Politecnico di Milano
School of Industrial and Information Engineering
Master of Science in Mathematical Engineering

TITLE: very interesting subject, ain't it?

Supervisors: Prof. Daniele Marazzina
Prof. Ferdinando M. Ametrano

Master thesis by:
Samuele Vianello
ID: *********

Academic year 2017-2018

The secret to happiness is freedom.
And the secret to freedom is courage.

Thucydides

Contents

List of Tables

List of Figures

List of Algorithms

Abstract

Including Bitcoin in an investment portfolio increases portfolio diversification.

Acknowledgements

add acknowledgements

Thank you.

Chapter 1

Introduction

1.1 Thesis structure

Chapter 2

Correlation Analysis

In order to get an initial insight on how Bitcoin is correlated with other assets, we will perform a correlation analysis based on the empirical time series of our data. We will focus our attention on the logarithmic returns it is the standard practice. We will often refer to logarithmic returns simply as returns, only specifying their nature when it is necessary to avoid confusion.

2.1 Empirical Correlation of Returns

We first start by performing some statistical analysis on the data in order to estimate the distribution from which they are sampled. For this part, we will consider our data as successive samples of a N-dimensional vector in $\mathbb{R}^N$, where N is the number of assets:

$$\mathbf{x}_j = \begin{pmatrix} x_{1,j} \\ x_{2,j} \\ \vdots \\ x_{N,j} \end{pmatrix}, j = 1 \dots N_{sample}$$

Each element i of the vector $\mathbf{x}_j$ represents the j^{th} realization of the returns for asset i.

Following basic statistics, we can now compute the *sample mean* of our vectors of returns as:

$$\bar{\mathbf{x}} = \frac{1}{N_{sample}} \sum_{j=1}^{N_{sample}} \mathbf{x}_j = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_N \end{pmatrix}$$

where $\bar{x}_i = \frac{1}{N_{sample}} \sum_{j=1}^{N_{sample}} x_{i,j}$ is the sample mean of component i.

Now we compute the *sample covariance matrix* through the following formula:

$$\bar{\Sigma} = \frac{1}{N_{sample} - 1} \sum_{j=1}^{N_{sample}} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T$$

where $\bar{\mathbf{x}}$ represent the sample mean of the returns just introduced.

All the information needed to obtain the *correlation matrix* C are already included in $\bar{\Sigma}$, we only need to perform some further calculations:

$$C_{i,j} = \frac{\bar{\Sigma}_{i,j}}{\sqrt{\bar{\Sigma}_{i,i}\bar{\Sigma}_{j,j}}} \tag{2.1}$$

We have thus obtained an empirical estimate of the correlation between our assets returns. The formula in (2.1) is often referred to as *Pearson correlation coefficient*, from the name of the English mathematician Karl Pearson who first formulated it.

Results are reported in the following tables.

******** ADD RESULT TABLES ********

We are mainly interested in the correlation between Bitcoin and other assets returns, so we will now focus on the first row (or equivalently column, by symmetry) of the correlation matrix.

All values are fairly close to zero, never exceeding 10% towards the positive or the negative side. One may thus wonder whether these correlations are *statistically significantly* different from zero. To answer this question, we will introduce two statistical tests to check the correlation significance.

2.2 Correlation Significance

The very core of Inferential Statistics, the branch of statistics that allows to draw conclusions from the information contained in a set of data, is hypothesis testing.

In our case, we are specifically interested in testing if the sample correlation coefficients are significantly different from zero or not. Both of the following tests are presented in the most general form for a sample of two variables, their distribution correlation ρ and their sample correlation $\hat{\rho}$.

Following standard testing procedure, we specify the *null hypothesis* and the *alternative hypothesis*:

$$\mathbf{H_0}: \quad \rho = 0 \quad vs. \quad \mathbf{H_1}: \quad \rho \neq 0$$

These will be common to both presented tests.

2.2.1 Pearson's *t*-test

Our first test is based on Student's t-distribution and the following t-statistic:

$$t = \hat{\rho}\sqrt{\frac{n-2}{1-\hat{\rho}^2}} \tag{2.2}$$

which under the null hypothesis is distributed as a Student's t with $n-2$ degrees of freedom, where n stands for the cardinality of the sample. We can thus proceed by computing the relative p-value and compare it to a given level of confidence α (usually $\alpha = 95\%$). The result of the test will be deduced as follows:

- $p - value < 1 - \alpha$: we have statistical evidence to state that the correlation is *significantly* different from zero;
- $p - value \geq 1 - \alpha$: there is *no statistical evidence* to state that the correlation is different from zero.

2.2.2 Permutation test

The permutation test is based on building an empirical distribution of values for the correlation by sampling different pairs of X and Y variable and then computing Pearson's correlation. If this is done a large enough number of times, we obtain an empirical distribution of possible values. From this distribution we can then obtain the p-value of the test and thus the final result in the same way as in the previous case.

2.2.3 Significance results

The values that we obtained for the correlation of Bitcoin with the other assets are reported in Table 2.1, including the resulting p-values for both of the tests that were introduced in the paragraph above.

Looking at the first line alone, we can see that the asset-Bitcoin correlation never surpasses 5% in absolute value. This is exactly what we would expect given that Bitcoin price seems to move on its and not really care about what is happening on the market (at least to some degree).

Moreover, if we also study the significance of the correlation level through Pearson's or the permutation test, we can see that the only asset that has a correlation that is *significantly different from zero*[1] is the Standard&Poor's

[1]Considering a confidence level of 5%.

	bric	sp500	eurostoxx	nasdaq	bond_europe	bond_us	bond_eur
Correlation	1,41%	4,37%	4,12%	3,59%	1,41%	-1,84%	1,92%
Pearson	52,70%	4,10%	5,40%	9,35%	50,35%	39,90%	37,95%
Permutation	51,21%	4,24%	5,55%	9,48%	51,29%	39,28%	37,20%

	eur	gbp	chf	jpy	gold	wti	grain	metal
Correlation	2,29%	0,73%	2,46%	-1,09%	-0,24%	0,77%	3,50%	2,69%
Pearson	29,00%	72,75%	24,75%	61,45%	91,00%	71,55%	10,10%	20,90%
Permutation	28,64%	73,33%	25,34%	61,32%	91,08%	72,00%	10,34%	21,13%

Table 2.1: Values of the correlation between Bitcoin and the other assets and their p-value using both Pearson's and the permutation test.

500 Index. Still, the correlation that we experience between S&P500 and Bitcoin returns is of 4,37 %, which is considerably low.

Thus, our results show that Bitcoin is fundamentally incorrelated to any of the asset that we are taking into consideration in our analysis.

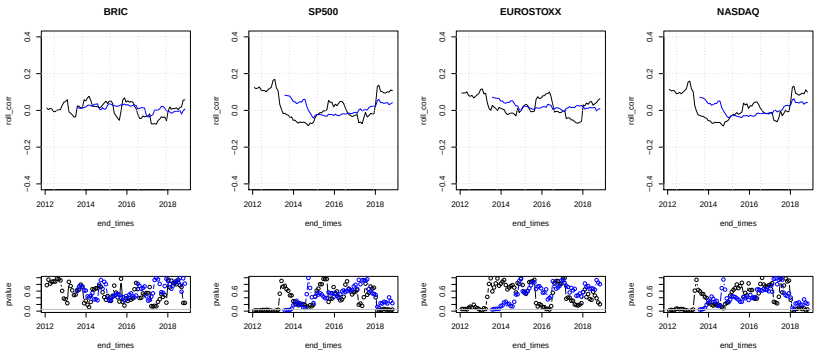
This result is what induced us to consider the possible diversification benefits of introducing Bitcoin in an investor's portfolio. We will see in Chapter 5 what great improvements in terms of increased return and decreased risk this addition brings to our reference portfolio.

2.3 Rolling Correlation

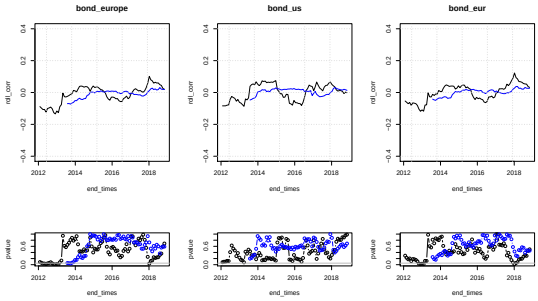
Our study so far has focused on the analysis of the dataset as a whole, with values spanning from July 2010 to December 2018. This is clearly important if we want to obtain a general overview of the period, but it is also interesting to see how the correlation between the assets has evolved through. Therefore, we present in Figure 2.1 the results obtained from calculating the correlation between Bitcoin and the other assets using rolling windows of 36 and 18 months, updated monthly.

There are two graphs for each asset: in the top plots levels of the rolling correlations are represented using two different colours, blue for the 3-year and black for the 18-month windows; in the bottom plots we included the significance of each rolling correlation through its p-value. The grey horizontal line represents the 5% level of significance[2].

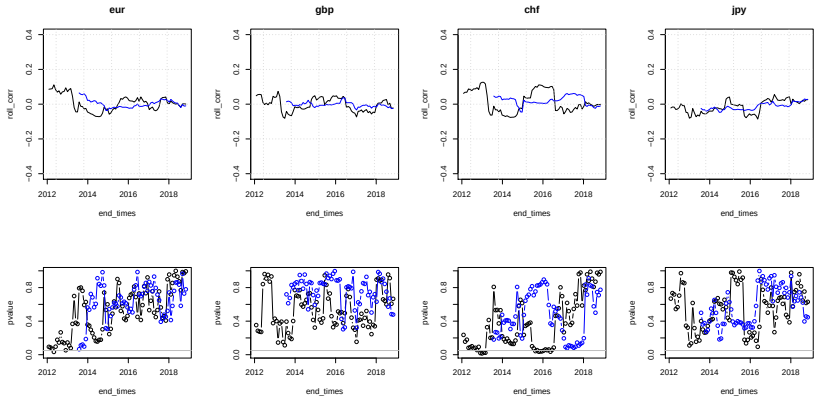
[2]As we explained in the previous section, to check that a sample correlation is *significantly* non zero, we compare the p-value of the test to a given level, here $1 - \alpha = 5\%$.



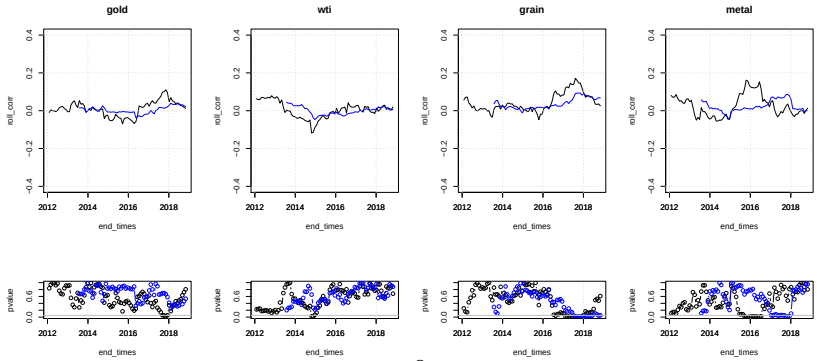
(a) Stocks



(b) Bonds



(c) Currency exchange



(d) Commodities

Figure 2.1: Plots of rolling correlation for the different asset classes (on top) and significance for each value (on bottom). Blue lines are the 3-year rolling correlations, while the black ones have a window of 18 months. Both computations are updated monthly.

The main conclusion we can draw from these images is that the correlation of any asset with Bitcoin is hardly ever significantly different from zero, and when it is, its absolute level is never more than greater than 20% for a small period of time.

