



Munich Personal RePEc Archive

Non-parametric Estimation of GARCH (2, 2) Volatility model: A new Algorithm

Cassim, Lucius

Economics Department, University of Malawi

18 May 2018

Online at <https://mpra.ub.uni-muenchen.de/86861/>

MPRA Paper No. 86861, posted 21 May 2018 09:36 UTC

Non-parametric Estimation of GARCH (2, 2) Volatility model: A new Algorithm

Lucius Cassim¹

May 18 2018

Abstract

The main objective of this paper is to provide an estimation approach for non-parametric GARCH (2, 2) volatility model. Specifically the paper, by combining the aspects of multivariate adaptive regression splines(MARS) model estimation algorithm proposed by Chung (2012) and an algorithm proposed by Buhlman and McNeil(200), develops an algorithm for non-parametrically estimating GARCH (2,2) volatility model. Just like the MARS algorithm, the algorithm that is developed in this paper takes a logarithmic transformation as a preliminary analysis to examine a nonparametric volatility model. The algorithm however differs from the MARS algorithm by assuming that the innovations are i.d.d. The algorithm developed follows similar steps to that of Buhlman and McNeil (200) but starts by semi parametric estimation of the GARCH model and not parametric while relaxing the dependency assumption of the innovations to avoid exposing the estimation procedure to risk of inconsistency in the event of misspecification errors.

Key Words: GARCH (2,2), MARS, Algorithm, Parametric, Semi parametric, Nonparametric

¹ Economics Department, Chancellor College, University of Malawi, P.O.Box 280, Zomba, Malawi :
Email: luciuscassim@gmail.com or lcassim@cc.ac.mw

1. Introduction

1.1 Approaches to volatility Modelling

There are two approaches to modelling time-dependent volatility in financial literature; deterministic and stochastic approach. The deterministic approach models volatility as conditional variance expressed as a function of lagged conditional variance and lagged squared innovations. Deterministic models come under parametric, semi-parametric or non-parametric sub- approaches depending on the assumptions about the structure of the volatility.

At one extreme, parametric models make explicit both the functional form of the volatility model, while assuming a specific probability distribution of the innovations (see Hansen & Lunde, 2001; many variants of GARCH type models that have been proposed in the literature (see Taylor, 1986; Engle & Ng., 1993; Zakoian, 1994; Glosten & Runkle, 1993; Geweke, 1986; Pantula, 1986; Higgings & Bera, 1992; Sentana, 1995; Hentshel, 1995; Duan, 1997). At the other extreme, nonparametric approach makes no specification of the volatility model and no explicit assumption of the probability distribution of the innovations. It lets the data guide the process (Buhlman & McNeil, 2000). Semi parametric estimation approaches are a hybrid of the two extremes. Under this, the volatility model is explicitly specified but the distribution of the innovations is left unspecified. Our emphasis in this paper is on non-parametric approach to GARCH modelling.

It must be mentioned here that there are two definitions of the term “non-parametric statistics”. The first definition incorporates methods that do not rely on

data belonging to any particular distribution (i.e. distribution free methods). The other definition incorporate techniques that do not assume that the structure of a model is fixed a priori. In such methods, variables are assumed to belong to parametric distributions (Chung, 2012). It should be noted here that in this definition, even though structural assumptions about the model are not made a priori, statistical assumptions about the variables are made. In either definition, it is clear that, generally, non-parametric models differ from parametric models in that the model structure is not specified a priori but is instead determined from data. In other words, the term non-parametric is not meant to insinuate that such models completely lack parameters but that the number and the nature of the parameters are flexible and not fixed a priori (Chung, 2012). To ease the structural assumptions in parametric models, nonparametric models make no structural assumptions. There are so many nonparametric approaches to modelling and estimating GARCH models. They, however, are of not without faults, as ably explained below.

