Nonparametric GARCH models for Financial Volatility

by

Siyan Hu

(Under the direction of Lily Wang)

Abstract

In this thesis, we investigate a variety of stochastic models for volatility prediction in financial time series. We compare two non-parametric volatility models with the standard GARCH(1,1) model. In the first nonparametric GARCH modeling, we consider the functional gradient descent (FGD) method in Audrino and Bühlmann (2009) to find out the optimal B-spline structure in order to get the maximum likelihood. In the second nonparametric GARCH modeling, we consider the additive autoregressive structure (aGARCH) with components linked together by a dynamic coefficient proposed in Wang, *et al.* (2011). B-spline smoothing method is adopted in both algorithms. The performance of both the parametric and non-parametric GARCH models is investigated by means of simulation studies and an application to S&P 500 index return study and Apple stock return study. They both demonstrate strong improvement in volatility prediction.

Index words: B-splines, Generalized auto-regressive conditional heteroscedasticity model, Time Series, Volatility

Nonparametric GARCH models for Financial Volatility

by

Siyan Hu

M.S., Shanghai University, Shanghai, China, 2007

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Master of Science

Athens, Georgia

2011

Nonparametric GARCH models for Financial Volatility

by

Siyan Hu

Approved:

Major Professor: Lily Wang

Committee: Abhyuday Mandal
Jaxk Reeves

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2011

Table of Contents

List of Tables

Dedication

This paper is dedicated to my family.

Acknowledgments

First and foremost, I'd like to express my sincere appreciation to my advisor, Professor Wang. I couldn't finish my research so smoothly and quickly without her insightful guidance and unlimited patience. She always gives me strong support and useful suggestions during my study.

I would like to thank two other members in my advisory committee, Professor Mandal and Professor Reeves. The final version of the dissertation has benefited from their proofreading.

I would also like to thank my family and friends for their continuous support, love and encouragement.

Thanks to Statistics Department for giving me an unforgettable and valuable graduate study.

Without all of these encouragement and help, it would be hard for me to finish this degree.

Chapter 1

Introduction

In recent decades, financial volatility has been viewed as an important feature in investment decision, derivative pricing and risk management, etc. When Engle (1982) introduced the auto-regressive conditional heteroscedasticity (ARCH) model, a large number of models were proposed to predict volatility afterward. One of the most popular models is the generalized auto-regressive conditional heteroscedasticity (GARCH) model. Most of the existing models only involve a small number of parameters and very few lags to keep the model parsimony. Recently more flexible models have been proposed through non-parametric and semi-parametric smoothing methods. Among all the smoothing methods, B-spline is simple and very powerful in terms of computing and also has explicit formula, thus it gains a lot of popularity in estimating the conditional variance functions.

In this thesis, we compare the non-parametric GARCH models in Audrino and Bühlmann (2009) and Wang, *et al.* (2011) with the classical GARCH(1,1) model. We aim to have some appropriate models that can fit the financial time series when it is nonlinear or if the conditional variance is highly persistent. One common feature of these two non-parametric models is that they are cater to high-dimensional non-parametric regression with many lags. We validate the goodness of fit in terms of volatility testing and prediction on both simulated data and real data. We compare results with the standard GARCH(1,1) model which is widely used as a benchmark in financial field.

The rest of the thesis is organized as follows. The background information of this work is given in Chapter 2. Specifically, Section 2.1 introduces the financial volatility; Sections 2.2, 2.3 and 2.4 describe some basic time series models and the standard ARCH/GARCH

models as well as their non-parametric extension; Section 2.5 presents the B-spline smoothing which is the key tool used in the non/semi-parametric methods. In Chapter 3, we elaborate two non-parametric GARCH models in depth with a functional gradient descent approach in Section 3.1 and a semi-parametric aGARCH method in Section 3.2. In Chapter 4, we compare them with the simple GARCH(1,1) model using different types of simulated data and also apply them to the real financial time series data. Conclusion is presented in Chapter 5.

