



NUMERICAL OPTIMIZATION METHODS FOR FINANCIAL TIME SERIES GARCH (P,Q) MODEL, A COMPARATIVE APPROACH

| Moulay El Mehdi Falloul ^{1*} | Ayoub Razouk ² | and | Youness Saoudi ² |

¹ Sultan Moulay Slimane University | department of Economics | Beni Mellal | Faculty polidisciplinary | Morocco |

² Mohamed V University | Decision aid and computing | ENSIAS | Morocco |

| Received February 22, 2020 |

| Accepted March 7, 2020 |

| Published March 14, 2020 |

| ID Article | Moulay-Ref.6-ajira250220 |

ABSTRACT

Background: Maximum likelihood estimation (MLE) is often used in econometric and other statistical models despite its computational considerations and because of its strong theoretical appeal. **Objectives:** Non-linear optimization discipline provides feasible alternative methods for calculating MLE's, especially when special structure may be exploited, as for example in probabilistic choice models. **Methods:** may be exploited, as for example in probabilistic choice models. This paper examines estimation of parameters of financial time series model named GARCH(p,q) using four numerical optimization methods and gives numerical comparisons of these methods. **Results:** Among the issues considered in this paper are theoretical background of MLE. Also methods of approximating the Hessian are presented. These include (DFP and BFGS) and statistical approximations (BHHH). **Conclusions:** In our case of GARCH (p, q) NR has approved to be the fastest in convergence according to the number of iterations followed by BHHH algorithm, BFGS and DFP is the last position rank.

Keywords: GARCH(p,q), Log-likelihood, Numerical optimization, BHHH, Newton-Raphson, BFGS, DFP.

1. INTRODUCTION

The maximum likelihood estimator is highly regarded for its excellent asymptotic properties, yet significant effort and ingenuity have nevertheless gone into the development of alternative statistical estimators. This is especially true in econometrics, where many other estimators are available for simultaneous equations. One major reason for this is the difficulty of computing MLE's, a task requiring the solution of a non-linear optimization problem [1]. This paper examines estimation of parameters of financial time series model named GARCH(p,q) using four numerical optimization methods and gives numerical comparisons of these methods. Among the issues considered in this paper are theoretical background of MLE. Also methods of approximating the Hessian are presented. These include (DFP and BFGS) and statistical approximations (BHHH).

2. Presentation of GARCH(p,q)

An observed time series y_t can be written as the sum of a predictable and an unpredictable part,

$$r_t = E(r_t / \Omega_{t-1}) + \varepsilon_t \quad (1)$$

Where Ω_{t-1} is the information set consisting of all relevant information up to and including time $t-1$, In particular, ε_t was assumed to be both unconditionally and conditionally homoscedastic – that is, $E(\varepsilon_t^2) = E(\varepsilon_t^2 / \Omega_{t-1}) = \sigma^2$ for all t. Here we relax part of this assumption and allow the *conditional variance* of t to vary over time – that is, $E(\varepsilon_t^2 / \Omega_{t-1}) = h_t$. Put differently, ε_t is conditionally *heteroscedastic* [2]. A convenient way to express this in general is

$$\varepsilon_t = z_t \sqrt{h_t} \quad (2)$$

Where $z_t \square i.i.d. N(0,1)$

Engle (1982) introduced the class of Autoregressive Conditionally heteroscedastic (ARCH) models to capture the volatility clustering of financial time series (even though the first empirical applications did not deal with high frequency financial data) [3]. In the basic ARCH model, the conditional variance of the shock that occurs at time t is a linear function of the squares of past shocks. For example, in the ARCH model of order 1, h_t is specified as

$$h_t = \omega + \varepsilon_{t-1}^2 + \nu_t \quad (3)$$

Where $v_t = \varepsilon_t^2 - h_t = h_t(z_t^2 - 1)$. Notice that $E(v_t / \Omega_{t-1}) = 0$. To cope with the extended persistence of the empirical autocorrelation function, one may consider generalizations of the ARCH (1) model. One possibility to allow for more persistent autocorrelations is to include additional lagged squared shocks in the conditional variance function. The general ARCH(q) model is given by