1.2 Objectives of the study

The main objective of this paper therefore is to develop a nonparametric approach to estimating GARCH (2, 2) model. Specifically, the study develops an algorithm that can be used to estimate nonparametric GARCH (2, 2) model.

2. Related Literature

2.1 Introduction

As explained above, there are three deterministic approaches to GARCH modelling. Since our emphasis in this paper is on one of them; the non-parametric approach, I will, in this section, review non-parametric models and algorithms that have been proposed in the literature. For a complete discussion on parametric and semi-parametric approaches see the following: Chung (2012); Yang & Song (2012); Dahl & Levine (2010); Drost & Klassenn (1996), Engle and Gonzale-Rivera (1991); Bollerslev & Woodridge (1992); Weiss (1986) Linton (2005); Yang (1998); Levine et al. (2012); Engle et al. (1993).

2.1.1. Basics of non-parametric estimation

Before going any further, let's look at some basics of non-parametric estimation techniques. The following derivations and formulae are due to Tschernig (2004). Assume equations (1.00) and (1.10) hold;

$$y_t = \mu(x_t) + \sigma(x_t)\varepsilon_t \quad (1.00)$$

$$x_t = (y_{t-i_1}, y_{t-i_2}, \dots, y_{t-i_n}) \quad (1.10)$$

Where x_t is the $(m \times 1)$ vector of all m current lagged values; $i_1 < i_2 < \dots < i_m$; $\varepsilon_t, t = i_m + 1, i_m + 2, \dots$, denotes a sequence of i.i.d random variables with zero mean and variance unity; $\mu(x_t)$ and $\sigma(x_t)$ denote the conditional mean and volatility function, respectively. Estimation of $\mu(x_t)$ and $\sigma(x_t)$ in equation (1.00) is mostly done locally, meaning it is estimated separately for each $(m \times 1)$ vector $x = (x_1, x_2, \dots, x_m)'$ of interest. Under this approach, although $\mu(x_t)$ is not observable it appears in a first order Taylor expansion of $\mu(x_t)$ taken at x as can be seen in equation (1.11) below.

$$\mu(x) = \mu(x_t) + \frac{\partial \mu(x_t)}{\partial x_t} (x_t - x) + R(x_t, x) \quad (1.11)$$

Where $R(x_t, x)$ denotes the remainder term. But equation (1.00) can now be written as equation (1.13) below.

$$y_t = \mu(x_t) + \sigma(x_t)\varepsilon_t \Rightarrow y_t = \mu(x_t)1 + \frac{\partial \mu(x_t)}{\partial x_t} (x_t - x) + R(x_t, x) + \varepsilon_t \quad (1.13)$$

From this, we observe that only 1 and $(x_t - x)$ are observable. This means that if $R(x_t, x) = 0$, one would estimate $\mu(x_t)$ by OLS, where $\mu(x_t)$ and $\frac{\partial \mu(x_t)}{\partial x_t}$ are parameters to be estimated. But, whenever the conditional mean is non-linear, the remainder term may not be zero and in such a case using standard OLS would give biased results for which the size of biasness depends on the sizes of all the remainder terms, $R(x_t, x), t = 1, 2, \dots, T$. One possibility to reduce the biasness is to use only those observations x_t that are in some sense close to x . That is to say, down weighing those observations that are not in local neighbourhood of x . If more data become available, it is possible to decrease the size of the local neighbourhood where the estimation variance and bias can reduce. The approximation error of the model can decline with sample size. This is the main idea behind non-parametric estimation approach. There are so many streams of non-parametric estimation techniques in the literature depending on the weighing scheme used. Technically, the weighing is controlled by the so-called kernel function $K(\cdot)$. A kernel function is a continuous function symmetric around zero that integrates to unity and satisfies the following additional boundness conditions (Cameron & Trivedi, 2005).