To confirm the fact that Bitcoin is not correlated with any asset, we can also take a look at the path of the rolling correlations: there is no line that is always above zero, nor below. This indicates that there is no underlying trend, whether positive or negative, and the correlation one might find is only temporary.

**** maybe add more comments on the results ****

Graphically, whenever the dots are above the grey line in Figure 2.1, the corresponding correlation is *not* significantly different from zero.

Chapter 3

Presentation of the Models

In this chapter we will present the stochastic frameworks in which we developed our analysis. We first introduce a *jump diffusion* (JD) model presented in 1976 by R.C. Merton: he added log-normal jumps to the simple B&S dynamics of the asset price. Then we move to the *stochastic volatility* (SV) model of Heston 1993. Heston introduced a new stochastic process that accounts for the variance of the underlying which evolves as a B&S with a stochastic volatility term. The last model we will present was introduced by Bates in 1996 and it is the combination of the former two: an asset dynamics which include jumps and is driven by a stochastic volatility. All models are first introduced in the one dimensional case and then generalised to the n dimensional case which was then implemented in our code.

3.1 Preliminary Notions

In this section we will briefly present the equation of a geometric brownian motion, introduce the notion of Poisson process and present the CIR process. All of these building blocks will be required to fully understand the models to follow.

3.1.1 Geometric Brownian Motion

The simplest continuous dynamics to describe the price of an asset is that of a geometric brownian motion:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \tag{3.1}$$

where S_t represents the price of the asset at time t, μ is the (constant) drift and σ is the (constant) volatility. W_t is a Wiener process. This is the standard and most widespread stichastic differential equation to model asset dynamics, so we will only present those results that will be later used in our study.

Applying Ito's lemma to the previous equation, we can also explicitly express the dynamics of the log-returns $X_t = log(S_t)$, obtaining:

$$dX_t = (\mu - \frac{\sigma^2}{2})dt + \sigma dW_t \tag{3.2}$$

This stochastic differential has a simple solution which can be computed via stochastic integrals and allows us to describe the dynamics of the log-returns at each instant t starting from $t = 0$:

$$X_t = X_0 + (\mu - \frac{\sigma^2}{2})t + \sigma W_t \tag{3.3}$$

Thanks to (3.3) we can now express the price dynamics of the asset by inverting the relation with the returns: $S_t = e^{X_t}$. We thus obtain the solution to (3.1):

$$S_t = S_0 \, e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t} \tag{3.4}$$

Given that $S_t = e^{X_t}$, that X_t by its equation is a generalized brownian motion and hence we have $X_t - X_0 \sim \mathcal{N}(\mu - \frac{\sigma^2}{2}, \sigma^2)$, the resulting distribution of prices at time t as units of the initial value is distibuted as a log-normal.

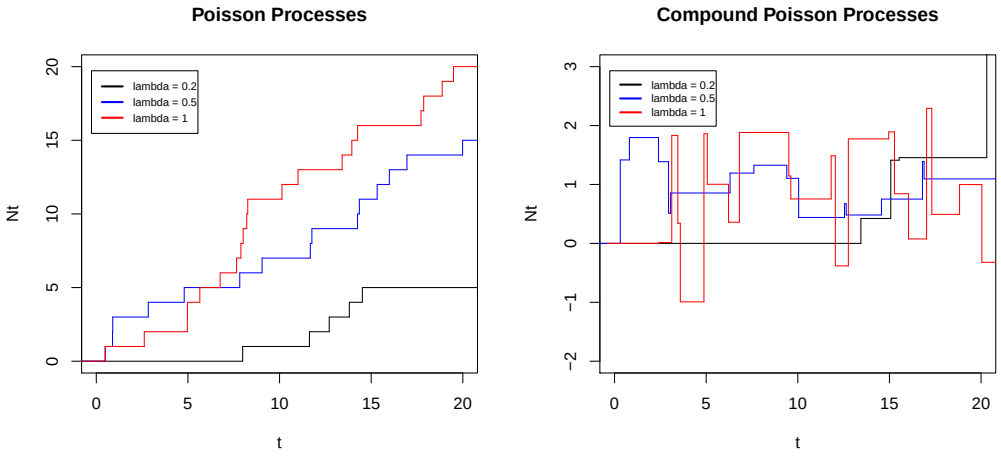
The great success of these framework comes from the simplicity of its dynamics. In particular, since the log-returns follow a Gaussian distribution, μ and σ are easy to calibrate from data and the formulas for pricing options are often explicit. As one can imagine, a simple model can only explain simple phenomena: that's why we have such a great deal of *new and improved* version of equation 3.1.

3.1.2 Poisson Process and Compound Poisson Process

Consider a sequence of *independent* exponential random variables $\{\tau_i\}_{i \geq 1}$ with parameter λ[1] and let $T_n = \sum_{i=1}^{n} \tau_i$. Then we can define the *Poisson process* N_t as

$$N_t = \sum_{n \geq 1} \mathbb{1}_{t \geq T_n} \tag{3.5}$$

[1] An exponential random variable τ of parameter λ has a cumulative distibution function of the form: $\mathbb{P}(\tau \geq y) = e^{-\lambda y}$



(a) Poisson processes.

(b) Compound with Gaussian jumps.

Figure 3.1: Poisson processes and compound Poisson processes with different λ.

where $\mathbb{1}_{condition}$ is 1 if the *condition* is true, 0 if it is false.

N_t is thus a piece-wise constant RCLL[2] process with jumps that happen at times T_n and are all of size 1, as we can see from Figure fig. 3.1a. An important property of Posson processes is that they have independent and stationary increments, meaning that the increment of $N_t - N_s$ (with $s \leq t$) is independent from the history of the process up to N_s and has the same law of N_{t-s}. At any time t, N_t is distributed as a Poisson of parameter λt, which means it is a discrete random variable on the integer set with

$$\mathbb{P}(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \tag{3.6}$$

When working with jump diffusion process, it is often the case that there is no explicit formula for its density, thus usually we resort to characteristic functions. The characteristic function of N_t is given by

$$\phi_{N_t}(u) = e^{\lambda t(e^{iu}-1)} \tag{3.7}$$

The computations to get 3.7 from 3.6 are carried out in Appendix A.

For financial applications, it is of little interest to have a process with a single possible jump size. The *compound* Poisson processes are a generalization of Poisson processes where the jump sizes can have an arbitrary

[2]RCLL is shorthand for right continuous with left limit.

distribution. More precisely, consider a Poisson process N_t with parameter λ and a sequence of i.i.d[3] variables $Y_{ii\geq 1}$ with law $f_Y(y)$. Then the process defined by

$$X_t = \sum_{n=1}^{N_t} Y_i \tag{3.8}$$

is a compound Poisson process. Examples of this kind of process are plotted in Figure fig. 3.1b.

As before, we developed the computations to obtain an expression for the characteristic function of X_t in Appendix A. The resulting expression depends on the distribution of Y, specifically from its characteristic function $\phi_Y(u)$:

$$\phi_{X_t}(u) = e^{\lambda t(\phi_Y(u)-1)} \tag{3.9}$$

Both in Merton's and in the Heston's models there will be a jump component driven by a compound Poisson process with Gaussian jump sizes, as we will see in the following paragraphs.

3.1.3 CIR Process

The CIR process was introduced in [5] in 1985 by Cox, Ingersoll and Ross (hence the name CIR) as a generalization of a Vasicek process to model the mean reverting dynamics of interest rates. Following their notation, the differential equation for the evolution of the rate is given by:

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t \tag{3.10}$$

where we have three parameters that characterize it: θ is the *long-term value* of the rate, the asymptotic level which it tends to settle at in the long run;κ is the *mean-reversion rate*, the speed at which the rate is pulled back to the θ value; finally σ accounts for the *volatility* of the stochastic component. When $\kappa, \theta > 0$, equation (3.10) represents a first order mean-reverting autoregressive process. Moreover, thanks to *******add citation*********, we know that the process will not hit zero if the following condition is satisfied:

$$2\kappa\theta > \sigma^2 \tag{3.11}$$

This condition is usually referred to as *Feller* condition, from the author of the cited paper in which this result was first presented.

An example of a CIR process is shown in Figure (3.2), where the mean-reverting effect is clearly visible.

[3] i.i.d stands for independent and identically distributed.

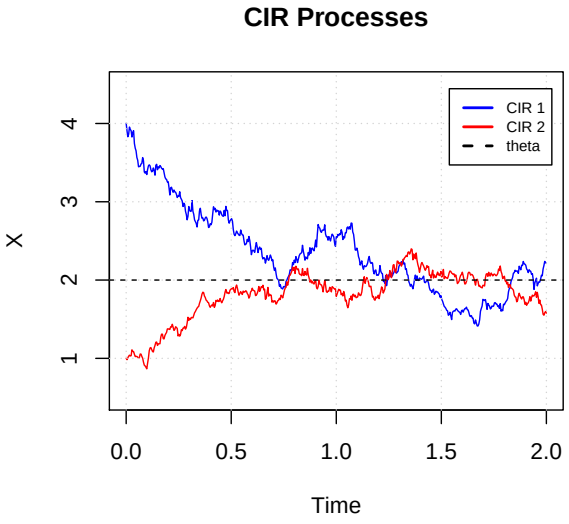


Figure 3.2: Two trajectories of the same CIR process: $dr_t = 3(2 - r_t)dt + 0.5\sqrt{r_t}dW_t$ with different starting point. We can see the mean-reverting effect that attracts both trajectories to the value $\theta = 2$.

When modelling interest rate in continuous time, having r_t hit zero is not an issue since when $r_t = 0$ equation (3.10) reduces to $dr_t = \kappa\theta dt$ which immediately brings the level back to positive values and hence the square root in the dynamics never loses meaning. Conversely, considering a *discretized* version of (3.10), as is the case in a simulation framework, one needs to pay attention on how he models the increments since a simple Euler discretization scheme may cause r_t to reach *negative* values[4] and invalidate the whole model representation.

The non negativity of the CIR process will become fundamental when we introduce Heston and Bates SV models, where the (stochastic) variance of the process driving the asset price will be model as a CIR process.

Since it will be useful later on in the paper, we also present the *stationary* distribution of r_t. Due to the mean reversion effect, r_t will approach a *gamma* distribution with density:

$$f_{r_\infty}(x) = \frac{\omega^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\omega x} \tag{3.12}$$

[4]This model was introduced having in mind the certainty that interest rates could never be negative, hence the introduction of a square root in the dynamics. Given recent years interest rates levels, this in no more the case.

where

$$\omega = \frac{2\kappa}{\sigma^2}\ ,\ \nu = \frac{2\kappa\theta}{\sigma^2}$$

Its *moment generating function*, which will be useful later as well, is defined as follows:

$$M_{r_\infty}(z) = \left(\frac{\omega}{\omega - z}\right)^{\nu} \tag{3.13}$$

3.2 Merton Model

Let us see now how the different building blocks that we just presented can be combined together by first taking a look at the jump diffusion model by Merton.