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \alpha_3 \varepsilon_{t-3}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 \tag{4}$$

To capture the dynamic patterns in conditional volatility adequately by means of an ARCH(q) model, q often needs to be taken quite large. It turns out that it can be quite cumbersome to estimate the parameters in such a model, because of the no negativity and stationarity conditions that need to be imposed. To reduce the computational problems, Bollerslev (1986) suggested adding lagged conditional variances to the ARCH model instead, this implied the Generalized GARCH model of order (p, q) structured as follows [4] :

$$h_t = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \tag{5}$$

Estimation of the parameters in the GARCH model can be done by the maximization of the likelihood function that would present in the next section.

3. Estimation of the parametrs of GARCH(p, q) model

Estimation of the parameters of the GARCH model can be done by ty maximum likelihood. The expression of likelihood function is as follows:

$$\begin{aligned} L(\alpha, \beta / \varepsilon_1, \dots, \varepsilon_n) &= f(\varepsilon_1, \dots, \varepsilon_n / \alpha, \beta) \\ &= f(\varepsilon_n / \Omega_{n-1}, \alpha, \beta) f(\varepsilon_{n-1} / \Omega_{n-2}, \alpha, \beta) \dots \\ &\dots f(\varepsilon_{m+1} / \Omega_m, \alpha, \beta) f(\varepsilon_1, \dots, \varepsilon_m / \alpha, \beta) \end{aligned} \tag{6}$$

Where n is the sample size, $m = \max(p, q)$, $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_q)^T$, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ and $f(\varepsilon_1, \dots, \varepsilon_m / \alpha, \beta)$ is the joint probability function of $\varepsilon_1, \dots, \varepsilon_m$. Since the exact form of $f(\varepsilon_1, \dots, \varepsilon_n / \alpha)$ is complicated, the conditional likelihood function is used instead, it is given as follows;

$$L(\alpha, \beta / \varepsilon_1, \dots, \varepsilon_n) = f(\varepsilon_{m+1}, \dots, \varepsilon_n / \alpha, \beta, \varepsilon_1, \dots, \varepsilon_m) \tag{7}$$

If we assume $z_t \square N(0, 1)$

When $z_t \square N(0, 1)$, $\varepsilon_t^2 / \Omega_{t-1} \square N(0, h_t)$, since $\varepsilon_t = z_t \sqrt{h_t}$. This could be calculated from the transformation of variables. Thus, the conditional likelihood function is:

$$\begin{aligned} L(\alpha, \beta / \varepsilon_{m+1}, \dots, \varepsilon_n) &= f(\varepsilon_{m+1}, \dots, \varepsilon_n / \alpha, \beta, \varepsilon_1, \dots, \varepsilon_m) \\ &= \prod_{t=1}^T \frac{1}{\sqrt{2\pi h_t}} \exp \left\{ -\frac{\varepsilon_t^2}{2h_t} \right\} \end{aligned}$$

Where $t = 1, \dots, T$ denotes the time points in the conditional likelihood function.

We have to notice that when maximizing the conditional likelihood function is equivalent to maximizing its logarithms because $\ln(\cdot)$ is a strictly increasing function [5].

Since the log-likelihood is easier to handle, it is described as follows:

$$\begin{aligned} \ln(L) &= l(\varepsilon_{m+1}, \dots, \varepsilon_n / \alpha, \beta, \varepsilon_1, \dots, \varepsilon_m) \\ &= -\frac{1}{2} \sum_{t=1}^T \left[\ln(2\pi) + \ln(h_t) + \frac{\varepsilon_t^2}{h_t} \right] \\ &= -\frac{1}{2} \sum_{t=1}^T \left[\ln(h_t) + \frac{\varepsilon_t^2}{h_t} \right] + \text{constan } t \end{aligned}$$

Where $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \dots + \beta_p h_{t-p}$ has to be evaluated recursively. The log-likelihood in eq. (8) can be maximized using numerical optimization methods that we would present in the next sections.

