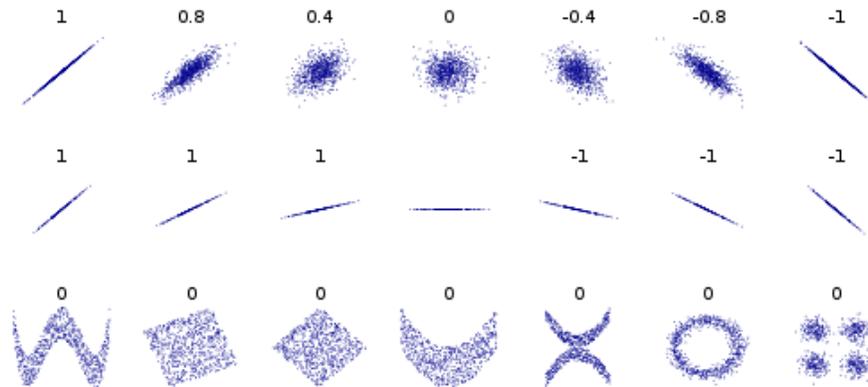


Figure 2.2: Scatterplots of outcomes from two random variables. The top row shows that different correlation coefficients mean different levels of noisiness, while the slope remains the same. The middle row shows that changing the slope can only change the sign of the correlation coefficient, but not its absolute value. In the middle plot on the middle row, the random variable on the Y-axis has variance 0, so the correlation coefficient is not defined. The bottom row shows that the correlation coefficient does not say anything about non-linear relationships between the two random variables. [6]

Like volatility, correlation is a variable which can change over time. Therefore, we also have **implied correlation** and **realized correlation**, just like we have implied and realized volatility. Their definitions are analogue: implied correlation is what we currently expect the correlation to be at a future point in time (an expected value using a model), while realized correlation is the true value of the correlation at that point in time (which is currently unknown).

2.1.4 Swaps

A **swap** can be seen as a bet on a market variable, such as the correlation coefficient between two stocks. Basically, a swap is a contract stating that one party will pay the other party a certain amount of money, and that amount of money depends on the changes in time of the underlying market variable.

Of course, the formal definition of a swap is quite different. A **swap** is a financial instrument traded between two companies. The swap has a **price**, a fixed predetermined payment from one side to the other. It also has a **payout**, based on a changing variable such as correlation, and on the predetermined **strike price**. This strike price is always tuned so that the up-front price of the swap is 0. The swaps we will be looking at will also have a predetermined **expiry date**. Another property of swaps is that they are traded **over the counter**, meaning that companies make the trade among themselves, without

the intervention of a stock exchange and its rules.

With this, we can define correlation swaps. Correlation can be described as a number using the correlation coefficient, which can change over time, so a swap can be applied to it. An amount of money is assigned for each point the correlation coefficient goes up or down; this is known as the **notional amount**.

For example, it might be a good idea to invest in a correlation swap between Volkswagen and McDonalds if Volkswagen announces it will only sell burgers instead of cars from now on, which probably means the correlation between their two stocks will rise. Conversely, the correlation between Volkswagen and Mercedes-Benz will then likely drop.

In the same way, **volatility** and **variance swaps** can be defined; just take the correlation swap and replace the correlation coefficient with volatility and variance rates respectively. These will later be used as a stepping stone to correlation swaps, because their complexity makes it hard to give a mathematical expression for their payoff straight away.

2.2 Why correlation swaps?

As with all financial products, correlation swaps can be traded for two purposes: **speculating** and **hedging**. Speculating can be seen as gambling on the future correlation: if you think correlation will go up, you can invest in a correlation swap. You can win money, or you can lose, similar to a casino. Hedging, however, is a bit more complicated.

2.2.1 Hedging

If you invest to speculate, you bear a risk. The purpose of hedging, however, is to reduce or even eliminate risks. Hedging always goes together with other investments, otherwise there would be no risks to reduce. For example, a bank might have tons of investments, and a net result of losing €200 per point the correlation between Apple and Google increases (and gaining €200 for every point it decreases). To get rid of this correlation-related risk, a correlation swap could be entered which makes the bank *gain* €200 per point the correlation between Apple and Google increases. This makes their investment safe for any changes in correlation; their **exposure** to correlation is zero.

This method is most often used by investment banks, who try to have no exposure and make marginal profits by selling financial products for slightly higher prices than what they are worth. Another use of hedging is to get rid of unwanted exposure, meaning e.g. that a financial product might come with exposure to both volatility and correlation when you only want to invest in

volatility. In this case, a correlation swap could cancel out profits or losses because of correlation changes, and you are left with only the desired investment in volatility.

Chapter 3

Bossu's toy model

In this chapter, we will look at the toy model described in [1] and the results derived from it. The first section will give some standard definitions of known statistical terms used in the article. Using these statistical terms, [1] defines its own (approximations of) variables. These are covered in section 3.2, and in section 3.3, these definitions are compared to the actual definitions. All of this is combined in section 3.4 to define the actual toy model, which is then used to calculate prices of variance swaps and correlation swaps. Finally, in section 3.5 it is explored if an assumption which in reality might not be true can be left out.

3.1 Statistical definitions used

3.1.1 Standard Brownian motion/Wiener process

A **Wiener process** W_t , also called **Standard Brownian motion**, is a random process characterized by the following properties:

1. $W_0 = 0$
2. W has independent increments: $W_{t+u} - W_t$ is independent of $W_s : s \leq t$ for $u \geq 0$
3. W has Gaussian increments: $W_{t+u} - W_t$ is normally distributed with mean 0 and variance u : $W_{t+u} - W_t \sim N(0, u)$
4. W has continuous paths: with probability 1, W_t is continuous in t .

3.1.2 Adapted process to a filtration

A **filtration** $(\mathcal{F}_i)_{i=1 \dots n}$ of the set of events \mathcal{F} in a probability space is an indexed set of subsets of \mathcal{F} . This index set \mathcal{I} is subject to the condition that if $i \leq j$ in \mathcal{I} , then $\mathcal{F}_i \subset \mathcal{F}_j$. The index is usually a time parameter, in this case

the filtration can be seen as including all historical data and information but not future data.

Then, a process that is **adapted** to a filtration can for the purposes of this thesis be interpreted as a process that cannot see into the future. This means that a process X is adapted if and only if for every realization and every n , X_n is known at time n ; it can be evaluated using the available information. In short, X only depends on past data and information and not on future data.

3.2 Definitions in Bossu (2007)

In this section, the definitions used in Bossu's article [1] will be presented. These definitions are used to derive the toy model for correlation swaps, so to understand the toy model, we must understand the definitions. These definitions will be looked at more in the next section to further improve our understanding of them.

3.2.1 Realised volatility and correlation

To define any sort of volatility and correlation, we must first define a stock market. We consider a universe of N stocks $S = (S_i)_{i=1\dots N}$. We take a probability space (Ω, \mathcal{E}, P) with P -filtration \mathcal{F} , and we assume that the vector S is an \mathcal{F} -adapted, positive Ito process.¹ We denote $S_i(t)$ the price of stock S_i at time t . To define the index, we first define the weights: a vector of positive real numbers $w = (w_i)_{i=1\dots N}$ that sum up to 1. We define the index as

$$I(t) \equiv \prod_{i=1}^N (S_i(t))^{w_i}.$$

This is a simplified model for the calculation of stock index I with constituent stocks S and weights w ; in practice, however, most stock indices are defined as an arithmetic weighted average.

Given a time period τ and a positive Ito process S , further definitions are given. For the length of a time period, a definition is given that also covers time periods consisting of multiple intervals. Let τ consist of n intervals, where the i -th interval ranges from a_i to b_i . The length of the time period τ is then given

¹An Itô process is a random process X such that

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

The first integral is a so-called Itô integral, which is defined in the same way as a Riemann-integral, only with a Brownian motion W replacing the integration variable. See also footnote 2.

so that it can be written as the sum of the lengths of the intervals it consists of:

$$\|\tau\| \equiv \int_{\tau} ds = \sum_{i=1}^n (b_i - a_i).$$

The continuously sampled realized volatility of a constituent stock or index is defined using a stochastic integral²:

$$\sigma^S(\tau) \equiv \sqrt{\|\tau\|^{-1} \int_{\tau} (d \ln S_x)^2}. \quad (3.1)$$

Because only variance has a liquid market, the volatilities of the individual stocks are squared before we take their weighted average. Finally the square root is taken to get to the continuously sampled average realized volatility of the constituent stocks, or constituent volatility:

$$\bar{\sigma}^S(\tau) \equiv \sqrt{\sum_{i=1}^N w_i (\sigma^{S_i}(\tau))^2}.$$

In the same way, we find the realized residual:

$$\epsilon(\tau) \equiv \sqrt{\sum_{i=1}^N w_i^2 (\sigma^{S_i}(\tau))^2}.$$

Using the above definitions, we define the continuously sampled average realized dispersion $d(\tau)$ between constituent stocks, and their continuously sampled average realized correlation $\rho(\tau)$, which will later be referred to as canonical realized correlation:

$$d(\tau) \equiv \sqrt{(\bar{\sigma}^S(\tau))^2 - (\sigma^I(\tau))^2} \quad (3.2)$$

$$\rho(\tau) \equiv \frac{(\sigma^I(\tau))^2 - (\epsilon(\tau))^2}{(\bar{\sigma}^S(\tau))^2 - (\epsilon(\tau))^2} \quad (3.3)$$

²Suppose that W is a Wiener process, H is a right-continuous, adapted and locally bounded process, and that $\{\pi_n\}$ is a sequence of partitions of $[0, t]$ with mesh going to zero. Then the Itô integral of H with respect to W up to time t is a random variable

$$\int_0^t HdW = \lim_{n \rightarrow \infty} \sum_{[t_{i-1}, t_i] \in \pi_n} H_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}).$$

For an Itô process X with ‘sub-processes’ σ and μ (see also footnote 1), the following stochastic integral is defined:

$$\int_0^t HdX = \int_0^t H_s \sigma_s dW_s + \int_0^t H_s \mu_s ds.$$

3.2.2 Implied volatility and correlation

Using conditional expectations, we can extend the above definitions to implied values. First we define implied volatility:

$$\sigma_t^{*S} := \sqrt{\mathbb{E} \left[(\sigma^S(\tau))^2 \mid \sigma^S([0, t]) \right]}, \quad (3.4)$$

where $\mathbb{E}[\cdot \mid \sigma^X([0, t])]$ denotes conditional expectation. P^* is a P -equivalent, \mathcal{F} -adapted measure, meaning that $P^*(X) = 0 \Leftrightarrow P(X) = 0$. In the same way, we define the implied constituent volatility and the implied residual as:

$$\bar{\sigma}_t^{*S}(\tau) \equiv \sqrt{\sum_{i=1}^N w_i (\sigma_t^{*S_i}(\tau))^2} = \sqrt{\mathbb{E} \left[(\bar{\sigma}^S(\tau))^2 \mid \bar{\sigma}^S([0, t]) \right]}, \quad (3.5)$$

$$\epsilon_t^*(\tau) \equiv \sqrt{\sum_{i=1}^N w_i^2 (\sigma_t^{*S_i}(\tau))^2} = \sqrt{\mathbb{E} \left[(\epsilon(\tau))^2 \mid \epsilon([0, t]) \right]}. \quad (3.6)$$

We consider a variance market on (S, w, I) , where agents can buy future realized variance against payment at maturity of a pre-agreed price. In this case, implied variance (the square of implied volatility) of S as defined above is the no-arbitrage price of future realized variance of S . In the definitions, t denotes time at which the variance is bought or sold, and $\tau = [0, T]$ denotes the period over which the variance is traded. The squares of implied constituent volatility and implied residual are the no-arbitrage prices of a portfolio of the N future realized variances of the constituent stocks with weights w and $(w_i^2)_{i=1 \dots N}$, respectively.

We can now define implied dispersion and canonical implied correlation:

$$d_t^*(\tau) \equiv \sqrt{(\bar{\sigma}_t^{*S}(\tau))^2 - (\sigma_t^{*I}(\tau))^2} = \sqrt{\mathbb{E} \left[(d(\tau))^2 \mid d([0, t]) \right]}, \quad (3.7)$$

$$\rho_t^*(\tau) \equiv \frac{(\sigma_t^{*I}(\tau))^2 - (\epsilon_t^*(\tau))^2}{(\bar{\sigma}_t^{*S}(\tau))^2 - (\epsilon_t^*(\tau))^2} \leq 1. \quad (3.8)$$

Here, d_t^* is the no-arbitrage price of realized dispersion as defined earlier, but in general, $\rho_t^*(\tau) \neq \mathbb{E}(\rho(\tau) \mid \rho([0, t]))$, so canonical implied correlation is not the no-arbitrage price for realized correlation.

3.2.3 Proxy formulas

If the amount of constituent stocks N goes to infinity, the residual terms $\epsilon(\tau)$ and $\epsilon_t^*(\tau)$ vanish, as will be shown later in this section. This yields the following

”proxy” formulas:

$$\rho(\tau) \xrightarrow{N \rightarrow +\infty} \left(\frac{\sigma^I(\tau)}{\bar{\sigma}^S(\tau)} \right)^2 \equiv \hat{\rho}(\tau) \quad (3.9)$$

$$\rho_t^*(\tau) \xrightarrow{N \rightarrow +\infty} \left(\frac{\sigma_t^* I(\tau)}{\bar{\sigma}_t^* S(\tau)} \right)^2 \equiv \hat{\rho}_t^*(\tau) \quad (3.10)$$

$\hat{\rho}(\tau)$ is called realized correlation, and $\hat{\rho}_t^*(\tau)$ is called implied correlation. For most indices, this is economically correct: fifty stocks is already enough to almost nullify the difference between canonical realized correlation and realized correlation, as shown in [1].