3.2.1 Original Univariate Model

The first jump diffusion model was originally introduced in [15] in order to account for the leptokurtic distribution of real market returns and to model sudden falls (or rises) in prices due to the arrival of new information. The asset price dynamics S_t is modelled as a GBM to which a jump component driven by a compound Poisson process is added:

$$\frac{dS_t}{S_t} = (\mu - \lambda\mu_J)dt + \sigma dW_t + Y_t dN, \tag{3.14}$$

where μ and σ are respectively the drift and the diffusion of the continuous part, Y_t is a process modelling the intensity of the jumps and $N(t)$ is the Poisson process driving the arrival of the jumps and has parameter λ. The jump intensity Y_t is distributed as a log-normal with parameters (μ_J, σ_J^2). This means that $y_t = \log(Y_t) \sim \mathcal{N}(\mu_J, \sigma_J^2)$

We can rewrite (3.14) in terms of the log-returns $X_t = log(S_t)$ and obtain, following the computations in [14] and using theory from [18]:

$$dX_t = (\mu - \frac{\sigma^2}{2} - \lambda\mu_J)dt + \sigma dW_t + \log(Y_t), \tag{3.15}$$

that has as solution:

$$X_t = X_0 + \mu t + \sigma W_t + \sum_{k=1}^{N(t)} y_k, \tag{3.16}$$

where X_0 is the initial value of the log-returns, $y_k = \log(Y_k) = \log(Y_{t_k})$ and t_k is the time when the k^{th} Poisson shock from $N(t)$ happens. Following

Merton in [15], we take y_k *i.i.d.* (independent and identically distributed) and Gaussian, as we already stated. Another choice for the distribution of y is given in [11], where an asymmetric distribution is used to account for positive and negative jump in two different ways.

It is often useful when dealing with market data that are by nature discrete, to consider a *discretized* version of (3.16) in which the values are sampled at intervals of Δt in $[0, T]$. We thus get that for $X_i = \log(\frac{S_{i+1}}{S_i})$:

$$X_i = \mu \Delta t + \sigma \sqrt{\Delta t}\, z + \sum_{k=1}^{N_{i+1}-N_i} Y_k \tag{3.17}$$

where we denote $X_i = X_{t_i}$, $N_i = N(t_i)$ and $t_i = i\Delta t$ with $i = 0 \dots N$, $t_N = N\Delta t = T$, z is distributed as a standard Gaussian $z \sim \mathcal{N}(0, 1)$.

The Poisson process $N(t)$ in (3.17) is computed at times t_{i+1} and t_i and these quantities are subtracted. Following basic stochastic analysis, one can prove that the resulting value $N_{i+1} - N_i$, is distributed as a Poisson random variable N of parameter $\lambda \Delta t$. This allows us to provide an explicit formulation for the transition density of the log-returns using the theorem of total probability:

$$f_{\Delta x}(x) = \sum_{k=0}^{\infty} \mathbb{P}(N = k) f_{\Delta x | N=k}(x) \tag{3.18}$$

This equation is a good analytical result but it is not so practical for calibration purposes, as we will explain more in depth in chapter 4.

3.2.2 Multivariate Model

Starting from the univariate model introduced in [15], we can developed a generalization to n assets including only idiosyncratic jumps:

$$\frac{dS_t^{(j)}}{S_t^{(j)}} = (\mu_j - \lambda_j \mu_{J_j}) dt + \sigma_j dW_t^{(j)} + Y_t^{(j)} dN_t^{(j)} \tag{3.19}$$

where $\mathbf{S}_t$ are the prices of the assets, $j = 1, ..., n$ represents the asset, μ_j are the drifts, σ_j are the diffusion coefficients, $W_t^{(j)}$ are the components of an n-dimensional Wiener process $\mathbf{W}_t$ with $dW^{(i)} dW^{(j)} = \rho_{i,j}$, Y_j represent the intensities of the jumps and $y^(j) = \log(Y^{(j)})$ are distributed as Gaussian: $y^(j) \sim \mathcal{N}(\mu_{J_j}, \sigma_{J_j}^2)$. Finally, $N^{(j)}(t)$ are Poisson processes with parameters λ_j, which are independent of $\mathbf{W}_t$ and of one another.

3.3 Heston Model

The Heston model was presented in 1993 in [9] as a new framework to model stochastic volatility in the asset price dynamics, which allows to better fit the skewness and kurtosis of the log-return distribution.

3.3.1 Univariate Heston Model

The Heston model belongs to the family of SD processes, generalizations of B&S model in which the volatility is no more constant but is itself stochastic. ***add examples?***

Starting with a GBM as in the B&S framework, we obtain the dynamics of the price process simply by allowing the volatility in (3.1) to evolve over time and specifying how this evolution takes place. In particular, the dynamics of the *instantaneous* variance $V_t = \sigma_t^2$ for the Heston model is described by a CIR process :

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{V_t} dW_t^S \tag{3.20}$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma_V \sqrt{V_t} dW_t^V \tag{3.21}$$

where μ represents the drift in the asset prices, $\kappa > 0$ and $\theta > 0$ are respectively the *mean-reversion* rate and the long-run level for the V_t process, $\sigma_V > 0$ is often referred to as the *volatility of volatility* parameter, in short vol-of-vol. The two brownian motions W_t^S and W_t^V are correlated with correlation coefficient equal to ρ. It can be proven that the variance process is always non-negative if the *Feller* condition

$$2\kappa\theta \geq \sigma_V \tag{3.22}$$

is satisfied. If the same inequalities holds strictly, we have that the variance process will always be strictly positive.

Let us now consider the dynamics of the log-return $x_t = \log(S_t)$, as we did in the Merton case. Unfortunately, given the increased complexity of the model due to the SV part, an explicit formula for the density of the log-return is not available and it has to be computed from the Fourier inversion of the characteristic function:

$$f_{x_t}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_{x_t}(iu) e^{iux} du \tag{3.23}$$

In (3.23) we omitted the dependence of $f_{x_t}(x)$ and $\phi_{x_t}(u)$ on model parameters and on the initial values of the log-returns x_0 and of the variance process V_0.

To derive the expression of the characteristic function $\phi_{x_t}(u)$, one has to solve a couple of *Fokker-Planck* partial differential equation as is shown in the Appendix of the reference paper by Heston [9]. This procedure is beyond the scope of this thesis and thus we will only report the final result, which is a log-affine equation on the information at time $t = 0$:

$$\begin{aligned}
\phi_{x_t}(u|x_0, V_0) &= \exp\{A(t,u) + B(t,u)x_0 + C(t,u)V_0\} \\
A(t,u) &= \mu u i t + \frac{\kappa\theta}{\sigma_V^2}\left((\kappa - \rho\sigma_V u i + d)t - 2\log\left[\frac{1-ge^{dt}}{1-g}\right]\right) \quad (3.24a) \\
B(t,u) &= iu \quad (3.24b) \\
C(t,u) &= \frac{\kappa - \rho\sigma_V u i + d}{\sigma_V^2}\left[\frac{1-e^{dt}}{1+ge^{dt}}\right] \quad (3.24c)
\end{aligned}$$

where :

$$\begin{aligned}
d &= \sqrt{(\rho\sigma_V u i - \kappa)^2 + \sigma_V^2(ui + u^2)} \\
g &= \frac{\kappa - \rho\sigma_V u i + d}{\kappa - \rho\sigma_V u i - d}
\end{aligned}$$

This formulation is the one proposed by Heston in his original paper [9] but it is shown in [1] that it has numerical issues when pricing Vanilla options using Fourier methods. We will not be pricing any instrument, however we may incur in the same errors when calibrating our model through the techniques explained in Chapter 4. For this reason, we will be using the alternative formulation presented in [1]:

$$\begin{aligned}
\phi^*_{x_t}(u|x_0, V_0) &= \exp\{A(t,u) + B(t,u)x_0 + C(t,u)V_0\} \\
A(t,u) &= \mu u i t + \frac{\kappa\theta}{\sigma_V^2}\left((\kappa - \rho\sigma_V u i - d)t - 2\log\left[\frac{1-g^*e^{-dt}}{1-g^*}\right]\right) \\
B(t,u) &= iu \\
C(t,u) &= \frac{\kappa - \rho\sigma_V u i - d}{\sigma_V^2}\left[\frac{1-e^{-dt}}{1-g^*e^{-dt}}\right]
\end{aligned} \quad (3.25)$$

where :

$$\begin{aligned}
d &= \sqrt{(\rho\sigma_V u i - \kappa)^2 + \sigma_V^2(ui + u^2)} \\
g^* &= \frac{\kappa - \rho\sigma_V u i - d}{\kappa - \rho\sigma_V u i + d} = \frac{1}{g}
\end{aligned}$$

The only difference is that the signs of the d terms are all flipped: the origin of the two representations for the characteristic function lies in the fact that the complex root d has two possible values and the second value is exactly minus the first value. A uniform choice between the two possibilities cannot be made over all the complex plane to obtain a continuous function due to the presence of a branch cut in the graph of $\sqrt{z}$. Selecting the sign of d as in $\phi^*_{x_t}$ and g^* avoids the discontinuity affecting our future computations.

For this reason, from now on we will be using the second formulation while leaving out the star * from our notation.

Our characteristic function still depends on the initial values x_0 and V_0 and is thus often called *conditional* characteristic function on x_0 and V_0. We can easily remove the dependence on x_0 by considering the distribution of incremental returns, namely $\Delta x_t = \log(S_t/S_0) = x_t - x_0$

Applying the definition of characteristic function:

$$\begin{aligned}
\phi_{\Delta x_t}(u|V_0) &= \mathbb{E}\big[e^{iu\Delta x_t}|V_0\big] \\
&= \mathbb{E}\big[e^{iu(x_t-x_0)}|V_0\big] \\
&= \mathbb{E}\big[e^{iux_t}|V_0\big]e^{-iux_0} \\
&= \phi_{x_t}(u)e^{-iux_0} \\
&= \exp\{A(t,u) + (B(t,u) - iu)x_0 + C(t,u)V_0\}
\end{aligned}$$

and since $B(t,u) = iu$, the second term in the exponential is equal to zero and we are left with:

$$\phi_{\Delta x_t}(u|V_0) = \exp\{A(t,u) + C(t,u)V_0\} \tag{3.26}$$

This equation is however still dependent on the initial value of the variance process. In a simulation framework, this would not be an issue, since we can define the level of V_0 ourselves and then generate all the different scenarios. However, if we need to calibrate the model parameters from asset prices, the market data for the variance process are not available. We will address more in depth different ways to solve this problem later on in Chapter 4. As for now, we show that we can obtain an *unconditional* expression for the characteristic function of a Heston process by approximating the distribution of V_t with its stationary distribution.

The idea of integrating out the variance was presented in [8] and here we follow a similar approach. The computations to obtain an unconditional

characteristic function are the following:

$$\begin{aligned}
\phi_{\Delta x_t}(u) &= \mathbb{E}\big[e^{iu\Delta x_t}\big] \\
&= \mathbb{E}\big[\mathbb{E}\big[e^{iu\Delta x_t}|V_0\big]\big] \\
&= \int_0^\infty \mathbb{E}\big[e^{iu(x_t-x_0)}|V_0 = v\big] f_{V_0}(v)dv \\
&= \int_0^\infty \exp\{A(t,u) + C(t,u)v\} f_{V_0}(v)dv \\
&= \exp\{A(t,u)\} \int_0^\infty \exp\{C(t,u)v\} f_{V_0}(v)dv \\
&= \exp\{A(t,u)\} M_{V_0}(C(t,u))
\end{aligned}$$

where we have used the Law of Total Expectation and $M_{V_0}(z)$ indicates the *moment generating function* of V_0 as considered stationary, so that of a Gamma distribution. In particular, it will have the same expression as in (3.13).

3.3.2 Parsimonious Multi-asset Heston Model

The problem with generalizing Heston framework to include more than one underlying is that the model becomes quickly very complex: we need 5 parameters for each asset ($\mu, \kappa, \theta, \sigma_V$ and ρ), and we can also model all types of correlation between the different stochastic drivers. In particular, we have 4 types of correlation:

- ρ^{S_i,V_i}: correlation that we already have in the unidimensional case, that models the way the price and the variance processes are linked ;
- ρ^{S_i,S_j}: correlation between the movements in prices of two different assets ;
- ρ^{V_i,V_j}: correlation between the variance processes of two different assets. One might expect that asset of similar nature have a higher correlation both in the price and in the variance processes;
- ρ^{S_i,V_j}: correlation between the price process of an asset and the variance process of a different one.