4. Maximum likelihood framework

Suppose that the vector of the dependent variable and the vector of the parameters are presented simultaneously as $y_i = (y_1, y_2, \dots, y_i)^T$ and $\theta_i = (\theta_1, \theta_2, \dots, \theta_i)^T$, the joint probability function density function can be written as a sequence of conditional distributions as follows:

$$f(y_1, y_2; \theta_1, \theta_2) = f(y_1 / y_1; \theta_2) f(y_1; \theta_1)$$

$$f(y_1, y_2, y_3; \theta_1, \theta_2, \theta_3) = f(y_3 / y_2, y_1; \theta_3) f(y_2 / y_1; \theta_2) f(y_1; \theta_1)$$

And hence for the full sample

$$f(y_1, y_2, \dots, y_T; \theta_1, \theta_2, \dots, \theta_T) = f(y_1; \theta_1) \prod_{i=2}^T f(y_i / y_{i-1}, y_{i-2}, \dots, y_1; \theta_i)$$

The important special case which simplify the probability function density is the case of independent and identically distributed probability density function (iid) as follows:

$$L(\theta) = f(y_1, y_2, \dots, y_T; \theta_1, \theta_2, \dots, \theta_T) = \prod_{i=1}^T f(y_i; \theta) \tag{9}$$

For many models in statistics and econometrics, it is often simpler to work with the log-likelihood function, that is written as follows in the case of independent and identically distributed probability density function (iid):

$$L(\theta) = f(y_1, y_2, \dots, y_T; \theta_1, \theta_2, \dots, \theta_T) = \sum_{i=1}^T \ln f(y_i; \theta) \tag{10}$$

The aim of maximum likelihood estimation is to find the value of θ that maximizes the log-likelihood function, a natural way to do this is to use the rule of calculus. This involves computing the first derivatives (gradient) and second derivatives (Hessian) of log-likelihood function with respect to the parameter vector θ [6].

4.1 Gradient: The first derivative of log-likelihood function with respect to the parameter vector θ is written as follows:

$$G(\theta) = \frac{\partial \ln L(\theta)}{\partial \theta} \tag{11}$$

This equality is known as the score or the gradient. In the iid case, where θ is a fixed (K*1) vector of parameters, the gradient is written as follows:

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial \ln L(\theta)}{\partial \theta_1} \\ \frac{\partial \ln L(\theta)}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ln L(\theta)}{\partial \theta_K} \end{bmatrix} \tag{12}$$

The maximum likelihood estimates of θ , namely $\hat{\theta}$ is obtained by solving the following set equalities, in other words, $\hat{\theta}$ satisfies:

$$G(\hat{\theta}) = \frac{\partial \ln L(\theta)}{\partial \theta} \Big|_{\theta = \hat{\theta}} = 0 \tag{13}$$

4.2 Hessian: The second derivative of log-likelihood function with respect to the parameter vector θ is known as Hessian and it is written as follows:

$$H(\theta) = \frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'} \tag{14}$$

This equality can be written as the following symmetric square matrix:

$$\frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'} = \begin{bmatrix} \frac{\partial^2 \ln L(\theta)}{\partial \theta_1 \theta_1} & \frac{\partial^2 \ln L(\theta)}{\partial \theta_1 \theta_2} & \dots & \frac{\partial^2 \ln L(\theta)}{\partial \theta_1 \theta_k} \\ \frac{\partial^2 \ln L(\theta)}{\partial \theta_2 \theta_1} & \frac{\partial^2 \ln L(\theta)}{\partial \theta_2 \theta_2} & \dots & \frac{\partial^2 \ln L(\theta)}{\partial \theta_2 \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L(\theta)}{\partial \theta_k \theta_1} & \frac{\partial^2 \ln L(\theta)}{\partial \theta_k \theta_2} & \dots & \frac{\partial^2 \ln L(\theta)}{\partial \theta_k \theta_k} \end{bmatrix}$$

In the case of the maximum likelihood estimate, the requirement is that the Hessian evaluated at $\hat{\theta}$

$$H(\hat{\theta}) = \frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta = \hat{\theta}} \tag{15}$$

is a negative definite matrix, in other words $x'Hx < 0$ for all non-zero vectors x .