We can achieve a better understanding of the reason why the ϵ terms are left out by looking at the definitions of all terms appearing in the original formula for ρ :

$$\begin{aligned} (\sigma^I)^2 &= |\tau|^{-1} \int_{\tau} (d \ln I)^2 \\ &= |\tau|^{-1} \int_{\tau} \left(d \sum_{i=1}^N w_i \ln S_i \right)^2 \end{aligned} \quad (3.11)$$

$$(\bar{\sigma}^S)^2 = \sum_{i=1}^N w_i (\sigma^{S_i})^2 \quad (3.12)$$

$$\epsilon^2 = \sum_{i=1}^N w_i^2 (\sigma^{S_i})^2 \quad (3.13)$$

We can see that in the definitions of $(\sigma^I)^2$ and $(\bar{\sigma}^S)^2$, the weight terms w_i appear without a power. In the definition for $(\sigma^I)^2$, the weight term is squared, but only after some other operations. In the definition for ϵ^2 , the weight term is immediately squared. This means that *if the weights would go to zero*, ϵ^2 would vanish in comparison to $(\sigma^I)^2$ and $(\bar{\sigma}^S)^2$. However, the weights do not necessarily go to zero if $N \rightarrow \infty$: if there are 1000 constituent stocks, it is possible for one stock to have a weight of 0.9 while the other 999 stocks have a total weight of 0.1. So [1] implicitly makes the assumption that this does not happen, and that $\lim_{N \rightarrow \infty} w_i = 0$ for all i . Since it is numerically shown in [1] that the canonical values and the proxy values are very close, this may be a reasonable assumption.

3.3 Comparison of the definitions in [1] to the regular definitions

The definitions for volatility and correlation in [1] are different from the common definitions. That is why we compare the definitions used in [1] with the standard

definitions in this section. First, some results will be derived from the textbook definitions of variance and correlation. These results are rewritten to be as close as possible to the definitions in [1]. It will turn out that there are disparities between the true values derived from the regular definitions and the formulas used in the article. The consequences of assuming these formulas to be close to the true values are discussed as well. Lastly, the case $\rho = 0$ is explored.

3.3.1 Using the definitions of variance and correlation

Take two random variables S_1 and S_2 . According to the standard definitions, the variance of S_1 is:

$$\text{var}(S_1) = \mathbb{E}(S_1^2) - \mathbb{E}(S_1)^2.$$

The correlation between S_1 and S_2 is:

$$\rho_{S_1, S_2} = \frac{\text{cov}(S_1, S_2)}{\sqrt{\text{var}(S_1)}\sqrt{\text{var}(S_2)}}.$$

$\sqrt{\text{var}(S_1)}$ may also be substituted by the volatility $\sigma_X = \sqrt{\text{var}(S_1)}$. This gives rise to the insight that the price of a correlation swap might also be a value divided by the price of the corresponding volatility swaps. This will later turn out to be true, where the price of the volatility swaps is represented by the price of constituent volatility.

If we take multiple S_i , $i = 1, 2, \dots, n$, we can take the average of the pairwise correlations as the correlation among all X_i :

$$\bar{\rho}_{S_i} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{\mathbb{E}(S_i S_j) - \mathbb{E}(S_i)\mathbb{E}(S_j)}{\sigma_{S_i} \sigma_{S_j}}.$$

This formula is not very informative. Another approach for finding a relation is looking at $\text{var}(S_1 + S_2)$. Expanding this gives:

$$\begin{aligned} \text{var}(S_1 + S_2) &= \text{var}(S_1) + \text{var}(S_2) + 2\text{cov}(S_1, S_2), \\ \sigma_{S_1+S_2} &= \sigma_{S_1}^2 + \sigma_{S_2}^2 + 2\sigma_{S_1}\sigma_{S_2}\rho_{S_1, S_2}. \end{aligned}$$

Bringing the correlation coefficient to the other side gives us an expression for the correlation in terms of the volatilities:

$$\rho_{S_1, S_2} = \frac{1}{2} \left(\frac{\sigma_{S_1+S_2}}{\sigma_{S_1}\sigma_{S_2}} - \frac{\sigma_{S_1}}{\sigma_{S_2}} - \frac{\sigma_{S_2}}{\sigma_{S_1}} \right). \quad (3.14)$$

3.3.2 Justifying the formulas in [1]

Going further into (3.14), we might be able to justify formula (3.3) for correlation given in [1], repeated here:

$$\rho(\tau) = \frac{(\sigma^I(\tau))^2 - (\epsilon(\tau))^2}{(\bar{\sigma}^S(\tau))^2 - (\epsilon(\tau))^2}, \quad (3.15)$$

where

$$\bar{\sigma}^S(\tau) = \sqrt{\sum_{i=1}^N w_i (\sigma^{S_i}(\tau))^2},$$

$$\epsilon(\tau) = \sqrt{\sum_{i=1}^N w_i^2 (\sigma^{S_i}(\tau))^2},$$

τ is a time period and w_i are the weights of the constituent stocks in the stock index. σ^{S_i} and σ^I also have other definitions than usual (using Itô integrals, see section 3.2), but we will work with the standard definitions here, since the difference between the definitions is small.

Looking at a portfolio I consisting of stocks S_1, S_2 with corresponding weights w_{S_1}, w_{S_2} and using the same reasoning as for (3.14), we obtain for the correlation:

$$\begin{aligned} \text{var}(I) &= w_{S_1}^2 \text{var}(S_1) + w_{S_2}^2 \text{var}(S_2) + 2w_{S_1} w_{S_2} \text{cov}(S_1, S_2) \\ \sigma_I^2 &= (w_{S_1} \sigma_{S_1})^2 + (w_{S_2} \sigma_{S_2})^2 + 2w_{S_1} \sigma_{S_1} w_{S_2} \sigma_{S_2} \rho_{S_1, S_2} \\ \rho_{S_1, S_2} &= \frac{\sigma_I^2 - (w_{S_1} \sigma_{S_1})^2 - (w_{S_2} \sigma_{S_2})^2}{2w_{S_1} \sigma_{S_1} w_{S_2} \sigma_{S_2}} \end{aligned}$$

This result can be extended for an index I consisting of N stocks S_i with weights w_i and volatilities σ^{S_i} . Define a correlation value that can be seen as the weighted average correlation in the index:

$$\rho = \frac{\sum_{i \neq j}^N w_i w_j \sigma^{S_i} \sigma^{S_j} \rho_{i,j}}{\sum_{i \neq j}^N w_i w_j \sigma^{S_i} \sigma^{S_j}}.$$

Using this definition and the same reasoning as for the case with 2 stocks, we lose the dependence on pairwise correlations and we gain a dependence on the index volatility:

$$\rho = \frac{(\sigma^I)^2 - \sum_{i=1}^N w_i^2 (\sigma^{S_i})^2}{\sum_{i \neq j}^N w_i w_j \sigma^{S_i} \sigma^{S_j}}$$

We can see that the numerator of this expression is exactly the same as the numerator in equation (3.3), but the denominator is different. We can rewrite the denominator to a term closer to the denominator in (3.3). Going back to the case with only two constituent stocks, we can substitute $2w_{S_1} w_{S_2} \sigma_{S_1} \sigma_{S_2}$ with

$(w_{S_1}\sigma_{S_1} + w_{S_2}\sigma_{S_2})^2 - \epsilon^2$, justified below:

$$\begin{aligned} & (w_{S_1}\sigma_{S_1} + w_{S_2}\sigma_{S_2})^2 - w_{S_1}^2\sigma_{S_1}^2 - w_{S_2}^2\sigma_{S_2}^2 \\ &= w_{S_1}^2\sigma_{S_1}^2 + w_{S_2}^2\sigma_{S_2}^2 + 2w_{S_1}w_{S_2}\sigma_{S_1}\sigma_{S_2} - w_{S_1}^2\sigma_{S_1}^2 - w_{S_2}^2\sigma_{S_2}^2 \\ &= 2w_{S_1}w_{S_2}\sigma_{S_1}\sigma_{S_2}. \end{aligned}$$

Using this substitution in the case with N constituent stocks, the formula for ρ becomes:

$$\rho = \frac{(\sigma^I)^2 - \epsilon^2}{\left(\sum_{i=1}^N w_i\sigma^{S_i}\right)^2 - \epsilon^2} \quad (3.16)$$

By this, the term $-\epsilon^2$ is incorporated, but we still have $\left(\sum_{i=1}^N w_i\sigma^{S_i}\right)^2 \neq (\bar{\sigma}^S)^2 = \sum_{i=1}^N w_i(\sigma^{S_i})^2$. This inequality is disregarded in [1]; it is assumed that the difference between these two terms is small. This becomes clear from the statement that constituent volatility is more frequently defined as the weighted arithmetic average of volatilities of constituent stocks, $\sum_{i=1}^N w_i\sigma^{S_i}$. By taking a different definition for constituent volatility, it is implicitly assumed in [1] that the difference between these two definitions remains small:

$$\left(\sum_{i=1}^N w_i\sigma^{S_i}\right)^2 \approx (\bar{\sigma}^S)^2. \quad (3.17)$$

The difference between these two terms is assumed not only to be close to zero, but also to always have the same sign. The following inequality is stated in [1]:

$$\rho = \frac{(\sigma^I)^2 - \epsilon^2}{(\bar{\sigma}^S)^2 - \epsilon^2} \leq \frac{(\sigma^I)^2 - \epsilon^2}{\left(\sum_{i=1}^N w_i\sigma^{S_i}\right)^2 - \epsilon^2},$$

which implies

$$\left(\sum_{i=1}^N w_i\sigma^{S_i}\right)^2 \geq (\bar{\sigma}^S)^2. \quad (3.18)$$

3.3.3 Implicit assumptions

In this subsection, assumptions (3.17) and (3.18) are looked into. The meaning of these assumptions on economical level is discussed, as well as how reasonable the assumptions and their economical equivalents are.

Consider the following:

$$\begin{aligned} \left(\sum_{i=1}^N w_i \sigma^{S_i} \right)^2 &= \sum_{i=1}^N w_i \sigma^{S_i} \left(\sum_{j=1}^N w_j \sigma^{S_j} \right) \\ (\bar{\sigma}^S)^2 &= \sum_{i=1}^N w_i \sigma^{S_i} (\sigma^{S_i}). \end{aligned}$$

This enables us to see that the assumption made at (3.17) is equivalent to

$$\sum_{i=1}^N w_i \sigma^{S_i} \approx \frac{1}{N} \sum_{i=1}^N \sigma^{S_i},$$

or in words: the weighted average of the volatilities of the constituent stocks is close to the non-weighted average. This means that the weights are more or less evenly distributed for constituent stocks with large and small volatilities, and that the constituent stocks with large volatilities generally do not have significantly higher or lower weights than those with small volatilities. This is a reasonable assumption, especially for a large number N of constituent stocks.

However, inequality (3.18) is not explained by this assumption. It is equivalent to:

$$\frac{1}{N} \sum_{i=1}^N \sigma^{S_i} - \sum_{i=1}^N w_i \sigma^{S_i} \geq 0,$$

while this value might just as easily be positive. This means that for the inequality to hold, the constituent stocks with low volatilities must have larger weights than those with high volatilities. When combined with the reasoning above, this means the following assumption is done in [1]: *Constituent stocks of the index with low volatilities have slightly larger weights than those with high volatilities.*

Economically, this assumption means that the relatively stable stocks (low volatility) are preferred when creating or updating the stock index. For many indices, this assumption is correct: the risk of high-volatility stocks is only accepted if in return there's a higher expected profit. Even when the index doesn't explicitly select stocks based on volatility levels, many indices give weights to constituent stocks corresponding to their market caps. Generally, stocks with high market caps, i.e. stocks of large companies, are less volatile than stocks of small companies. This means that in such indices, stocks with low volatilities have higher weights than stocks with high volatilities, in accordance with the assumption.

3.3.4 The case $\rho = 0$

Whether we use the formula in [1] or the true formula for ρ , the result for taking $\rho = 0$ remains the same, since the numerator of the fraction needs to be 0 in

that case, and the numerators are the same in both formulas. So both formulas yield the following:

$$(\sigma^I)^2 - \epsilon^2 = 0;$$

$$(\sigma^I)^2 = \sum_{i=1}^N w_i^2 \cdot (\sigma^{S_i})^2.$$

This means that the index variance equals the ‘weighted’ (squared weights, so they do not add to 1) arithmetic average of the variances of constituent stocks if and only if the average correlation equals 0. This conclusion agrees with our intuition: if there is a correlation, it would alter the index variance with regard to the average constituent variance. Since we are only considering averages, there may be some pairwise correlations, but if they cancel each other out, their effects on the index variance apparently cancel out as well. The squared weights make sense as well, since a scalar taken out of a variance must be squared.

3.4 The toy model

3.4.1 One factor: variance swaps

Definition

Now the ‘toy model’ in [1] for derivatives on realized variance will be defined using the definitions discussed in the previous sections. These derivatives can be seen as variance swaps. We start simple, by taking a market on a single asset S , where agents can trade the asset’s realized variance $(\sigma^S(\tau))^2$ over a fixed time period $\tau = [0, T]$. For $t \in [0, T]$, we take the variance price v_t^* to be the best estimate of the variance we can give at time t : we take the realized variance until time t , and from time t until time T , we substitute it with implied variance. This is defined mathematically as

$$v_t^* = \frac{t}{T} (\sigma^S([0, t]))^2 + \frac{T-t}{T} (\sigma_t^{*S}([t, T]))^2, \quad (3.19)$$

where σ^S and σ_t^{*S} are volatility and implied volatility of S respectively, defined according to (3.1) and (3.4). The forward-neutral dynamics of v^* - the way in which v^* changes over time - are not determined by (3.19). In [1] it is assumed that these forward dynamics have the following structure:

$$dv_t^* = 2\omega \frac{T-t}{T} v_t^* dW_t^*, \quad (3.20)$$

where ω is a positive ‘volatility of volatility’ parameter, and W^* is a standard Brownian motion under P^* . The flexibility of this structure lies in the parameter ω , which can be chosen using past data and a maximum likelihood

estimate. The term $\frac{T-t}{T}$ gets closer to 0 as $t \rightarrow T$, meaning the price of the variance swap becomes less volatile as the time approaches maturity, which agrees with our intuition. As expected, the volatility of the swap price is proportional to the swap price v_t^* itself. The Brownian motion W^* gives the variance process its randomness.