To overcome this increase in both mathematical and computation complexity, we will use the *parsimonious* model introduced by Szimayer, Dimitroff and Lorenz in [7]. Their framework has two main characteristics: each

single-asset sub-model forms a traditional Heston model and the parameters are composed by n single-asset Heston parameter sets and $n(n-1)/2$ *asset-asset* correlations.

The n-dimensional model hence becomes:

$$\begin{pmatrix} dS_i(t)/S_i(t) \\ dV_i(t) \end{pmatrix} = \begin{pmatrix} \mu_i \\ \kappa_i(\theta_i - V_i(t)) \end{pmatrix} dt + \begin{pmatrix} \sqrt{V_i(t)} & 0 \\ 0 & \sigma_{V_i}\sqrt{V_i(t)} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \rho_i & \sqrt{1-\rho_i^2} \end{pmatrix} \begin{pmatrix} dW^{S_i}(t) \\ dW^{V_i}(t) \end{pmatrix} \quad (3.27)$$

where $dW^{S_i}(t)$ and $dW^{V_i}(t)$ are independent processes and we have written explicitly the dependence between the price and the variance processes.

To fully characterize the model, we need to describe how the different W^{S_i} and W^{V_i} are correlated. In accordance with what was stated earlier, the correlation structure is the following:

$$\Sigma^{(S,V)} = cor(\boldsymbol{W}^S, \boldsymbol{W}^V) = \begin{pmatrix} \Sigma^S & 0 \\ 0 & I_n \end{pmatrix} \quad (3.28)$$

in which $\Sigma^S = cor(\boldsymbol{W}^S)$. Equation (3.28) mathematically represents what we defined as the correlation structure for our model: we only explicitly define the *asset-asset* correlations through matrix Σ^S, the dependence of every variance on other processes is carried over via Σ^S and the correlations ρ_i. More clearly, let $(\Sigma^S)_{i,j} = \rho_{i,j}$[5]:

- $dW^{S_i}(t)dW^{S_j}(t) = \rho_{i,j}dt$
- $dW^{S_i}(t)dW^{V_j}(t) = \rho_{i,j}\rho_j dt$
- $dW^{V_i}(t)dW^{V_j}(t) = \rho_i\rho_{i,j}\rho_j dt$ if $i \neq j$, dt otherwise

A detailed representation of a 2-asset *parsimonious* Heston model can be found in the reference paper [7].

3.4 Bates Model

Bates model is a way of combining both of the characteristics of Merton (price jumps) and Heston (stochastic volatility) in a single framework. It was introduced in 1996 by David Bates, an American professor at University of Iowa.

[5]Of course we will have $\rho_{i,j} = 1$ whenever $i = j$.

3.4.1 Univariate Model

In his paper [3], Bates proposes his model as a way of capturing the leptokurtosis in the distribution of log-differences of the USD/DeutscheMark exchange rate. He suggested to improve the versatility of Heston model by including *log-normal, Poisson driven* jumps in the price process, borrowing this addition from Merton model.

Thus, we will have to deal with a total of 8 parameters: $\mu, \kappa, \theta, \sigma_V$ and ρ for the stochastic volatility part, and μ_J, σ_J and λ for the jump component.

$$\frac{dS_t}{S_t} = (\mu - \lambda\mu_J)dt + \sqrt{V_t}dW_t^S + Y_t dN \tag{3.29}$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma_V\sqrt{V_t}dW_t^V \tag{3.30}$$

Of course, as these equations are directly obtained from (3.20) and (3.21), we have that the variance process is strictly positive whenever the Feller condition $2\kappa\theta \geq \sigma_V$ is satisfied. As in (3.14), Y_t is the jump process and is such that $y_t = \log(Y_t) \sim \mathcal{N}(\mu_J, \sigma_J^2)$, while $N(t)$ is a Poisson process with parameter λ.

Unfortunately, for the same reason as in previous section, an explicit formula for the distribution of the log-returns $x_t = log(S_t)$ is not available and thus we have to make reference to their characteristic function. The silver lining is, though, that to obtain expression of $\phi_{x_t}(u)$, we only have to include an extra additive term to the exponential in (3.31) to account for the jumps:

******* controllare jump component***********

$$\begin{aligned}
\phi^*_{x_t}(u|x_0, V_0) &= \exp\{A(t,u) + B(t,u)x_0 + C(t,u)V_0 + D(t,u)\} \\
A(t,u) &= \mu uit + \frac{\kappa\theta}{\sigma_V^2}\left((\kappa - \rho\sigma_V ui - d)t - 2\log\left[\frac{1 - g^*e^{-dt}}{1 - g^*}\right]\right) \\
B(t,u) &= iu \\
C(t,u) &= \frac{\kappa - \rho\sigma_V ui - d}{\sigma_V^2}\left[\frac{1 - e^{-dt}}{1 - g^*e^{-dt}}\right] \\
D(t,u) &= -\lambda\mu_J uit + \lambda t\left[(1+\mu_J)^{ui}e^{\sigma_J^2(ui/2)(ui-1)} - 1\right]
\end{aligned} \tag{3.31}$$

where :

$$\begin{aligned}
d &= \sqrt{(\rho\sigma_V ui - \kappa)^2 + \sigma_V^2(ui + u^2)} \\
g^* &= \frac{\kappa - \rho\sigma_V ui - d}{\kappa - \rho\sigma_V ui + d} = \frac{1}{g}
\end{aligned}$$

We will used the * formulation for the same desirable numerical properties as already stated, while leaving out the * symbol from our notation from now on.

3.4.2 Parsimonious Multi-asset Bates Model

We will generalize the single-asset Bates model to a multi-asset model in the same way that we did with the Heston: we only model the asset-asset correlation ρ^{S_i,S_j} for each pair of asset.

The model definition becomes very similar to (3.27), with the addition of the jump part for the price processes:

$$\begin{pmatrix} dS_i(t)/S_i(t) \\ dV_i(t) \end{pmatrix} = \begin{pmatrix} \mu_i \\ \kappa_i(\theta_i - V_i(t)) \end{pmatrix} dt + \begin{pmatrix} \sqrt{V_i(t)} & 0 \\ 0 & \sigma_{V_i}\sqrt{V_i(t)} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \rho_i & \sqrt{1-\rho_i^2} \end{pmatrix} \begin{pmatrix} dW^{S_i}(t) \\ dW^{V_i}(t) \end{pmatrix} + \begin{pmatrix} Y_i(t)dN_t^i \\ 0 \end{pmatrix}. \quad (3.32)$$

T

Chapter 4

Calibration of the Models

In this chapter we will explain how the different models were calibrated and what difficulties were overcome. Empirical results are included for each section.

*******aggiungere commento su diversi tipi di calibrazione e in particolare sul fatto che calibriamo su timeseries*******

4.1 Maximum Likelihood Calibration

The calibration method that we implemented is the *maximum likelihood* approach, which is a statistical method to obtain the parameters of a target distribution family that best approximate the unknown distribution of the observed data.

Let us call $X = \{x_1, x_2, ..., x_N\}$ the set of available observations and let $\psi = \{\alpha_1, \alpha_2, ...\}$ the set of parameters of the distribution $\mathcal{D}$ that we want to calibrate. Moreover, define $f_{\mathcal{D}}(x; \psi)$ to be the probability density function (pdf) of distribution $\mathcal{D}$ with parameters ψ computed at $x \in \mathbb{D}$, where $\mathbb{D}$ is the domain of the density function.

Our aim is to find the best set of parameters ψ such that we can describe X as samples taken from $\mathcal{D}$:

$$x_i \sim \mathcal{D}(\psi),\ i = 1, \ldots, N. \tag{4.1}$$

The estimation of the parameter set ψ can then be obtained by:

$$\hat{\psi} = \arg\max_{\psi \in \Psi}\ \mathcal{L}(\psi|X), \tag{4.2}$$

where $\mathcal{L}(\psi|X)$ is the likelihood function for distribution $\mathcal{D}(\psi)$ given the observed data X.

In the case that the data in X are i.i.d. we have that the likelihood function can be computed as the product of the probability density function computed at each observation x_i:

$$\mathcal{L}(\psi|X) = \prod_{i=1}^{N} f_{\mathcal{D}}(x_i;\psi). \tag{4.3}$$

In practice, it is often common to consider the *log-likelihood* function $\ell(\psi|X) = \log \mathcal{L}(\psi|X)$. The natural logarithm is a monotonic function so the maximum of the log-likelihood function is achieved in the same point as the basic likelihood function.[1] Taking the logarithm of (4.4) also simplifies the expression in that the product now becomes a sum:

$$\ell(\psi|X) = \mathcal{L}(\psi|X) = \sum_{i=1}^{N} \log f_{\mathcal{D}}(x_i;\psi). \tag{4.4}$$

The issues that arise from the definition of a maximum likelihood estimator are mainly two. The first problem is that to use equation (4.4) we need to know the pdf of the distribution: this might not always be the case with complicated models that only have an explicit expression for their characteristic function. This is what happens for Heston and Bates model, and, as we will see, we are going to need to invert the chf via Fourier inversion and obtain the pdf numerically.

The second issue is that we need to solve a maximization problem in order to obtain an estimate for the parameters of the distribution: it is well known how optimization routines might perform well for some special circumstance and badly under other, especially when the number of parameters increases. The trade-off is, as usual, between fast computations and robustness of the optimization. To solve this, we will opt for a combination of global and local optimizers.

4.2 Calibration of Merton Model

Let us start by considering how we calibrated the jump diffusion model by Merton. We will first present how to obtain the parameters for a single asset model and then move to a multi-asset one.

[1] By how the likelihood function is defined, it can attain only positive values so that the logarithm of $\mathcal{L}(\psi|X)$ is always well defined.

4.2.1 Single Asset Merton Calibration

As we already introduced, the calibration method that we used is the maximum likelihood approach.

We are going to calibrate the values of the 5 parameters $\psi = \{\mu, \sigma, \mu_J, \sigma_J, \lambda\}$ of the jump diffusion model based on the observations of the log-returns $\Delta x_i = \log \frac{S_i}{S_{i-1}}$ for each time interval Δt. In our analysis, since we are going to use *daily* log-returns, $\Delta t = 1/255$.

Considering daily log-returns allows us to assume that the different samples of Δx are independent and identically distributed according to the distribution of log-returns in Merton model given in (3.18). For ease of reading, we present it again here:

$$f_{\Delta x}(x;\psi) = \sum_{k=0}^{\infty} \mathbb{P}(N=k) f_{\Delta x|N=k}(x;\psi). \tag{4.5}$$

The formula represents an infinite mixture of Gaussian distributions, due to the infinite possible realization of the Poisson variable that accounts for the arrival of jumps. This fact has a great downside in terms of maximum likelihood estimation. As discussed in [10], the infinite Gaussian mixture causes the problem of maximizing the (log-)likelihood to be unbounded and thus intractable.