The log-likelihood in (8) and (9) can be maximized using numerical optimization methods that we would present in the next section.

5. Numerical optimization methods of mle

Of the numerous maximization algorithms that have been developed by scholars, we would describe only the most prominent that has been implemented in statistical and econometrics softwares.

5.1 Newton-Raphson: The log-likelihood in for GRACH(p,q) (8) could be written and alternative way as follow:

$$\ln L(\beta) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^T \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^T \left(\frac{\varepsilon_i^2}{\sigma_i^2} \right) \tag{16}$$

By taking first derivatives of function (16), and after rearrangement:

$$G(\theta) = \frac{\partial \ln L(\theta)}{\partial \theta} = \frac{1}{2} \sum_{i=1}^T \left(\frac{\varepsilon_i^2}{\sigma_i^2} - 1 \right) \frac{1}{\sigma_i^2} * \frac{\partial \sigma^2}{\partial \beta} \tag{17}$$

and setting the system of equations in (16) equal to zero, becomes complex to solve, it is difficult to solve it analytically, thus a numerical approach is needed.

It can be noticed that $\ln L(\beta)$ is always negative, since the likelihood is a probability by definition that is always between 0 and 1 and the ln of any number between 0 and 1 is negative. Numerically, the maximum can be found by "walking up" the likelihood function until no further increase can be found. The researcher specifies starting values of β . Each iteration moves to a new value of the parameters at which $\ln L(\beta)$ is higher than at the previous step. If we denote the current value at iteration k by β_k , the question is what is the best step we can take value for next, i.e what is the best β_{k+1} [7].

To determine the best value of β_{t+1} , we use a second -order Taylor's approximation of $LL(\beta_{t+1})$ around $LL(\beta_t)$:

$$LL(\beta_{t+1}) = \ln L(\beta_t) + (\beta_{t+1} - \beta_t)^T g_t + \frac{1}{2} (\beta_{t+1} - \beta_t)^T H_t (\beta_{t+1} - \beta_t) \tag{16}$$

Now we find the value of β_{t+1} that maximizes this approximation to $LL(\beta_{t+1})$:

$$\begin{aligned} \frac{\partial LL(\beta_{t+1})}{\partial \beta_{t+1}} &= g_t + H_t (\beta_{t+1} - \beta_t) = 0 \\ \beta_{t+1} &= \beta_t + (-H_t^{-1}) g_t \end{aligned} \tag{18}$$

The Newton-Raphson method uses this formula. The step from the current value of β_t to the new value is $(-H_t^{-1})g_t$, that is the gradient vector multiplied by the inverse of the Hessian matrix. The negative of this negative Hessian matrix is positive and represents the degree of curvature. That means that, $-H_t^{-1}$ is the positive curvature, assuming

that the log-likelihood function is globally concave. Each step is the slope of likelihood function divided by its curvature. If the slope is positive, β is raised and if the slope is negative, β is lowered, see "Figure. 1". The curvature determines how large a step is made. If the curvature is great, meaning that the slope changes quickly, and the maximum is likely to be close and vice versa as specified in "Figure. 2".