This model is used to define the fair strike of a volatility swap, which we will call fair volatility. The payoff is $\mathcal{V}_T \equiv \sqrt{v_T^*}$, quite logically the square root of the price of realized variance. The fair value of the swap at any time t is calculated in [1] using (3.20) and some properties of Itô integrals:

$$\mathcal{V}_t = \sqrt{v_t^*} \exp \left[-\frac{1}{6} \omega^2 T \left(\frac{T-t}{T} \right)^3 \right].$$

Setting $t = 0$, we find for the fair volatility \mathcal{V}_0 :

$$\mathcal{V}_0 = \sqrt{v_0^*} \exp \left[-\frac{1}{6} \omega^2 T \right]$$

v_0^* is known as the fair strike of a variance swap, or fair variance in this thesis. Notice that the term strike is used instead of the term price: this is the fair price for a variance swap with strike 0, but as is the rule with swaps, the strike is adjusted so that the price becomes 0. In our case, this means the strike must be v_0^* . The term $\exp \left[-\frac{1}{6} \omega^2 T \right]$ is known as the **convexity adjustment**; a term that compensates for the fact that variance is convex in volatility.

Simulation

This toy model can be simulated in MATLAB using the Euler discretization scheme on its forward-neutral dynamics. For details about the Euler discretization scheme, see chapter 4 of this thesis or [4].

An example of several variance paths simulated using this model can be seen in figure 3.1. ω is taken as 0.61: this is a reasonable value for a maturity of 1 year according to an estimation done in [1] using real-world data. v_0 has been arbitrarily picked, because there are many different stocks with many different variances. It can clearly be seen from this plot that the volatility of a path is lower when v is lower (because of the term v in the forward dynamics) and when t is higher (because of the term $\frac{T-t}{T}$). Both for $\frac{T-t}{T} \rightarrow 0$ and for $v \rightarrow 0$ the volatility should approach 0, since both these terms appear as a scaling factor in (3.20). The situation $\frac{T-t}{T} \rightarrow 0$ occurs in the simulation, at the right side of figure 3.1, and we see that the volatility does indeed approach 0 in that part. To illustrate what happens for $v \rightarrow 0$, the blue path has been modeled with $v_0 = 0.02$. We can see that the volatility for this path is indeed very low, as expected, and the value of v barely changes.

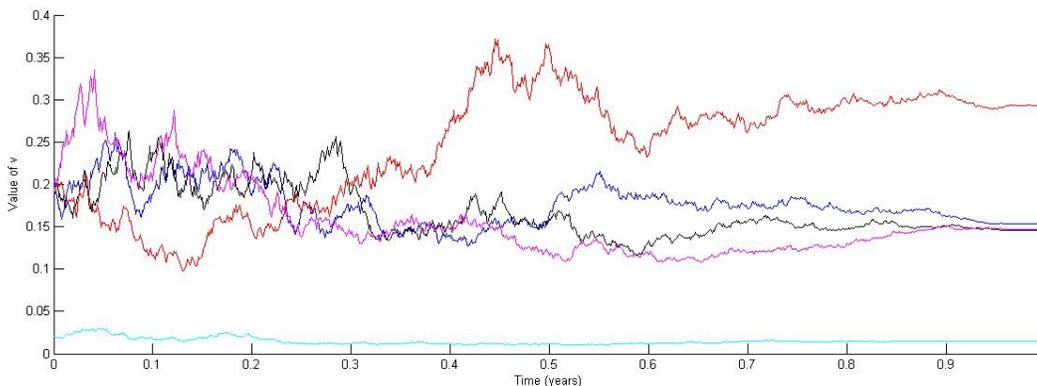


Figure 3.1: 5 paths of the toy model using Euler discretization. $\omega = 0.61$, $\Delta t = 0.001$, $T = 1$, $v_0 = 0.02$ for the bottom (blue) path, $v_0 = 0.2$ for the other paths. It can be seen that the volatility of these paths is lower for low values of v and for t approaching maturity T .

3.4.2 Two factors: correlation swaps

In this subsection, the fair value of a correlation claim is modeled by extending the toy model to two factors, being index variance and constituent variance. This correlation claim is inspired by the proxy formulas discussed in 3.2.3. Its payoff is

$$c_T \equiv \hat{\rho}(\tau) = \left(\frac{\sigma^I(\tau)}{\bar{\sigma}^S(\tau)} \right)^2.$$

This is different from a standard correlation swap with strike 0, which has payoff

$$\rho(\tau) = \frac{\sum_{i < j} w_i w_j \rho^{S_i S_j}(\tau)}{\sum_{i < j} w_i w_j},$$

where $\rho^{S_i S_j}$ is the pairwise correlation between stocks S_i and S_j . The difference in payoff between this correlation claim and the standard correlation swap is almost zero (shown in [1] using real-world data), so their prices will also be close to each other. In reality, all swaps (including correlation swaps) have a strike such that the up-front price of the swap becomes zero. This is not considered in [1], but in chapter 5 we use the price of a correlation swap with strike 0 to find the value the strike should have to make the price equal to 0.

Definition

The toy model is extended to two factors; those factors are the index variance and the constituent variance. We denote with v_t^{*I} the market price of realized index variance at time t , with $v_t^{*S_i}$ the market price of the realized variance of stock S_i at time t , and we define the no-arbitrage price of realized constituent variance at time t as:

$$\bar{v}_t^{*S} \equiv \sum_{i=1}^N w_i v_t^{*S_i}. \quad (3.21)$$

In the same way as for the one-factor toy model, we define a structure for the forward-neutral dynamics of both paths:

$$dv_t^{*I} = 2\omega_I \frac{T-t}{T} v_t^{*I} dW_t^{*I}, \quad (3.22)$$

$$d\bar{v}_t^{*S} = 2\bar{\omega}_S \frac{T-t}{T} \bar{v}_t^{*S} dW_t^{*S}, \quad (3.23)$$

where ω_I and $\bar{\omega}_S$ are constants volatility of volatility parameters and W^{*I} and W^{*S} are standard Brownian motions under P^* . We assume that W^{*I} and W^{*S} have a constant correlation χ , i.e. $(dW_t^{*I})(dW_t^{*S}) = \chi dt$. Note that this value denotes the correlation between *changes* in index and constituent volatilities, rather than the correlation between the absolute levels.

Equation (3.23) is only an approximation: it is assumed that each v^{*S_i} follows this type of dynamics, but their arithmetic average \bar{v}_t^{*S} does not necessarily have to follow the same dynamics. Assuming that

$$d\bar{v}_t^{*S} \approx \bar{v}_{t_{i+1}}^{*S} - \bar{v}_{t_i}^{*S} \quad (3.24)$$

for some points in time t_{i+1} and t_i close to each other, we can justify formula (3.23), using that each v^{*S_i} follows this type of dynamics. First we substitute definition (3.21) in (3.24):

$$\begin{aligned} d\bar{v}_t^{*S} &\approx \sum_{j=1}^N w_j v_{t_{i+1}}^{*S_j} - \sum_{j=1}^N w_j v_{t_i}^{*S_j} \\ &= \sum_{j=1}^N w_j \left(v_{t_{i+1}}^{*S_j} - v_{t_i}^{*S_j} \right) \\ &\approx \sum_{j=1}^N w_j dv_t^{*S_j}, \end{aligned}$$

using approximation (3.24) (with $dv_t^{*S_j}$ instead of $d\bar{v}_t^{*S}$) in the last step.

Now we can use that each $dv_t^{*S_j}$ follows the dynamics in (3.23):

$$d\bar{v}_t^{*S} \approx \sum_{j=1}^N 2w_j\omega_{S_j} \frac{T-t}{T} v_t^{*S_j} dz_t^{*S_j} \quad (3.25)$$

$$= 2 \frac{T-t}{T} \sum_{j=1}^N w_j\omega_{S_j} v_t^{*S_j} dz_t^{*S_j}. \quad (3.26)$$

Each volatility of volatility parameter ω_{S_j} of the individual constituent stocks is much higher than the volatility of constituent volatility parameter $\bar{\omega}_S$, so the (weighted) mean of the ω_{S_j} will also be much higher. On the other hand, the (weighted) mean of the Wiener process increments $dz_t^{*S_j}$ is much lower than dz_t^{*S} in absolute value, which is a Wiener process as well: because of the law of large numbers, the weighted mean of Wiener process increments converges to zero. It is a reasonable assumption that these two disparities cancel each other out, i.e.

$$\sum_{j=1}^N w_j\omega_{S_j} dz_t^{*S_j} \approx \bar{\omega}_S dz_t^{*S}.$$

This means that $\omega_{S_j} dz_t^{*S_j}$ can be approximated by $\bar{\omega}_S dz_t^{*S}$ for each j . Making this substitution in (3.26), we get

$$\begin{aligned} d\bar{v}_t^{*S} &\approx 2 \frac{T-t}{T} \bar{\omega}_S \left(\sum_{j=1}^N w_j v_t^{*S_j} \right) dz_t^{*S} \\ &= 2 \frac{T-t}{T} \bar{\omega}_S \bar{v}_t^{*S} dz_t^{*S}, \end{aligned}$$

which is exactly how it is defined in (3.23).

Now the analytically derived formula for the fair value of the correlation claim will be introduced. The payoff of the correlation claim is

$$c_T \equiv \frac{v_T^{*I}}{\bar{v}_T^{*S}}.$$

Using the definitions for index volatility and constituent volatility given earlier in this chapter and some properties of Itô integrals, the following fair price for the correlation claim at time t is obtained in [1]:

$$c_t \equiv \mathbb{E}(c_T \mid v_t^{*I}, \bar{v}_t^{*S}) = \frac{v_t^{*I}}{\bar{v}_t^{*S}} \exp \left[\frac{4}{3} (\bar{\omega}_S^2 - \chi \bar{\omega}_S \omega_I) T \left(\frac{T-t}{T} \right)^3 \right].$$

In words: the fair value of the correlation claim is equal to the ratio of fair index variance to fair constituent variance, multiplied by an adjustment factor which depends on the volatility of index volatility, the volatility of constituent

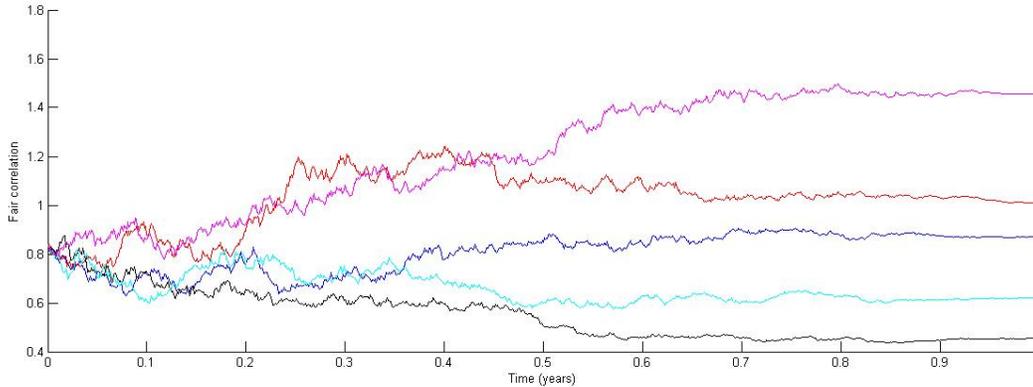


Figure 3.2: 5 paths of the fair value of the correlation claim. $\omega_i = 0.61$, $\omega_s = 0.54$, $\chi = 0.9$, $v_0^{*I} = 0.5$, $\bar{v}_0^{*S} = 0.6$, $\Delta t = 0.001$, $T = 1$. The top paths (purple and red) do not stay below their boundary of 1 because there is no mathematical condition enforcing the boundary.

volatility, and the correlation between index volatility and constituent volatility. Taking $t = 0$ and rewriting the equation, we obtain for the **fair correlation adjustment**, the ratio between implied correlation and the trading price of the correlation claim:

$$\frac{\hat{\rho}_0^*}{c_0} = \exp\left(\frac{4}{3}(\chi\bar{\omega}_S\omega_I - \bar{\omega}_S^2)T\right).$$

Historical data shows that the fair correlation adjustment is close to one for most maturities, meaning implied correlation and fair correlation are very close. [1]

Simulation

The two-factor toy model can be simulated as a simple extension of the simulation of the one-factor toy model. Both v_t^{*I} and \bar{v}_t^{*S} are modeled in the same way as v_t^* in the one-factor toy model. Using these values, the formula for the fair price of a correlation claim can be filled in.