- i. $K(z)$ is symmetric around 0 and is continuous
- ii. $\int K(z) dz = 1, \int zK(z) dz = 0, \&, \int |K(z)| dz < \infty$
- iii. $K(z) \rightarrow 0$, as $|z| \rightarrow \infty$
- iv. $\int z^2 K(z) dz = \delta$ where δ is a constant

To adjust the size of the neighbourhood one introduces a band-width h , such that for a scalar x , the kernel function becomes $\frac{1}{h}K\left(\frac{x_i-x}{h}\right)$. If $m > 1$ and $x = (x_1, x_2, \dots, x_m)'$ is a vector, one uses a product kernel in equation (2.24) below;

$$K_h(x_i - x) = \prod_{i=1}^m \frac{1}{h^2} K\left(\frac{x_{ii} - x}{h}\right) \quad (1.14)$$

Here x_{ii} denotes the i -th component of x . The larger the h , the larger the neighbourhood and the larger the estimation bias. The band-width is also called a smoothing parameter. Since the observations in the local neighbourhood of x are the most important, this estimation is also called local estimation.

Owing to the introduction of a kernel function, one now has to solve a weighted least-squares problem as shown in equation (1.15);

$$(c_1, c_2, \dots, c_m) = \arg \min \sum_{t=i_m+1}^T \left(y_t - c - \sum_{i=1}^m c_i (x_{ii} - x) \right)^2 K_h(x_t - x) \quad (1.15)$$

This delivers the local linear function estimate $\hat{\mu}(x, h)$ at the point x . Technically, with matrices;

$$e = (1 \quad 0_{1 \times m})'; \quad z(x) = \begin{pmatrix} 1 & \dots & 1 \\ x_{i_m+1} & \dots & x_T - x \end{pmatrix}'; \quad W(x, h) = \text{diag} \left(\frac{K_h(x_t - x)}{T} \right)_{t=i_m+1}^T$$

$$y = (y_{i_m+1}, \dots, y_T)' \Rightarrow \hat{\mu}(x, h) = e'(z'(x)W(x, h)z(x))^{-1} z'(x)W(x, h)y \quad (1.16)$$

The most popular weighting scheme is that of Watson (1964) and Nadaraya (1964) given in equation. According to Nadaraya-Watson assume the following regression model;

$$Y_i = m(x_i) + \varepsilon_i, \forall i = 1, 2, \dots, n \quad (1.17)$$

Where $m(\cdot)$ is unknown. One notices that $m(x_i)$ can be expressed in terms of $f(x, y)$ as follows;

$$m(x) = E[Y | X = x] = \int yf(y | x)dy = \frac{\int yf(x, y)dy}{\int f(x, y)dy} \quad (1.18)$$

The goal is to estimate the numerator and the denominator separately using kernel estimators. One notices that;

$$\begin{aligned} \hat{f}(x, y) &= \frac{1}{nh_x h_y} \sum_{\forall i} K\left(\frac{x - x_i}{h_x}\right) K\left(\frac{y - y_i}{h_y}\right) = \frac{1}{n} \sum_{\forall i} K_{h_x}(x - x_i) K_{h_y}(y - y_i) \\ &\Rightarrow \int y \hat{f}(x, y) dy = \frac{1}{n} \int y \sum_{\forall i} K_{h_x}(x - x_i) K_{h_y}(y - y_i) \end{aligned}$$

$$\text{But } \int y K_{h_y}(y - y_i) dy = y \Rightarrow \int y \hat{f}(x, y) dy = \frac{1}{n} \sum_{\forall i} K_{h_x}(x - x_i) y_i$$

$$\text{Analogously, } \int \hat{f}(x, y) dy = \frac{1}{n} \sum_{\forall i} K_{h_x}(x - x_i) \int K_{h_y}(y - y_i) dy = \frac{1}{n} \sum_{\forall i} K_{h_x}(x - x_i) = \hat{f}(x)$$

$$\therefore \hat{m}(x) = \frac{\sum_{\forall i} K_{h_x}(x - x_i) y_i}{\sum_{\forall i} K_{h_x}(x - x_i)} = \sum_{\forall i} W_{h_x}(x, x_i) y_i, W_{h_x}(x, x_i) = \frac{K_{h_x}(x - x_i)}{\sum_{\forall i} K_{h_x}(x - x_i)}$$