To solve this issue we can introduce a first order approximation, as it has been proposed in [2]. Since we are considering a small time interval Δt, only the first terms of the series in (4.5) become significant. In particular, we have that the probabilities to have $k = 0, 1, 2$ jumps in a single time step, as presented in (3.6), are:

$$\mathbb{P}(N=0) = e^{-\lambda\Delta t}, \tag{4.6a}$$
$$\mathbb{P}(N=1) = \lambda\Delta t\, e^{-\lambda\Delta t}, \tag{4.6b}$$
$$\mathbb{P}(N=2) = \frac{(\lambda\Delta t)^2}{2}\, e^{-\lambda\Delta t} \tag{4.6c}$$

since $N \sim Pois(\lambda\Delta t)$, which is the Poisson process counting jumps that happen in a Δt interval. Considering $\lambda\Delta t$ as small, we can approximate to the first order $e^{-\lambda\Delta t} = 1 - \lambda\Delta t + o((\lambda\Delta t)^2)$. We thus obtain that:

$$\mathbb{P}(N=0) = 1 - \lambda\Delta t, \tag{4.7a}$$
$$\mathbb{P}(N=1) = \lambda\Delta t\,(1 - \lambda\Delta t) = \lambda\Delta t + o((\lambda\Delta t)^2), \tag{4.7b}$$
$$\mathbb{P}(N=2) = \frac{(\lambda\Delta t)^2}{2}\,(1 - \lambda\Delta t) = o((\lambda\Delta t)^2). \tag{4.7c}$$

The result is that the only relevant terms in (4.5) are the ones for $k = 0, 1$. The formula for the transition density thus becomes:

$$f_{\Delta x}(x;\psi) = \mathbb{P}(N=0) f_{\Delta X|N=0}(x;\psi) + \mathbb{P}(N=1) f_{\Delta x|N=1}(x;\psi) \tag{4.8}$$

We can then write this equation explicitly by considering the different distributions of the log-returns in the case of no jumps and a single jump:

$$f_{\Delta x}(x;\psi) = (1-\lambda\Delta t)\ f_{\mathcal{N}}\Big(x; \widetilde{\mu}\,\Delta t, \sigma^2\Delta t\Big) + (\lambda\Delta t)\ f_{\mathcal{N}}(x; \widetilde{\mu}\,\Delta t + \theta, \sigma^2\Delta t + \delta^2) \tag{4.9}$$

where $f_{\mathcal{N}}(x;\mu,\sigma^2)$ indicates the pdf of a Gaussian with parameters $\mathcal{N}(\mu,\sigma^2)$ compute at x. We used $\widetilde{\mu} = \mu - \sigma^2/2 - \lambda\mu_J$ as a simpler notation to indicate the drift term in the return dynamics.

The distribution of the sample log-returns is thus given by equation (4.9) and we can plug it into (4.3) to obtain the log-likelihood function that we are going to maximize:

$$\ell(\psi|\Delta x) = \sum_{i=1}^{N} \log f_{\Delta x}(\Delta x_i;\psi). \tag{4.10}$$

One can then proceed to maximize equation (4.10) to obtain the optimal parameter set ψ:

$$\hat{\psi} = \arg\max_{\psi\in\Psi}\ \ell(\psi|\Delta x) \tag{4.11}$$

$$\Psi = \{(\mu,\sigma,\mu_J,\sigma_J,\lambda) \in \mathbb{R}^5 \mid \sigma,\sigma_J,\lambda > 0\} \tag{4.12}$$

4.2.2 Multi-asset Merton Calibration

We now know how to calibrate the Merton parameters when the underlying asset is one. In the case of multiple assets, we decided to adapt [7] both for the model definition and the calibration procedure. In the referenced paper, the authors present a *parsimonious approach* to the extension of Heston Model to a multi-asset framework. Since this is the approach that we will follow for the multivariate Heston calibration and for Bates as well, we decided to also perform in the same way the calibration of the parameters for Merton. Using a common approach will allow us to better compare the results, especially in terms of the correlation structure that is what we are ultimately interested in.

The first step in the calibration algorithm is to obtain the parameters of all the single asset models that we are studying. The general way to do this was explained in the previous section.

The next step involves computing the correlation matrix between the different Brownian motions that drive each single-asset model. The way this computation is carried out in [7] is by asset pairs: we consider each possible couple of assets and try to obtain the model correlation that best approximates the observed correlation. Let us see how this is achieved more in detail in the case of only two assets.

The problem we are trying to solve is, given the observed sample correlation ρ_{obs} between two assets' returns, find the best model correlation ρ_{model} that solves $\bar{\rho}(\rho_{model}) = \rho_{obs}$. $\bar{\rho}$ is the expectation of the correlation that we find when computing the empirical correlation of the returns simulated using ρ_{model}.

More precisely, we have to simulate N_{sim} different realisations of the 2-asset model in a given time period $[0, T]$ with the same Δt that we have in the observed data, using ρ_{model} as the correlation between the two driving Brownian motions. We then proceed to compute the correlation of the simulated log-returns ρ_{scen} in each scenario and then we take the average of this values. What we obtain is $\bar{\rho} = \mathbb{E}[\rho_{scen}]$.

In order to simulate a two-asset Merton model, we used the following discretization for the log-returns:

$$x_i^{(1)} = x_{i-1}^{(1)} + (\mu_1 - \sigma_1^2/2 - \lambda_1 \mu_{J_1})\Delta t + \sigma_1 \sqrt{\Delta t}\, z_1 + N_1 Y_1, \tag{4.13a}$$

$$x_i^{(2)} = x_{i-1}^{(2)} + (\mu_2 - \sigma_2^2/2 - \lambda_2 \mu_{J_2})\Delta t + \sigma_2 \sqrt{\Delta t}\, z_2 + N_2 Y_2, \tag{4.13b}$$

where:

- $i = 1, \ldots, N_{step}$ represents the i-th time-step in our simulation: $x_i = x_{t_i}$ and $N_{step} = T/\Delta t$ so that $t_i = i\Delta t$. The value of the initial return x_{t_0} does not affects the final correlation so we can simply set it to zero.

- z_1 and z_2 are the first and second component of a Gaussian vector $\mathbf{z} = (z_1, z_2)^T$ that is distributed as $\mathbf{z} \sim \mathcal{N}_2(\mathbf{0}, Corr)$. $Corr$ is the 2×2 matrix composed of ones in the main diagonal and ρ_{model} in the remaining two spots. More simply, $\mathbf{z}$ is a two dimensional Gaussian with zero mean with covariance matrix equal to the correlation matrix of the model.

- N_1 and N_2 are in general the realizations of two Poisson random variables with parameters $\lambda_1 \Delta t$ and $\lambda_2 \Delta t$ respectively. Given our first order approximation in the density and considering small Δt, N_1 and N_2 are actually two Bernoulli with probability of success (i.e. a jump happens) equal to $\lambda_1 \Delta t$ and $\lambda_2 \Delta t$.

- Finally, Y_1 and Y_2 are the jump intensities and are realisations of Gaussian variables with parameters $\mathcal{N}(\mu_{J_1}, \sigma_{J_1}^2)$ and $\mathcal{N}(\mu_{J_1}, \sigma_{J_1}^2)$.

Through (4.13) we can thus simulate $x_i^{(1,k)}$ and $x_i^{(2,k)}$ for $k = 1, ..., N_{sim}$ and compute for each scenario the correlation observed from the simulated returns $\rho_{scen}^{(k)} = corr(x_i^{(1,k)}, x_i^{(2,k)})$. From these values we can easily compute $\bar{\rho}$:

$$\bar{\rho} = \mathbb{E}[\rho_{scen}] = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \rho_{scen}^{(k)} \tag{4.14}$$

We have that implicitly the value of the *expected model correlation* $\bar{\rho}$ is a function of the model correlation ρ_{model}. The equation we need to solve is then represented by:

$$\bar{\rho}(\rho_{model}) = \rho_{obs}, \tag{4.15}$$

as we have already stated earlier in this section.

In the case of an n-asset model, we have to repeat the previous passages for all possible $n(n-1)/2$ different pairs of assets. We then can build a $n \times n$ matrix M that has ones in the main diagonal and in each element $(l, m), l \neq m$ stores the corresponding model correlation $\rho_{model}^{(l,m)}$. This matrix may not be a well defined correlation matrix: it may happen that M is not positive definite. The last step to obtain a proper correlation matrix for our n-asset model is to perform some form of *regularization* on M that transforms it into a well defined correlation matrix: positive definite, with elements in the range $[-1, 1]$ and ones in the main diagonal.

Through empirical study performed in the reference paper [7], it turns out the best algorithm between the one proposed by the authors is the regularization by Jäckel, that we briefly present here.

Regularization algorithm by Jäckel:

Input Model correlation matrix M (not necessarily positive definite).

1. Perform an eigenvalue decomposition on $M = S\Lambda S^T$, where $\Lambda = diag(\lambda_l)$ and λ_l are the eigenvalues of matrix M.
2. Define $\Lambda' = diag(\lambda_l'$ with $\lambda_l' = \max(\lambda_l, 0))$ as the diagonal matrix that contains the positive part of each eigenvalue.
3. Create the diagonal matrix $T = diag(t_l)$ where $t_l = 1/\sum_m (S_{l,m})^2 \lambda_l'$.
4. Define $B = \sqrt{T} S \sqrt{\Lambda'}$.

5. Set $\widehat{M} = BB^T$.

Output Positive definite correlation matrix $\widehat{M}$.

Hence, $\widehat{M}$ is the model correlation matrix that best represents the observed correlation values and that forms the structure of dependence between the Brownian motions that drive each asset in a multivariate Merton model.

4.3 Calibration of Heston Model

Calibrating the Heston model on time-series data presents an added difficulty since we only have data on the price process, while the model also takes into consideration the dynamics of the variance as a stochastic process.

The problem of deducing the parameter of a *non-observable* process from the observable states of a system is called *filtering* and is in general a complicated numerical procedure. There are nonetheless a few studies in the literature on the subject of calibrating the Heston model through filtering. *****ADD REFERENCE*****

In our analysis we preferred to keep the same approach to the calibration for all the three models. We will thus proceed to explain how we can perform a maximum likelihood calibration in the case of stochastic volatility and, in the next section, of both jumps and stochastic volatility.

4.3.1 Single Asset Heston Calibration

The first issue that we have to overcome is that an explicit formula for the probability density function of the log-returns is not available and we have to resort to performing a Fourier inversion on the characteristic function, as we already showed in equation (3.23)

This allows us to have a pdf for the log-return $x_t = \log S_t$ but in order to be able to perform a maximum likelihood calibration in the same way that we did in the previous section, we need to have the distribution for $\Delta x_t = \log(S_{t+\Delta t}/S_t)$ and make sure that they are i.i.d. so that we can apply (4.3).

In order to obtain the pdf for the incremental log-returns Δx_t, we first need the distribution of $\Delta x_{[0,t]} = \log(S_t/S_0)$ we can simply proceed as we explained in Chapter 3 in the section about Heston model to get the characteristic function for $\Delta x_{[0,t]}$ conditioned on the values of V_0:

$$\phi_{\Delta x_{[0,t]}}(u|V_0) = \exp\{A(t,u) + C(t,u)V_0\}. \tag{4.16}$$

The full expressions for $A(t,u)$ and $C(t,u)$ are given in (3.24a). Notice that we omitted the dependence on the model parameters $\psi = \{\mu, \kappa, \theta, \sigma_V, \rho\}$ for ease of notation.

We are now left with the distribution of $\Delta x_{[0,t]}$ as a function of V_0: this represents an issue since moving from $\Delta x_{[0,t]}$ to $\Delta x_t = \Delta x_{[t,t+\Delta t]}$ we have a dependence on the value of V_t which is not available. Moreover, even if data for V_t was indeed available, considering different time-steps t_i for Δx_t, the data samples of Δx_{t_i} would not be identically distributed as they depend on levels of the variance V_{t_i} which are different.