In figure 3.2, 5 of these paths are shown. ω_i, ω_s and χ are taken to be equal to the estimates in [1] using real-world data. v_0^{*I} and \bar{v}_0^{*S} are arbitrarily picked, but still satisfying $v_0^{*I} \leq \bar{v}_0^{*S}$. This inequality cannot be maintained for the rest of the paths, as can be seen in figure 3.3, since there is no condition built into the model that prevents v_t^{*I} from getting higher than \bar{v}_t^{*S} . There is only the fair correlation adjustment, which is lower than 1 for most maturities when using the parameter estimates from [1], since $\bar{\omega}_S \leq \chi\omega_I$. However, this adjustment is

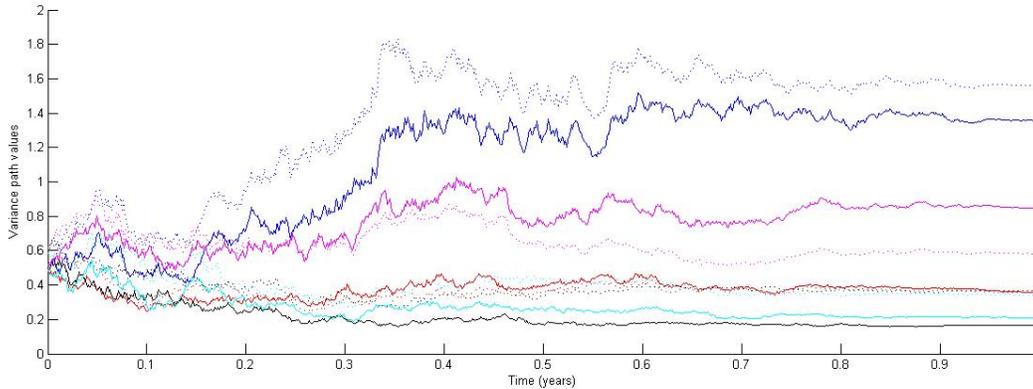


Figure 3.3: The corresponding variance paths for the correlation paths in figure 3.2. Each color represents the variance paths corresponding to the correlation path of the same color. The solid lines are index variance paths v_t^{*I} , the dotted lines are constituent variance paths \bar{v}_t^{*S} . $\omega_i = 0.61$, $\omega_s = 0.54$, $\chi = 0.9$, $v_0^{*I} = 0.5$, $\bar{v}_0^{*S} = 0.6$, $\Delta t = 0.001$, $T = 1$. It can be seen that whenever the correlation in figure 3.2 is above 1, the index variance path is above the constituent variance path, as expected.

quite small, especially for larger t .

This is what causes the problem that the fair correlation becomes higher than 1. When comparing figure 3.2 with figure 3.3, we see that the correlation being higher than 1 corresponds to the index variance being higher than the constituent variance. This problem does not appear if v_0^{*I} is much lower than \bar{v}_0^{*S} and χ stays high. That way, the high correlation prevents v_t^{*I} from growing too fast without \bar{v}_t^{*S} growing as well. This problem is also illustrated in appendix G in [1], where the probability of the payoff being higher than 1 is shown as a function of χ for different values of ρ_0 . The result of this appendix can be seen in figure 3.4.

In our example in figures 3.2 and 3.3, we used correlation between volatilities $\chi = 0.9$ and $\hat{\rho}_0^* = 0.5/0.6 \approx 0.83$. If we look at the plot in figure 3.4, following the line for $\rho = 0.8$, we find a probability of around 0.25 for $c_T > 1$. This agrees quite well with our simulations in figure 3.2, where at maturity T , the top path (purple) clearly exceeds 1 and the second path from the top (red) is close to 1.

To make sure the correlation does not exceed 1, we apparently need high χ and low $\hat{\rho}_0^*$. Since the χ we used was already quite high (0.9), we simulate correlation paths for lower $\hat{\rho}_0^*$. According to figure 3.4, the probability for $c_T > 1$ is almost 0 for $\hat{\rho}_0^* = 0.5$. To achieve this, we keep $\bar{v}_0^{*S} = 0.6$, meaning we have to set $v_t^{*I} = 0.3$. The results of this simulation can be seen in figure 3.5. We

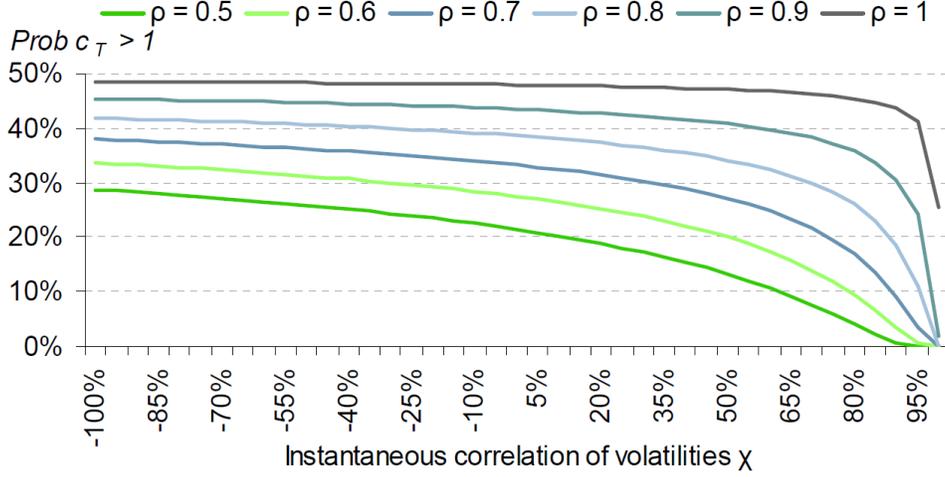


Figure 3.4: The result of appendix G in [1], comparing the probability of the correlation being higher than 1 at maturity T (for $T = 1$ year) as a function of the correlation between volatilities χ with different values of the starting value of correlation $\hat{\rho}_0^* = v_0^{*I}/\bar{v}_t^{*S}$ (denoted with ρ in this image). This image shows that the probability for $c_T > 1$ goes down for higher χ , but for higher starting correlations ρ the effect is diminished, only having a real impact for $\chi \rightarrow 100\%$.

can see from the left plot that the correlation paths stay well below zero. The right plot shows that the condition $v_t^{*I} < \bar{v}_t^{*S}$ is well satisfied. We can conclude that this model may give unrealistic values of $c_T > 1$, but if the parameters are right, i.e. high χ and low $\hat{\rho}_0^*$, the model does work.

Furthermore, the diminishing volatility of the fair correlation over time is visible again, as in figure 3.1. This is because of the terms $\frac{T-t}{T}$ in the paths for the implied and constituent volatilities. The term $(\frac{T-t}{T})^3$ in the fair correlation adjustment makes this term go to 1 for $t \rightarrow T$, as it should, since the payoff of the correlation swap is $c_T = v_T^{*I}/\bar{v}_T^{*S}$, without any adjustment factor.

3.5 Without the assumption $\sigma^I \leq \bar{\sigma}^S$

From formula (3.16) for the true, unapproximated correlation value, it is clear that

$$(\sigma^I)^2 \leq \left(\sum_{i=1}^N w_i \sigma^{S_i} \right)^2,$$

because the correlation value needs to stay below 1. By making approximation (3.17) and substituting it in the formula for the correlation, [1] implies that

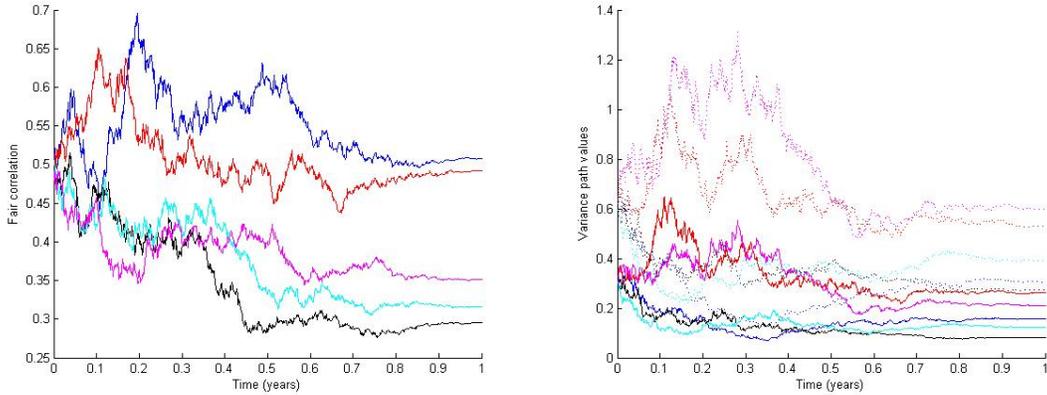


Figure 3.5: 5 simulations of fair correlation paths (left) and in the same color their corresponding variance paths (right). Dotted lines are constituent variance paths \bar{v}_t^{*S} , solid lines are index variance paths v^{*I} . $\omega_i = 0.61$, $\omega_s = 0.54$, $\chi = 0.9$, $v_0^{*I} = 0.3$, $\bar{v}_0^{*S} = 0.6$, $\Delta t = 0.001$, $T = 1$. The parameters are chosen such that χ is high and $\hat{\rho}_0^* = v_0^{*I}/\bar{v}_0^{*S}$ is low, to avoid any correlation paths going over 1. This was succesful, since all correlation paths stay well below 1.

the correlation value remains below 1, so:

$$\sigma^I \leq \bar{\sigma}^S.$$

To simplify notation, we use the following definition for the weighted arithmetic constituent volatility for this section:

$$\sigma^A := \sum_{i=1}^N w_i \sigma^{S_i}.$$

This gives a shorter formula for the true correlation:

$$\rho = \frac{(\sigma^I)^2 - \epsilon^2}{(\sigma^A)^2 - \epsilon^2},$$

and for the proxy formula:

$$\hat{\rho} = \left(\frac{\sigma^I}{\sigma^A} \right)^2.$$

This all works fine without approximation (3.17). The part where the assumption simplifies the calculations is in the two-factor toy model. Definition (3.21) is the logical definition for a variance process corresponding to $(\bar{\sigma}^S)^2$. Below the definition, it is explained how this process can be approximated by

the forward dynamics in (3.23). If we look at the variance process corresponding to σ^A :

$$v_t^{*A} = \left(\sum_{i=1}^N w_i \sqrt{v_t^{*S_i}} \right)^2, \quad (3.27)$$

we can expect that this process is further away from the type of dynamics in (3.23) than \bar{v}_t^{*S} . To find out what kind of process approximates (3.27), we start rewriting the process from its definition. Again, we use the approximation

$$dv_t^{*A} \approx v_{t_{k+1}}^{*A} - v_{t_k}^{*A} \quad (3.28)$$

for some points in time t_{k+1} and t_k close to each other, to allow us to use definition (3.27). Substituting this:

$$dv_t^{*A} \approx \left(\sum_{i=1}^N w_i \sqrt{v_{t_{k+1}}^{*S_i}} \right)^2 - \left(\sum_{i=1}^N w_i \sqrt{v_{t_k}^{*S_i}} \right)^2 \quad (3.29)$$

$$= \sum_{i=1}^N \left(w_i \sqrt{v_{t_{k+1}}^{*S_i}} \right)^2 + \sum_{i=1}^N \sum_{j \neq i}^N w_i \sqrt{v_{t_{k+1}}^{*S_i}} w_j \sqrt{v_{t_{k+1}}^{*S_j}} \quad (3.30)$$

$$- \sum_{i=1}^N \left(w_i \sqrt{v_{t_k}^{*S_i}} \right)^2 - \sum_{i=1}^N \sum_{j \neq i}^N w_i \sqrt{v_{t_k}^{*S_i}} w_j \sqrt{v_{t_k}^{*S_j}} \quad (3.31)$$

$$= \sum_{i=1}^N w_i^2 \left(v_{t_{k+1}}^{*S_i} - v_{t_k}^{*S_i} \right) \quad (3.32)$$

$$+ \sum_{i=1}^N \sum_{j \neq i}^N w_i w_j \left(\sqrt{v_{t_{k+1}}^{*S_i}} \sqrt{v_{t_{k+1}}^{*S_j}} - \sqrt{v_{t_k}^{*S_i}} \sqrt{v_{t_k}^{*S_j}} \right). \quad (3.33)$$

The first step we took here was writing out the square in the first term in (3.29) to $\left(\sum_{i=1}^N w_i \sqrt{v_{t_{k+1}}^{*S_i}} \right) \left(\sum_{j=1}^N w_j \sqrt{v_{t_{k+1}}^{*S_j}} \right)$. For $i = j$, this is equal to the first term in (3.30) and for $i \neq j$, it is equal to the second term. We do the same for the second term in (3.29), which becomes the two terms in (3.31). Next, we subtract the first term in (3.31) from the first term in (3.30) to get to (3.32) and we subtract the second term in (3.31) from the second term in (3.30) to get to (3.33).

Term (3.32) can be approximated by $\sum_{i=1}^N w_i^2 dv_t^{*S_i}$, but term (3.33) is a bit more involved. It can be a substantial part of the full value of dv_t^{*A} , so we cannot neglect it. We can take $\sqrt{v_{t_{k+1}}^{*S_i}}$ and $\sqrt{v_{t_{k+1}}^{*S_j}}$ together, as well as $\sqrt{v_{t_k}^{*S_i}}$ and $\sqrt{v_{t_k}^{*S_j}}$, but there is no way to express this part in terms of $dv_t^{*S_i}$. So to be able to model (3.33), we need to define a model for

$$\sqrt{v_{t_{k+1}}^{*S_i} v_{t_{k+1}}^{*S_j}} - \sqrt{v_{t_k}^{*S_i} v_{t_k}^{*S_j}}. \quad (3.34)$$

We expect that if we have a good model for (3.34), we can substitute that in the toy model and use v_t^{*A} instead of \bar{v}_t^{*S} . The model would most likely become more accurate from this, and depending on the accuracy of the model for (3.34), the modeled value of the correlation claim might not exceed 1 anymore. Of course this is a good thing, but it is hard to find a satisfactory approximate for (3.34). The model would have to contain an expression for the correlation in the portfolio I , but we are trying to model the correlation ourselves, so if we had had an expression for the correlation, we would not be doing all these calculations. This also explains why [1] suggested approximation (3.17); even though the accuracy of the model suffers, the model becomes so much more complex when not making the assumption that the additional accuracy is not worth it.