Chapter 2

Time Series

2.1 Financial Volatility

Volatility forecasting in financial markets is a key task to study during the recent decades. There are around 100 published and working papers (Poon and Granger, 2003) on forecasting performance of various volatility models, and even more papers about the volatility topic but without considering forecasting. Volatility can be viewed as a vital role in the financial performance, especially in the investment strategy, bond valuation, risk management and asset pricing. For example, we need to estimate the volatility of some options before they expire in order to do the pricing strategy. Recently, there are a lot of derivatives priced as volatility and therefore, volatility becomes more and more important and has attracted more people's attention. If investors can understand and predict the underlying assets' volatility quite well, there's no doubt that they will benefit greatly during the investment.

However, many investors get confused about the difference between volatility, standard deviation, and risk. It is worth making a clarification here. Generally, we only say *risk* for short. But sometimes *risk* is different from volatility as we just mentioned above. The risk we are talking about has a lot of levels in portfolio managements while volatility is the key input among them when investors make their decision. Thus, a good estimation of volatility of financial prices is crucial in the investment decision. Usually, in finance, volatility is defined as the standard deviation, σ, or its variance, σ^2. In statistics, we've know the σ is computed as the following:

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{t=1}^{N}(Y_t - \bar{Y})^2,$$

where Y_t is the return and $\bar{Y}$ is the mean return.

Usually, if σ follows a standard normal distribution, or a t distribution, it is easy to derive analytically its probability density and cumulative probability density. In real life, more cases are that σ has irregular shape and it is difficult to do any calculation and can only be derived empirically; see Poon and Granger (2003).

2.2 Time Series Introduction

2.2.1 Background

In real life, observations are made sequentially over time, instead of independent on each other. It means in more cases, values we are interested in the future may rely on the past observations or observation at present. This provides us the opportunity to predict the future from the past. Indeed, we are able to describe the underlying dynamics from what we've already had and possibly control the future events. Time Series analysis has been an important part of statistics; see Chatfield (2009). Time series has a long history and some of the branches cross the control theory, econometrics and even some fields of physics and engineering.

Time series analysis deals with records that are collected over time. Examples can be daily mortality counts, particulate air pollution measurements, and temperature data, etc. The sequence of data plays important role. A fundamental feature in time series is that the records are usually dependent on its previous values. The application of time series is very diverse, while data may be collected hourly, daily, weekly, monthly, or yearly, and so on. We use notation such as X_t or Y_t $(t = 1, ..., T)$ to denote a time series of length T. The unit of the time scale is usually implicit in the notation above. In this thesis, we use Y_t as the observations.

As pointed out in Fan and Yao (2003), a fundamental task for time series analysis is to reveal the probability law that governs the observed time series. Hence, we can understand

the underlying dynamics, forecast future events, and control future events via intervention, which are the three main objectives of time series analysis.

Financial time series analysis is related with the theory and practice of valuation over time. It is a highly empirical discipline, but like other scientific fields theory forms the foundation for making inference. To represent the empirical properties observed in real prices and to estimate and test financial models are the two main motivations to analyze financial data; see Balbás, *et al.* (2003). However, one key distinction between financial time series analysis and other time series analysis is that financial theory and its empirical time series contain an element of uncertainty. For example, there are various definitions of asset volatility, and for a stock return series, the volatility is not directly observable. As a result of the added uncertainty, statistical theory and methods play an import role in financial time series analysis.

2.2.2 Stationary Time Series

In time series analysis, a variety of techniques can be applied to extract features of interest. The base line is to find the characteristic of interest that fits a general form which provides prediction precisely with growing sample size; see Härdel, *et al.* (1997).

We usually generate a *stochastic* process in order to analyze a time series. Such stochastic process can be described as a statistical phenomenon that involves in time. However, statistical problems concerned with estimating properties of a population from a sample is a different situation. Since it is impossible to make multiple observations at any single time this makes the conventional statistical procedures, based on large sample estimates, inappropriate. Therefore, stationarity is a convenient assumption that allows us to describe the statistical properties of a time series.