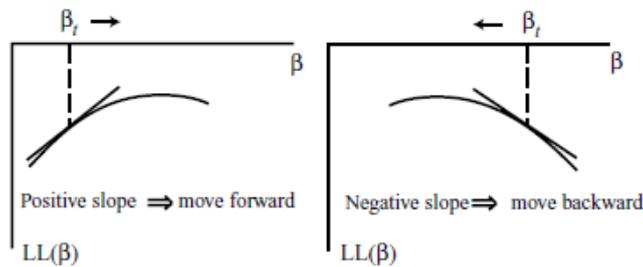


Figure 1: Direction of the step follows the slope.

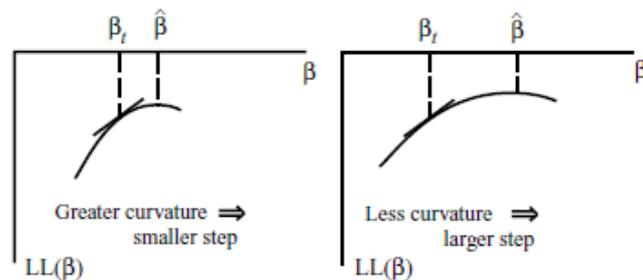


Figure 2: Step size is inversely related to curvature.

Three main issues are relevant to the Newton-Raphson method:

5.2 Quadratics: If $LL(\beta)$ were exactly quadratic in β , then the Newton-Raphson procedure would reach the maximum from any starting value. Let's take for example $LL(\beta)$ quadratic, then it can be written as follows:

$$LL(\beta) = a + b\beta + c\beta^2 \tag{19}$$

The maximum is

$$LL(\beta) = a + b\beta + c\beta^2$$

$$\frac{\partial LL(\beta)}{\partial \beta} = b + 2c\beta = 0 \rightarrow g_t$$

$$\beta = -\frac{b}{2c} \text{ and } H_t = 2c$$

Knowing the gradient g_t and the Hessian H_t and so Newton-Raphson give us:

$$\beta_{t+1} = \beta_t - H_t^{-1} g_t$$

$$= \beta_t - \frac{1}{2c} (b + 2c\beta)$$

$$= -\frac{b}{2c} = \hat{\beta} \tag{20}$$

Most of log-likelihood functions are not quadratic, and so the Newton-Raphson procedure takes more than one step to reach the maximum.

5.3 Step Size: It is possible for the Newton-Raphson procedure to step past the maximum and move to a lower $LL(\beta)$ as specified in "Figure. 3".

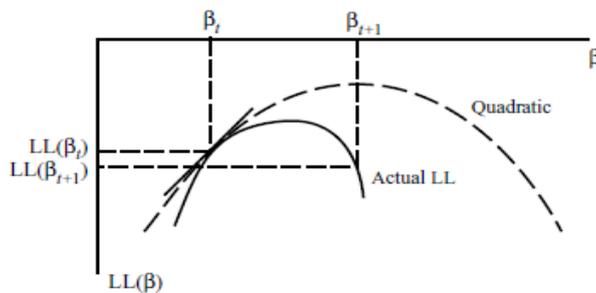


Figure 3: Step may go beyond maximum to lower LL.

To avoid this possibility, the step is multiplied by a scalar λ :

$$\beta_{t+1} = \beta_t + \lambda(-H_t)^{-1} g_t \tag{21}$$

The vector $(-H_t)g_t$ is called the direction, and λ is called step size.

5.4 Concavity: We have to mention that if the Log-Likelihood function (LL) is globally concave, then the Newton-Raphson procedure provides an increase in the Log-Likelihood at each iteration.

Moreover, if the LL is concave then the Hessian is definite negative at all values of β . The slope is declining and the second derivative is negative.

