Internship report (CONFIDENTIAL)

Simulation of Equity Index under historic probability

Eric Va
Université Pierre et Marie Curie
Master 2 Probabilité et Finance

Internship supervisor :
Nicolas Grima
Model Review Lead for EMEA Equity
HSBC Global Banking & Markets

Educational supervisor :
Stefano de Marco
Assistant Professor
CMAP, Ecole Polytechnique

21 septembre 2018

Introduction : What is a good historical volatility model ?

Volatility is an abstract concept that captures a feature in the stock market returns behavior. It measures somehow the variance of the returns. Unfortunately, it cannot be directly observed, even though we can think about known volatility indexes, which use in any case an underlying model (like the CBOE VIX which uses implied volatility of vanilla options).
One has to formulate a clear definition of volatility before analyzing it. Usually, it is not model-free : we have to formulate precise mathematical hypothesis to define volatility. Many methods exist in the literature to estimate volatility. Parametric estimations are more common in the option pricing industry. It consists first in making assumptions on the dynamic of the asset we want to model. And the dynamic is usually governed by a volatility that is thus defined by the dynamic. For instance, volatility can be defined in a Brownian semi-martingale framework, as the diffusion coefficient in front of a Brownian motion. It can be a constant (like in the Black-Scholes model), or a deterministic function of time and the underlying current price (like in local volatility model).
In pricing/hedging, under these parametric assumptions, one uses the current market data to estimate all the parameters of the volatility model. This estimation (or calibration) must respect some important properties in order to be useful for sell side business, like no free lunch conditions. It must particularly also match the current option prices given by the market. This calibration to the current option price is fundamental because it is the starting point for defining hedging strategies in order to mitigate market risk when dealing with these options. One cannot give an option price if it is not able to define a feasible hedging strategy for this option. The option price is interpreted as the premium one has to pay for hedging it. This vision is called "risk-neutral" as one is only concerned by volatility calibrating current market data and its future evolution. Indeed, market makers can directly hedge options themselves by buying/selling assets on the market at the current price. Therefore, they practically use models that are completely independent of the past (using martingale process for instance, with no drift).

As far as we are concerned, we adopt here a "historical" point of view. It means that we will focus on defining and estimating models for volatility using past market data. This historical vision is more appropriate in the buy side business and risk management. If we consider an asset manager having a portfolio of assets, we want to calculate some risk metrics like the 95% quantile of the portfolio profit/loss distribution (Value-at-Risk), no matter the hedging strategies deployed. In this case, we assume that we cannot hedge this portfolio and we are completely exposed to the variations of its value. Then, we are interested in how it can possibly go worse in the future. To assess that, we would like to take into account the history of the portfolio assets. Here again, we have to set up statistical model in order to define volatility. The difference with the risk neutral vision is that we calibrate the models on the past market data.
To that end, many empirical studies have been made on financial time series that can be characterized by so called styled facts (see for instance [6]). We list here the principal ones on daily data : fat tailed and skewed distribution, long memory volatility, volatility clustering, time scaling properties. Then, we consider that a historical volatility model will be acceptable if it captures all these stylizes facts.
The report is organized as follow. In the first part, we introduce succinctly the main stylized fact on Equity returns. In the second part, we introduce different specifications of GARCH type models. In the third part, we describe the calibration and the results of the GARCH simulations. And in the last part, we present Multifractal modeling.

Table des matières

Chapitre 1

Stylized fact on realized volatility

1.1 Fat tail

Fat tail distributed return is illustrated in the empirical distribution of Equity indexes daily returns on figure 1.1 with high kurtosis. High kurtosis means superior to three which corresponds to the normal distribution. We recall that kurtosis is the statistical name for the fourth order moment. When it is superior to three, it means that the distribution profile is more peaked around its mean and that it decays slower to zero than the Gaussian distribution. We say that the distribution has fat tail (or is leptokurtic). We insist on the fact that, in whole that follow, we are interested by the distribution of the yield over a time period. For instance, if we observe a single path of the yield at N given dates $y_{t_1}, y_{t_2}, ..., y_{t_N}$, we will focus on the historical distribution of these values and compute risk metrics on that distribution. Therefore, it is the empirical distribution of $y_{t_1}, y_{t_2}, ..., y_{t_N}$ that is fat tailed. Note that we could have used a more specific definition called "heavy tail" distribution, which focus on the tail distribution that decay hyperbolically. In all this report, we will use the kurtosis definition.

1.2 Volatility clustering and Long term memory

Volatility clustering is the observed phenomenon that high movement in returns fluctuations are usually followed by other high movements fluctuations of these same returns (and similarly, low movements tend to be followed by low movements). It means that "volatility" is somehow packed through time. In order to give some quantitative intuition on volatility clustering, one can observe the autocorrelation function [1] (ACF) of the absolute value of returns and sees that it

1. The ACF f_X of a process X is the function associating a lag to the Pearson correlation between X and its corresponding lag process : $f_X(\tau) = Cor((X_t)_t, (X_{t+\tau})_t)$.

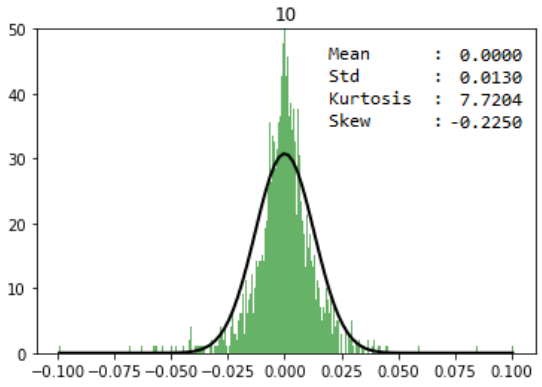

FIGURE 1.1 – Empirical distribution of Equity index 10 daily returns. In blue, fit Gaussian density.

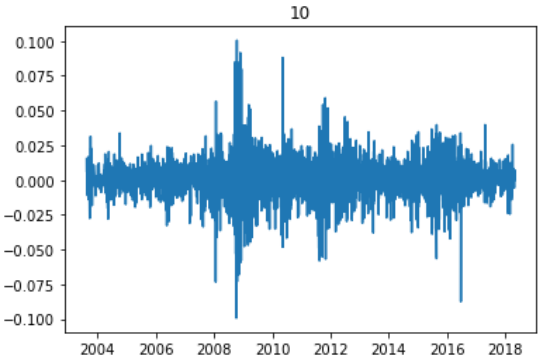

FIGURE 1.2 – Historical values of equity index 10 daily returns.

is significant for days (see figure 1.4). Also, if we plot the ACF of the square of the returns, we can see a slow decay over the lags : volatility is said to have long term memory. More precisely, a process is said to have long term memory if its autocorrelation function decay hyperbolically. In practical terms, it means that the squared returns present significant autocorrelation over months. On the contrary, returns themselves exhibits no autocorrelation on low frequency lags, meaning that we cannot significantly predict the future returns with a linear predictor, which is consistent with the efficient market hypothesis (see figure 1.4). Long term memory obviously imply volatility clustering, but the reciproque is not true (see for instance GARCH models presenting volatility clustering but with an exponential decay of its ACF).

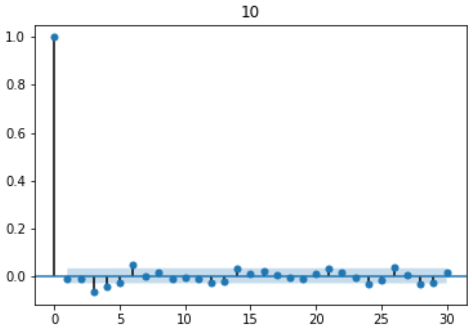

FIGURE 1.3 – Autocorrelation function of Equity10 daily returns.

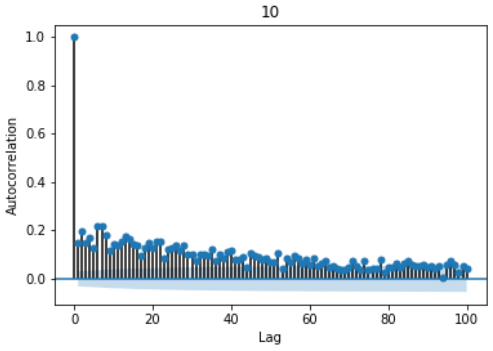

FIGURE 1.4 – Autocorrelation function of Equity10 absolute daily returns

1.3 Skewness and leverage effect

The empirical distributions of returns display a significant negative skew (i.e. negative third order moment). That means upward movement in stocks appear more often than downward movements. Leverage effect can partly accounts for skewness. Leverage effect is the empirical negative correlation observed between measures of volatility and the returns. Indeed, the more a stock value of a company go down, the more likely this company will have to leverage itself, i.e. use debt to finance itself. As leverage increases, perception of risk for this company on the markets increases, and consequently volatility increases.

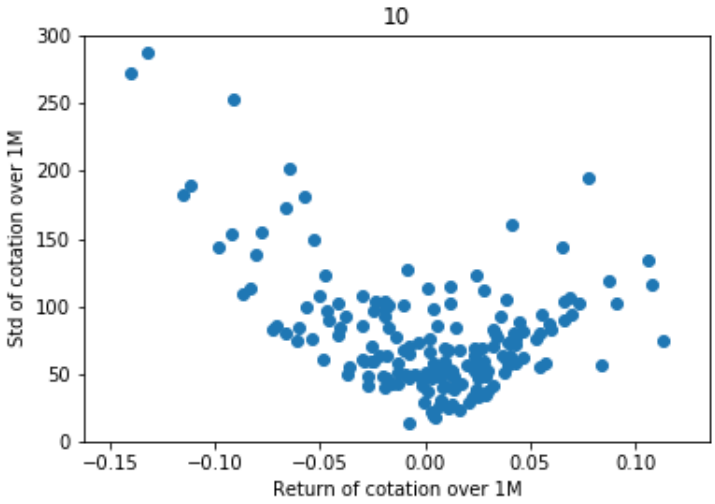

FIGURE 1.5 – Empirical Standard Deviation plotted against average daily return of Equity index 10. One point corresponds to a distinct month

1.4 Multifractality

A process is said to have a time scale invariance property if its "forms" (statistical behavior, distribution) do not change by time dilatation of contraction up to a certain factor. The Brownian motion B for example is scale invariant because of the scaling property $\forall a > 0, \sqrt{a}B_{at} \overset{law}{=} B_t$. For the Brownian motion, the scaling factor $\sqrt{a}$ is deterministic and depends only on the dilatation coefficient a. We say then that it is a monofractal scale invariance. If the scaling factor is random (but always depending only of a), we have a multifractal scale invariance. Scale invariance is illustrated on the historical yields whose distribution is fat tailed on a daily scale, but converges in a regular manner toward a normal distribution for

very low annual frequencies (see graph 1.6 for the N225 index). Multi fractality is a property on the distribution.