Chapter 4

The Jacobi Process

To see if our calculations actually work, it is useful to perform simulations. Simulations can generate plots to help us explain certain phenomena or they can provide estimates for certain parameters by doing a large number of simulations and looking at the properties of this set of simulations, such as their mean. The Jacobi process is a very useful process, since it can be easily simulated and its parameters can be chosen such that it approximates a correlation process.

4.1 What is the Jacobi process?

The **bounded Jacobi process**, which we will call simply the **Jacobi process** in this thesis, has several properties that make it ideal to model correlation paths. The dynamics of the process are given by

$$d\rho(t) = \kappa_\rho (\mu_\rho - \rho(t)) dt + \gamma_\rho \sqrt{1 - \rho^2(t)} dW(t). \quad (4.1)$$

The term $(\mu_\rho - \rho(t))$ makes the process mean-reverting. If $\rho(t) < \mu_\rho$ this term makes the process go up over time and if $\rho(t) > \mu_\rho$ it makes the process go down over time, so the process will always be inclined to move in the direction of μ_ρ . How strong this effect is can be controlled with κ_ρ . For κ_ρ , the condition $0 \leq \kappa_\rho < 1$ must hold, since for negative κ_ρ the process would move away from its mean and for $\kappa_\rho > 1$ the process would move past its mean instead of towards it. The mean μ_ρ must be between -1 and 1, since a correlation value cannot exceed these values. The mean-reverting property can also be seen in reality on the stock market, with the correlation between stocks usually being around the same level for a period of time.

The term $\gamma_\rho \sqrt{1 - \rho^2(t)} dW(t)$ in (4.1) is the random part of the process, with W denoting a Wiener process. The term $\sqrt{1 - \rho^2(t)}$ makes sure the process does not cross its boundaries 1 and -1, because $\sqrt{1 - \rho^2(t)} \rightarrow 0$ for both $\rho \rightarrow 1$ and $\rho \rightarrow -1$. This means that whenever ρ approaches one of its boundaries, the random part of the process will start contributing less and less to the

total shift in ρ and the mean-reverting part will take the process back away from the boundary. The square root makes sure that this term does not contribute too much to the process when ρ is far from its boundaries. The parameter γ_ρ can be seen as a volatility parameter of the entire process, since it scales the term with the Wiener process in it. Since volatility must always be positive, we have $\gamma_\rho > 0$.

To ensure that the process will stay within its boundaries of -1 and 1, the following parameter constraint is given in [4]:

$$\kappa_\rho > \max\left(\frac{\gamma_\rho^2}{1 - \mu_\rho}, \frac{\gamma_\rho^2}{1 + \mu_\rho}\right). \quad (4.2)$$

If we substitute (4.2) into (4.1), we find

$$\lim_{\rho \rightarrow 1} d\rho(t) < -\gamma_\rho^2 dt, \quad \lim_{\rho \rightarrow -1} d\rho(t) > \gamma_\rho^2 dt.$$

Since γ_ρ^2 is always positive, this makes sure the path will go down if it approaches 1 and up if it approaches -1.

All parameters (i.e. κ_ρ , μ_ρ and γ_ρ) have ρ in their subscripts to denote that they can be functions of ρ . This gives even more versatility to the Jacobi process, but for simplicity reasons they will be modeled as constant values in this thesis.

4.2 Discretization schemes

The Jacobi process defined in (4.1) is an example of a so-called **stochastic differential equation**: an equation where a path is specified implicitly by specifying the shift in its value using the value of the path itself. The general formula for a stochastic differential equation for a path X is as follows:

$$dX(t) = \alpha(t, X(t))dt + \sigma(t, X(t))dW(t). \quad (4.3)$$

Every stochastic differential equation has a term dependent on the shift in time dt , and a random term dependent on the increment $dW(t)$ of a Wiener process $W(t)$. The solution X of a stochastic differential equation is a continuous path. However, we cannot simulate true continuity, so we have to use a discretization scheme to divide the time period into small intervals $[t_{i-1}, t_i]$ and change the terms appearing in (4.3) accordingly.

Two potentially useful discretization schemes for the Jacobi process are the Euler discretization and the Milstein discretization. In this section, we check how well both discretization schemes perform with the Jacobi process. At the end, we make a choice for which discretization scheme to use in the rest of the thesis.

4.2.1 Euler discretization

The simplest of the discretization schemes is the Euler discretization. It turns the time variable t into a finite number of equidistant points in time t_i . The path $X(t)$ is made discrete by taking their values at time t_i :

$$X_i := X(t_i),$$

and the shift in its value becomes $X_{i+1} - X_i$. The functions α and σ are evaluated for t_i, X_i and the shift in time dt is replaced by $\Delta t := t_{i+1} - t_i$. Finally, since Wiener increments $W(t_{i+1}) - W(t_i)$ are normally distributed with mean 0 and variance Δt , $dW(t)$ is replaced with $\sqrt{\Delta t}Z$, where Z is a draw from a standard normal distribution. Applying all this, the Euler discretization for a general stochastic differential equation (4.3) becomes

$$X_{i+1} = X_i + \alpha(t_i, X_i)\Delta t + \sigma(t_i, X_i)\Delta tZ. \quad (4.4)$$

The Jacobi process can be written as a general stochastic differential equation by taking $X = \rho$, $\alpha(t, \rho(t)) = \kappa(\mu - \rho(t))$ and $\sigma(t, \rho(t)) = \gamma\sqrt{1 - \rho^2(t)}$. Substituting this into (4.4), we get a discrete version of the Jacobi process:

$$\rho_{i+1} = \rho_i + \kappa(\mu - \rho_i)\Delta t + \gamma\sqrt{1 - \rho_i^2}\sqrt{\Delta t}Z. \quad (4.5)$$

Since computer programs like MATLAB can take draws from a standard normal distribution, this is a formula that can be used to numerically simulate a correlation path. An example of a simulation of a Jacobi process is shown in figure 4.1. In this example, there were values of the Jacobi path crossing the boundary of 1; these have been **truncated**, meaning any values higher than 1 were replaced by 1. For the continuous Jacobi process, it should not be possible to cross the boundaries because of parameter constraint (4.2). This is because for the path to become higher than 1, we must first have $\rho(t) = 1$. In this case, the parameter constraint makes sure the path goes down immediately afterwards. After discretization however, we can have $\rho_i < 1$ and $\rho_{i+1} > 1$ without the path ever being exactly equal to 1. Possible solutions to this problem are explored in section 4.3.

4.2.2 Milstein discretization

A slightly more complicated way to turn the Jacobi process into a discrete path is by using the Milstein discretization. This discretization takes one additional term from the Itô-Taylor expansion, the stochastic equivalent of the Taylor expansion. For a general stochastic differential equation as described in (4.3), the discretization looks like this:

$$X_{i+1} = X_i + \alpha(t_i, x_i)\Delta t + \sigma(t_i, x_i)\sqrt{\Delta t}Z + \frac{1}{2}\sigma(t_i, x_i)(\Delta tZ^2 - \Delta t)\frac{\partial\sigma}{\partial x}(t_i, x_i), \quad (4.6)$$

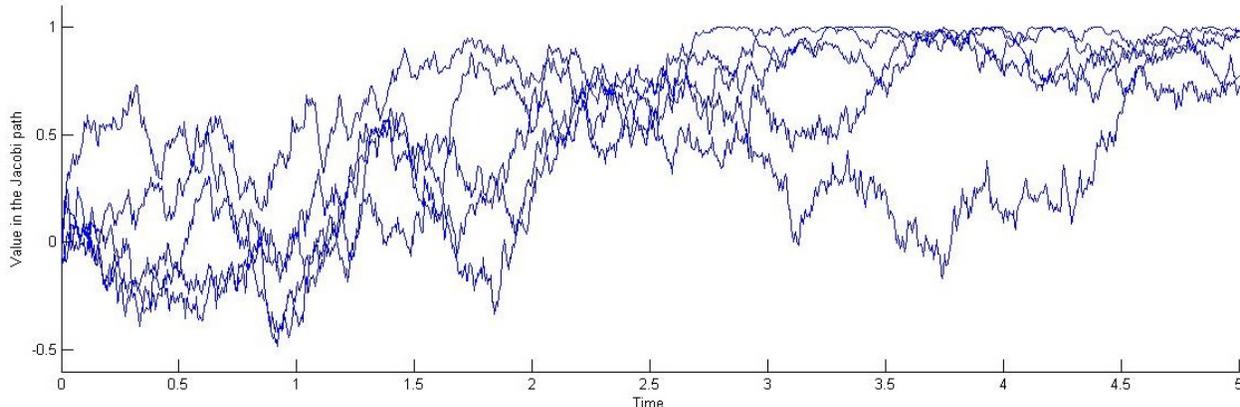


Figure 4.1: 5 simulations of Jacobi paths with values above 1 truncated to 1. $\rho_0 = 0$, $\kappa = 0.5$, $\mu = 0.6$, $\gamma = 0.6$, $\Delta t = 0.005$.

where Z is a draw from a standard normal distribution. For the Jacobi process

$$d\rho(t) = \kappa(\mu - \rho(t))dt + \gamma\sqrt{1 - \rho^2(t)}dW(t),$$

we have for the partial derivative:

$$\frac{\partial\sigma}{\partial x} = \frac{-\gamma\rho(t)}{\sqrt{1 - \rho^2(t)}}.$$

Substituting this and the other necessary functions into the Milstein discretization, equation (4.6):

$$\rho_{i+1} = \rho_i + \Delta t\kappa(\mu - \rho_i) + \sqrt{\Delta t}Z\gamma\sqrt{1 - \rho_i^2} + \frac{1}{2}\gamma^2\rho_i(\Delta tZ^2 - \Delta t).$$

Notice that in the last term, the part $\sqrt{1 - \rho_i^2}$ in σ is canceled out because it also appears in the denominator of $\frac{\partial\sigma}{\partial x}(t_i, \rho_i)$. Using this formula, we can simulate the Jacobi process in Matlab.

In figure 4.2, 5 paths simulated using the Milstein discretization are shown. The parameters are exactly the same as in figure 4.1. The figures obviously look very similar, since they model the same process. An interesting observation is that the amount of exceedances of the boundary $\rho = 1$ does not appear to be much lower than for the Euler discretization. This will be looked into in section 4.2.3.

4.2.3 Comparison of the discretization schemes

In this section, we will compare the Euler discretization scheme with the Milstein discretization scheme. A good way to do this is by comparing the amount

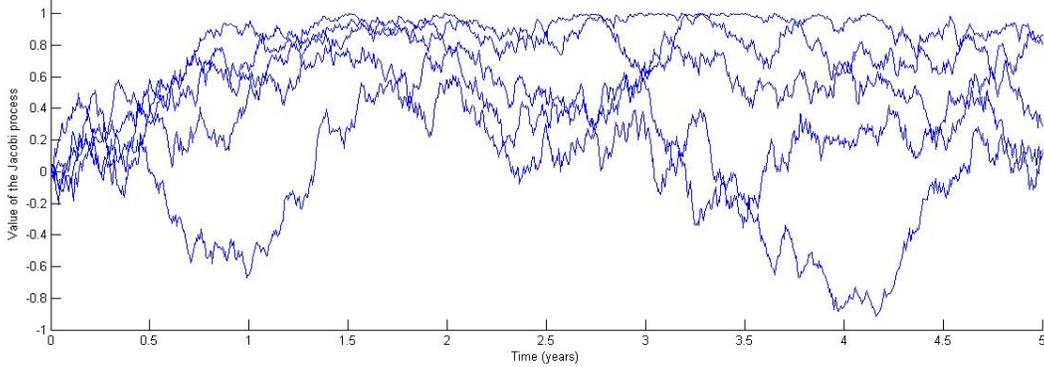


Figure 4.2: 5 simulations of Jacobi paths simulated using the Milstein discretization with values above 1 truncated to 1. $\rho_0 = 0$, $\kappa = 0.5$, $\mu = 0.6$, $\gamma = 0.6$, $\Delta t = 0.005$.

N	T = 1	Δt	T = 5	Δt	T = 10	Δt	T = 20	Δt	T = 50	Δt
100	0.019	0.01	0.482	0.05	0.804	0.1	0.971	0.2	1	0.5
1000	0.014	0.001	0.464	0.005	0.771	0.01	0.976	0.02	1	0.05
10000	0.009	0.0001	0.444	0.0005	0.792	0.001	0.958	0.002	1	0.005
100000	0.014	$1 * 10^{-5}$	0.433	$5 * 10^{-5}$	0.786	0.0001	0.961	0.0002	1	0.0005

Table 4.1: The proportion of exceedances for different amounts of timesteps N and different end times T . Δt represents the timestep size. The paths are modeled using Euler discretization. $\rho_0 = 0$, $\kappa = 0.5$, $\mu = 0.6$, $\gamma = 0.6$, $M = 1000$.

of boundary exceedances for both discretization schemes. These amounts will be calculated and with them, a decision will be made about which discretization scheme best fits our needs and will be used for the rest of the thesis.

If we look at figures 4.1 and 4.2, our first idea is that the amount of exceedances do not differ much between the two discretization schemes. This is not completely unexpected, since both discretization schemes have the same order of weak convergence: the convergence of the simulated process value to the true value for $\Delta t \rightarrow 0$. (See [4] for details). To investigate this further, a table comparing the timestep sizes to the proportion of paths exceeding 1 or -1 is shown in table 4.1 for Jacobi discretization and in 4.2 for Milstein discretization.