By definition a symmetric matrix M is positive definite if $x^T M x > 0$ for any $x \neq 0$. Let's consider a first Taylor's approximation of $LL(\beta_{t+1})$ around $LL(\beta_t)$:

$$\begin{aligned} LL(\beta_{t+1}) &= LL(\beta_t) + (\beta_{t+1} - \beta_t)^T g_t \rightarrow \beta_{t+1} - \beta_t = \lambda(-H_t^{-1}) g_t \\ LL(\beta_{t+1}) &= LL(\beta_t) + \lambda g_t^T (-H_t) g_t \end{aligned} \tag{22}$$

Since $-H_t^{-1}$ is positive definite, we have $\lambda g_t^T (-H_t) g_t > 0$ and $LL(\beta_{t+1}) > LL(\beta_t)$

If the LL is not concave, the Newton-Raphson procedure fails to find an increase because the step is on the opposite direction of the slope as specified in "Figure. 4":

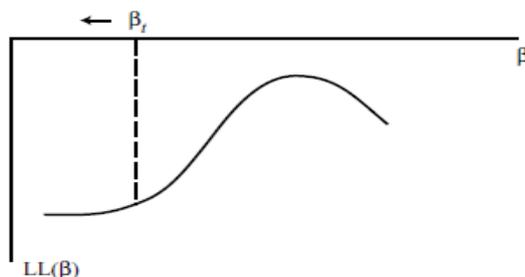


Figure 4: NR in the convex portion LL.

The NR has two drawbacks, in one side calculation of Hessian is usually computation-intensive, in the other side NR doesn't guarantee an increase in each step if LL is not globally concave. Other approaches use approximations to the Hessian that addresses this two issues. These methods differ in the form of the approximation. Each procedure defines a step as:

$$\beta_{t+1} = \beta_t + \lambda M_t g_t \tag{23}$$

Where M_t is $K \times K$ matrix. For NR, $M_t = -H^{-1}$. Other procedures use M_t that easier to calculate than the Hessian and are positive definite, so as to guarantee an increase in each iteration even in the convex region of LL.

5.2 Berndt, Hall, Hall, and Hausman (BHHH)

Maximization can be faster if we utilize the fact the function being maximized is a sum of terms in a sample. The score of an observation is the derivative of that observation's log-likelihood with respect to the parameters:

$$S_n(\beta_t) = \frac{\partial P_n(\beta)}{\partial \beta} \text{ evaluated at } \beta_t \tag{24}$$

The gradient is the average score

$$g_t = \frac{\sum_n S_n(\beta_n)}{N} \tag{25}$$

The outer product of observation n's score is the K*K matrix:

$$S_n(\beta_t)S_n(\beta_t)' = \begin{pmatrix} s_n^1 s_n^1 & s_n^1 s_n^2 & \dots & s_n^1 s_n^K \\ s_n^1 s_n^2 & s_n^2 s_n^2 & \dots & s_n^2 s_n^K \\ \dots & \dots & \dots & \dots \\ s_n^1 s_n^K & s_n^2 s_n^K & \dots & s_n^K s_n^K \end{pmatrix} \tag{26}$$

The average outer product in the sample is related to the covariance matrix.

The maximum occurs where the slope is zero, which means that the gradient (i.e, the average score) is zero and β_t is the variance of scores in the sample. The variance of the scores provides a measure of the curvature of the log-likelihood function, similar to the Hessian. The curvature is great when the variance of the scores is high as illustrated in the second panel of "Figure. 5":

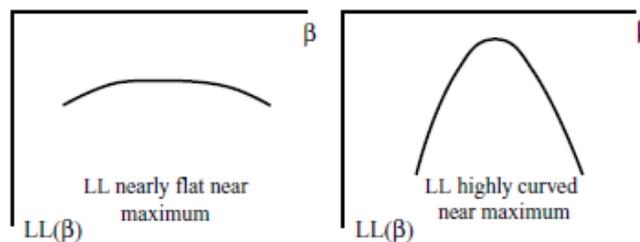


Figure 5: Shape of Log-likelihood function near maximum.

BHHH uses B_t in the optimization routine in place of $-H_t$

$$\beta_{t+1} = \beta_t - \lambda_t (B_t^{-1}) g_t \tag{27}$$

This yields two advantages over N, firstly; B_t is faster to calculate than $-H_t$ and we don't need to calculate the second derivative, Secondly, B_t is necessarily positive definite. The BHHH procedure is therefore guaranteed to provide an increase in $LL(\beta)$, in each iteration even in convex portions of the function.