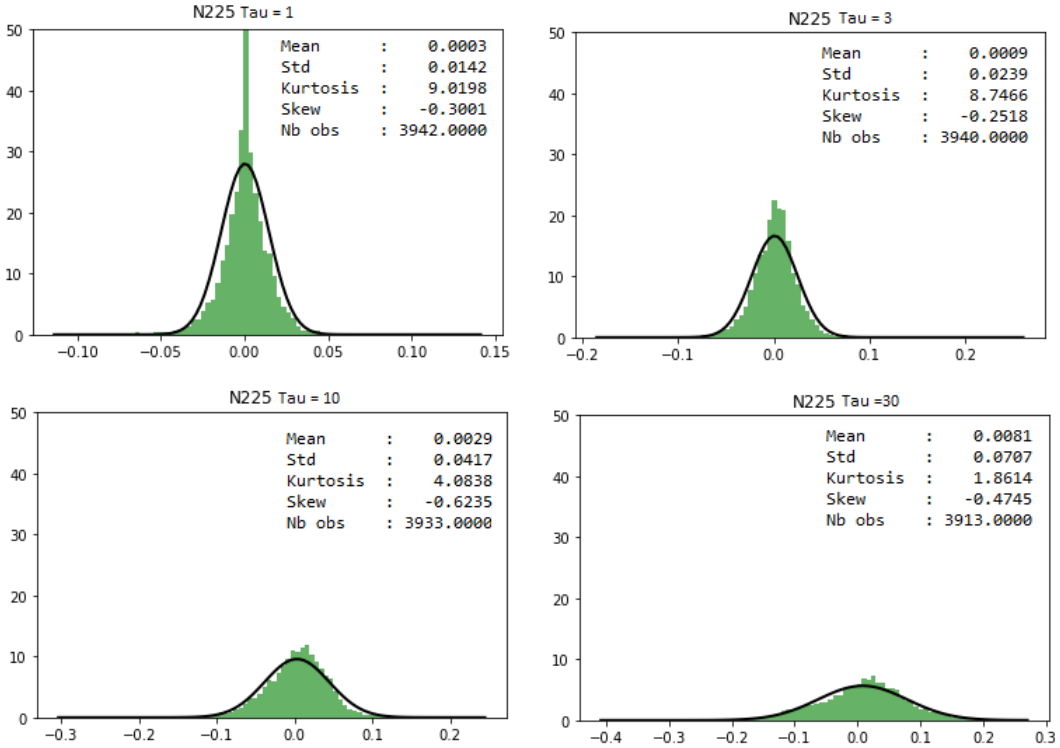

FIGURE 1.6 – Empirical of distribution of $\frac{P_{t+\tau}}{P_t} - 1$ for $\tau = 1, 3, 10, 30$ with $(P_t)_t$ being the N225 daily values. In black line the Gaussian pdf fit.

1.5 Literature review in historical modeling of financial returns

In the financial literature in historical modeling, we are usually interested in simulating returns paths that have the same properties as their historical evolution (stylized facts). These simulations aim at calculating risk metrics of portfolios of options written on these underying returns. Several methods exist to simulate and compute these risk metrics. Some of them are directly proposed by the regulators and are widely spread in banks [2] . We can distinguish parametric and non parametric approaches. In non parametric simulations, the common practice is to re

2. See [10] for more details

sample independently from the past values of returns. One has to fix a rolling period on which the data can be sampled (the two last years at HSBC for instance). The regulator also defined the "stressed Value at Risk" where the simulations come from a sampling on a fix period of turmoil (2007-2009 for instance). The main advantage of historical simulation is that it is simple to implement (particularly on a whole banking portfolio with thousands of risk factors), and that its non parametric feature can reproduce historical fat tail and the complex correlation structure between the assets. The main drawbacks are that it cannot reproduce volatility clustering and thus cannot simulate change of risk regime.
Concerning the parametric methods, either in continuous or discrete time, returns are commonly modeled by a predictable part and a pure noise part. Depending on how the noise, and thus the volatility, is defined, we have two historical types of models : ARCH type models (first presented by Engle and then developed by Bollerslev in the 80's) and Stochastic Volatility models (see Bauwens and al. (2006) and Asai and al. (2006) for more details). In ARCH models, we consider only a unique source of randomness, meaning the returns. In stochastic volatility models, we consider two sources of randomness, one from the return, and one from the volatility itself. Conditionally to the past values of the returns, the ARCH volatility is thus a deterministic, while the Stochastic volatility is... stochastic. In section 2, we will be focused on parametric ARCH type models.
In the 90's and early 2000, researchers developed a new class of models called multi-fractal process to account for multi-scaling invariance property of the historical yields. Section 4 will be devoted to these processes.

Chapitre 2

GARCH

We recall here the specification of GARCH models. GARCH stands for "Generalized Auto-Regressive Conditionally Heteroskedastic" models. They are commonly used in the financial literature in modeling historical financial yields time series. GARCH models present the remarkable propriety of taking into account time-varying volatility and volatility clustering that are often observed in financial data. GARCH model is also known to produce fat tail distributions. In this paragraph, we introduce various specifications of multivariate GARCH that we will use to make our simulations.

2.1 Original form

We specify the model in a discrete framework. Let $T \in \mathbb{N}$ be the length of the time interval of historical data. For a given date $t \in \mathbb{N}$, we observe data at each anterior date $t-1$, $t-2$,..., $t-T$. Let us consider N underlyings. For any date t, we denote their yields by $y_t = (y_t^1, ..., y_t^N)^T \in \mathbb{R}^N$. Let $p \in \mathbb{N}$ and $q \in \mathbb{N}$. In a purely statistical point of view, the specification of a N multivariate GARCH(p,q) is as follow :

$$\forall t, \begin{cases} y_t = f(y_{t-1}, ..., y_{t-T}, \epsilon_t, ..., \epsilon_{t-T}) + \epsilon_t \\ \sigma_t^2 = \gamma + \sum_{s=1}^{q} \alpha_s * \epsilon_{t-s} \epsilon_{t-s}^T + \sum_{s=1}^{p} \beta_s * \sigma_{t-s}^2 \end{cases} \tag{2.1}$$

where $\epsilon_t = (\epsilon_t^1, ..., \epsilon_t^N)^T \in \mathbb{R}^N$ is a realization of a random vector whose variance is given by a symmetric positive semi definite matrix $\sigma_t \in \mathcal{M}_n(\mathbb{R})$. The function f corresponds to a predictor of the value y_t given the data up to $t-1$. Thus, ϵ can be statistically interpreted as a historical path of the residuals of a regression of y_t on the past values of ϵ and y. If f is linear, we recognize the classical ARMA specification. in our simulations, we do not model f and we consider the yields

to be a pure centered noise. The variance matrix also depends on its past values in a linear way, meaning that the model has time-varying (in statistic we say heteroskedatic), and auto-regressive volatility. Note that GARCH models are a generalization of ARCH models which does not contain the moving average term. From a probabilistic point of view, if we set a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathbb{P}$ can be seen as the historical probability, y, ϵ and σ are processes verifying :

$$\forall t, \begin{cases} y_t = \mathbb{E}^{\mathbb{P}}[y_t|\mathcal{F}_{t-1}] + \epsilon_t \\ \sigma_t^2 = \mathbb{V}^{\mathbb{P}}[y_t|\mathcal{F}_{t-1}] \\ \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_t) \end{cases} \tag{2.2}$$

where we conditioned by the filtration $\mathcal{F}_t = \sigma(\epsilon_t, \epsilon_{t-1}, ...)$ (the model is said to be Conditionally Heteroskedastic). The parameters of the N multivariate GARCH(p,q) that must be chosen are $\gamma \in \mathcal{M}_n(\mathbb{R})$, $\forall i \in \{1, ..., q\}, \alpha_i \in \mathcal{M}_n(\mathbb{R})$ and $\forall i \in \{1, ..., p\}, \beta_i \in \mathcal{M}_n(\mathbb{R})$.
The choice of the matrix parameters is tricky as they can contain a lot of coefficients, and that the whole process σ must be symmetric positive definite. Therefore, some practical specifications can be used to guarantee semi definite positiveness (see for instance the Engle and Kroner parameterization in [9]).
We can enforce the matrix parameters to be symmetric positive definite themselves, because, if γ and all the α_i and the β_i are symmetric definite positive, then so will be the variance/covariance matrices σ_t^2 along all the path. That will be the case in our estimations. The estimation technique that we used is a parametric one. We make here a strong assumption on the distribution of the residuals ϵ_t, considering it belongs to a continuous Gaussian density family. Under this assumption, we will estimate the parameters of the model by maximizing the Gaussian likelihood. The choice of the parametric family has been discussed in [15] where she tried other continuous distribution such as Student distributions.

2.2 Flexible Gaussian Multivariate GARCH

For N underlyings, assuming the matrix parameters symmetric, one has to estimate $3 * N * (N + 1)/2$ parameters, which makes classical optimization algorithms impossible to converge even for simple GARCH(1,1) models with relatively small number of assets. In [13], the authors proposed a method to fit multivariate GARCH that seems to give good results. They called it *Flexible Multivariate GARCH*. The idea is to replace the initial problem of maximizing the N dimensional Gaussian likelihood by a set of suboptimal (but easier) univariate and bivariate GARCH estimations. It is a kind of Quasi-maximum likelihood estimate. More pre-

cisely, we estimate the N independent univariate GARCH models as :

$$\forall i \in \{1, ..., N\}, \begin{cases} \max\limits_{\gamma_{ii},\alpha_{ii},\beta_{ii} \in \mathbb{R}} \prod\limits_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_{ii,t}^2}} \exp(-\frac{\epsilon_{i,t}^2}{2\sigma_{ii,t}^2}) \\ \forall t, \sigma_{ii,t}^2 = \gamma_{ii} + \sum\limits_{s=1}^{q} \alpha_{ii,s}\epsilon_{i,t-s}^2 + \sum\limits_{s=1}^{p} \beta_{ii,s}\sigma_{ii,t-s}^2 \\ \sum\limits_{s=1}^{q} \alpha_{ii,s} + \beta_{ii,s} < 1, \alpha_{ii,s} > 0, \beta_{ii,s} > 0, \gamma_{ii} > 0 \end{cases} \tag{2.3}$$

where σ_{t-T} and (ϵ_t) are given. The constraints on $\alpha_{ii} + \beta_{ii} < 1$ guarantee that the real process (σ_t) is stationnary. We denote by $\hat{\alpha}_{ii}$, $\hat{\beta}_{ii}$, $\hat{\gamma}_{ii}$ and $(\hat{\sigma}_{ii,t})$ the resulting estimators. Alongside with these variances estimation, we compute the covariances couple by couple of assets using bivariate GARCH models :

$$\forall i, j \in \{1, ..., N\}, i \neq j, \begin{cases} \max\limits_{\gamma_{ij},\alpha_{ij},\beta_{ij} \in \mathbb{R}} \prod\limits_{t=1}^{T} \frac{1}{\sqrt{2\pi \det(\Sigma_{ij,t})}} \exp(-\frac{1}{2}\Xi_{ij,t}^T \Sigma_{ij,t}^{-1} \Xi_{ij,t}) \\ \forall t, \Sigma_{ij,t} = \begin{pmatrix} \hat{\sigma}_{ii,t}^2 & \sigma_{ij,t}^2 \\ \sigma_{ij,t}^2 & \hat{\sigma}_{jj,t}^2 \end{pmatrix}, \Xi_{ij,t} = \begin{pmatrix} \epsilon_{i,t} \\ \epsilon_{j,t} \end{pmatrix} \\ \sigma_{ij,t}^2 = \gamma_{ij} + \sum\limits_{s=1}^{q} \alpha_{ij,s}\epsilon_{i,t-s}\epsilon_{j,t-s} + \sum\limits_{s=1}^{p} \beta_{ij,s}\sigma_{ij,t-s}^2 \end{cases} \tag{2.4}$$