A theoretical formulation of Nadaraya -Watson conditional variance estimator using squared residuals obtained from the conditional mean function is proposed by Fan and Yao (1998) and Fan and Gijbels (1995) and we sketch their formalization below. The explanation below is due to Chung (2012). Assume that a strictly stationary process $\{x_t; t = 1, 2, \dots, T\}$ is generated by equation (1.19) below.

$$x_t = m(x_{t-1}) + \sqrt{\lambda}(x_{t-1})\varepsilon_t \quad (1.19)$$

Where ε_t is i.i.d, $E(\varepsilon_t | \Xi_{t-1}) = 0$, $Var(\varepsilon_t | \Xi_{t-1}) = 1$, and Ξ_{t-1} is a sigma algebra generated by x_{t-1} (or some past information). Fan and Yao (1998) proposed a two stage method to obtain a local linear estimator for conditional variance (Chung, 2012; Dahl & Levine, 2010).

- i. Obtain local linear estimator $\hat{m}(x_t) = \hat{a}$ which is the minimization intercept in the following weighted least-squares problem;

$$(\hat{a}, \hat{b}) = \arg \min \sum_{t=i_m+1}^T (x_t - a - b(x - x)) K_h \left(\frac{x_{t-1} - x}{h_1} \right) \quad (1.20)$$

- ii. Obtain the squared residuals $\hat{r}_t = [x_t - \hat{m}(x_{t-1})]^2$ to use in equation (2.30) below.

$$(\hat{\alpha}, \hat{\beta}) = \arg \min \sum_{t=i_m+1}^T (\hat{r}_t - \alpha - \beta(x - x)) K_h \left(\frac{x_{t-1} - x}{h_2} \right) \quad (1.21)$$

Where the bandwidth $h_2 > 0$ is different from h_1 (Chung, 2012; Dahl & Levine, 2010). As we have seen in equation (1.17) above, the non-parametric mean estimator is;

$$\hat{\mu}_{NW}(x, h) = e' \left(z'_{WN}(x) W(x, h) z(x) \right)^{-1} z'_{WN}(x) W(x, h) y$$

$$\left(\sum_{t=i_m+1}^T K_h(x_t - x) \right)^{-1} \left(\sum_{t=i_m+1}^T K_h(x_t - x) \right) y_t$$

According to Tschernig (2004);

$$\sqrt{Th^m} (\hat{\mu}(x, h) - \mu(x) - b(x)h^2) \rightarrow N(0, v(x))$$

Where the asymptotic bias $b(x)$ and asymptotic variance $v(x)$ are as given respectively in equations (1.21) and (1.22) below

$$b(x) = \frac{\sigma_k^2}{2} \text{tr} \left(\frac{\partial \mu(x)}{\partial x \partial x'} \right) \quad (1.22)$$

$$v(x) = \frac{\sigma^2 \|K\|_2^{2m}}{f(x)} \quad (1.23)$$

$$\Rightarrow \hat{\mu}(x, h) \rightarrow N \left(\mu(x) + \frac{\sigma_k^2}{2} \text{tr} \left(\frac{\partial \mu(x)}{\partial x \partial x'} \right) h^2, \frac{1}{Th^m} \frac{\sigma^2 \|K\|_2^{2m}}{f(x)} \right)$$

Thus, if we denote any positive constant β , any optimal band-width for which $h = \beta T^{\left(-\frac{1}{m+4}\right)}$ holds has an optimal rate of decline of bias (Tschernig, 2004).