Hence the need for an *unconditional* characteristic function to get rid of the dependence on initial variance. We followed the approach that was introduced in [8] of integrating out the dependence on V_0. The computations to obtain it were presented in Chapter 3. The unconditional chf for the Heston model has the following expression:

$$\phi_{\Delta x_t}(u) = \exp\{A(\Delta t, u)\} M_\Gamma(C(\Delta t, u)) \tag{4.17}$$

with M_Γ as the moment generating function of a Gamma distribution:

$$M_\Gamma(z) = \left(\frac{\omega}{\omega - z}\right)^\nu$$
$$\omega = \frac{2\kappa}{\sigma_V^2}$$
$$\nu = \frac{2\kappa\theta}{\sigma_V^2}$$

We now have a specific expression for the unconditional characteristic function of Δx_t and we can obtain the corresponding density by Fourier inversion:

$$f_{\Delta x_t}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_{\Delta x_t}(iu) e^{iux} du \tag{4.19}$$

Another advantage of using the uncoditional chf is that since all the observed log-returns $\Delta x_i = \Delta x_{t_i}$ are considered to be sampled from the same distribution, we can numerically perform the inversion in (4.19) using the Fast Fourier Transform (FFT). This allows for a great increase in the speed of the numerical computations since we can obtain the value of the pdf for all different Δx_i with a single inversion. The main ideas behind the FFT algorithm are explained in Appendix B.

Now that we have the density function of Δx with parameters $\psi = \{\mu, \kappa, \theta, \sigma_V, \rho\}$ we are able to compute the log-likelihood function by substituting (4.19) in (4.4).

In order to perform the maximization of the log-lokelihood function to calibrate the parameters, we have to make sure that the Feller condition (3.22) is satisfied.

The MLE procedure will thus need to be performed including an additional constraint in the optimization given by Feller condition:

$$\hat{\psi} = \arg\max_{\psi \in \Psi} \ \ell(\psi|\Delta x) \tag{4.20}$$

$$\Psi = \{(\mu, \kappa, \theta, \sigma_V, \rho) \in \mathbb{R}^5 \mid \kappa, \theta, \sigma_V > 0, \rho \in [-1, 1], 2\kappa\theta \geq \sigma_V\} \tag{4.21}$$

4.3.2 Multi-asset Heston Calibration

As we already stated in this paper, the approach to extending the Heston model to N assets that we are taking into consideration is the parsimonious multi-asset model introduced by Dimitroff et al. in [7]. It's called *parsimonious* since, as we explained in chapter 3, we are only modelling the asset-asset correlations.

The procedure to calibrating a N asset model is the same that we presented in the multi-asset Merton section and consists of three parts.

Firstly, we have to calibrate all the single asset parameters by themselves. Then we can proceed to compute each asset-asset correlation by solving (4.15) through the simulation of the realization of the log-returns in different scenarios. Finally, we perform Jäckel regularization algorithm to make that the final output matrix is indeed a correlation matrix.

The main steps were all presented in section 4.2.2, here we will only give the details for the simulation of a two-asset Heston model.

Since we now have two processes for each asset, we have to simulate the path for both, in each time-step. The discretized dynamics of the log-returns and the variance for a single asset are the following:

$$x_i = x_{i-1} + (\mu - \frac{V_{i-1}}{2})\Delta t + \sqrt{V_{i-1}\Delta t}\, z_x, \tag{4.22a}$$

$$V_i = V_{i-1} + \kappa(\theta - V_{i-1})\Delta t + \sigma_V \sqrt{V_{i-1}\Delta t}\, z_V \tag{4.22b}$$

with z_x and z_V that are the two components of a Gaussian vector with mean zero and correlation ρ[2].

The issue that arises when moving from a continuous dynamics to a discrete one as in (4.22) is that the variance V_i might assume negative values

[2]In a more extensive form: $\mathbf{z} = \begin{pmatrix} z_x \\ z_V \end{pmatrix} \sim \mathcal{N}_2\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \Big)$

even if the Feller condition is verified. This happens because the second and third term in the equation of the variance process may be so negative that combined with the first term V_{i-1} the result is less than zero. Having $V_i < 0$ causes the next update at $i+1$ to be undefined since we would have to compute the square root of a negative value in a real framework.

To solve this issue, a number of possible solutions have been proposed and are collected in [12]. Among those we have: *absorption*, which consists in setting $V_i = 0$ when it is negative, *reflection*, taking the absolute value $V_i = |V_i|$, and finally *full truncation*, which is the solution proposed in [12] and the one we will implement. It is obtained modifying (4.22) as following:

$$x_i = x_{i-1} + (\mu - \frac{V_{i-1}^+}{2})\Delta t + \sqrt{V_{i-1}^+ \Delta t}\, z_x, \tag{4.23a}$$

$$V_i = V_{i-1} + \kappa(\theta - V_{i-1}^+)\Delta t + \sigma_V \sqrt{V_{i-1}^+ \Delta t}\, z_V. \tag{4.23b}$$

The notation $y^+ = \max(y, 0)$ indicates the positive part of y.

Equations (4.23) represent the single asset simulation scheme but in order to calibrate the asset-asset correlations as already stated we need to simulate the time-series for a pair of assets.

The updating computations only amount to repeating the single scheme twice:

$$x_i^{(1)} = x_{i-1}^{(1)} + (\mu_1 - \frac{\left(V_{i-1}^{(1)}\right)^+}{2})\Delta t + \sqrt{\left(V_{i-1}^{(1)}\right)^+ \Delta t}\, z_{x^{(1)}}, \tag{4.24a}$$

$$V_i^{(1)} = V_{i-1}^{(1)} + \kappa_1(\theta_1 - \left(V_{i-1}^{(1)}\right)^+)\Delta t + \sigma_{V^{(1)}} \sqrt{\left(V_{i-1}^{(1)}\right)^+ \Delta t}\, z_{V^(1)}, \tag{4.24b}$$

$$x_i^{(2)} = x_{i-1}^{(2)} + (\mu_2 - \frac{\left(V_{i-1}^{(2)}\right)^+}{2})\Delta t + \sqrt{\left(V_{i-1}^{(2)}\right)^+ \Delta t}\, z_{x^{(2)}}, \tag{4.24c}$$

$$V_i^{(2)} = V_{i-1}^{(2)} + \kappa_2(\theta_2 - \left(V_{i-1}^{(2)}\right)^+)\Delta t + \sigma_{V^{(2)}} \sqrt{\left(V_{i-1}^{(2)}\right)^+ \Delta t}\, z_{V^{(2)}}. \tag{4.24d}$$

The important difference with the single asset case is the correlation structure. In order to simulate the random vector $\mathbf{z} = (z_{x^{(1)}}, z_{V^{(1)}}, z_{x^{(2)}}, z_{V^{(2)}})^T$ we have to extract it from a 4-dimensional Gaussian with zero mean and covariance given by Σ where:

$$\Sigma = \begin{pmatrix} 1 & \rho_1 & \rho_{1,2} & \rho_{1,2}\rho_2 \\ \rho_1 & 1 & \rho_1\rho_{1,2} & \rho_1\rho_{1,2}\rho_2 \\ \rho_{1,2} & \rho_1\rho_{1,2} & 1 & \rho_2 \\ \rho_{1,2}\rho_2 & \rho_1\rho_{1,2}\rho_2 & \rho_2 & 1 \end{pmatrix} \tag{4.25}$$

We could also use the Cholesky decomposition of $\Sigma = LL^T$ in order to obtain $\mathbf{z} = L\tilde{\mathbf{z}}$ from a standard 4-dimensional Gaussian vector $\tilde{\mathbf{z}}$. In this case matrix L would be lower triangular and defined as:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \rho_1 & \sqrt{1-\rho_1^2} & 0 & 0 \\ \rho_{1,2} & 0 & \sqrt{1-\rho_{1,2}^2} & 0 \\ \rho_{1,2}\rho_2 & 0 & \rho_2\sqrt{1-\rho_{1,2}^2} & \sqrt{1-\rho_2^2} \end{pmatrix}. \tag{4.26}$$

From here on, the procedure is the same as for Merton: we need to solve (4.15) for every possible asset pair to obtain the model correlation M.

The final step is to apply Jäckel regularization algorithm to obtain a valid correlation matrix $\widehat{M}$.

Chapter 5

Optimal Portfolio Allocation

In this chapter we will explore what the optimal allocation is for our portfolio of assets. We will study the *efficient frontier* using two different risk measures, volatility and expected shortfall. In all our analyses, we will be comparing the effects that including Bitcoin in our portfolio has on the optimal allocation.

5.1 Markowitz Mean-Variance Portfolio Optimization

Modern Portfolio Theory (MPT) is a mathematical framework for creating a portfolio of asset by maximizing the expected return for a given level of risk or by minizing the risk while maintaining the same expected gain. Before the article [13] by Harry Markowitz in 1952, the concept of *diversification* (the old warning *not to put all your eggs in one basket*) was only driven by the experience of how markets behave. Moreover, investors used to base their decisions on expected return alone and thus when given a choice between two assets with different expected returns, they would put all their money on the top performing one.

With his article, that would later grant him the Nobel Prize in Economics, Markowitz introduced a more rigorous and mathematically sound framework to assembly a portfolio of assets. His key insight is that an asset's return and risk risk should not be assessed by itself, but rather by how it affects the overall portfolio risk and return. To do so, the *variance* is used as a proxy for risk. Hence the name *mean-variance* analysis that is often used as a substitute for MPT.

Let's introduce the assumption underlying the MPT:

1. Investors are *risk averse*: they will always choose the less risky asset, when two assets offer the same return. At the same time, an investor wanting a higher return has to be willing to accept a higher risk. This equally holds for portfolios as a whole: given two portfolio with the different risk profiles, he will choose the less risky in case of same return and the most remunerating in case of same risk.

2. Portfolio return is the weighted sum of the single assets' returns: in general $\mathbb{E}[R_{ptf}] = \sum_{i=1}^{N} w_i \mathbb{E}[R_i]$.

3. Portfolio variance is a function of both the assets variances and their correlations: $V_{ptf} = \sum_{i=1}^{N} w_i \sigma_i^2 + \sum_{i=1}^{N} \sum_{j \neq i, j=1}^{N} w_i w_j \rho_{i,j} \sigma_i \sigma_j$

Items 2 and 3 above can be more compactly stated using matrix notation, which will come in handy later on in our analysis:

$$r_{ptf}(\mathbf{w}) = \mathbf{w}^T \mathbf{r} \tag{5.1}$$

$$\sigma_{ptf}^2(\mathbf{w}) = \mathbf{w}^T \Sigma \mathbf{w} \tag{5.2}$$

where we have the weights vector $\mathbf{w} = [w_1, w_2, ..., w_N]^T$, $\mathbf{r} = [r_1, r_2, ..., r_N]^T$, using the shorthand $r_i = \mathbb{E}[R_i]$ and finally Σ is the $N \times N$ covariance matrix of the assets.

We can now state the *optimization problem* involving the minimization of the portfolio risk for a specified expected portfolio return in terms of the variable we have just introduced.

$$\begin{aligned}
&\min_{\mathbf{w} \in \mathbb{R}^N} && \sigma_{ptf}^2(\mathbf{w}) && \text{(5.3a)}\\
&\text{subject to} && \mathbf{e}^T \mathbf{w} = 1, && \text{(5.3b)}\\
& && \mathbf{r}^T \mathbf{w} = r_{target}, && \text{(5.3c)}\\
& && w_i \geq 0, \text{for } i = 1 \dots N. && \text{(5.3d)}
\end{aligned}$$

where $\mathbf{e}$ indicates a vector of ones and the first constraint makes sure that the sum of the weights always equals to one. This is to represent a portfolio in which all the money available is allocated in the assets we are taking into consideration. The second constraint ensures that the portfolio allocation $\mathbf{w}$ produces the target expected return r_{target}. Finally, the last constraint is in fact optional and is only used to exclude the possibility to go short on any asset.