As mentioned, to insure that Σ is positive definite, we impose that the 2x2 matrices α, β, γ are positive definite. Concerning its stationarity, if we have definite positiveness on the parameters, then the conditions $\forall i \in \{1, ..., N\}, a_{ii} + b_{ii} < 1$ are sufficient to achieve stationarity for bi and univariate GARCH(1,1) (see [13] for the proof). Concerning bivariate models, we will manually ensure that the matrices α, β and γ are semi definite positive by verifying that $\forall s, \alpha_{ij,s}^2 < \hat{\alpha}_{ii,s} * \hat{\alpha}_{jj,s}$ (idem for β and γ).

By fitting the bi and univariate GARCH models independently, we can more easily construct over times covariance/variance matrices of dimension $N * N$. The problem here is that we do not know if the resulting aggregated matrix is positive semi definite. [13] suggested to simply take the nearest positive definite matrix in the norm sense (the Frobenius norm is proposed to facilitate computation). To do so, we use the procedure developed in [11]. More precisely, for a given square symmetric matrix A we want to solve

$$\begin{cases} \min\limits_{B \in \mathcal{S}_n^{++}} \|A - B\|_F \\ \forall B = (b_{ij}) \in \mathcal{M}_N(\mathbb{R}), \|B\|_F = \sqrt{\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N} b_{ij}^2} \end{cases} \tag{2.5}$$

with $\mathcal{S}_n^{++}$ being the set of definite positive matrices. One introduces $B = \frac{1}{2}(A+A^T)$ and its polar decomposition $B = UH$ with $UU^T = I$ and $H = H^T$. Then, it was proved in [11] that the matrix $\frac{1}{2}(B + H)$ is the unique solution of 2.5.

2.3 Dynamic conditional correlation GARCH model

2.3.1 Changing the distribution family : Student, Gaussian mixtures

In order to simulate fat tailed distribution of yields, we decided to change the distribution assumption made on the innovation ϵ as we could not achieve that with Gaussian distribution. Let us first introduce the Multivariate Student distribution, known to have fat tail marginals. We recall its density :

$$f(x) = \frac{\Gamma(\frac{\nu+n}{2})}{\sqrt{(\nu-2)\pi}^n \Gamma(\frac{\nu}{2})\sqrt{|\Sigma|}} * (1 + \frac{1}{\nu-2}(x-\mu)^T\Sigma^{-1}(x-\mu))^{-\frac{\nu+n}{2}} \tag{2.6}$$

where ν is the number of degree of freedom, Σ the covariance matrix and μ the mean vector. The second moment order is defined if and only if $\nu > 2$. And the fourth moment order if finite if and only if $\nu > 4$. The bigger ν is, the thiner the tail is. The t Student distribution converges to a normal distribution point-wise as $\nu \to +\infty$.
To simulate returns under this distribution, we need to estimate the covariance matrix Σ and the parameter ν. To calibrate Σ, we use the Dynamic Conditional Correlation (DCC) specification first presented in [8]. Indeed, the Flexible Multivariate GARCH for variance matrix estimation framework cannot be applied here as we cannot easily unify several univariate and bivariate Student GARCH as it was the case with the normal distributions (particularly, we would have had to estimate a number of degree of freedom for each couple of assets, that would make the estimation instable and the simulation difficult). We decided to choose DCC t Student specification because of its sparsity (only three parameters) and the fact that it enables to have a dynamic correlation matrix having the features of a GARCH model. The DCC t Student has been very well described in [15]. With the same notation as above, we assume that the $N*N$ covariance matrix Σ^2 of the innovations ϵ is given by $\Sigma_t^2 = D_t R_t D_t$ where D_t is a $N*N$ diagonal matrix containing the conditional standard deviation of ϵ at time t : $D_t = \text{diag}(\sigma_{11,t}, ..., \sigma_{nn,t})$ with $\sigma_{ii,t}$ like in equation 2.1. R_t contains the conditional correlation of ϵ at time

t. R_t is symmetric semi definite positive and follows the DCC specification :

$$\begin{cases} R_t = Q_t^{*-1} Q_t Q_t^{*-1} \\ Q_t = (1-a-b)\bar{Q} + a * \epsilon_{t-1}\epsilon_{t-1}^T + b * Q_{t-1} \\ Q_t^* = \text{diag}(\sqrt{q_{11,t}}, ..., \sqrt{q_{nn,t}}) \\ \bar{Q} = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t \epsilon_t^T \end{cases} \tag{2.7}$$

To ensure that the variances are positive, we also must have $a + b < 1$, $a \leq 0$ and $b \leq 0$.
The DCC student has been used only to estimate a dynamic correlation matrix. However, the problem of unifying several univariate Student into a multivariate framework still remains. Thus, for simulation purposes, we decided to use a t-Student copula parametrization along side with univariate estimations of variances. The reason is that the classical multivariate Student distribution given by the density 2.6 has only one parameter ν for the degree of freedom, which makes all its marginals having the same degree of freedom. When we estimated our DCC model with this density for our quasi maximum likelihood method, we found that the estimated $\hat{\nu}$ was closed to 4, indicating a priori that the distributions will have very fat tail.

2.3.2 The t-Student copula

We used the parametrization proposed in [5]. This parametrization is practical as it gives enough flexibility to simulate multivariate Student processes with marginals having different degree of freedom. It also allows skew parameters. Besides, it gives a stochastic representation of such processes, so that it gives clear instructions for simulations. More precisely, if we define the random vector $X = (\sqrt{W_1}Z_1, ..., \sqrt{W_N}Z_N)$ with $(Z_i)_{i=1...N} \sim \mathcal{N}(0_N, \Sigma)$ and $\forall i \in \{1, ..., N\}$, $W_i = f^{-1}_{IG(\nu_i/2, \nu_i/2)}(U)$, where $U \sim \mathcal{U}(0,1)$ is the same for all i, and where IG stands for the inverse gamma law whose probability distribution function is $f_{IG(\alpha,\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\frac{\beta}{x})$. Then, it was proved in [5] that each marginal X_i has a Student distribution with ν_i degree of freedom, and that the renormalized vector $u = (t_1(X_1), ..., t_N(X_d))$ is distributed according to the copula introduced in [5] paragraph 2.4. The calibration of Σ has been made with the DCC Student specification, and each degree of freedom has been estimated by fitting univariate GARCH(1,1) with Student likelihood.

2.4 Fractionally Integrated GARCH Model

As we said, stationary GARCH models has no long term memory. Stationary GARCH model is defined by equation 2.1 with $\sum_{s=1}^{q} \alpha_s + \sum_{s=1}^{p} \beta_s < 1$. As we saw, these stationary GARCH models have exponentially decaying autocorrelation function, and thus do not have long term memory. When the latter quantity is equal to one, we say that the process is integrated of order one. Such process is not stationary and has persistent memory, meaning its autocorrelation function is equal to one. Between these two extrema, [4] introduced Fractionnaly Integrated GARCH Mode (FIGARCH) in order to find a GARCH like volatility that have hyperbolic decaying autocorrelation. Note that we did not manage to produce acceptable calibrations and simulations with this specification, as multivariate generalization was very tricky. We just recall in this paragraph the definition of univariate FIGARCH process for information purpose only. It can be an interesting model to calibrate for the continuation of this study.

To make the presentation more intuitive, we start by the integrated GARCH model. By rearranging the terms in the GARCH specification 2.1, we have $(1-\beta(L)-\alpha(L))\epsilon_t^2 = \omega + (1-\beta(L))(\epsilon_t^2 - \sigma_t^2)$ where we introduced the polynomials $\beta(x) = \sum_{s=1}^{p} \beta_s x^s$ and $\alpha(x) = \sum_{s=1}^{q} \alpha_s x^s$. L^k is the lag operator of order $k \in \mathbb{N}$ on the real sequences space (with the convention $L^0 = Id$). If the polynomial $1 - \alpha - \beta$ has a unit root, then we can rewrite 2.1 as $(1-L)\Phi(L)\epsilon_t^2 = \omega + (1-\beta(L))(\epsilon_t^2 - \sigma_t^2)$ with Φ being a polynomial of order $p + q - 1$. This is the volatility equation of an integrated GARCH. The idea of FIGARCH model is to replace the integrated term $1 - L$ by a fractionally one, meaning $(1 - L)^d$ with $d \in [0, 1]$. The fractional integration operator has a power serie development given by $(1-L)^d = \sum_n^{+\infty} \frac{\Gamma(n-d)}{\Gamma(-d)\Gamma(n+1)} L^n$. For estimation purpose, we can consider a truncated version of the infinite sum. In order to produce almost surely positive conditional variances, we can impose positiveness for all the coefficients of the polynomial in front of ϵ_t^2. We can first consider the most basic FIGARCH model with $p = q = 1$, i.e. $\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1 L - (1 - \phi_1 L)(1 - L)^d)\epsilon_t^2$ with $\phi_1 = \alpha_1 + \beta_1$. To ensure positiveness of σ_t at any time, we must enforce $0 < \omega$, $0 < d < 1 - 2 * \phi_1$ and $0 < \beta_1 < \phi_1 + d$. And in order to have stationary returns, we must have $d < \frac{1}{2}$.