$$\begin{aligned} &\Rightarrow \sqrt{Th^m} (\hat{\mu}(x, h) - \mu(x) - b(x)h^2) \rightarrow N(0, v(x)) \\ &\Rightarrow T^{\frac{2}{m+4}} (\hat{\mu}(x, h) - \mu(x)) \rightarrow N \left(\frac{\sigma_k^2}{2} \text{tr} \left(\frac{\partial \mu(x)}{\partial x \partial x'} \right) h^2, \frac{1}{h^m} \frac{\sigma^2 \|K\|_2^{2m}}{f(x)} \right) \\ &\log \{\sigma_t^2\} = k(X_{t-1}; \beta) \end{aligned}$$

The equation above can be written as: $Y_t = h(X_{t-1}; \beta) + \mu_t$; where

$Y_t = \log \{X_t^2\}, h(X_{t-1}; \beta) = E[\log \{\varepsilon_t^2\}] + k(X_{t-1}; \beta)$ and μ_t is an i.i.d sequence such that $E[\mu_t] = 0$.

2.1.2 Proposed non-parametric GARCH models and estimation algorithms in the literature

There are a lot of non-parametric GARCH models and associated estimation algorithms that have been proposed in the literature. In this paper we will review just two of them; Buhlamann and McNeil (200) Model and algorithm and the Multivariate Adaptive Regression Splines (MARS) model and algorithm. These are

popular and mostly used (except the MARS) approaches in the literature. For a complete discussion on the other models and algorithms consult Chung (2012).

Buhlamann and McNeil algorithm: their paper, Buhlamanna and McNeil (200) considered the following model; assume a stationary stochastic process

$$\begin{aligned} & \{X_t; t \in Z\} \\ & \left\{ \begin{array}{l} X_t = \sigma_t Z_t, \{Z_t; t \in Z \sim i.i.d(0,1)\}, Z_t \perp X_t \\ \sigma_t^2 = f(X_{t-1}, \sigma_{t-1}^2) \end{array} \right. \\ & \Rightarrow X_t^2 = f(X_{t-1}, \sigma_{t-1}^2) + V_t \\ V_t = f(X_{t-1}, \sigma_{t-1}^2)(Z_t^2 - 1) & \Rightarrow E[X_t^2 | \Omega_{t-1}] = f(X_{t-1}, \sigma_{t-1}^2) \& Var[X_t^2 | \Omega_{t-1}] = f^2(X_{t-1}, \sigma_{t-1}^2)(E[Z_t^4] - 1) \end{aligned}$$

This, as rightly put by Buhlamanna and McNeil (200), suggests one can estimate f by regressing X^2 on X_{t-1} and the lagged variances of innovations. However the problem is that the lagged variance of innovations is unobserved. This is where one needs to develop an algorithm for estimating it. Buhlamanna and McNeil (200) proposed the following algorithm;

- i. Estimate volatility $\{\hat{\sigma}_{t,0}; 1 \leq t \leq n\}$ by fitting a GARCH (1,1) parametrically by standard maximum likelihood. Set $m=1$
- ii. Regress $\{X^2_t; 2 \leq t \leq n\}$ against $\{X^2_{t-1}; 2 \leq t \leq n\}$, and $\{\sigma^2_{t-1,m-1}; 2 \leq t \leq n\}$ using a non-parametric procedure
- iii. Calculate $\{\hat{\sigma}_{t,m}^2 = \hat{\xi}_m(X_{t-1}, X_{t-2}, \hat{\sigma}_{t-1,m-1}^2, \hat{\sigma}_{t-2,m-2}^2)\}$ and get a value of $\hat{\sigma}_{t,m}^2$
- iv. Increase m and return to (2) if $m < M$
- v. Then $\hat{\sigma}_{t,*} = \left(\frac{1}{K}\right) \sum_{m=M-k+2}^M \hat{\sigma}_{t,m}$
- vi. Regress $\{X^2_t; 2 \leq t \leq n\}$ on $\{X^2_{t-1}; 2 \leq t \leq n\}$, and $\hat{\sigma}_{t-1}^2$ to get $\hat{\xi}_m(X_{t-1}, \hat{\sigma}_{t-1,m-1}^2)$.