The optimization problem in (5.3) has a quadratic objective function given by (5.2) and only has linear constraints[1]. Thanks to this property, the optimization can be carried out numerically by any of the quadratic/linear optimizers that are available for most programming languages.

As we are going to explain in the following sections, we will be mainly focusing on the case where there is no short selling, as indeed so far there are no instruments on the market that allow an investor to go short on Bitcoin and our analysis shows that the main diversification advantage comes from including Bitcoin in our portfolio. Allowing short-selling improves our diversification capability only so slightly.

5.2 Markowitz Efficient Frontier

It is interesting to study the set of optimal allocation as a whole, rather than simply focus on one target return and minimizing the portfolio risk. To do so, we can consider a set of target returns and compute for each of them the respective minimum variance. We thus get a set of pairs (σ^2, r) that represents the best allocation in terms of the minimum risk.

We can thus plot those pairs on an X-Y graph and obtain a curve, the *portfolio frontier*, that intrinsically represents our portfolio of N assets. As a usual practice in finance, we will be plotting on the X-axis the volatility σ instead of the variance σ^2.

In Figure 5.1 we can see what the portfolio frontier looks like for our portfolio of assets. It is interesting to notice how the curve divides the plane in two region: the area to the left of the line includes all those pairs (σ, r) that are not attainable with our assets, since they have a volatility that is too low for that level of expected return. On the other hand, the region to the right of the portfolio frontier is made of all the pairs that are possible to obtain with a specific allocation $\mathbf{w}$ but that will never be chosen by an investor: moving to the left on the same level of return we eventually reach a point on the frontier. The portfolio represented by this point will dominate the one we started from in terms of risk, so it will always be a better choice.

We can proceed with the same argument arguing in terms of best return for a given level of risk: we can thus introduce the *efficient* frontier. For every level of volatility that has two corresponding points on the portfolio frontier, only the one with the higher expected return will be chosen by an investor in our reference framework: hence only the top half of the curve (from the vertex and up) will form the *efficient portfolio frontier*.

[1]The last positivity constraint can be easily expressed in matrix form by writing $\mathbf{I}_N\mathbf{w} \geq \mathbf{0}_N$ where $\mathbf{I}_N$ is the identity matrix of order N and $\mathbf{0}_N$ is N-dimensional vector of zeros.

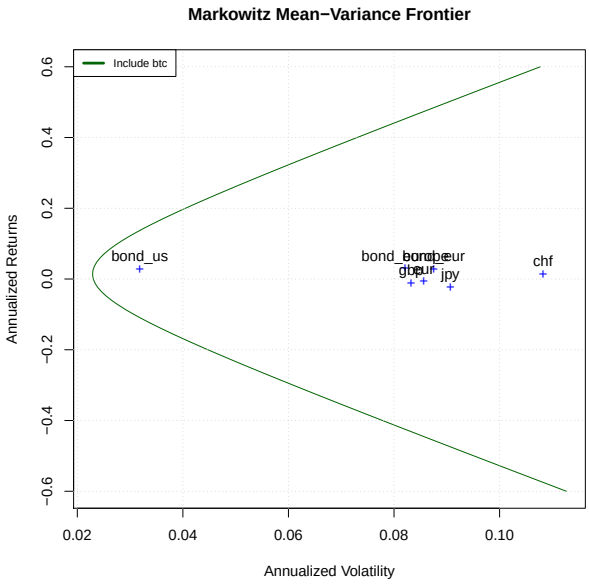


Figure 5.1: The Markowitz Mean-Variance frontier obtained from our portfolio of assets, that includes Bitcoin and allows short-selling.

As a summary, let us keep in mind that any point in the volatility-expected return plane *dominates* all the other portfolio allocations that are represented by points situated below and to the right of it. Conversely, that same point will be dominated by all other allocations above and to the left of it.

Another mathematical perk of considering only the efficient top section is that it renders the frontier a *bijective* function. Hence, for each (attainable) level of risk there is one and only one corresponding expected portfolio return, and vice-versa.

5.2.1 Efficient Frontier with and without Bitcoin

Let us now study how including Bitcoin in our portfolio of assets can help increase the diversification and obtain a higher (expected) return with a lower risk.

In Figure 5.2 we have plotted the efficient Markowitz frontier in all possible cases: including and excluding Bitcoin while allowing short-selling, and the same two curves when short-selling is not allowed.

Comparing the two green lines, we can clearly see that the exclusion of short-selling does not penalize our efficient allocation by much. The same

Volatility Level	Return without Bitcoin	Return including Bitcoin
2,61%	3,00%	3,00%
2,75%	3,89%	7,70%
3,00%	4,59%	10,94%
3,25%	5,05%	13,37%
3,50%	5,43%	15,48%
3,75%	5,76%	17,41%
4,00%	6,06%	19,21%
4,25%	6,34%	20,94%
4,50%	6,61%	22,60%
4,75%	6,87%	24,21%
5,00%	7,12%	25,79%
5,25%	7,37%	27,34%
5,50%	7,61%	28,86%
5,75%	7,85%	30,37%
6,00%	8,08%	31,85%

Table 5.1: Expected return for different levels of volatility, both including and excluding Bitcoin.

can be stated for the red and orange curves, which represent our portfolio when excluding the digital asset. Thus, given our particular set of assets, allowing for short-selling does very little to improve the diversification of our portfolio.

Let us now take a look of what happens when we include Bitcoin in the reference portfolio: as we can see from Figure 5.2, we get a significant improvement in the expected return when considering each level of risk . Equivalently, for the same level of return we have a noticeable decrease in the volatility of our portfolio.

*****AGGIUNGERE COMMENTO INCISIVO*****

We can see some numerical proof of the diversification properties of adding Bitcoin to our portfolio in Table 5.1 and Table 5.2.

5.2.2 Portfolio Allocation

We have so far seen the implications of introducing the digital asset in our portfolio in terms of improvement in the expected return and of lowering the overall portfolio risk. Let us now take a look at how Markowitz MPT allocates the money in the different assets.

Return Level	Volatiliy without Bitcoin	Volatility including Bitcoin
3,00%	2,61%	2,61%
3,50%	2,61%	2,67%
4,00%	2,61%	2,78%
4,50%	2,62%	2,96%
5,00%	2,63%	3,22%
5,50%	2,65%	3,55%
6,00%	2,67%	3,95%
6,50%	2,69%	4,40%
7,00%	2,71%	4,88%
7,50%	2,74%	5,39%
8,00%	2,77%	5,91%
8,50%	2,80%	6,45%
9,00%	2,84%	7,00%
9,50%	2,88%	7,56%
10,00%	2,92%	8,12%
10,50%	2,96%	8,69%
11,00%	3,01%	9,27%
11,50%	3,05%	9,85%
12,00%	3,10%	10,43%
12,50%	3,16%	11,01%
13,00%	3,21%	11,60%
13,50%	3,26%	12,19%
14,00%	3,32%	12,78%
14,50%	3,38%	13,37%
15,00%	3,44%	13,96%

Table 5.2: Volatility for different level of expected portfolio return, both including and excluding Bitcoin.

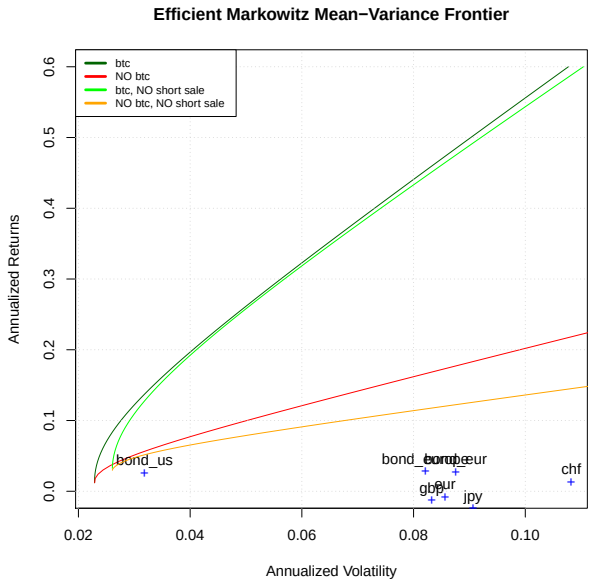


Figure 5.2: The *Efficient* Markowitz Mean-Variance frontier obtained from our portfolio of assets, both including and excluding Bitcoin and with short-selling or not.

To do so, we can plot the values of $\mathbf{w}$ as resulting from (5.3) for different levels of volatilities (and hence returns).

5.3 Portfolio Optimization with CVaR as a Risk Measure

We have so far studied the problem of optimal portfolio allocation through Markowitz MPT, considering the portfolio volatility as a proxy for its risk. This is only one of the possible choices in a more general framework in which the optimization problem to be solved maintains the same constraints but has a different objective function. We can reformulate (5.3) as follows:

$$\min_{\mathbf{w}\in\mathbb{R}^N} \quad PtfRisk(\mathbf{w}) \tag{5.4a}$$

$$\text{subject to} \quad \mathbf{e}^T\mathbf{w} = 1, \tag{5.4b}$$

$$\mathbf{r}^T\mathbf{w} = r_{target}, \tag{5.4c}$$

$$w_i \geq 0, \text{for } i = 1 \dots N. \tag{5.4d}$$

where we have substituted the portfolio volatility with a general measure of the portfolio risk as a function of the weights $\mathbf{w}$.

Using the portfolio volatility as risk measure has its perks and downsides: it is an intuitive and simple way to evaluate how the possible returns vary from the expected value but it takes into account both the positive and the negative deviation from the mean. Thus, a high volatility may be caused by a few extremely high returns that are anything but something to avoid and penalize. For this reason, the notion of *semi-volatility* was introduced to only consider variations towards lower returns than the one expected. This solution however is not very popular.

A more common approach is to measure risk based on the quantile of the loss distribution [2]. The *Value-at-Risk* of level α for a loss distribution is precisely defined as the quantile of order α [3]. In formulas, VaR_α is the value such that:

$$\mathbb{P}(Loss \leq VaR_\alpha) = \alpha \tag{5.5}$$

VaR is a vastly popular and common way to measure the so called "tail-risk": the intrinsic risk contained in loss events that happen very rarely. Its advantage is that it considers only the downside of the expected return as opposed to what the volatility does.

However, VaR has drawbacks as its mathematical definition does not make it a *coherent measure of risk*. Specifically, it lacks the property of sub-additivity: the VaR of two different portfolios considered as one may be greater than the sum of the two single VaRs. This is in direct contradiction with the principle of diversification.

aggiungere appendice dove si definisce una misura coerente di rischio con le 3/4 proprieties

An improved version of VaR is the *Conditional Value-at-Risk* (CVaR), also referred to as *expected shortfall* since indeed it is defined as the average loss among the values of the loss distribution that exceed the corresponding VaR level. Thus, for a continuous loss distribution is defined as:

$$CVaR_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 VaR_\gamma d\gamma \tag{5.6}$$

[2] Depending on how the returns are expressed, we have different ways to compute the loss. In particular, considering the time interval $[0, T]$, for log-returns $r_{log} = log(S_T/S_0)$ the loss is simply the opposite $Loss_{log} = -r_{log}$ while for $r_\% = S_T/S_0$ the loss is $Loss_\% = 1 - r_\%$. As usual we indicate by S_t the price of the asset or the value of the portfolio at time t.