Chapitre 3

Calibration Procedure : maximization of a likelihood function

The calibration of GARCH models in the literature is exclusively based on the Maximum Likelihood Estimator (MLE). To maximize the log likelihood of the GARCH model in its parameters, we used the Python standard library Scipy, through its optimize function whose code source is publicly available on github. At the moment we wrote this note, we used the current version Scipy 0.19.0. This function proposes several solvers. The models were calibrated on 13 equity indexes (called 10,102, 110, 130, 20, 40, 50E, 60, 80, 90, 901, N225, SMI), with 15 years of daily observations (4000 days).

3.1 Gradient descent method

When analyzing the MLE problem for GARCH models, we can see that the likelihood is a very smooth function $\mathcal{C}^{+\infty}$ of its parameters inside its polyedric domain of definition (but is not defined at its frontiers). Gradient descent methods are too sensitive to the starting points because of very ill conditioned Hessian matrices. This is usually inevitable when we reach high dimensions. In practice, we tried some well known solver of Scipy that accept constrains like BFGS method but it seems not to work very well on our data, always giving degenerate solutions.

3.2 Nelder-Mead method

To avoid costly gradient calculations, we used the Nelder-Mead method. We recall here the principles of this algorithm. It is a method used for solving non

linear objective function which is our case, and based on a simplex search.

Simplex is a generalization of the triangle concept in any dimension. More precisely, if we consider $\mathbb{R}^M$ where $M = p+q+1$ is the number of the likelihood parameters, a simplex $\bar{S}$ is defined as the convex hull of $M+1$ points of $\mathbb{R}^M$, i.e.

$$\bar{S} = \{x \in \mathbb{R}^M | x = \sum_{i=1}^{M} \lambda_i x_i, \sum_{i=1}^{M} \lambda_i = 1, \lambda_i \geq 0\} \tag{3.1}$$

We present in detail the version of the method implemented in the current version of Python Scipy. The idea behind a simplex optimizing algorithm is to create a simplex in the space of the parameters where we can evaluate the objective function on its summits. This simplex will consequently be transformed so that it will move progressively toward a local minimum of the objective function. We denote by f_i the objective function we want to minimize in the univariate case for a given asset $i \in \{1, ..., N\}$:

$$f_i : \begin{cases} \mathbb{R}^{p+q+1} \to \mathbb{R} \\ f(\gamma_{ii}, \alpha_{ii,s}, \beta_{ii,s'}, s \in \{1, ..., p\}, s' \in \{1, ..., q\}) = -\log(\mathcal{L}(\alpha_{ii}, \beta_{ii}, \gamma_{ii} | \mathcal{F}_{t-1})) \end{cases} \tag{3.2}$$

where $\mathcal{L}$ is the univariate GARCH(p,q) likelihood. The algorithm is as follow :

1. Choose $M+1$ sets of parameters $(x_1, ..., x_{M+1})$ in $\mathbb{R}^M$ making a non degenerate simplex.
2. Evaluate f_i on these sets of parameters and sort x by re indexation such that $f_i(x_1) \leq ... \leq f_i(x_{M+1})$.
3. Calculate x_G the gravity center of $(x_1, ..., x_M)$ (all the current points except x_{M+1}).
4. Calculate the reflexion of x_{M+1} with respect to x_G, i.e. $x_r := x_G + (x_G - x_{M+1})$
5. If $f_i(x_r) < f_i(x_G)$ (i.e the reflexion of the worst individual x_{M+1} does better than the average summits, leading to an expansion of the simplex toward x_{M+1}) :
   — Calculate $x_e = x_G + 2 * (x_G - x_{M+1})$. x_e is a possible extension point of the current simplex toward the worst set x_{M+1}.
   — We replace the worst set :
   If $f_i(x_e) < f_i(x_r) : x_{M+1} := x_e$
   Else : $x_{M+1} := x_r$
   — Return to step 2
6. Else : (i.e. the reflexion is not better than the average summit, leading to a contraction of the simplex on its gravity center)

— If $f_i(x_r) \leq f_i(x_M) : x_{M+1} := x_r$
Else :
— If $f_i(x_r) \leq f_i(x_{M+1})$: Calculate $x_c = x_{M+1} + \frac{1}{2}(x_G - x_{M+1})$. x_c is a possible contraction point of the current simplex toward the best set x_G.
Else : Calculate $x_c = \frac{1}{2}(x_G + x_{M+1})$
— If $f_i(x_c) \leq f_i(x_{M+1}) : x_{M+1} := x_c$
Else : $\forall i \geq 2$, $x_i := x_1 + \frac{1}{2}(x_i - x_1)$. We perform here an homothetie of the simplex with coefficient 0.5 toward the best set x_1. Thus, all the simplex is contracted inside towards x_1.
— return to step 2

7. Continue until $\max_{k\in\{2,..,M\}} ||x_1 - x_k||_\infty > \epsilon_1$ and $\max_{k\in\{2,..,M\}} |f(x_1) - f(x_k)| > \epsilon_2$ where ϵ_1 and ϵ_2 are given tolerances. The returned result is obviously the x_1 summit of the last simplex.

In Scipy, the constants in the transformation may depend on the dimension of the space parameters. Also, by default, at the initialization, one just gives x_1 as an argument, the other summits of the initial simplex being obtained by small perturbation on x_1 along different dimension. The Nelder-Mead method is known to be a simple algorithm that can optimize non smooth functions. The main drawback of the method is that it does not have convergence theory. Therefore, it can have difficulties to find the minimum even after a lot of iterations, particularly in high dimension and at the frontiers of the objective function space definition. Concerning the constrains, the Scipy function does not allow to specify them directly. Therefore, the function f_i put in argument is set at $+\infty$ for parameters sets that do not satisfy these constraints.

Another important drawback is that it is essentially a local minimum finder. It means that it can be very sensitive to the first given x_1 argument if f_i present several local minima. First, in order to stabilize the minimum proposed by the Nelder-Mead solver, we heuristically restart the algorithm with an argument x_1 being the previous result of the minimization. Second, we decide to couple the Nelder-Mead method with a more robust algorithm for searching the global minimum.

Thus, to calibrate a GARCH model, we first use a heuristic that may find a proper area where the global minimum of the objective function might be, and then we use this first solution as a starting point for the local Nelder-Mead solver.

3.3 Global optimizer : Differential evolution

Like the Nelder-Mead method, we present here some basics of methods for minimizing complex functions that are heuristic. There is no theoretical guarantee of convergence. The class of method used is called Differential Evolution which was introduced by [17]. Given an initial population of possible parameter sets, the algorithm seeks to improve them with respect to the objective function to minimize, by using transformations inspired by biological evolution processes. We detail the algorithm in this paragraph. We denote by $x_G = (x_{1,G}, ..., x_{K,G}) \in (\mathbb{R}^{p+q+1})^K$ the population of K candidate sets of parameters at the G^{th} iteration of the algorithm called generation. We describe here the current implementation of Scipy.

1. We initialize a population of K sets of parameter $(x_{1,0}, ..., x_{K,0})$ by a latin hypercube sampling over the constrained domain. It consists in drawing uniformly over the unit-hypercube the K initial sets of parameters, the drawing taking into account the previous drawing so that we obtain systematically an homogeneous population over the hypercube.
2. The mutation of the G^{th} generation. For a given set $x_{k,G}$ in the population of generation G, we create a corresponding mutant calculated according to a given function. For instance, if we denote by $x_{best,G} = \arg\min_{l\in\{1,..,K\}} f_i(x_{l,G})$ the best set of the generation G, we can define a mutant $v_{k,G+1} := x_{best,G} + M * (x_{r_1,G} - x_{r_2,G})$ where $M \in [0, 2[$ is the mutation constant, and $x_{r_1,G}, x_{r_2,G}$ two randoms sets of generation G out of the best. It means that the mutant comes systematically from the best individual of the generation. This strategy is only an example. We used this one but many variants can be figured out.
3. Recombination. We create a new candidate set for the next generation $G+1$ as $u_{k,G+1} = (u_{j,k,G+1})_{j=1...p+q+1}$ where

$$u_{j,k,G+1} := \begin{cases} v_{j,k,G+1} & \text{if } R_{j,k} \geq C \text{ or } j = p+q+1 \\ x_{j,k,G} & \text{otherwise} \end{cases}$$

   $C \in [0, 1]$ is the cross-over probability parameter, $R_{j,k} \sim \mathcal{U}[0, 1]$. Note that the last parameter is forced to come from the mutant to insure strict distinction between u and x. This type of recombination is called binomial. Scipy implemented also exponential crossover where $u_{k,G+1}$ exclusively inherits its parameters from the mutant $v_{k,G+1}$ while $R_{j,k} < C$, and then exclusively inherits from $x_{k,G}$.
4. Selection. We set

$$x_{k,G+1} := \begin{cases} u_{k,G+1} & \text{if } f_i(u_{k,G+1}) \geq f_i(x_{k,G}) \\ x_{k,G} & \text{otherwise} \end{cases}$$

The mutant replace its predecessor only if it is better than it.

5. We loopback to step 2 until the convergence criterion is met i.e. $\hat{\sigma}_{G+1}/\hat{m}_{G+1} < \epsilon$ where $\hat{\sigma}_{G+1}$ and $\hat{m}_{G+1}$ are respectively the empirical standard deviation and mean of $(f_i(x_{k,G+1}))_k$, and ϵ a tolerance level. This criterion measures somehow the homogeneity of the G^{th} generated population around its mean.

Differential evolution has been applied to many field of science and seems to give good results, but we do not really know why yet. Finally, one noticeable drawback is its slowness.

3.4 Gaussian calibration

We extracted data of 13 equity indexes. We want to simulate the joint distribution at an horizon T, given these 13 time series containing the historic over the past 15 years. We first consider the simplest GARCH(1,1) model. Using the above methods to calibrate the parameters, we calculated a GARCH trajectory for the matrix time series $\hat{\sigma}_t$. Using these predicted variance matrices, we simulated a multivariate trajectory of these 13 equity indexes by sampling $\epsilon_t \sim \mathcal{N}(0, \hat{\sigma}_t)$, with a daily step. Unfortunately, the predicted distribution does not present enough fat tails (see the simulations moments in table 3.1). The simulation produce in average too much standard deviation too. In the next paragraphs, we present attempts to correct this issue.