Tables 4.1 and 4.2 show the amounts of exceedances for different timestep sizes Δt and end times T for Euler discretization and Milstein discretization. The reaction of the amount of exceedances to the timestep size and the end time will be discussed in section 4.3. Here we note that the tables show almost exactly the same amounts of exceedances for the Euler discretization as for the

N	T = 1	Δt	T = 5	Δt	T = 10	Δt	T = 20	Δt	T = 50	Δt
100	0.024	0.01	0.504	0.05	0.839	0.1	0.975	0.2	1	0.5
1000	0.014	0.001	0.494	0.005	0.813	0.01	0.977	0.02	1	0.05
10000	0.017	0.0001	0.440	0.0005	0.801	0.001	0.974	0.002	1	0.005
100000	0.016	$1 * 10^{-5}$	0.434	$5 * 10^{-5}$	0.786	0.0001	0.960	0.0002	1	0.0005

Table 4.2: The proportion of paths modeled using Milstein discretization exceeding 1 or -1, for different amounts of timesteps N and different end times T . Δt represents the timestep size. $\rho_0 = 0, \kappa = 0.5, \mu = 0.6, \gamma = 0.6, M = 1000$.

Milstein discretization. Even though the Milstein discretization is slightly more complicated than the Euler discretization, adding an additional term with a partial derivative, this apparently does not result in a lower amount of exceedances. This is because they have the same order of weak convergence, as pointed out earlier. The increased complexity of the Milstein discretization turns out not to have any effects for modeling the Jacobi process, so we use the Euler discretization in the rest of this thesis.

4.3 Exceeding the boundaries

Even though the truncated path shown in figure 4.1 is theoretically a viable correlation path, the values have been changed. That means that this is no longer a true Jacobi path, so it is not a good simulation of the Jacobi process anymore. To prevent this from happening, we need to make sure that we do not have to truncate. For that goal, we simulate 1000 paths at once and look at the proportion of paths for which at least one value has been truncated. First, we compare this proportion to a value derived from the parameter constraint (4.2) on the Jacobi process. The parameter constraint is repeated here:

$$\kappa > \max\left(\frac{\gamma^2}{1-\mu}, \frac{\gamma^2}{1+\mu}\right).$$

Rewriting this constraint to be described by a single variable, we get:

$$A := \max\left(\kappa - \frac{\gamma^2}{1-\mu}, \kappa - \frac{\gamma^2}{1+\mu}\right) > 0.$$

In table 4.3, this value A is compared to the proportion of exceedances while moving around the parameters γ , κ and μ . It appears that A is not representative for the proportion of exceedances. For example, for $\kappa = 0.5, \gamma = 0.6, \mu = 0.4$: $A = 0.2429$ and $\#exc = 0.2630$, or 0.2470 in a different set of simulations with the same parameters. However, for $\kappa = 0.30, \gamma = 0.30, \mu = 0.40$: $A = 0.2357$, but $\#exc = 0.0000$. In these two cases, A is almost the same, but the number of exceedances vastly differs. This means that the number of exceedances and the value of A react differently to changes in the underlying

#exc	A	κ	γ	μ
0,1240	0,4429	0,70	0,60	0,40
0,1830	0,3429	0,60	0,60	0,40
0,2630	0,2429	0,50	0,60	0,40
0,3380	0,1429	0,40	0,60	0,40
0,3580	0,0429	0,30	0,60	0,40
0,0000	0,2357	0,30	0,30	0,40
0,0010	0,1357	0,20	0,30	0,40
0,0170	0,0357	0,10	0,30	0,40
0,0000	0,6357	0,70	0,30	0,40
0,0000	0,4357	0,50	0,30	0,40
0,2470	0,2429	0,50	0,60	0,40
0,3180	0,2600	0,50	0,60	0,50
0,4860	0,2750	0,50	0,60	0,60

Table 4.3: The proportion of paths exceeding 1 or -1 at least once, with 1000 simulations for every combination of parameters. For every simulation, $\rho_0 = 0$, $T = 5$, $\Delta t = 0.005$. The Jacobi discretization is used.

parameters.

Note that, as mentioned above for the case of $\kappa = 0.5, \gamma = 0.6, \mu = 0.4$, there may be a difference in the number of exceedances per simulation even when the underlying variables are kept the same. This means the amount of simulations (1000) is not enough to provide an accurate value for the proportion of exceedances. It is accurate enough, however, to show that A is not representative for the proportion of exceedances: a difference of around 0.25 is too big to be accounted for by the inaccuracy of the simulations.

To find a quantity that is representative of the amount of exceedances, we try looking at the process itself and the probability that a path crosses 1 or -1. Unfortunately, this probability cannot be written in terms of the underlying parameters. This is firstly because we only have an implicit formula for the path ρ , where ρ_{i+1} depends on ρ_i in several ways, and secondly because there is no explicit cumulative distribution function for the normal distribution, which Z follows.

Intuitively, we expect the amount of exceedances to be related to the size of the timesteps. To check this, the proportion of exceedances is compared with the amount of timesteps and the end times in table 4.1. The amount of exceedances does not appear to be connected to the amount of timesteps N . I personally think this is very strange at first sight. Usually, lowering the size of the timesteps makes a numerical simulation more accurate. However, a high amount of exceedances means inaccuracy of the simulation, because there

cannot be any exceedances in the continuous Jacobi process. Because this phenomenon seems so strange, it will be discussed further below.

The end time T does matter for the amount of exceedances in table 4.1. For $T = 1$, less than 2% of all paths exceeds 1 or -1 for each N . This proportion is already much higher for $T = 5$, and continues to rise for higher T , until at $T = 50$ all paths exceed 1 or -1.

The unimportance of the timestep size

In table 4.1, it appears as if decreasing Δt for a certain end time T has little to no effect on the proportion of paths exceeding 1 or -1. However, it must be noted that since the number of timesteps N is increased as well, there are more ‘opportunities’ (discrete points in time) where ρ can cross these values. Since lowering Δt without increasing N , namely by decreasing T , does have effect, the most logical explanation would be this: the positive effect in decreasing the amount of exceedances gained from lowering Δt is exactly canceled out by the negative effect gained from increasing N . In this part, it will be checked if this idea is in accordance with the theory.

First, we look at the formula used to simulate the Jacobi process, obtained with Euler discretization:

$$\rho_{i+1} = \rho_i + \Delta t \kappa(\mu - \rho_i) + \sqrt{\Delta t} Z \gamma \sqrt{1 - \rho_i^2}.$$

The significance of Δt here is a scaling factor for both the mean reverting part and, in a lesser manner, the random part. Now, let us look at a random value ρ_i and the probability that it crosses 1. We will be comparing this probability for different Δt and different amounts of timesteps, so we can disregard the probability that it crosses -1 (since it is calculated in exactly the same way and thus has the exact same relationship for different Δt and N) and we can assume without loss of generality that $\rho_i \geq 0$. For ρ_{i+1} to be larger than 1, we need the following:

$$\rho_{i+1} - \rho_i = \Delta t \kappa(\mu - \rho_i) + \sqrt{\Delta t} Z \gamma \sqrt{1 - \rho_i^2} > 1 - \rho_i.$$

Since we keep ρ_i and all parameters fixed for this argumentation, we can substitute $\kappa(\mu - \rho_i)$ with a constant c_1 and $\gamma \sqrt{1 - \rho_i^2}$ with another constant c_2 . Then we get

$$c_1 \Delta t + c_2 \sqrt{\Delta t} Z > 1 - \rho_i. \quad (4.7)$$

If we know $c_1, c_2, \Delta t$ and ρ_i , a probability can be calculated from this formula. For now, we will compare it to a case with $\delta t = \frac{\Delta t}{10}$ and look at the probability that $\rho_{i+10} > 1$. Looking from ρ_i , we need to look 10 timesteps ahead to calculate this probability:

$$\sum_{n=1}^{10} \left(c_1 \delta t + c_2 \sqrt{\delta t} Z_n \right) > 1 - \rho_i,$$

	M = 10	M = 100	M = 1 000	M = 10 000	M = 100 000
Mean	0.10624	0.11024	0.15608	0.15342	0.15275
Variance	0.046396	0.053783	0.045237	0.046241	0.046427

Table 4.4: The mean and variance of all simulated values $\rho(T)$, for different amounts M of simulated paths. There are 0 paths exceeding 1 or -1 for each M .

where each Z_n is an independent draw from a standard normal distribution. Defining $Z := \sqrt{\frac{1}{10}} \sum_{n=1}^{10} Z_n$, we get

$$c_1 \Delta t + c_2 \sqrt{\Delta t} Z > 1 - \rho_i.$$

The random variable $\sum_{n=1}^{10} Z_n$ is normally distributed with mean 0 and variance 10, so Z follows the standard normal distribution. This means the above equation is exactly equal to equation (4.7). This shows that the probability for ρ crossing 1 or -1 in one timestep is exactly the same as with ten timesteps that are ten times smaller. This explains why the proportion of exceedances stays the same if we decrease the timestep size by only increasing the amount of timesteps with the same factor.

4.4 Correlation swaps

To model correlation swaps using a Jacobi process, we will use the mean of many simulated processes as the expected value of the process. To be able to use this as an expected value, we have two requirements: the amount of exceedances should be as low as possible, and the mean must converge to a single value for large numbers of simulations M . How to keep the exceedances low has been discussed in the previous section: we will take T , γ and μ low, and κ high. It appears that it does not matter what we choose for N and whether we use Euler or Milstein discretization schemes, so we will take $N = 100$ and use the Euler discretization scheme. For $T = 1$, $\gamma = 0.3$, $\mu = 0.3$ and $\kappa = 0.7$, some test runs of the simulation all yield 0 exceedances.

First, we will check if the mean of the results of the simulation at $t = T$ converges. Looking at table 4.4, we can see that for M higher than 1000, the mean is stable. The variance is quite low here, because we have T and γ low. But the mean does converge, so we can at least use it as the expected value of the correlation with these parameters. (If the variance is higher, the mean will still converge as long as there are little to no exceedances, although M may need to be higher.) In figure 4.3, it can be seen that the distribution followed by the simulated correlation values at time T is a normal distribution.

The Jacobi process can be used to model the correlation path used in [1], if we take the right values (or functions) for γ, μ, κ and ρ_0 . This is an inter-

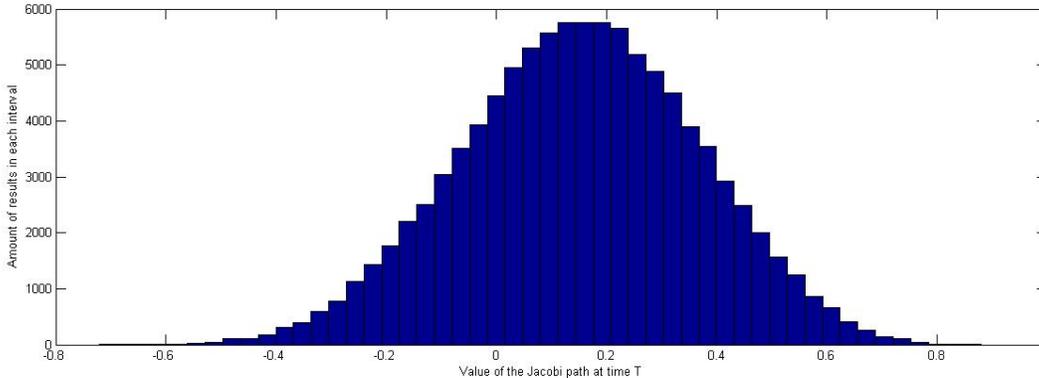


Figure 4.3: A histogram showing how the results at time $t = T$ of the simulations of the Jacobi process are distributed. $M = 100000$, $T = 1$, $\gamma = 0.3$, $\mu = 0.3$ and $\kappa = 0.7$. The values are normally distributed around their mean.

esting topic for further research. The forward dynamics of the correlation path $\hat{\rho}_t^* = v_t^{*I}/\bar{v}_t^{*S}$ will first have to be defined. This definition of $\hat{\rho}_t^*$ will have to be rewritten in order to be able to use the forward dynamics of v_t^{*I} and \bar{v}_t^{*S} , perhaps some approximations will have to be made. When a formula for the forward dynamics of $\hat{\rho}_t^*$ is obtained, it can be compared with the forward dynamics of the Jacobi process, as defined in equation (4.1). This will show what the best definitions for κ_ρ , μ_ρ and γ_ρ would be; they would probably have to be functions of ρ and t . Even with the right definitions, the Jacobi process would still be different from the correlation path in [1], because of the term $\sqrt{1 - \rho^2(t)}$ in the forward dynamics of the Jacobi process. This is what makes it interesting, however. This term aims to keep the value of the process below 1, so it may solve the problem of the correlation path defined in [1] getting higher than 1. Inspecting this difference may even lead to an improvement of the model in [1].

Another use of the Jacobi process would be to model the correlation swap in a different way. It could be used to model the correlation defined in [1] as described above, but instead of inspecting the differences in the correlation paths themselves, a swap on the correlation path could be inspected. The fair value of this swap could be obtained by running a large number of simulations of the correlation path and taking the mean of the swap's payoff as the swap price (discounting to compensate for the interest rate as necessary). This price could then be compared to the price given in [1], and an explanation for any difference could be found. Although these ideas are interesting, there is no room to explore them in this thesis.

Chapter 5

Hedging

5.1 Introduction

In this section, we will look at how to hedge the correlation swap as described in [1]. In [1], a hedging strategy is described using so-called variance dispersion trades. Hedging the correlation swap here means that a portfolio consisting of the swap and the hedging products should be **vega-neutral**: independent of changes in both index variance and constituent variance. We go deeper into dispersion trades and check if it cancels out the dependence on index variance and constituent variance (the **vegas**) of the correlation swap. Using this vega-neutral portfolio, we find derive a partial differential equation for the no-arbitrage price of a correlation swap with strike 0. We confirm that the value given in [1] satisfies this equation. Finally, we use this no-arbitrage price of a correlation swap with strike 0 to find the fair strike of a correlation swap to make its price 0.