[3] There are two different notations when it comes to what the number α indicates. In this work, α is considered as the percentage of losses that are lower or equal to the value of VaR_α. The other notation is to consider α as the percentage of losses that exceed the VaR_α.

There are two advantages in the usage of CVaR as opposed to VaR. Firstly, it can be proven that CVaR is a *coherent* measure of risk, so we gain an important mathematical property that before was lacking. Secondly, we now take into account also the very extreme values that a loss distribution might have. The latter is especially true for discrete loss distributions (e.g. losses obtained from daily assets returns, which will is our main focus) since there might be a very few truly high losses that nonetheless will be considered in the computation of CVaR. On the other hand, the VaR for discrete distributions just amounts to sorting all the possible losses in increasing order and taking the element in position $100 * \alpha / N_{sample}$ as our VaR_α.

These are the reasons that induced us to consider the CVaR as an alternative measure of portfolio risk to be inserted as the objective function to be minimized in the optimization problem (5.4). Considering the literature, there are already a few studies in which the conditional value at risk is used in the optimal portfolio allocation problem, above all the paper by Rockafellar and Uryasev in [17]. Following studies include for instance in [6] and [16]. Most of these approaches are based on continuous loss functions, and as a further study beyond this work it could be of interest to study the diversification effect of the inclusion of Bitcoin in the assets portfolio in those frameworks. As we are about to see, our approach focuses on optimizing the empirical historical CVaR from daily return data.

5.4 CVaR Efficient Frontier

Following a similar approach to the one we developed in previous sections regarding Markowitz optimization, we will now present the CVaR efficient frontier.

5.4.1 Efficient Frontier with and without Bitcoin

In order to perform an analysis similar to what we have done using Markowitz approach in the case of the daily CVaR as the portfolio risk measure, we have to solve the following optimization problem:

$$\begin{aligned}
\min_{\mathbf{w}\in\mathbb{R}^N} \quad & PtfCVaR_\alpha^{daily}(\mathbf{w}) && \text{(5.7a)}\\
\text{subject to} \quad & \mathbf{e}^T\mathbf{w} = 1, && \text{(5.7b)}\\
& \mathbf{r}^T\mathbf{w} = r_{target}, && \text{(5.7c)}\\
& w_i \geq 0, \text{for } i = 1 \dots N. && \text{(5.7d)}
\end{aligned}$$

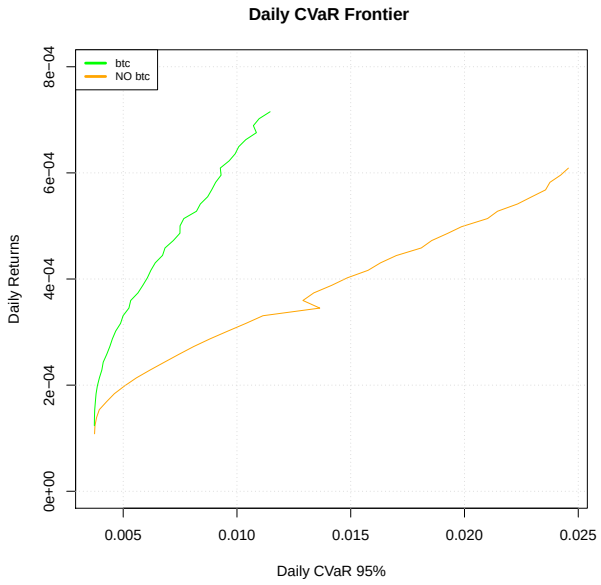


Figure 5.3: The *Efficient* Markowitz Mean-Variance frontier obtained from our portfolio of assets, both including and excluding Bitcoin and without short-selling.

To compute the daily CVaR of the portfolio, we considered the daily percentage returns for all the assets and all the time-steps available, and stored them in matrix R. R has dimension $N_{timesteps} \times N_{assets}$. We then obtained the distribution of returns for a portfolio with weights $\mathbf{w}$ by doing $R_{ptf}(\mathbf{w}) = R\mathbf{w}$. The CVaR is now computed as explained in the previous section.

We computed the efficient frontier for the two usual portfolios: one that includes Bitcoin and the other that only contains *standard* assets, both without short-selling. The graphs for both frontiers are plotted in Figure 5.3.

At first glance, we notice that both curves are not perfectly smooth: this is due to the fact that we are using the *discrete* historical distribution for the returns of the assets. On the contrary, when we set our analysis in the MPT framework it can be shown that the optimal frontier has the shape of a hyperbole when allowing short-selling. Without short sales, the frontier is no longer a hyperbole but it is still a smooth function.

The improvement in portfolio diversification when including Bitcoin in our basket is clearly visible in how the two efficient frontiers part from each other, in the same way that they did in Figure 5.2. We can easily tell from the picture that for the same level of return an investor has to be willing to

risk losing more money (higher CVaR) when he does not include Bitcoin in his portfolio of asset. Similarly, for a given level of risk, including the digital asset in our basket allows for a significant increase in the daily expected return, hence also for a higher annual profit.

*******ADD TABLE OF RESULTS***********

5.4.2 Portfolio Allocation

We have seen again the implications of introducing the digital asset in our portfolio in terms of improvement in the expected return and of lowering the daily CVaR of the portfolio. Let us now take a look at how this second type of optimization allocates the money in the different assets.

To do so, we will plot the values of $\mathbf{w}$ as resulting from (5.7) for different levels of daily CVaR (and hence returns).

Chapter 6

Conclusions

Appendix A

Characteristic Function of a Compound Poisson Process

By definition, the characteristic function of a random variable X is given by:

$$\phi_X(u) = \mathbb{E}[e^{iuX}] \tag{A.1}$$

Let's first consider a simple Poisson process N_t of parameter λ. At each time $t > 0$, N_t has a discrete distribution that follows:

$$\mathbb{P}(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!},\ n = 0, 1, 2, ... \tag{A.2}$$

Since N_t is a *discrete* random variable, the expectation in A.1 amounts to a sum over all the possible values of N_t:

$$\begin{aligned}\phi_{N_t}(u) = \mathbb{E}[e^{iuN_t}] &= \sum_{n=0}^{\infty} e^{iun}\mathbb{P}(N_t = n) \\ &= \sum_{n=0}^{\infty} e^{iun}e^{-\lambda t}\frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t e^{iu})^n}{n!}\end{aligned}$$

Using the definition of exponential $e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}$ we then get the final result:

$$\begin{aligned}\phi_{N_t}(u) &= e^{-\lambda t}e^{\lambda t e^{iu})} \\ &= e^{\lambda t(e^{iu}-1)}\end{aligned}$$

Let's consider now a compound Poisson process defined by

$$X_t = \sum_{i=1}^{N_t} Y_i \tag{A.3}$$

where Y_i are i.i.d. and have density expressed by the function $f_Y(y)$.

To compute the characteristic function of X_t we can follow the same steps as in the simple Poisson case, but in addition we use the theorem of total expectation to first simplify the expression and then proceed exploiting the i.i.d property:

$$\begin{aligned}
\phi_{X_t}(u) &= \mathbb{E}[e^{iuX_t}] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[e^{iuX_t}|N_t = n]\mathbb{P}(N_t = n) \\
&= \sum_{n=0}^{\infty} \mathbb{E}[\prod_{i=1}^{N_t} e^{iuY_i}|N_t = n]\mathbb{P}(N_t = n) \\
&= \sum_{n=0}^{\infty} \prod_{i=1}^{n} \mathbb{E}[e^{iuY_i}|N_t = n]\mathbb{P}(N_t = n) \\
&= \sum_{n=0}^{\infty} (\mathbb{E}[e^{iuY}])^n \mathbb{P}(N_t = n) \\
&= \sum_{n=0}^{\infty} (\phi_Y(u))^n e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\
&= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t\, \phi_Y(u))^n}{n!} \\
&= e^{\lambda t(\phi_Y(u)-1)}
\end{aligned}$$

Appendix B

Characteristic Function of a Compound Poisson Process

The Fast Fourier Transform is an algorithm that enables us to compute the Discrete Fourier Transform faster. In particular, a *naïve* implementation of the DFT requires a number of operations on the order of $\mathcal{O}(N^2)$, where N is the number of points that we use for the discrete transform. An FFT algorithm performs the same computation using $\mathcal{O}(N \log N)$ operations.

The discrete Fourier transform for a set $x = x_0, ..., x_{N-1}$ is represented by N values u_k, $k = 0, ..., N-1$:

$$u_k = \sum_{n=0}^{N-1} x_n e^{-ikn2\pi/N}, \; k = 0, \ldots, N-1. \tag{B.1}$$

As we can see, we have to perform $\mathcal{O}(N)$ computations for each u_k, for a total of $\mathcal{O}(N^2)$ operations. The FFT is a smart way to compute the same quantities by dividing $N = N_1 N_2$ into two factors N_1 and N_2, computing the DFT for the two smaller samples and then aggregating the results to return the final N values of u_k.

The factorization of N can be performed recursively on N_1 and N_2 as long as they are not primes, so the usual approach is to take $N = 2^n$ as the n-th power of 2. The most common algorithm is the one by Cooley and Tuckey in [4] and is the one that is usually implemented in most computational tools (in our case R).

Bibliography

[1] Albrecher, H., Mayer, P., Schoutens, W., and Tistaert, J. The little heston trap. *Wilmott* (03 2007), 83–92.

[2] Ball, C. A., and Torous, W. N. A simplified jump process for common stock returns. *Journal of Financial and Quantitative Analysis 18*, 01 (1983), 53–65.

[3] Bates, D. S. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies 9*, 1 (1996), 69–107.

[4] Cooley, J. W., and Tukey, J. W. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation 19*, 90 (1965), 297–301.

[5] Cox, J. C., Ingersoll, J. E., and Ross, S. A. A theory of the term structure of interest rates. *Econometrica 53*, 2 (1985), 385–407.

[6] DiClemente, A. The empirical value at risk-expected return frontier: A useful tool of market risk managing. *Centro Interdipartimentale sul Diritto e l?Economia dei Mercati*, 11 (2002).

[7] Dimitroff, G., Lorenz, S., and Szimayer, A. A parsimonious multi-asset heston model: Calibration and derivative pricing. *International Journal of Theoretical and Applied Finance (IJTAF) 14*, 08 (2011), 1299–1333.

[8] Dragulescu, A. A., and Yakovenko, V. M. Probability distribution of returns in the heston model with stochastic volatility. *Quantitative Finance 2*, 6 (2002), 443–453.

[9] Heston, S. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6* (02 1993), 327–43.

[10] Honoré, P. Pitfalls in estimating jump-diffusion models. *SSRN Electronic Journal* (01 1998).

[11] Kou, S. G. A jump-diffusion model for option pricing. *Management Science 48*, 8 (2002), 1086–1101.

[12] Lord, R., Koekkoek, R., and Dijk, D. V. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance 10*, 2 (2010), 177–194.

[13] Markowitz, H. Portfolio selection. *The Journal of Finance 7*, 1 (1952), 77–91.

[14] Martin, M. A two-asset jump diffusion model with correlation. Master's thesis, University of Oxford, 2007.

[15] Merton, R. Option prices when underlying stock returns are discontinuous. *Journal of Financial Economics 3* (01 1976), 125–144.

[16] Quaranta, A. G., and Zaffaroni, A. Robust optimization of conditional value at risk and portfolio selection. *Journal of Banking & Finance 32*, 10 (2008), 2046–2056.

[17] Rockafellar, R. T., Uryasev, S., et al. Optimization of conditional value-at-risk. *Journal of risk 2* (2000), 21–42.

[18] Tankov, P., and Cont, R. *Financial Modelling with Jump Processes, Second Edition*. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis, 2015.