Equity	Excess Kurtosis		Standard Deviation (in %)	
Index	Simulated	Historic	Simulated	Historic
10	3.5 (3.8)	7.7	2.7 (2.9)	1.3
102	4.1 (4.4)	15	2.6 (2.7)	1.1
110	3.2 (3.6)	12	3.1 (2.9)	1.4
130	3.6 (3.9)	8.3	2.9 (2.8)	1.2
20	4.4 (6.1)	7.9	2.6 (2.4)	1.3
40	4.8 (5.3)	16	3.0 (3.3)	1.1
50E	3.7 (3.5)	8.0	2.5 (1.9)	1.3
60	5.1 (11)	8.4	2.5 (3.0)	1.1
80	4.0 (4.4)	10	3.2 (3.0)	1.4
90	4.3 (5.2)	11	2.9 (3.0)	1.2
901	4.5 (4.3)	6.1	3.3 (3.1)	0.98
N225	4.3 (4.6)	10	2.9 (3.1)	1.4
SMI	5.5 (5.9)	11	2.9 (3.0)	1.0

TABLE 3.1 – Moments on 100 simulated path of the Gaussian GARCH(1,1) with Flexible multivariate GARCH specification for correlation. Simulation length = 1000

Concerning the confidence interval, it was proven that the MLE behaves asymptotically as a $\mathcal{N}(\theta_0, I^{-1}(\theta_0))$ where θ_0 is the true parameter that governs the GARCH model. $I(\theta) = \mathbb{E}[-\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta^T}]$ is the Fisher information. The Fisher information is easily computed for a Gaussian GARCH(1,1) model and n observations : $I(\theta) = \frac{1}{2}\sum_{t=2}^{n}\mathbb{E}[\frac{1}{\sigma_t^4}\frac{\partial\sigma^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta^T}]$, with $\frac{\partial\sigma_t^2}{\partial\theta^T} = (1+\alpha(1+\alpha(...(1+\alpha\sigma_0)...)), \epsilon_{t-1}^2, \sigma_{t-1}^2 + \alpha(\sigma_{t-2}^2 + \alpha(... + (\alpha\sigma_0^2)...)))$. We can give an estimation $\hat{I}(\hat{\theta})$ calculated at the MLE value of θ_0 by replacing the $\mathbb{E}[\sigma_t^4]$ with the mean of the σ_t^4 calculated along the historical path. As the parameter is $\theta \in \mathbb{R}^3$, we need to generalize the confidence interval for a multi parameters case. We used the "simultaneous confidence interval" proposed by Hardle and Simar (2012) where $\forall a \in \mathbb{R}^3$, we have the asymptotical confidence interval $a^T\hat{\theta} \pm \sqrt{h(\alpha)a^T\hat{I}(\hat{\theta})^{-1}a}$ with $h(\alpha) = \frac{n-1}{n}\frac{3}{n-3}F_{1-\alpha,3,n-3}$. $F_{1-\alpha,3,n-3}$ is the $1-\alpha$ percentile of a Fisher distribution. We recall that an 95% confidence interval is not an interval in which the true parameter belongs to with 95% probability, because this true parameter (even though unknown) is not random. One must say that if the true parameter does not belong to the confidence interval, then the MLE estimator would have had only 5% a priori to have the value it actually has (a priori in the sense assuming the specified GARCH model is the good one). In other terms, among all the data sample of length N possible,

95% of the associated MLE confidence interval would have contained the true parameter.
We put in the table 3.2 the result of this MLE estimation of the GARCH(1,1). The estimation is quite difficult, as it gives relatively large confidence interval. We also note that the estimated GARCH are at the verge of being non stationary.

Equity	MLE Estimators		
Index	γ (10^{-6})	α (%)	β (%)
10	2.1 ($\pm$ 11)	8.2 ($\pm$ 5.0)	91 ($\pm$ 13)
102	1.3 ($\pm$ 5.8)	8.3 ($\pm$ 4.9)	90 ($\pm$ 12)
110	1.2 ($\pm$ 12)	5.2 ($\pm$ 4.4)	94 ($\pm$ 13)
130	1.1 ($\pm$ 10)	5.7 ($\pm$ 4.4)	94 ($\pm$ 13)
20	2.2 ($\pm$ 12)	7.2 ($\pm$ 4.8)	91 ($\pm$ 14)
40	1.3 ($\pm$ 7.2)	8.3 ($\pm$ 5.0)	91 ($\pm$ 13)
50E	2.2 ($\pm$ 11)	7.9 ($\pm$ 4.9)	91 ($\pm$ 13)
60	2.2 ($\pm$ 9.0)	8.5 ($\pm$ 5.0)	90 ($\pm$ 13)
80	2.4 ($\pm$ 13)	8.0 ($\pm$ 5.0)	91 ($\pm$ 13)
90	1.4 ($\pm$ 8.5)	7.8 ($\pm$ 4.9)	91 ($\pm$ 13)
901	0.88 ($\pm$ 6.1)	7.8($\pm$ 5.0)	91 ($\pm$ 12)
N225	3.7 ($\pm$ 16)	8.5 ($\pm$ 5.0)	90 ($\pm$ 14)
SMI	1.9 ($\pm$ 6.6)	10 ($\pm$ 5.4)	88 ($\pm$ 13)

TABLE 3.2 – Estimation of the GARCH(1,1). Confidence interval length in parenthesis.

3.4.1 Fitting GARCH(2,2)

Theoretically speaking, as σ_t is varying over time, we should sample an ϵ_t with a different Gaussian law, so that the resulting distribution over the whole projection path should have a fatter tail than the Gaussian law. It seems that it is not the case, maybe because σ_t does not change a lot from one day to another. An idea to address this problem was to increase the orders of the GARCH model. We considered GARCH(2,2). We note (see table 3.3) that it gives in average higher kurtosis than the (1,1). But unfortunately, GARCH (2,2) did not give fat tail neither. And high orders GARCH become rapidly intractable because of the dimensions.

Equity Index	Excess Kurtosis		Standard Deviation (in %)	
	Simulated	Historic	Simulated	Historic
10	4.3 (3.8)	7.7	3.0 (3.9)	1.3
102	5.7 (6.0)	15	2.6 (3.2)	1.1
110	3.7 (3.0)	12	2.8 (1.6)	1.4
130	4.5 (4.3)	8.3	2.7 (1.8)	1.2
20	4.7 (3.7)	7.9	2.3 (1.5)	1.3
40	5.5 (4.8)	16	3.1 (2.5)	1.1
50E	6.2 (7.2)	8.0	2.4 (1.5)	1.3
60	5.6 (7.8)	8.4	2.1 (1.2)	1.1
80	5.3 (5.2)	10	3.3 (3.5)	1.4
90	6.3 (4.8)	11	2.8 (2.2)	1.2
901	5.2 (4.0)	6.1	3.4 (4.8)	0.98
N225	5.6 (6.6)	10	2.4 (1.4)	1.4
SMI	7.7 (7.6)	11	2.4 (2.1)	1.0

TABLE 3.3 – Moments on 100 simulated path of the Gaussian GARCH(2,2) with Flexible multivariate GARCH specification for correlation. Simulation length = 1000

3.4.2 Trying normal mixture law

One difficulty in multivariate GARCH model estimation is to ensure that the process of variance/covariance matrices Σ you obtain is positive semi-definite. That is why, [9] proposed the following specification for a multivariate GARCH(p,q) :

$$\forall t, \Sigma^2 = \gamma\gamma^T + \sum_{k=1}^{K}\sum_{i=1}^{q} \alpha_{ik}\epsilon_{t-i}\epsilon_{t-i}^T\alpha_{ik} + \sum_{k=1}^{K}\sum_{i=1}^{p} \beta_{ik}\Sigma_{t-i}\beta_{ik}^T \tag{3.3}$$

where γ is a triangular $N \times N$ matrix, α and β $N \times N$ matrices. It was proven by Engle and Kroner that if γ has a positive diagonal, if $\alpha_{11}, \beta_{11} \geq 0$ and if the first σ_i^2 for $i \leq \max(p, q)$ are positive semi-definite, then σ^2 will be positive semi-definite for any ϵ. One has to estimate each coefficients of each matrix of the BEKK specification, meaning hundreds of parameters for a dozen of assets. In practice, we simplify the problem by considering $K = 1$, α and β diagonals. Even so, we will have to estimate a few hundreds of parameters for a dozen of assets.
The reason for introducing this BEKK specification and for looking directly to this high dimension problem is that we will have a well defined normal mixture GARCH likelihood function for multivariate processes, studied in [1]. We hope to reproduce fat tailed with normal mixtures law as simple normal assumption seemed to fail.

We considered here the multivariate centered normal mixture likelihood whose density writes :

$$\forall x \in \mathbb{R}^{\mathbb{N}}, \mathcal{L}(x, \lambda, \Sigma, \alpha, \beta) = \prod_{t=1}^{T}(\sum_{j=1}^{J} \lambda_j * \frac{1}{\sqrt{2\pi \det \Sigma_{k,t}}} \exp -\frac{1}{2}\epsilon_t^T \Sigma_{k,t}^{-1} \epsilon_t) \tag{3.4}$$

where each Σ_k is defined in 3.3. The number J of mixtures should be fixed a priori (usually 2 or 3). The maximization of this likelihood can be derived analytically and is sufficiently smooth to use gradient descent algorithm. These explicit calculations have been made in [1] in the bivariate GARCH(1,1) cases. However, it is not a tractable way to estimate a multivariate GARCH because of the tremendous amount of time it requires when we have at least a dozen of assets, and because of the instability of the gradient descent as the Hessian matrix seems to be ill conditioned on our data.

3.5 Calibrating the DCC multivariate GARCH Student

The log likelihood of the DCC Student model can be written as

$$\ln(\mathcal{L}) = \sum_{t=1}^{T}(\ln[\Gamma(\frac{\nu+n}{2})] - \ln[\Gamma(\frac{\nu}{2})] - \frac{n}{2}\ln(\nu) - \frac{1}{2}\ln[\det(D_t R_t D_t)] - \frac{\nu+n}{2}\ln[1 + \frac{\epsilon_t^T D_t^{-1} R_t^{-1} D_t^{-1} \epsilon_t}{\nu-2}]) \tag{3.5}$$

It has to be optimized in a, b, ν, α, β and γ. It is a tricky problem that can be tackled in a quasi MLE style like for the previous Flexible MGARCH. Meaning, the problem can be divided in two steps. The first one consists in estimating the conditional variances D_t assuming there is no correlation between assets, i.e. replacing R_t by I_N. Then, injecting the estimated D_t into the likelihood function, so that we can consider the D_t terms as constants and thus simplify the objective function of the MLE, and finally estimate R_t. To do this two steps calibration, we followed the procedures presented in [15]. Particularly, for the estimation of the univariate conditional variances, it is argued that we can reasonably use a Gaussian calibration (like the one we did previously), as the student specification has not important impact on the GARCH parameters compared to the Gaussian one. However, as we will use a Student copula for simulations, we need to calibrate individual degree of freedom on each asset. We decide consequently to estimate the univariate GARCH variances parameters with student likelihood maximization. For the confidence interval of the MLE estimate, we use the Fisher Information introduced in the previous paragraph. However, it is more difficult to estimate in

this Student law context. To estimate the Fisher Information, we use the procedure detail in [16] and that we sum up here :

1. We first simulate $\Delta \in \mathbb{R}^p$ (where p is the number of parameters), whose components are in a set $\{c, -c\}$ with c small, following a Bernoulli sampling.