5.3 Quasi-Newton iterative procedures

The quasi-Newton methods that build up an approximation of the inverse Hessian are often regarded as the most sophisticated for solving unconstrained problems. Even so, taking expectation of the inverse Hessian is essential for variance and covariance estimates in econometric modeling. Let be $p_t = \beta_{t+1} - \beta_t$ and $q_t = g_{t+1} - g_t$. There are many solutions to the quasi-Newton condition described above. Initial Hessian matrix is usually chosen as identity matrix which is updated by update formula. In classical modified quasi-Newton iterative procedure, assuming minimization problem:

$$\beta_{t+1} = \beta_t - \lambda_t (H_t^{-1}) g_t \tag{28}$$

rank two updates are the most widely used. Earliest update formula for constructing the inverse Hessian was originally proposed by Davidson (1959) and later developed by Fletcher and Powell (1963). DFP update formula has nice property: for a quadratic objective function, it simultaneously generates the directions of the conjugate gradient method while constructing the inverse Hessian. DFP update formula for inverse Hessian is given by:

$$H_{t+1}^{-1} = H_t^{-1} + \frac{p_t p_t^T}{p_t^T q_t} - \frac{H_t^{-1} q_t q_t^T H_t^{-1}}{q_t^T H_t^{-1} q_t} \tag{29}$$

According to Broyden-Fletcher-Goldfarb-Shanno (1970), update formula is given by:

$$H_{t+1}^{-1} = H_t^{-1} + \left(\frac{1 + q_t^T H_t q_t}{q_t^T p_t} \right) \frac{p_t p_t^T}{p_t^T q_t} - \frac{p_t q_t^T H_t + H_t q_t^T p_t}{q_t^T p_t} \tag{30}$$

Weighted combinations of these formulas leads to a whole collection of updates:

$$H_t = (1 - \alpha)H_k^{DFP} + \alpha.H_k^{BFGS} \quad \text{for } 0 \leq \alpha \leq 1 \quad (31)$$

In practice the calculated gradient vector is never exactly zero, but can be very close. Therefore $g_t^T(-H_t)^{-1}g_t$ is often used to evaluate convergence. If inequality: $g_t^T(-H_t)^{-1}g_t < 0.0001$ is satisfied, the iterative process stops and the parameters at current iteration are considered as estimated.

6. Estimation of GARCH(1,1) model

In this section, we present the parameter estimation of our econometric model GARCH (1,1), the number of lags that has been used are specified according to Akaike and Schwarz information criteria, the purpose of this section is not only the estimation of the GARCH(p,q) model but also make a comparison of the numerical method of maximization of the Log-likelihood function.

We first begin by plotting the graph of the MASI Index (Moroccan all share index) time series, which represents the main stock index of Casablanca Stock Exchange in Morocco. It is compound of all stocks of Casablanca stock exchange as specified in "Figure 6".