Chung (2012) Model and algorithm:

$$\left\{ \begin{array}{l} X_t = \sigma_t \varepsilon_t, \{\varepsilon \sim i.i.d(0,1)\} \\ \log(\sigma_t^2) = k(X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p}, \sigma_{t-1}, \sigma_{t-2}, \sigma_{t-3}, \dots, \sigma_{t-q}) \end{array} \right.$$

For simplicity purposes, Chung (2012) omitted the lagged volatility to have;

One notices that
$$\begin{cases} X_t = \sigma_t \varepsilon_t, \{\varepsilon \sim i.i.d(0,1)\} \\ \log(\sigma_t^2) = k(X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p}) \end{cases}$$
 can be written as;

$$\begin{cases} X_t = \sigma_t \varepsilon_t, \{\varepsilon \sim i.i.d(0,1)\} \\ Y_t = h(X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p}) + \mu_t \end{cases}$$

$$Y_t = \log\{X_t^2\}, h(X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p}) = E[\log\{\mu_t^2\}] + k(X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p})$$

The MARS algorithm, therefore, involves the problem of minimizing the residual

sum of squares; $\sum_{t=2}^T [Y_t - h(X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p})]^2$ with respect to β , lagged variables

and knot locations at each iteration where $h(X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-p})$ takes the basis form.

Limitations of the algorithm

The following criticisms can be directed at this model and algorithm;

- i. It assumes that the innovations are i.i.d. We have already explained the negative implications of making this assumption.
- ii. It imposes an assumption on the transformed distribution of innovations μ_t to be normal. This strongly violates the assumption if we specify the original innovation distribution to be normal or t-distribution, which is a reasonable assumption most financial application.

Critique of the current literature

It should be mentioned here that the reviewed approaches are some of the most popular approaches to modelling and estimating nonparametric GARCH models. Much as they have been major advances in providing estimates without imposing structural assumptions of the model they are, as we have seen above, not without faults. As we have mentioned above, the Buhlmann's algorithm not only restricts the parameter space to be positive but also assumes that the innovations are i.i.d. On top of that, the algorithm starts by estimating the GARCH model using

parametric approach. It is a known fact that the parametric estimation of GARCH models exposes the parameter estimators to high risk of inconsistency due to its restrictive structural and statistical assumptions. The MARS algorithm also makes an i.i.d assumption and imposes a distributional assumption on the transformed innovations to be normal. Both of these expose the estimators to high risk of inconsistency. One improvement is therefore clear from the current non-parametric approaches;

There is a need for an algorithm that has very low chance of exposing the estimators to inconsistency by solving limitations of the two approaches explained above.

This is the literature gap that the paper is trying to fill. This can be done by combining the good aspects of the two approaches to come up with one algorithm that is free of the problems outlined above.

3. The Proposed Algorithm

3.1 The Model and Notations

In this subsection, I present the proposed model and algorithm. I present a GARCH (2, 2) model, a slightly generalized version of Buhlamann's (2000) GARCH (1,1) model. Unlike Buhlamann's, however, I follow Chung (2012) by taking log of the volatility. This will help remove the positivity parameter restrictions. On top of that I relax the i.i.d assumption of innovations. Specifically, I relax the statistical independence (within the i.i.d assumption) of innovations. This will not only reduce the exposure of the estimators to inconsistency risk but it's also a better assumption to make of financial asset returns' innovations. Assume a stochastic process $\{z_t : t \in Z\}$ adapted to filtration $\{\Omega_t : t \in Z\}$ where $\Omega_t = \sigma(\{z_s : s \leq t\})$;

$$\begin{cases} X_t = \sigma_t Z_t \\ \log(\sigma_t^2) = \xi(X_{t-1}, X_{t-2}, \sigma_{t-1}, \sigma_{t-2}) \end{cases}$$