2. We already have the MLE estimator $\hat{\theta} \in \mathbb{R}^p$. We define the gradient vector of the log likelihood student function $G(\theta|(\epsilon_t), (\sigma_t)) = (\frac{\partial l}{\partial \gamma}, \frac{\partial l}{\partial \alpha}, \frac{\partial l}{\partial \beta}, \frac{\partial l}{\partial \nu})$ where the partial derivative can be directly computed from the log likelihood expression above. ϵ is a realization of the GARCH student, and $sigma$ is its associated GARCH volatility.

3. We compute an approximation of the Hessian matrix of the log likelihood by $\hat{H}(\Delta, \epsilon, \sigma, \theta) = \frac{1}{2}(\frac{G(\theta+\Delta|\epsilon)-G(\theta-\Delta|\epsilon)}{2} * \Delta^{-1} + (\frac{G(\theta+\Delta|\epsilon)-G(\theta-\Delta|\epsilon)}{2} * \Delta^{-1})^T)$ where $\Delta^{-1} = (1/\Delta_i)$.

4. We compute the Fisher information matrix $\hat{F}(\hat{\theta})$ as an estimation of the expectation of $\hat{H}(\Delta, \epsilon, \sigma, \hat{\theta})$, i.e. $\hat{F}(\hat{\theta}) = -\sum_{k=1}^{K}\sum_{m=1}^{M} \hat{H}(\Delta^k, \epsilon^m, \sigma^m, \hat{\theta})$ where (Δ^k, ϵ^m) are an iid sampling. To speed up computations, we decided to use only one trajectory of ϵ (M=1), this trajectory being the historical one. The confidence interval obtained is thus wider a priori.

All the MLE estimators of the univariate Student GARCH(1,1) are listed in the table 3.4. Concerning the estimation of the Multivariate DCC specification parameters (meaning, the parameters a and b introduced in the equations 2.7), we calculated the following MLE for our 13 indexes : $\hat{a} = 0.988$ and $\hat{b} = 0.005$. At the time we wrote this note, we do not have estimate the confidence intervals or these parameters. To do it, we can use the previous method which require to compute the gradient G. This computation for the DCC parameters is basic algebra. But the formulas are quite long... The MLE estimators are summarized in 3.4. With our confidence interval estimation methods, as in the normal case, we find that the estimation is not obvious as the confidence interval are wide.

Equity Index	$\gamma(10^{-6})$	α (%)	β (%)	ν
10	1 (± 62)	4.4 (± 4.4)	91 (± 6.0)	4.40 (± 0.05)
102	1 (± 63)	5.0 (± 4.3)	89 (± 4.4)	4.00 (± 0.02)
110	1 (± 83)	2.8 (± 2.9)	94 (± 4.1)	4.00 (± 0.02)
130	1 (± 109)	2.9 (± 11)	94 (± 15)	4.00 (± 0.46)
20	1 (± 65)	3.8 (± 4.2)	92 (± 5.7)	4.00 (± 0.04)
40	1 (± 92)	5.6 (± 9.8)	87 (± 6.4)	4.00 (± 0.30)
50E	1 (± 65)	4.1 (± 6.9)	91 (± 6.3)	4.16 (± 0.14)
60	1 (± 68)	5.3 (± 3.2)	89 (± 5.2)	4.28 (± 0.13)
80	66 (± 133)	12 (± 5.0)	82 (± 4.0)	4.00 (± 0.41)
90	1 (± 103)	4.1 (± 9.2)	92 (± 2.8)	4.34 (± 0.12)
901	1 (± 77)	4.5 (± 4.4)	92 (± 4.2)	5.65 (± 0.02)
N225	2 (± 65)	3.8 (± 7.1)	91 (± 6.8)	4.00 (± 0.13)
SMI	1 (± 68)	7.0 (± 17)	87 (± 19)	4.04 (± 0.31)

TABLE 3.4 – MLE Estimation of univariate GARCH(1,1) student. 95% Confidence interval in parenthesis.

The DCC specification gives simulated paths of returns that exhibit fat tailed distribution with volatility clustering and dynamic correlations. We put the moments over 100 simulations in table 3.5. The DCC GARCH seems to reproduce well the historical kurtosis. However, theoretically speaking, it does not account for the long term memory feature. In fact, it can be proven that GARCH models volatility has exponentially decaying autocorrelation. On the graphics 3.1 and 3.2, we plotted this hyperbolic decay property empirically observed on the historic ACF of absolute return of equity index "10". The dot at the up left corner of 3.1 is not an outlier because as the ACF must be equal to one at zero lag : it is thus on the contrary a dot that should have a bigger weight than the others in the regression. The R^2 statistic is very sensitive to outliers and cannot be relevant to assess the quality of the regression. Therefore, historical ACF decay is best described by a power law rather than an exponential. However, on many equity indexes, the line is not so straight ; on the log log scatter plot 3.1 for instance, we can already notice the beginning of a curvature, which would mean an exponential decay of the ACF, and thus no long term memory. To sum up, long term memory property defined by a hyperbolic decay of the ACF function is fuzzy on our geometrical historical daily returns, meaning that the DCC GARCH with non long memory volatility can well describe them at first approximation. We will see in the next paragraph that this long term memory property is more relevant for log returns in a continuous framework (at least much more than the geometrical returns in discrete time we have been modeling so far).

Equity Index	Excess Kurtosis		Standard Deviation (in %)	
	Simulated	Historic	Simulated	Historic
10	7.6 (7.3)	7.7	1.2 (0.3)	1.3
102	8.2 (4.8)	15	1.1 (0.1)	1.1
110	6.6 (4.6)	12	1.3 (0.3)	1.4
130	9.6 (10)	8.3	0.98 (0.3)	1.2
20	9.0 (7.0)	7.9	1.1 (0.4)	1.3
40	9.2 (10)	16	1.3 (0.2)	1.1
50E	8.3 (6.9)	8.0	1.2 (0.3)	1.3
60	9.5 (10)	8.4	1.1 (0.3)	1.1
80	9.4 (15)	10	0.8 (0.3)	1.4
90	7.9 (6.5)	11	1.1 (0.3)	1.2
901	4.4 (5.4)	6.1	1.3 (0.2)	0.98
N225	8.1 (6.7)	10	1.3 (0.3)	1.4
SMI	13 (14)	11	1.4 (0.2)	1.0

TABLE 3.5 – Moments on 100 simulated path of the Student GARCH(1,1) with DCC specification for correlation. Simulation length = 1000

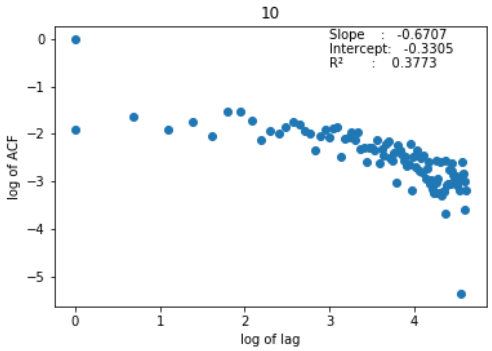

FIGURE 3.1 – Log log regression of historical absolute returns ACF of Index "10" against its time lags

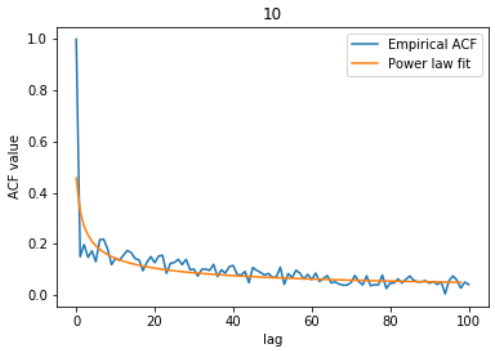

FIGURE 3.2 – Power law fit of historical absolute returns ACF of Index "10"

Chapitre 4

Multi-fractal models

4.1 Time Scale Invariance

All the results and the precise construction of multi-fractal processes are exposed in [12]. Multi-fractal models have been introduced by ([14],[3]) and applied in FX returns modeling in a continuous time framework. This class of models have been proposed in order to reproduce the main stylized facts (particularly long memory volatility) and the time scale invariance properties.
Formally, a process X_t is said to have a scale invariance property if its absolute increments moments are governed by a power law scaling i.e.

$$\forall l, q > 0, m(q, l) = \mathbb{E}[|X(t + l) - X(t)|^q] = K_q l^{\zeta_q} \tag{4.1}$$

where K_q and ζ_q only depends on the moment order q. When ζ_q is linear in q, X is said to be mono-fractal. The most important example of such process is the fractional Brownian Motion, whose Hurst coefficient corresponds to the slope of ζ_q. If ζ_q is not linear, X is said to be multi-fractal. Then, it can be shown that equation 4.1 can be only satisfied for $l \in]0, T]$ with $T > 0$ called "integral time", and that necessarily, ζ_q is concave. The figure 4.1 illustrates this property for several moment orders on the log price of the 50E equity index. Linear behavior are straightforward although we can spot some statistical disturbance for big value of l and big moment order.

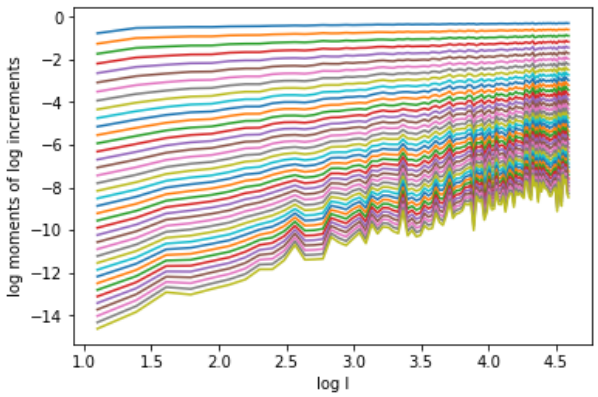

FIGURE 4.1 – Linearity of log moments of log increments of historic 50E to log lag at which the increments are calculated, from $p = 0.1$ (top line) to $p = 4$ (bottom line)

To produce mono-fractal processes, one can simply uses the so called auto similar process, like the Fractionnal Brownian motion. Auto similar processes are processes that have a distribution (varying in time) that has a scaling property. For example, we have for the fractional Brownian motion B^F the property

$$\forall t > 0, a > 0, B^F_{at} \sim a^H B^F_t$$

with $0 < H < 1$. In that case, one can show that $\zeta_q = Hq$
The production of multi-fractal processes are more complicated. To do that, we define a process X with a "stochastic auto similarity" property by

$$\forall t > 0, a > 0, X_{at} \sim W_a X_t$$

where W_a is a random variable independent from X whose law only depends on the dilatation coefficient a. This generalization of auto-similarity is practical has one as only has to know the density of W_a to simulate X_{at} (knowing how to simulate X_t). We can then specify a particular shape for the distribution of W_a in order to have a concave ζ_q. In [3], the authors introduce a continuous and stationary process they called "multi-fractal random walk" that matches this time invariance property, and that they used for log price modeling. We now present succinctly its construction.