Loosely speaking, a *stationary* time series is the one whose statistics properties stay constant over time, (Nason, 2008) or say, the strictly periodic variations or seasonality do not exist.

Suppose a series of observations $Y(t_1), ..., Y(t_n)$ is stationary, then, the joint distribution of $Y(t_1), ..., Y(t_n)$ is the same as that of $Y(t_1 + h), ...Y(t_n + h)$ for all $t_1, ..., t_n$ and h. This implies that the expected value and covariance structure of any two components $Y(t_i)$ and $Y(t_j)$ are constant in time, like the following:

$$\begin{aligned} E\{Y(t)\} &= \mu_a, \\ var\{Y(t)\} &= \sigma_a^2, \\ corr\{Y(t), Y(t+h)\} &= \gamma(h). \end{aligned}$$

If $h \neq 0$, the above function $\gamma(h)$ is defined as the cross-correlation function, while if $h = 0$, it is called the autocorrelation function.

In general, people often adopt stationarity in a less restricted way. In many cases, the statistical processes can be completely described with the second-order properties of the above three equations.

2.2.3 AR/MA/ARMA Model

A stochastic process Y_t is called *white noise*, denoted as $Y_t \sim WN(0, \sigma^2)$, if

$$\begin{aligned} E(Y_t) &= 0, \\ Var(Y_t) &= \sigma^2, \\ Cov(Y_i, Y_j) &= 0, \quad i \neq j. \end{aligned}$$

White noise is only defined as the first two statements. And it is easy to figure out that a sequence of *independent and identically distributed* (i.i.d) random variables with mean 0 and finite variance σ^2 can be a special white noise process.

A model in which future values are forecast purely on the basis of past values of the time series is called an *autoregressive* (AR) process. An AR model of order $p \geq 1$ is defined as the following:

$$Y_t = b_1 Y_{t-1} + ... + b_p Y_{t-p} + \epsilon_t,$$

where $\{\epsilon_t\} \sim WN(0, \sigma^2)$. By convention, we denote $\{Y_t\} \sim \text{AR}(p)$, and the time series $\{Y_t\}$ generated from this model is called the AR(p) process. If we take a closer look at this model, it is easy to see that the current Y_t is a linear regression form through its p past values $Y_{t-1}, ..., Y_{t-p}$. Because it is easy to implement, it is the most popular time series model in practice.

Similarly, a model in which future values are forecast purely on the basis of past shocks (or noise or random disturbances) is called a *moving average* (MA) precess. An MA process with order $q \geq 1$ is defined as:

$$Y_t = \epsilon_t + a_1\epsilon_{t-1} + ... + a_q\epsilon_{t-q},$$

where $\{\epsilon_t\} \sim \text{WN}(0, \sigma^2)$. By convention, we write $\{Y_t\} \sim \text{MA}(q)$. An MA model indicates a time series as a moving average of a white noise process, while Y_t and Y_{t-h} are uncorrelated to each other.

We can see AR(p) can be stated as an infinite-order MA process and vice versa. Therefore, a model that combines AR and MA terms together will be a more parsimonious model; see Cressie and Wikle (2011). It is called an *autoregressive-moving average* (ARMA) process. The ARMA model is defined as:

$$Y_t = b_1Y_{t-1} + ... + b_pY_{t-p} + \epsilon_t + a_1\epsilon_{t-1} + ... + a_q\epsilon_{t-q},$$

where $\{\epsilon_t\} \sim \text{WN}(0, \sigma^2)$, while p, q are integers and (p, q) is viewed as the order of the model. We denote $\{Y_t\} \sim \text{ARMA}(p, q)$.

ARMA models are commonly used in the time series analysis because of their flexibility in estimating many stationary processes. But, they don't have advantages in nonlinear phenomena; see Fan and Yao (2003).