I make the following assumptions

- i. $Z_t \perp \{X_s; s < t\}$
- ii. $\xi : \mathfrak{R} \times \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ is a strictly positive valued function
- iii. μ_t is a martingale difference
- iv. $E[\log\{\mu_t\}] = 0$

Notice that $X_t = \sigma_t \mu \Rightarrow X_t^2 = \sigma_t^2 + Z_t^2 \Rightarrow \log(X_t^2) = \log(\sigma_t^2) + \log(Z_t^2)$.

$$\therefore \log(X_t^2) = \xi(\mu_{t-1}, \mu_{t-2}, \sigma_{t-1}, \sigma_{t-2}) + E[\log(\mu_t^2)] \text{ since } \xi(\mu_{t-1}, \mu_{t-2}, \sigma_{t-1}, \sigma_{t-2}) = \log(\sigma_t^2)$$

But $E[\log\{\mu_t\}] = 0 \Rightarrow \therefore \log(X_t^2) = \xi(\mu_{t-1}, \mu_{t-2}, \sigma_{t-1}, \sigma_{t-2})$. Therefore estimation of the model involves regressing $\log(X_t^2)$ on $\mu_{t-1}, \mu_{t-2}, \sigma_{t-1}, \sigma_{t-2}$. But, $\sigma_{t-1}, \sigma_{t-2}$ are unobserved

latent variables .I therefore adapt Buhlmann's (2000) algorithm in estimating $\sigma_{t-1}, \sigma_{t-2}$ s explained below.

3.2 The Proposed Algorithm:

- i. Estimate volatility $\{\hat{\sigma}_{t,0}; 1 \leq t \leq n\}$ by fitting a GARCH (2,2) model semi parametrically by assuming that the innovations have a generalized error distribution, while relaxing the independency assumption, as proposed by Cassim(2018).
- ii. Regress $\{X^2_t; 2 \leq t \leq n\}$ against $\{X^2_{t-1}; 2 \leq t \leq n\}$, $\{X^2_{t-2}; 2 \leq t \leq n\}$, $\{\sigma^2_{t-1,m-1}; 2 \leq t \leq n\}$ and $\{\sigma^2_{t-2,m-2}; 2 \leq t \leq n\}$ using Nadaraya non-parametric procedure
- iii. Calculate $\{\hat{\sigma}_{t,m}^2 = \hat{\xi}_m(X_{t-1}, X_{t-2}, \hat{\sigma}_{t-1,m-1}^2, \hat{\sigma}_{t-2,m-2}^2)\}$ and get a value of $\hat{\sigma}_{t,m}^2$
- iv. Increase m and return to (2) if $m < M$
- v. Then $\hat{\sigma}_{t^*} = \left(\frac{1}{K}\right) \sum_{m=M-k+2}^M \hat{\sigma}_{t,m}^2$
- vi. Regress $\{X^2_t; 2 \leq t \leq n\}$ on $\{X^2_{t-1}; 2 \leq t \leq n\}$, $\{X^2_{t-2}; 2 \leq t \leq n\}$, $\hat{\sigma}_{t-1}^2$ and $\hat{\sigma}_{t-2}^2$ to get $\hat{\xi}_m(X_{t-1}, X_{t-2}, \hat{\sigma}_{t-1,m-1}^2, \hat{\sigma}_{t-2,m-2}^2)$.

4. Conclusion

The main objective of this paper was to develop a nonparametric approach to estimating GARCH (2, 2) model. Specifically, the study aimed at developing an algorithm that can be used to estimate nonparametric GARCH (2, 2) model. By combining the MARS algorithm and an algorithm developed by Buhlman and McNeil(2000), the study has developed an algorithm for estimating nonparametric GARCH(2,2) model with very small risk of exposing the estimates to inconsistency.