4.2 Multifractal Random Walk : univariate Case

4.2.1 Construction of the Multifractal Random Walk

We consider a positive equity stock price P_t. We want to model the quantity $X_t = \log(P_t)$. To do that, we will have to define the log returns at scale τ as

$\delta_\tau X_t = X_t - X_{t-\tau} = \log \frac{P_t}{P_{t-\tau}}$.
Bacry-Muzy proposed to model X_t with a Log Normal Multifractal Random Walk. To define this object, we first introduce a particular Gaussian process $\omega_{l,T}(t)$ with :

$$\forall T, t, \lambda^2 > 0 \begin{cases} \mathbb{E}[\omega_{l,T}(t)] = -\lambda^2(\log(\frac{T}{l}) + 1) \\ Cov(\omega_{l,T}(t), \omega_{l,T}(t+k)) = \begin{cases} \lambda^2(\log(T/l) + 1 - k/l), & \text{if } 0 \leq k < l \\ \lambda^2 \log(T/k), & \text{if } l \leq k < T \\ 0, & \text{otherwise} \end{cases} \end{cases} \tag{4.2}$$

The reason behind this shape of the autocorrelation is to capture the long memory property of asset price log increment, meaning an autocorrelation decreasing at a power law of the lag. This long memory should be limited in time, up to the integral time T. The l parameter corresponds to a discretization step for simulation purpose, but that is meant to converge to zero in theoretical calculation. We then define the following sequence of measures :

$$\forall I \in \mathcal{B}(\mathbb{R}), M_{l,T}(I) = \int_I \exp(2\omega_{l,T}(t))dt \tag{4.3}$$

Bacry-Muzy established the weak convergence of $M_{l,T}$ toward a non degenerate limit measure M_T when l goes to 0. M_T is called Multi-fractal Random Measure (MRM).
A log normal multi-fractal random walk $Y_T(t)$ can then be defined as the subordination of an independent Brownian motion with respect to the MRM, i.e.

$$\forall t \geq 0, Y_T(t) = B_{M_T([0,t])} \tag{4.4}$$

where B is a Brownian motion independent from M and with volatility σ. The idea of subordinating a Brownian motion can be found in many other model in finance (Variance-Gamma, CGMY...). This technic works well in the multifractal process construction as Bacry-Muzy showed that Y_T was indeed a stochastic auto similar process that was multifractal with stationary increments (see theorem 3.15 p 121 in [12]). And they used the multifractal process Y to model $\delta_\tau X$.

4.2.2 Estimation of the MRW parameters

The reason for such specification lies on empirical observations made on the behavior of the magnitude of the auto covariance of the logarithmic returns. The magnitude of X_t is defined as the logarithm of its absolute increments and seems to behave like

$$Cov(\log|\delta_\tau X_t|, \log|\delta_\tau X_{t+h}|) \sim \begin{cases} -\lambda^2 \log(T/h), & \text{if } h \leq T \\ 0, & \text{if } h > T \end{cases} \tag{4.5}$$

We can see that the equation 4.5 is very similar to the equation 4.2. The λ coefficient (called intermittent coefficient) should be the same in the two equations. For illustration, we considered an Equity Index 10 at our disposal, we fixed the sampling step at $l = 2^{-4}$ and we plot the covariance of the magnitudes against the log lag for different value of τ (see figure 4.2)

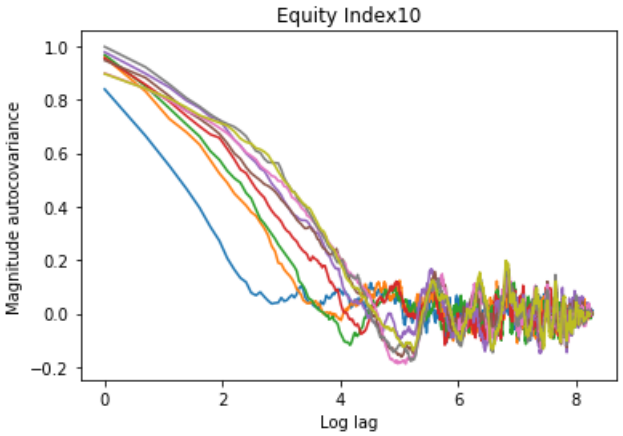

FIGURE 4.2 – Autocorrelation of Magnitude for lag between 1 and 4000. Each line corresponds to a value of $\tau = 20, 40, ..., 200$

We can thus estimate this λ coefficient with the slope of the linear part, and the constant T (called integral time, which corresponds to the maximal time from which the process does not have autocorrelation anymore) with the intercept. For the Multifractal Random Walk, the autocovariance of the magnitudes does not theoretically depend on τ. Thus the estimation we present here for our Equity Index 10 is a mean over the different values of $\tau \in \{1, ...500\}$. We did the estimation using ordinary least squares and calculating the moments over τ an estimation of λ and T for each asset. Here, we must aware of some limit to this approach. First, we checked that for small value of $\tau < 3$, the empirical autocorrelation of the magnitudes does not show a linear part and seems to be instable. Second, the estimators $\hat{\lambda}$ and $\hat{T}$ are not constant in τ at all. And third, variance of $\hat{T}$ is huge, giving large confidence interval, which is problematic as the autocorrelation function is quite sensitive to T. We can also check the multifractality property 4.1 by estimating ζ_q. We make a regression between an estimator of $\log(m(l, q))$ (empirical mean) and $\log(l)$ for several q (we use for each regression 100 lags l to draw figure 4.3). The slope of this regression gives us an estimation of ζ_q. For the log-normal case (i.e. using equations 4.2, 4.3 and 4.4 for X), one can show that

$$\zeta_q = \frac{1}{2} * (q - q(q-2)\lambda^2)$$

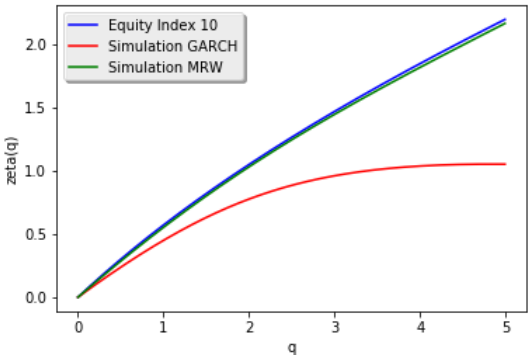

FIGURE 4.3 – Estimation of ζ_q for historical and simulated path for $q \in [0, 5]$

Thus, fitting a quadratic curve on an estimated ζ_q in graph 4.3 give an estimation of λ. However, these estimations of λ are not very precise, but above all, they do not give simulated path that match the historical moments. These estimation technics described by [3] are not really satisfactory and we decide to discard these estimations. Instead, we decide to calibrate freely the parameters in order to have simulated paths that match the historical moments. This kind of calibration is tricky because we are trying to fit a stochastic volatility model on a single historical path. Here we present our method for the estimation of λ and T :

— We first fix the volatility σ of the Brownian B at the empirical standard deviation of the historical path.
— For each values of T and λ on a homogeneous grid, we evaluate the Mean of the Excess of Kurtosis and the Autocovariance function on a set of simulated $MRW(T, \lambda)$ path. To be more precise on the autocorrelation function, we calculate the mean of the euclidean distance between the autocovariance of the simulated path and the autocovariance of the historical path.
— Take the value of λ and T that gives MRW simulated path with a kurtosis and an autocovariance function that is the closest to the historical ones on average, up to a confidence interval.

The drawback of this method is the that it is very slow (greedy search method), all the more that it needs a lot of simulations in order to reduce the confidence interval. We also make the implicit assumption that the average of the moments are continuous in the parameters on our grid search. For example, on figures 4.4 and 4.5, we plot the result of the grid search for the equity index 10. The heat maps were plotted with 80 simulated path. These heat maps are up to a confidence interval. Also, this kind of Generalized Moment Method estimation can be more practical if we have access to theoretical formulas for moments. For the MRW

model, the theoretical moments are given in [12], but we did not manage to find proper values of kurtosis with them. Finally, it would be by the way interesting to apply it to the GARCH method to see if we can have more excess of Kurtosis.
Knowing that the historical excess of kurtosis for this index was 7.7, we found that setting the parameters to $T = 180$ and $\lambda = 0.125$ gives an average excess of kurtosis (on the 80 simulations) of 7.5 (with a confidence interval at 95 % given by the central limit theorem of 0.7). We found that these estimations are not even close to those obtained by the previous regressions technics . However, they give more realistic simulated path in average. Therefore, we will use them for our simulations. Concerning the ACF, we empirically observed that an euclidean distance of 0.3 between the simulations and the historical was fairly enough to see the right long memory behavior.

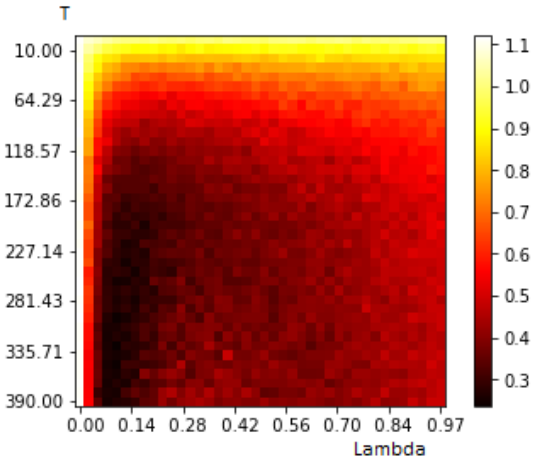

FIGURE 4.4 – Average excess of kurtosis on 80 simulated path, function of T and λ

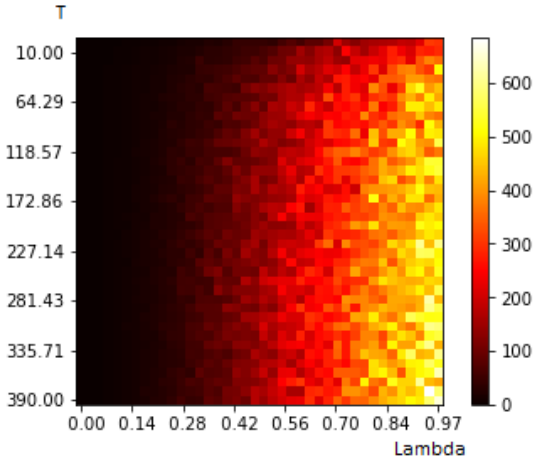

FIGURE 4.5 – Average euclidean distance between the ACF of the historical path and the ACF of the simulated ones, function of T and λ

4.2.3 Simulation of the Multifractal Random Walk

For simulation purpose, we will have to use the approximation given in equation 4.3 of the MRM, i.e. $Y_T(t) = \lim_{l \to 0^+} B_{M_{l,T}([0,t])}$. The idea is to model the log returns $\delta_\tau X_t$ with Y_T. We now give the different steps to simulate an approximation of Y_T. The simulation at any time t is as follow :

— Fix l as small as possible. T, σ and λ^2 are given parameters
— Simulate the autocorrelated Gaussian process $\omega_{l,T}$ at discrete time jl with $j \in \mathbb{N}$.