2.2.4 Non-linear Time Series

Based on what we have discussed before, if $\{\epsilon_t\}$ is a normal distribution, and conditional mean of Y_t given past Y's will be linear, the conditional variance of Y_t given old Y's will

be a constant and essentially, the process will be time-reversible (Lawrance, 1991). Many situations can be realized by that, however, there are more that fail to be fitted. It has been noticed that in many cases we need to fit a non-constant conditional variance, particularly, in the finance field, for instance, when econometricians describe the risk in the financial time series; see Bera and Higgins (1993).

Some researchers proposed extracting non-linear features by non-normal ARMA process, and these models are not reversible and have a non-linear conditional mean and variance (Breidt and Davis, 1991). But the drawback is that it is too restrictive to describe different case of interest.

Based on various thoughts, there are two common agreed objectives while dealing with the non-linear models for univariate series. The first and also the most important is to have some idea about the structure of the dataset that are being used. A common situation is when data shows non-linearity, we would like to characterize it by fitting an appropriate non-linear model. The second objective is to do optimal forecasts. But we need to be careful since we attempt to use a non-linear model to minimize forecasts errors for the given data but in a lot of cases there are only a limited number of observations available. Moreover, the danger of over-fitting also exists. See Tjostheim (1994)

There is no exact definition of non-linear model, people would like to define it into the following classes: (1) parametric models, (2) non-parametric models, (3) restricted non-parametric and semi-parametric models. In the parametric models class, it can be subdivided as parametric models for the conditional mean, parametric models for the conditional variance, mixed models which means a general model containing both a conditional mean and a conditional variance component and the regime models. In the restricted non-parametric and semi-parametric models, people developed additive models, projection pursuit, which are additive models of linear combinations of past values, and regression trees, splines and MARS (multivariate adaptive regression splines) methodology. Another way of trying to eliminate the difficulties in evaluating high-dimensional conditional quantities is to assume

non-linear and non-parametric dependence in some of the predictors and parametric and usually linear in others. These are called semi-parametric models; see Hárdle and Lütkepohl *et al* (1997).

2.3 GARCH Model

A huge amount of econometrics work use the least squares methods, which, assumes that the expected value of all error terms, when squared is the same at any given point. This is called *homoscedasticity* and however, such assumption will not always hold. We call it *heteroscedasticity* if the data's variances of the error terms are not constant. The coefficients from the regression for the ordinary least squares are still unbiased in that case, however, it will be too narrow to get the standard errors and confidence intervals by the conventional procedures. Therefore, ARCH and GARCH model are specially designed to treat heteroscedasticity as a variance to be modeled. Their applications in finance have been particularly successful, since the financial decisions are generally based upon the tradeoff between risk and return, and the econometric analysis of risk is therefore an integral part of asset pricing, portfolio optimization, option pricing and risk management. Many economic time series exhibit non-constant variance, thus, ARCH/GARCH are widely applied in the finance field.

The basic idea of ARCH model is that the conditional variance of the returns at time t, Y_t, depends on those returns before t. For example, if the conditional variance depends on various q time values of Y^2, then it is called ARCH(q) process and can be written as the following regression model:

$$Var(Y_t) = \sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \alpha_2 Y_{t-2}^2 + ... + \alpha_q Y_{t-q}^2.$$

In real life, we often use the variance as the effects of shocks on the variance of stock market returns, or the effects of increases in the variance of excess returns of bonds on risk premiums.

Furthermore, if we include lags of σ_t^2 to ARCH process, we will get:

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

This is widely known as the generalized ARCH process, i.e. GARCH$(1,1)$; see Bollerslev (1986).

We can extend GARCH(1,1) to GARCH(p,q) through the following:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2, \tag{2.1}$$

where p refers to the lag on σ_t^2 and q to the lag on Y_t^2. In the Equation 2.1 , $p = 1$, and $q = 1$ is known as the GARCH(1,1). As discussed in Engle (2001), GARCH(p,q) also has interpretation for its lag terms. The higher-order models are often useful when it includes a long span of data, like several decades of daily data or a year of hourly data. Such models allow both fast and slow decay of information with these lags.