REFERENCES

- Bollerslev. (1987). A conditional Heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69, 542-547.
- Bollerslev, T. (1986). Generalized Autoregressive Heteroskedasticity. *Journal of Econometrics*, 307-327.
- Bollerslev, J., & Woodridge, J. (1992). Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews*, 11(2), 143-172.
- Buhlman, P., & McNeil, A. J. (2000). *Non-parametric GARCH Models*. Zurich, Switzerland: Seminar Fur Statistik:CH-8092.
- Cameron, C. A., & Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. New York, USA: Cambridge University Press.
- Choi, E. J. (2004). Estimation of Stochastic Volatility Models by Simulated Maximum Likelihood Method. *University of Waterloo*.
- Chung, S. S. (2012). A class of non-parametric volatility models: Application to financial time series. *Journal of Econometrics*.
- Dahl, C. M., & Levine, M. (2010). Non-parametric estimation of volatility models under serially dependent innovations. *Econometrica*.
- Davidson, J. (2000). *Econometric Theory*. Blackwell: Oxford University Press.
- Davidson, R., & Mackinnon, J. (1993). *Estimation and inference in Econometrics*. London: Oxford University Press.
- Drost, F. C., & Klassens, C. (1996). Efficient estimation in Semi-parametric GARCH Models. *Discussion paper; vol(1996-38), Tilburg*.

- Duan, J. (1997). Augmented GARCH(p,q) process and its diffusion limit. *Journal of Econometrics*,79(1), 97-127.
- Engle, R. (1982). Autoregressive Conditional Heteroskedasticity with estimation of the variance of U.K inflation. *Econometrica*, 987-1008.
- Engle, R. F., & Gonzale-Rivera, G. (October,1991). Semiparametric ARCH Models. *Journal of Business and Economic Statistics*, 9(4), 345-359.
- Engle, R. F., & Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of finance*,48, 1747-1778.
- Fan, & Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting:variable bandwidth and spatial adaptation. *Journal of Royal Statistical Society,B*,57, 371-394.
- Fan, J., & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*,85, 645-660.
- Gallant, A. R., & Hsieh, D. (1989). Fitting a Recalcitrant series:The Pound/Dollar Exchange Rate,1974-83. *Econometrica*.
- Geweke, J. (1986). Modelling Persistence in Conditional Variances: A Comment. *Econometric Review*,5, 56-61.
- Glosten, L. R., & Runkle, D. (1993). On the relation between the expected value and volatility of the nominal excess return of stocks. *Journal of Finance*,48, 1779-1801.
- Gourieroux, C., & Trognon, A. (1984). Pseudo Maximum Likelihood Methods:Theory,52(1). *Econometrica*, 681-700.

- Haafner, C. M. (2003). Analytical quasi maximum likelihood inference in BEKK-GARCH models. *Econometric Institution, Erasmus University, Rotterdam.*
- Hafner, C. M., & Rombouts, J. (2002). Semiparametric multivariate GARCH models . *Discussion paper, 2002/XX, CORE.*
- Hansen, B. C. (2006). *Econometrics*. New York: Cambridge University Press.
- Hansen, P., & Heyde, C. (1980). *Martingale limit theory and its applications*. New York: Academic Press.
- Hansen, R. P., & Lunde, A. (2001). A comparison of volatility models: Does anything beat GARCH(1,1)? *Centre for analytical finance: University of AARHUS.*
- Hentshel, L. (1995). All in the family: Nesting Symetric and Asymmetric GARCH Models. *Journal of Financial Economics, 39, 71-104.*
- Herwartz, H. (2004). Conditional Heteroskedasticity. In H. Lutkepoh, & M. Kratzig (Eds.) *Themes in Modern Econometrics*, pp. 197-220.
- Higging, M. L., & Bera, A. (1992). A Class of non-linear ARCH Models. *International Economic Review.*
- Holly. (2009). *Modelling Risk using fourth order Pseudo Maximum Likelihood Methods*. University of Lausanne, Institute of Healthy Economics.
- Holly, A., & Montifort, A. (2010). Fourth Order Pseudo Maximum Likelihood Methods. *Econometrica.*