— Calculate an approximation of the integral 4.3 with
$\tilde{M}_{l,T}([0,t]) = \sum_{j=1}^{t/l} l * \exp(2\omega_{T,l}(jl)))$ at different time t multiple of l

— Simulate $\tilde{Y}_{l,T}(t) = \sqrt{\tilde{M}_{l,T}([t-\tau,t])}\epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0,\sigma^2)$ at different time t multiple of l. We used here the scaling property of the Brownian motion (proposition 3.14 p120 in [12]).

The simulation of the autocorrelated Gaussian is explained in detail in [12] in Annex C. We can simulate $\tilde{Y}_{l,T}$ on a set of discrete time $t_k = k * \tau$ with $\tau \in \mathbb{N}$ by sampling independently the ϵ_{t_k}. But if we want to sample k points on the $\tilde{Y}_{l,T}$ trajectory, we will have to simulate $k\tau/l$ points on the trajectory of $\omega_{l,T}$.

4.3 Multifractal Random Walk : multivariate Case

One idea to generalize this MRM framework to multi assets modeling, is to introduce an equivalent representation of a multi-fractal process. First, let us define the Wiener process $Y_{l,T}(t) = \int_0^t \exp(\omega_{l,T})(u)dB_u$. Bacry and Muzy proved that, under some technical assumption (see [12] Theorem 3.13 p 120), $Y_T(t) = \lim_{l\to 0^+} Y_{l,T}(t)$ where Y_T is the same as introduced in 4.4. Thanks to this representation, we can simulate a discrete version of the univariate multifractale random walk (first introduced in [2]), defining at different time t multiple of l :

$$\tilde{Y}_{l,T} = \sum_{j=1}^{t/l} \epsilon_t^l * e^{\omega_{l,T}(jl)} \tag{4.6}$$

with $\epsilon^l \sim \mathcal{N}(0,\sigma^2 l)$. Then, [2] propose a natural multivariate generalization, with the gaussian vector $\omega_{l,T}(t) = (\omega^i_{l,T}(t))_i$ following

$$\forall i,j>0, Cov(\omega^i_{l,T}(t), \omega^j_{l,T}(t+k)) = \begin{cases} \lambda^2_{i,j}(\log(T_{i,j}/l) + 1 - k/l), & \text{if } 0 \le k < l \\ \lambda^2_{i,j}\log(T_{i,j}/k), & \text{if } l \le k < T_{i,j} \\ 0, & \text{otherwise} \end{cases} \tag{4.7}$$

with $\lambda = (\lambda_{i,j})_{i,j}$ and $T = (T_{i,j})_{i,j}$ being matrices of the multivariate MRW parameters. The generalization of 4.6 is then straightforward :

$$\tilde{Y}_{l,T} = \sum_{j=1}^{t/l} \epsilon_t^l \otimes e^{\omega_{l,T}(jl)} \tag{4.8}$$

with $\epsilon^l \sim \mathcal{N}(0_K, \sigma^2 l)$, $\sigma^2 \in \mathcal{M}_K(\mathbb{R})$, K is the number of assets and $\otimes$ stands for the element wise product between two vectors, and the exponential is applied element wise. The most important difficulty in this multivariate simulation is the simulation of the multivariate Gaussian $\omega_{l,T}$. For the moment, we failed to have a proper generator of such vector process.

Another idea is to keep the the subordinated Brownian representation in equation 4.4 and to introduce some specification of the correlation matrix. We can either try Lévy Copula to bind the subordinators (see Panov,Sirotkin), or just calibrate univariate and independent subordinators and use a correlation matrix to bind the univariate Brownian motions (see [7]). We try this latter solution for a fix correlation matrix (the easier and the more naive way). We used the empirical correlation matrix calculated over all the historic period. Note that to avoid asynchronicity issue between time series, we used an empirical correlation matrix calculated on a weekly basis.

4.4 Results

To describe the simulated path, we put in the table 4.1 the estimated moments. Note that these results are highly sensitive to the λ, σ and T parameters. We used the methods described above to have an estimation of these parameters (estimators are put in table 4.2). The results were calculated on 80 simulations of path of length 4000. Note that the model does not parametrize the volatility and the correlation locally, as the MRW is a kind of stochastic volatility model. Thus, we cannot define the volatility on the historical data in this framework, like we did for the conditionally deterministic GARCH volatility. The MRW simulations manage to fit quite well the historical moments. Concerning the euclidean distance between the ACF of the average simulated path and the historical one, it is difficult to give a proper interpretation. Qualitatively, a distance between 0.3 and 0.4 is enough to reproduce the historical behavior with a good fit. To assess the proximity between the ACF, we also could have used regression to a power law function and compare the slopes/intercepts.

Equity	Excess Kurtosis		Std Dev. (in %)		ACF euclidean
Index	Simulated	Historic	Simulated	Historic	distance
10	8.7 (6.7)	7.7	1.0 (0.09)	1.3	0.32 (0.13)
102	7.3 (3.2)	15	0.86 (0.08)	1.1	0.86 (0.35)
110	13 (8.9)	12	1.3 (0.1)	1.4	0.76 (0.23)
130	7.7 (3.0)	8.3	1.0 (0.08)	1.2	0.69 (0.21)
20	9.5 (7.1)	7.9	1.0 (0.08)	1.3	0.34 (0.13)
40	10 (5.8)	16	0.92 (0.07)	1.1	0.30 (0.12)
50E	9.4 (4.1)	8.0	1.1 (0.07)	1.3	0.41 (0.15)
60	7.1 (2.9)	8.4	0.91 (0.06)	1.1	0.46 (0.15)
80	13 (12)	10	1.3 (0.09)	1.4	0.34 (0.14)
90	8.4 (4.9)	11	1.0 (0.07)	1.2	0.55 (0.20)
901	8.3 (4.1)	6.1	0.81 (0.05)	0.98	0.65 (0.19)
N225	11 (8.1)	10	1.3 (0.1)	1.4	0.27 (0.077)
SMI	15 (19)	11	0.93 (0.08)	1.0	0.25 (0.090)

TABLE 4.1 – Moments on 80 simulated path (length 4000) of the multivariate multifractal random walk

Equity Index	λ	T
10	0.125	300
102	0.188	300
110	0.155	300
130	0.125	300
20	0.125	300
40	0.188	300
50E	0.125	300
60	0.125	300
80	0.155	300
90	0.155	300
901	0.125	300
N225	0.155	300
SMI	0.188	300

TABLE 4.2 – Parameters used for the MRW simulations

Conclusion

In this report, we used parametric models to simulate equity indexes paths that satisfy main stylized fact currently accepted in the literature. On that point, these models are more interesting than historic re sampling. We reviewed the classical GARCH modeling and the multifractal modeling. We consider generalizations to multivariate cases. The main difficulties were issues concerning the calibration to the historical path that can be very costly in calculation time. In this work, we found that the DCC t Student GARCH model reproduced realistic kurtosis and volatility clustering. However, from a theoretical point of view, it should not give long term memory volatility. But this fact was not so clear on our historical data and on our simulations when we plotted the ACF.
Multi Fractal process seems to be promising as it reproduced in average long memory, fat tail, volatility clustering and scale invariance property. Beside, it is a kind of stochastic volatility model that has few parameters.
There are some tasks that remain to be done, like calibrating model that produce simulated skewed distribution yields. Or calculate risk measures on real portfolio in order to validate the simulation methods. In terms of coding, I believe that many improvements have to be done to speed up the calculations. For instance, in calibrating MRW models, one should try a gradient descent method in the calculation of the heat maps. From a theoretical point of view, it would be interesting to continue the studies on fat tailed distribution using an other definition than the kurtosis. Particularly, there is an important literature about extreme value theory which models precisely the distribution tail, using the power law decay of the density distribution as a definition for fat tail.

Bibliographie

[1] Carol Alexander and Emese Lazar. Normal mixture garch (1, 1) : Applications to exchange rate modelling. *Journal of Applied Econometrics*, 21(3) :307–336, 2006.

[2] E Bacry, J Delour, and JF Muzy. A multivariate multifractal model for return fluctuations. *arXiv preprint cond-mat/0009260*, 2000.

[3] Emmanuel Bacry, Jean Delour, and Jean-François Muzy. Multifractal random walk. *Physical Review E*, 64(2) :026103, 2001.

[4] Richard T Baillie, Tim Bollerslev, and Hans Ole Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 74(1) :3–30, 1996.

[5] Christ Church. The asymmetric t-copula with individual degrees of freedom, 2012.

[6] Rama Cont. Empirical properties of asset returns : stylized facts and statistical issues. 2001.

[7] Ernst Eberlein and Dilip B Madan. On correlating lévy processes. 2009.

[8] Robert Engle. Dynamic conditional correlation : A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3) :339–350, 2002.

[9] Robert F Engle and Kenneth F Kroner. Multivariate simultaneous generalized arch. *Econometric theory*, 11(1) :122–150, 1995.

[10] Olivier Gueant. Computing the value at risk of a portfolio : Academic literature and practionners response.

[11] Nicholas J Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear algebra and its applications*, 103 :103–118, 1988.

[12] Alexey Kozhemyak. Modélisation de séries financières à l'aide de processus invariants d'échelle. application à la prédiction du risque. 2006.

[13] Olivier Ledoit, Pedro Santa-Clara, and Michael Wolf. Flexible multivariate garch modeling with an application to international stock markets. *Review of Economics and Statistics*, 85(3) :735–747, 2003.

[14] Benoit Mandelbrot, Adlai Fisher, and Laurent Calvet. A multifractal model of asset returns. 1997.

[15] Elisabeth Orskaug. Multivariate dcc-garch model :-with various error distributions. Master's thesis, Institutt for matematiske fag, 2009.

[16] James C Spall. Monte carlo computation of the fisher information matrix in nonstandard settings. *Journal of Computational and Graphical Statistics*, 14(4) :889–909, 2005.

[17] Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4) :341–359, 1997.