In brief, ARCH/GARCH models don't consider information on the direction of returns, instead, they only care the magnitude. However, people have believed that the direction does affect volatility as well, especially for broad-based equity indices and bond market indices; see Poon and Granger (2003). It appears that market declines forecast higher volatility than comparable market increases do. This arises a variety of asymmetric GARCH models, including EGARCH, TGARCH, etc. These are beyond our topic in this thesis.

2.4 Non-parametric GARCH model

There are a lot of difficulties in modelling the nonparametric GARCH models. The following sections introduce several methods briefly to overcome some of these limits of the parametric assumptions in GARCH models and we refer to Linton (2006) for more details.

2.4.1 Error Density

The restrictiveness of the parametric assumptions in Gaussian strong GARCH models is to define the error density ϵ_t is standard normal and then maximizing the (conditional on initial values) Gaussian likelihood function. Under that assumption, the resulting estimators are supposed to be consistent and asymptotically normal. However, the error terms don't

actually need to be normal or i.i.d, but the resulting estimator will not be efficient without this condition.

Evidence so that the standardized residuals from estimated GARCH models are not normally distributed, especially for high frequency financial time series. Thus, study of semi-parametric models in which ϵ_t is i.i.d. with some density f that may be non-normal is developed:

$$\begin{aligned} Y_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \epsilon_{t-1}^2, \end{aligned}$$

where ϵ_t is i.i.d. with density f as unknown function. This provides the standardized residuals $\epsilon_t = Y_t/\sigma_t$ is non-Gaussian.

We can estimate f non-parametrically to improve the efficiency of the parameter estimation. Kernel based estimates is therefore proposed and the semi-parametric efficiency bounds for parameters are estimated.

2.4.2 Functional Form of Volatility Function

There are other ways to question the form of the volatility function. We define the news impact curve as the relationship between σ_t^2 and $y_{t-1} = y$, with past values σ_t^2 constant at some level. New impact curve is important in the finance study since it indicates how the volatility is affected by the news information.

So far, there are different ways to model the news impact curve and provide answers to different practitioners. This is anther motivation of modeling the non-parametric ARCH model, which has greater flexibility to various data. Some of them consider the case where $\sigma_t^2 = \sigma^2(y_{t-1})$, where $\sigma(\cdot)$ is a smooth but unknown function, like $\sigma_t^2 = \sigma^2(y_{t-1}, y_{t-2}, ..., y_{t-d})$. It is necessary to include many lagged variables in $\sigma^2(\cdot)$ to match the data, however, it also brings out the well-known "curse of dimensionality". In addition, it is also hard to interpret the dimension more than two.

Additive models offer a more flexible but parsimonious alternative to nonparametric models. Suppose that

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{d} \sigma^2(Y_{t-j}),$$

where σ_t^2 is some unknown function, which are allowed to be of general functional form but only depend on Y_{t-j}. These $\sigma^2(Y_{t-j})$ can be estimated by special kernel regression techniques.

2.4.3 Mean and Variance

Some of the views focus on the relationship between risk and return. We consider:

$$y_t = g(\sigma_t^2; \beta) + \sigma_t \epsilon_t \tag{2.2}$$

for linear or log-linear functional forms of g, and β in the Equation 2.2 are parameters to be estimated with the parameters of the error variances. This has been applied to stock index return data and the corresponding estimated g function is non-monotonic.

2.4.4 Long Memory

It has been found that traditional models adopt a dependent structure that doesn't fit the data well enough. The GARCH model(1,1) with the form:

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j Y_{t-j}^2$$

for the constants α_j satisfying $\alpha_j = \gamma\beta^{j-1}$. These coefficients decay dramatically but some empirical evidences show that the autocorrelation function function of Y_t^2 for high frequency returns data that suggests a slower decay rate than these coefficients in the model. Therefore, a single parameter called d which determines the memory properties of the series is proposed:

$$(1-L)^d \sigma_t^2 = \alpha_0 + \beta\sigma_{t-1}^2(\epsilon_{t-1}^2 - 1),$$

where $(1-L)^d$ is known as the fractional differencing operator. When $d = 1$, we have the IGARCH model.