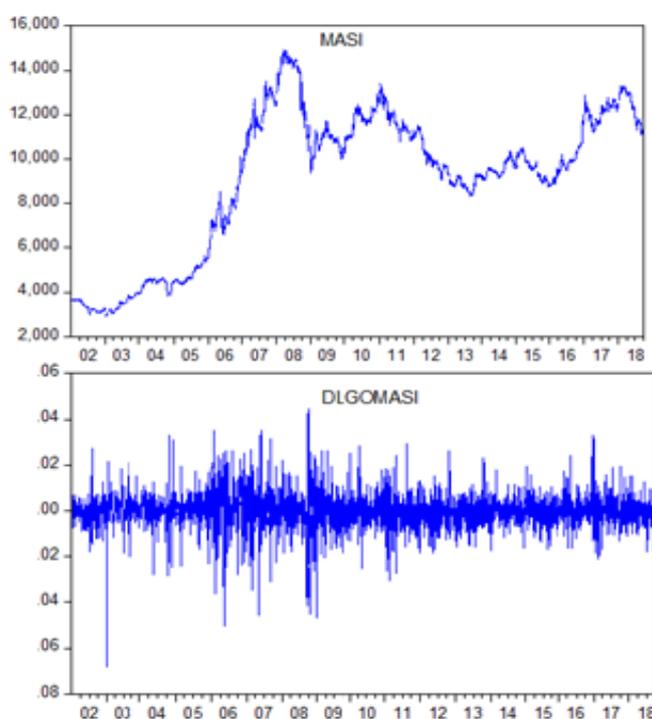


Figure 6: MASI closing prices and the logarithm of stock returns from 3 January 2002 to 18 October 2018.

In the table 1 and table 2 estimation of GARCH (1,1) parameters according to NR, BFGS, DFP and BHHH numerical algorithm are presented for different convergence criteria respectively $g_t^T(-H_t)^{-1}g_t < 0.0001$ and less than 0.00001.

Table 1: GARCH (1,1) parameter estimations when convergence criteria is satisfied the relative change in maximum likelihood between two successive iterations is less than 0.0001.

Parameter	BHHH	NR	BFGS	DFP
μ	0.000644	0.000644	0.000644	0.000644
α_0	0.46288864	0.4628887	0.4628914	0.4628897
α_1	0.3176352	0.3176343	0.3176338	0.3176353
β_1	2.78e-06	2.78e-06	2.78e-06	2.78e-06
Iterations	16	7	25	26
Log-likelihood	18404.79	18404.79	18404.79	18404.79

Table 2: GARCH(1,1) parameter estimations when convergence criteria is satisfied the relative change in maximum likelihood between two successive iterations is less than 0.00001.

Parameter	BHHH	NR	BFGS	DFP
μ	0.000644	0.000644	0.000644	0.000644
α_0	0.4628746	0.4628887	0.4628914	0.4628897
α_1	0.3176347	0.3176343	0.3176338	0.3176353
β_1	2.78e-06	2.78e-06	2.78e-06	2.78e-06
Iterations	16	7	25	26
Log-likelihood	18404.79	18404.79	18404.79	18404.79

7. CONCLUSION

In our case of GARCH(p,q) NR has approved to be the fastest in convergence according to the number of iterations followed by BHHH algorithm, BFGS and DFP is the last position rank. Even so, convergence problem may arise, because the more parameters in the model are entered the "flatter" the log-likelihood function becomes, and therefore the more difficult it is to maximize.

8. REFERENCES

- [1] David S. Bunch. a comparison of algorithms for maximum likelihood estimation of choice models. *Journal of Econometrics*.1988; 145-167.
- [2] Franses, Philip Hans, Dijk, Dick van. Non-Linear Time Series Models in Empirical Finance," **Cambridge Books**, Cambridge University Press, 2000.
- [3] Engle, R.F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*. 1982; 50:, 987–1007
- [4] Bollerslev, Tim. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*. 1986; 31(3): 307-327.
- [5] Elisabeth Orskaug, Multivariate DCC-GARCH Model -With Various Error Distributions, Master of Science in Physics and Mathematics, Norwegian university of science and technology, June 2009.
- [6] Stan Hum. Likelihood methods in financial econometrics, School of Economics and Finance, Queensland University of Technology, April 2009.
- [7] Josep Armeric, Zoran Barbic, Blanka Skrabic, Maximization of the likelihood function in financial time series models, "UNPUBLISHED"



Cite this article: Moulay, El Mehdi Falloul, Ayoub Razouk, and Youness Saoudi. NUMERICAL OPTIMIZATION METHODS FOR FINANCIAL TIME SERIES GARCH (P,Q) MODEL, A COMPARATIVE APPROACH. *Am. J. innov. res. appl. sci.* 2020; 10(3): 121-129.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>