5.2 The method of variance dispersion trades

5.2.1 What are variance dispersion trades?

Variance dispersion is a financial term describing the difference between constituent variance and index variance in a portfolio of stocks. It follows that variance dispersion trades are spread trades between constituent variance and index variance, with payoff:

$$D(\beta, \tau) \equiv \beta (\bar{\sigma}^S(\tau))^2 - (\sigma^I(\tau))^2,$$

where β is a positive constant. The motivation for entering such a trade is usually to trade correlation between constituent stocks. This is contained in the index volatility, and the unwanted volatility exposure is hedged away through the opposite position in constituent volatility. This makes it an interesting op-

tion for hedging a correlation swap.

There are several choices for β . One of them is the **fundamental** approach: $\beta = \left(\frac{\sigma_{t_0}^{*I}}{\bar{\sigma}_{t_0}^{*S}}\right)^2 = \hat{\rho}_{t_0}^*$. Substituting this into the formula of the variance dispersion payoff, we get:

$$D(\beta, \tau) = \left(\frac{\sigma_{t_0}^{*I}}{\bar{\sigma}_{t_0}^{*S}}\right)^2 (\bar{\sigma}^S(\tau))^2 - (\sigma^I(\tau))^2 \quad (5.1)$$

$$= \left(\frac{\sigma_{t_0}^{*I}}{\bar{\sigma}_{t_0}^{*S}}\right)^2 (\bar{\sigma}^S)^2 - \left(\frac{\sigma^I}{\bar{\sigma}^S}\right)^2 (\bar{\sigma}^S)^2 \quad (5.2)$$

$$= (\hat{\rho}_{t_0}^* - \hat{\rho}) (\bar{\sigma}^S)^2 \quad (5.3)$$

So we can rewrite the variance dispersion payoff as: $D(\beta) = (\hat{\rho}_{t_0}^* - \hat{\rho}) (\bar{\sigma}^S)^2$. This essentially turns the dispersion trade into a correlation swap, with realized constituent variance playing the role of a scaling factor. For the fundamental approach, the variance dispersion trade is vega-neutral at t_0 , meaning the position is not sensitive to fluctuations in either constituent or index volatility.

5.2.2 The hedging property

In [1] it is stated that the correlation swap described in the article can be replicated by dynamically trading variance dispersion trades. The word *dynamically* here means the hedging portfolio is updated over time. The beta for this dispersion trade is

$$\beta_t = \frac{v_t^{*I}}{\bar{v}_t^{*S}}.$$

This beta is similar to the one used in the fundamental approach described in section 5.2.1. The differences are that the theoretical volatilities $(\sigma_{t_0}^{*I})^2$ and $(\bar{\sigma}_{t_0}^{*S})^2$ are replaced by the modeled variance paths v_t^{*I} and \bar{v}_t^{*S} , and that to make the hedge time dependent, the β is updated by taking values at time t instead of t_0 . This section will show that if one owns a correlation swap as described in [1] and the correct amount of variance dispersion trades with beta as described above, their total investment is not affected by small changes in volatility. What this ‘correct amount’ of variance dispersion trades is will also be calculated in this section.

We first look at the definition of vega: the derivative of the price of a financial product with respect to the variance. First, we look at the correlation swap. Its value at time t according to [1] is:

$$c_t = \frac{v_t^{*I}}{\bar{v}_t^{*S}} \exp \left[\frac{4}{3} (\bar{\omega}_S^2 - \chi \bar{\omega}_S \omega_I) T \left(\frac{T-t}{T} \right)^3 \right].$$

It has two underlying variances (v_t^{*I} and \bar{v}_t^{*S}), so it also has two vegas:

$$\nu_{c_t}^I = \frac{\partial c_t}{\partial v_t^{*I}} = \frac{c_t}{v_t^{*I}}, \quad (5.4)$$

$$\nu_{c_t}^S = \frac{\partial c_t}{\partial \bar{v}_t^{*S}} = -\frac{c_t}{\bar{v}_t^{*S}}. \quad (5.5)$$

A variance dispersion trade entered at time t_0 has the following value at time t :

$$D_t^* = \frac{v_{t_0}^{*I}}{\bar{v}_{t_0}^{*S}} \bar{v}_t^{*S} - v_t^{*I}.$$

Its vegas are:

$$\nu_{D_t}^I = \frac{\partial D_t^*}{\partial v_t^{*I}} = -1,$$

$$\nu_{D_t}^S = \frac{\partial D_t^*}{\partial \bar{v}_t^{*S}} = \frac{v_{t_0}^{*I}}{\bar{v}_{t_0}^{*S}}.$$

These are both constants, while the vegas for c_t are dependent on v_t^{*I} (only $\nu_{c_t}^S$) and on \bar{v}_t^{*S} . To hedge the correlation swap, the vegas of the correlation swap and the dispersion trade should cancel each other out. This is impossible if we try a ‘static hedge’, meaning we enter a position at t_0 and do not update it. In this case, the vegas of the dispersion trade are not time-dependent, as shown above, so they can never cancel out the vegas of the correlation swap, which are time-dependent. This means that a static hedge is impossible.

So we look at dynamic hedging. For this, we continuously update our position in the variance dispersion trade, meaning we can also change our beta. Bossu suggests we use a beta of $\beta_t = v_t^{*I}/\bar{v}_t^{*S}$. To get this beta, we substitute t_0 in the vegas of the variance dispersion trade $\nu_{D_t}^I$ and $\nu_{D_t}^S$ with t . This is because t_0 represents the time when the dispersion trade is entered, and for dynamic hedging, new dispersion trades are entered continuously at time t . Doing this, we get

$$\nu_{D_t}^S = \frac{v_t^{*I}}{\bar{v}_t^{*S}},$$

and $\nu_{D_t}^I$ remains -1 . Since we can choose freely how much of the dispersion trade we buy, depending on how much of the correlation swap we have, we only need to make sure the ratio between the vegas is the same for the correlation swap and the dispersion trade. This appears to be the case:

$$\frac{\nu_{c_t}^I}{\nu_{c_t}^S} = -\frac{\bar{v}_t^{*S}}{v_t^{*I}} = \frac{\nu_{D_t}^I}{\nu_{D_t}^S}. \quad (5.6)$$

This means that dynamically trading a variance dispersion trade with

$$\beta_t = \frac{v_t^{*I}}{\bar{v}_t^{*S}}$$

can indeed hedge the correlation swap described in [1].

Now we will find a formula for how much of the dispersion trade we should own for each unit of the correlation swap owned. This ratio may depend on v_t^{*I} and \bar{v}_t^{*S} . First, let us see what the changes in c_t and D_t are when v_t^{*I} goes up by 1. This is equal to the vega values:

$$\begin{aligned}\nu_{c_t}^I &= \frac{c_t}{v_t^{*I}}, \\ \nu_{D_t}^I &= -1.\end{aligned}$$

For these changes to cancel each other out, we need $\nu_{c_t}^I = -\nu_{D_t}^I$. This means we need to own c_t/v_t^{*I} of dispersion trades. Because of (5.6), this should also cancel out the changes in c_t and D_t when \bar{v}_t^{*S} goes up by 1. We can check this in the following way:

$$\frac{c_t}{v_t^{*I}} \nu_{D_t}^S = \frac{c_t}{v_t^{*I}} \frac{v_t^{*I}}{\bar{v}_t^{*S}} = \frac{c_t}{\bar{v}_t^{*S}} = -\nu_{c_t}^S.$$

This means that changes in \bar{v}_t^{*S} will also cancel each other out if we own c_t/v_t^{*I} of dispersion trades, so the portfolio consisting of a correlation swap and this amount of dispersion trades is vega-neutral.

Dynamic hedging

As stated before, dynamic hedging means the hedge is updated continuously. In reality however, this is impossible. Here we will make a discretization to show more clearly what happens in reality with dynamic hedging.

Since all of the variables in the dispersion trade are time-dependent, we will first make clear which variables change automatically over time and which have to be updated. By definition, the dispersion trade consists of a long position of size β in constituent variance and a short position of size 1 in index variance:

$$D_t^* = \beta \bar{v}_t^{*S} - v_t^{*I} \tag{5.7}$$

This means that if the dispersion trade is not updated, only these variables \bar{v}_t^{*S} and v_t^{*I} will change over time. The ratio parameter β , which we take to be v_t^{*I}/\bar{v}_t^{*S} , as well as the amount c_t/v_t^{*I} of dispersion trades will need to be updated.

We can approximate continuous updating by updating the trade very often. For this, we define $0 = t_0 < t_1 < t_2 < \dots < t_n = T$ (i.e. a partition of $[0, T]$) such that $t_i - t_{i-1}$ are equal for all $1 < i < n$. We update the dispersion trade at each time t_i . This means we sell the old dispersion trade we had, and enter a new one.

Recall that the payoff of a dispersion trade with $\beta = \hat{\rho}_{t_0}^*$ is described in (5.3), so its value at any time t is given by taking the conditional expectation:

$$D_t^*(\tau) = \mathbb{E} \left[(\hat{\rho}_{t_0}^*(\tau) - \hat{\rho}(\tau)) (\bar{\sigma}^S(\tau))^2 \mid \hat{\rho}([0, t]), \bar{\sigma}^S([0, t]) \right]. \quad (5.8)$$

Here, t_0 indicates the time at which the dispersion trade was entered, and t indicates the current time. If we look at a point where the dispersion trade is updated, the current time becomes t_i and the time at which the dispersion trade was entered becomes t_{i-1} . Since the position in this dispersion trade will be terminated, we get a payoff out of it, which we will call D_i . Filling in (5.8) gives us the following expression for D_i :

$$\begin{aligned} D_i &= \mathbb{E} \left[(\hat{\rho}_{t_{i-1}}^* - \hat{\rho}) (\bar{\sigma}^S)^2 \mid \hat{\rho}_{t_i}, \bar{\sigma}_{t_i}^S \right] \\ &= (\hat{\rho}_{t_{i-1}}^* - \hat{\rho}_{t_i}^*) (\bar{\sigma}_{t_i}^{S*})^2. \end{aligned}$$

If we take $\delta t = t_i - t_{i-1}$ small enough, this kind of dynamic trading approximates continuous updating quite well. Note that at the time when a (new) dispersion trade is entered, it always has price zero. This can be seen by looking at the price of the dispersion trade as the expected value of its payoff at time t_i when the trade is entered:

$$\begin{aligned} D_i^*(t_i) &= \mathbb{E} \left[(\hat{\rho}_{t_i}^* - \hat{\rho}) (\bar{\sigma}^S)^2 \mid \hat{\rho}_{t_i}, \bar{\sigma}_{t_i}^S \right] \\ &= (\hat{\rho}_{t_i}^* - \hat{\rho}_{t_i}^*) (\bar{\sigma}_{t_i}^{S*})^2 \\ &= 0. \end{aligned}$$

This means that entering the dispersion trade is free, so the change in time of the value of a dynamically hedged dispersion trade consists of the payoffs D_i the dispersion trades generate in between updates.

5.3 Fair strike of a correlation swap

For swaps, strike K is always chosen such that the up-front price of the swap is zero. Because of this, we're not interested in the price in the swap, but in this fair strike K . To find it, we look at a correlation swap with strike 0 and determine its no-arbitrage price. With this price, we should be able to find the fair strike K such that the price becomes 0.

To achieve this, we use the hedging products discussed above. We look at the portfolio consisting of a correlation swap and the right amount of dispersion trades, such that the portfolio is vega-neutral. We take steps similar to the derivation of the Black-Scholes partial differential equation in [2] to reach a partial differential equation for the value of a correlation swap. Next, we show that the solution given in [1] satisfies this partial differential equation. Finally,

we use the pricing method used in [1] to get from this no-arbitrage price to the fair strike of the correlation swap.

5.3.1 The vega-neutral portfolio

We start by setting up a hedging position; we have shown in section 5.2 that a dynamically traded long position of c_t/v_t^{*I} in a variance dispersion trade with $\beta_t = v_t^{*I}/\bar{v}_t^{*S}$ continuously updated hedges a correlation swap. So we take this as our hedging position, so that we have a total portfolio Π consisting of a correlation swap and a position in dispersion trades:

$$\Pi_t = c_t + \frac{c_t}{v_t^{*I}} D_t^*. \quad (5.9)$$

Note that because we set strike $K = 0$, the value of the correlation swap is simply the value of the correlation c_t . Since this is a hedged portfolio, its changes over time $\Delta\Pi$ should be non-random. No arbitrage means that a non-random investment should always have a payoff equal to the payoff you would get from investing an amount of cash equal to the price of the investment in a bank account. Mathematically, this means the growth of the portfolio over time should be equal to the interest rate:

$$\delta\Pi = r\delta t\Pi, \quad (5.10)$$

where r is the interest rate. This is already a PDE for the value of the portfolio, but it will have to be filled in further to get an expression for c_t . First we can write down the change in value, or P&L (profit and loss) variation of the portfolio at time t . We regard a very small shift in time at first, so that the portfolio is not updated during this time.

$$\delta\Pi_t = \delta c_t + \frac{c_t}{v_t^{*I}} \delta D_t^*. \quad (5.11)$$

We now use a Taylor expansion of δc with respect to the time, the implied volatility and the constituent volatility:

$$\begin{aligned} \delta c_t = & \frac{\partial c_t}{\partial t} dt + \frac{\partial c_t}{\partial v^{*I}} dv_t^{*I} + \frac{\partial c_t}{\partial \bar{v}^{*S}} d\bar{v}_t^{*S} \\ & + \frac{1}{2} \left(\frac{\partial^2 c_t}{(\partial v^{*I})^2} (dv_t^{*I})^2 + \frac{\partial^2 c_t}{(\partial \bar{v}^{*S})^2} (d\bar{v}_t^{*S})^2 + 2 \frac{\partial^2 c_t}{\partial v^{*I} \partial \bar{v}^{*S}} dv_t^{*I} d\bar{v}_t^{*S} \right). \end{aligned} \quad (5.12)$$

Note that the squared terms of variance shift (the part between brackets) are being kept, because these will make a contribution of size proportional to dt . [2] We also need a more extensive representation of δD_t^* . We can start by

simply filling in its definition:

$$\begin{aligned}\delta D_t^* &= \delta \left[\frac{v_t^{*I}}{\bar{v}_t^{*S}} \bar{v}_t^{*S} - v_t^{*S} \right] \\ &= \frac{v_t^{*I}}{\bar{v}_t^{*S}} \delta \bar{v}_t^{*S} - \delta v_t^{*I}\end{aligned}\quad (5.13)$$

We know from (5.4) and (5.5) that

$$\frac{\partial c_t}{\partial v^{*I}} = \frac{c_t}{v_t^{*I}}, \quad \frac{\partial c_t}{\partial \bar{v}^{*S}} = -\frac{c_t}{\bar{v}_t^{*S}}.$$

This means that the second and third terms from (5.12) are canceled out by $(c_t/v_t^{*I}) \delta D_t^*$ when substituting (5.12) and (5.13) into (5.11). This leaves us with

$$\delta \Pi_t = \frac{\partial c_t}{\partial t} dt + \frac{1}{2} \left(\frac{\partial^2 c_t}{(\partial v^{*I})^2} (dv_t^{*I})^2 + \frac{\partial^2 c_t}{(\partial \bar{v}^{*S})^2} (d\bar{v}_t^{*S})^2 + 2 \frac{\partial^2 c_t}{\partial v^{*I} \partial \bar{v}^{*S}} dv_t^{*I} d\bar{v}_t^{*S} \right). \quad (5.14)$$

We can combine (3.22) and (3.23) with the quadratic variation t of a standard Brownian motion and the correlation between the two Brownian motions $(dW_t^{*I})(dW_t^{*S}) = \chi dt$ to substitute dv_t^{*I} and $d\bar{v}_t^{*S}$ in (5.14):

$$\begin{aligned}\Delta \Pi_t &= \frac{\partial c_t}{\partial t} \Delta t + 2 \left(\frac{\partial^2 c_t}{(\partial v^{*I})^2} \omega_I^2 (v_t^{*I})^2 + \frac{\partial^2 c_t}{(\partial \bar{v}^{*S})^2} \bar{\omega}_S^2 (\bar{v}_t^{*S})^2 \right) \left(\frac{T-t}{T} \right)^2 \Delta t \\ &\quad + 4 \frac{\partial^2 c_t}{\partial v^{*I} \partial \bar{v}^{*S}} \omega_I \bar{\omega}_S \chi v_t^{*I} \bar{v}_t^{*S} \left(\frac{T-t}{T} \right)^2 \Delta t.\end{aligned}\quad (5.15)$$

Note that the δ and d have been substituted by Δ : this is to indicate that we've summed over all small time intervals so that we can approximate the squared Brownian motion increments $(dv_t^{*I})^2$ and $(d\bar{v}_t^{*S})^2$ with their quadratic variation dt , as well as approximate the product of the two Brownian motion increments $dv_t^{*I} d\bar{v}_t^{*S}$ with their correlation χdt .

Now let us take a moment to look at (5.15) and analyze what it means. The portfolio is hedged, but the shift in value is not zero, so it will still change over time. This is because we have only vega-hedged the correlation swap: the vega is the derivative with respect to variance, and we can see that these derivatives do not appear in (5.15) for neither constituent variance nor index variance. These were canceled out by the contribution (5.13) of the dispersion trade. The shift in portfolio value is still dependent on the derivative with respect to time, called the **theta**, and the second derivatives with respect to both types of variance, called the **volgas**. Further research may be done to make the portfolio value independent from these values as well.

If we assume the volatility of volatility parameters ω_I and $\bar{\omega}_S$ and the correlation between volatilities χ to be constant over time, then (5.15) contains only non-random values, so we can apply (5.10). We can apply the formula to a large time shift, since (5.15) is non-random for any time period. We choose an initial time t when the dispersion trade has just been updated, and we view a time period long enough for (5.15) to hold. The left hand side of (5.10) is of course given by (5.15) and for the right hand side we can use the composition of the portfolio given by (5.9). Since the dispersion trade has just been updated, its value is (approximately) zero, so the portfolio value is $\Pi_t = c_t$. Substituting this and (5.15) into (5.10) gives

$$rc_t = \frac{\partial c_t}{\partial t} + 2 \left(\frac{\partial^2 c_t}{(\partial v^{*I})^2} \omega_I^2 (v_t^{*I})^2 + \frac{\partial^2 c_t}{(\partial \bar{v}^{*S})^2} \bar{\omega}_S^2 (\bar{v}_t^{*S})^2 \right) \left(\frac{T-t}{T} \right)^2 + 4 \frac{\partial^2 c_t}{\partial v^{*I} \partial \bar{v}^{*S}} \omega_I \bar{\omega}_S \chi v_t^{*I} \bar{v}_t^{*S} \left(\frac{T-t}{T} \right)^2. \quad (5.16)$$

The solution c_t of this PDE is the fair, no-arbitrage price of a correlation swap with strike zero.

5.3.2 Comparison to the solution in [1]

Instead of trying to solve (5.16), we check if the fair value for c_t given in [1] satisfies this PDE. For ease of notation, we set

$$u(t) := \exp \left[\frac{4}{3} (\bar{\omega}_S^2 - \chi \bar{\omega}_S \omega_I) T \left(\frac{T-t}{T} \right)^3 \right].$$

Using this, we will state the solution in [1] and all its derivatives appearing in (5.16):

$$c_t = \frac{v_t^{*I}}{\bar{v}_t^{*S}} u(t) \quad (5.17)$$

$$\frac{\partial c_t}{\partial t} = \left(\frac{d}{dt} \left(\frac{v_t^{*I}}{\bar{v}_t^{*S}} \right) - 4 \frac{v_t^{*I}}{\bar{v}_t^{*S}} (\bar{\omega}_S^2 - \chi \bar{\omega}_S \omega_I) \left(\frac{T-t}{T} \right)^2 \right) u(t) \quad (5.18)$$

$$\frac{\partial^2 c_t}{(\partial v^{*I})^2} = 0 \quad (5.19)$$

$$\frac{\partial^2 c_t}{(\partial \bar{v}^{*S})^2} = \frac{2v_t^{*I}}{(\bar{v}_t^{*S})^3} u(t) \quad (5.20)$$

$$\frac{\partial^2 c_t}{\partial v^{*I} \partial \bar{v}^{*S}} = \frac{-1}{(\bar{v}_t^{*S})^2} u(t). \quad (5.21)$$

Substituting (5.18), (5.19), (5.20) and (5.21) into (5.16), most terms cancel each other out, and we are left with

$$rc_t = u(t) \frac{d}{dt} \left(\frac{v_t^{*I}}{\bar{v}_t^{*S}} \right). \quad (5.22)$$

In [1], the interest rate is disregarded, which means r is assumed to be 0. So there is a spread between both sides of equation (5.22). This may be because the portfolio is only vega-hedged, but still has exposure to large changes in volatility and changes in other variables. Going deeper into this is a topic for further research. More information about the spread between a correlation swap and a dispersion trade can be found in [3].

For the remainder of this thesis, we will assume that this spread is small. This is motivated by the reasoning that the remaining exposure of the vega-neutral portfolio is small, so if the spread is indeed caused by this, it will not be a substantial spread. This means that the solution c_t given in [1] can be regarded as a solution of (5.16). We will not prove uniqueness of the solution mathematically, but if we keep in mind that the solution of (5.16) is the fair price of a correlation swap, it makes sense that there can only be one solution. That means (5.17) is the only solution of (5.16) and thus (5.17) is the fair value of a correlation swap with strike 0 in a world where there is no interest rate.

5.3.3 From no-arbitrage price to fair strike

We take the solution c_0 of (5.16) for time t_0 at which the correlation swap is entered, so that c_0 indicates the no-arbitrage price of the correlation swap. Using this solution, we can quickly derive what the strike K should be to make the correlation swap have price 0. The payoff of the correlation swap at maturity T is $c_T - K$. Another way to calculate the price c_0 of a correlation swap with strike 0 (different from the method we used here) is the method used in [1], which is:

$$c_0 = \mathbb{E}[c_T]. \quad (5.23)$$

In the same way, we get for a swap with price zero

$$0 = \mathbb{E}[c_T - K] = \mathbb{E}[c_T] - \mathbb{E}[K] = c_0 - K, \quad (5.24)$$

using (5.23) and the fact that K is known at time t_0 for the last step. Here we keep the assumption that there is no interest rate. This equation implies that $K = c_0$; the fair strike of the correlation swap to make the price 0 is equal to the fair price of a correlation swap with strike 0.

If we consider a non-zero interest rate r , the solution becomes a bit different. We no longer have $\mathbb{E}[K] = K$, but $\mathbb{E}[K] = e^{-r(T-t_0)}K$. The term $e^{-r(T-t_0)}$ is to account for the fact that if you put an amount K of cash in a bank account from time t_0 until time T , it will grow with interest rate r over that time. This is called the **discount factor** in financial terms. Substituting this into (5.24), we get

$$c_0 - e^{-r(T-t_0)}K = 0,$$

which yields for the fair strike

$$K = e^{r(T-t_0)}c_0. \quad (5.25)$$

This is the final solution to the problem of this thesis. Equation (5.25) gives the fair strike of a correlation swap to make the price equal to zero, as it should be for a swap. It depends on the solution c_0 of PDE (5.16), which is obtained by looking at a vega-hedged correlation swap and using the law of no arbitrage. This solution c_0 turns out to be equal to the solution (5.17) given in [1].

Chapter 6

Conclusion

In this thesis, we have thoroughly examined the toy model in [1]. All its definitions and assumptions were looked into and compared to the regular definitions of variance and correlation. Using these regular definitions, we came to a formula for correlation close to the one in [1], but there was still a difference between the weighted arithmetic average of volatilities and the constituent volatility as it was defined in [1]. The assumption that these definitions are very close to each other turned out to be equivalent to the assumption that stable stocks with low volatility are preferred over stocks with high volatility when creating a stock index, which is a reasonable assumption.

Next, we looked into Bossu's toy model for correlation. Its main flaw is that the correlation does not always stay below 1, which is caused by the lack of a mathematical condition enforcing the boundary. This inaccuracy in the model was caused by the assumption that the index volatility is lower than the constituent volatility, which was suggested because the model would otherwise become too complicated. We found that if we did simulations with a high correlation between volatilities and a low implied correlation at the starting time, the simulated index variance would stay below the simulated constituent variance, satisfying the theoretical assumption that this would hold. The correlation value modeled in this way could be multiplied with an adjustment factor to reach the fair correlation value suggested in [1].

In chapter 4 we looked at the Jacobi process. We compared the Euler discretization scheme with the Milstein discretization scheme for this process and found that the more complicated Milstein discretization scheme did not improve the accuracy of the simulations. We found that although the continuous Jacobi process cannot exceed its boundaries of 1 or -1, the discretization caused it to sometimes cross these boundaries. The proportion of paths within a simulation that exceeded these boundaries turned out not to depend on the timestep size, but on the end time and the parameters of the Jacobi process.

Furthermore, we showed that a correlation swap can be hedged by dynamically trading variance dispersion swaps. Using this hedged position, we derived a partial differential equation for the fair price of a correlation swap with strike 0, similar to the Black-Scholes equation. The main result of the thesis is that the solution given in [1] satisfies this partial differential equation and thus is the no-arbitrage price of a correlation swap with strike 0. Finally, this no-arbitrage price was used to find the fair strike of a correlation swap so that it would have price 0, as all swaps should.

6.1 Further research

The next step could be to connect the Jacobi process further to the toy model in [1] and its results. The parameters of the Jacobi process could be chosen so that the process would approximate the fair correlation path as given by [1]. This model could be used either to solve the problem that the correlation path in [1] does not stay under 1, or to model a correlation swap and check if the simulation results agree with the theory in [1].

A different way to improve the model in [1] would be to remove the assumption that index volatility is lower than constituent volatility. This would make the model considerably more complex, but it should also remove the problem of the correlation paths going above 1.

Something that could also be expanded into is the behaviour of the time derivative of the (non-adjusted) correlation path as described in [1]. While checking if the fair correlation given in [1] was a solution of our partial differential equation for fair correlation, we assumed this derivative to be close to zero, but of course the result would become more accurate if this assumption is not made.

Another assumption that could be removed is the assumption in [1] that there is no interest rate. In reality there is always interest rate, so taking this into consideration would make the model more widely applicable.

Bibliography

- [1] Sebastien Bossu, A New Approach For Modelling and Pricing Correlation Swaps, Dresdner Kleinwort, version of May 2007.
- [2] Desmond J. Higham, An Introduction to Financial Option Valuation, Cambridge University Press, version of April 2004.
- [3] Antoine Jacquier and Saad Slaoui, Variance Dispersion and Correlation Swaps, Department of Mathematics, Imperial College London and Zeliade Systems, Paris, version of April 2010.
- [4] C. W. Oosterlee and L. A. Grzelak, Models and Computation in Quantitative Finance, Manuscript to be published in 2017.
- [5] <https://vixforthepeople.com/2012/11/27/what-is-the-vix/>, VIX For The People, What is the VIX?, November 2012.
- [6] https://en.wikipedia.org/wiki/Correlation_and_dependence, Wikipedia, uploaded by user DenisBoigelot.