2.5 B-spline Smoothing Method

The word "spline" has been originated from the ship building industry, which was a thin strip of wood which draftsmen would use like a flexible French curve. Metal weights were placed on the drawing surface and the spline was threaded between the ducks. Thus, a spline curve in industry is referred to a sequence of curve segments connected to each other to form a single continuous curve. And Basic-Spline, as known as B-Spline, is spline functions that has minimal support with respect to a give degree, smoothness, and domain partition.

A curve $s(u)$ is called a *spline of degree* n with the *knots* $a_0, ...a_m$, where $a_i \leqslant a_{i+1}$ and $a_i < a_{i+n+1}$ for all possible i. and $s(u)$ is $n-r$ times differentiable at any $r-fold\ knot$. (We call a knot a_{i+1} the *r-fold knot* if $a_i < a_{i+1}... = a_{i+r} < a_{i+r+1}$). It is also common to refer to a spline of degree n as a spline of order $n+1$.

2.5.1 B-spline knots and basis functions

In order to define B-splines, let $(a_j)_{j=-1}^{N}$ be a series of knots on interval $[a, b]$, which satisfies

$$a < a_1 < ... < a_j < a_{j+1}... < a_N < b.$$

Suppose we have N interior knots and the B-spline is of order d (or degree $d-1$), we can set the following boundary knots

$$a_{-(d-1)} = ... = a_{-1} = a_0 = a < a_1 < ... < a_N < b = a_{N+1} = ... = a_{N+d},$$

thus we have $N+d$ knots in total. With the above knots, we define B-spline of order d (or degree $d-1$), $B_{j,d}$, using the following recursion formula:

$$B_{j,1}(x) = \begin{cases} 1 & \text{if } x \in [a_j, a_{j+1}); \\ 0 & \text{otherwise.} \end{cases}$$

and

$$B_{j,d}(x) = \frac{(x-a_j)B_{j,d-1}(x)}{a_{j+d-1}-a_j} + \frac{(a_{j+d}-x)B_{j+1,d-1}(x)}{a_{j+d}-a_{j+1}}.$$

For example, suppose we have 5 equally spaced interior knots on the interval $[0, 6]$, then we have the following constant spline basis functions:

$$
\begin{aligned}
B_{0,1}(x) &= \begin{cases} 1 & \text{if } 0 \le x < 1; \\ 0 & \text{otherwise.} \end{cases} \\
B_{1,1}(x) &= \begin{cases} 1 & \text{if } 1 \le x < 2; \\ 0 & \text{otherwise.} \end{cases} \\
&\vdots \\
B_{5,1}(x) &= \begin{cases} 1 & \text{if } 5 \le x < 6; \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

Now we can deduce the linear basis functions $B_{j,2}(x)$ as:

$$
\begin{aligned}
B_{0,2}(x) &= (1-x)B_{1,1}(x) \\
B_{1,2}(x) &= xB_{1,1}(x) + (2-x)B_{2,1}(x) \\
B_{2,2}(x) &= (x-1)B_{2,1}(x) + (3-x)B_{3,1}(x) \\
B_{3,2}(x) &= (x-2)B_{3,1}(x) + (4-x)B_{4,1}(x) \\
B_{4,2}(x) &= (x-3)B_{4,1}(x) + (5-x)B_{5,1}(x) \\
B_{5,2}(x) &= (x-4)B_{5,1}(x)
\end{aligned}
$$

We can keep doing this until the desired degree. Usually, degree of 3 (cubic spline) is often used in most cases, since low degree may cause unsmooth curve while too high degree may result in overfitting.

Next we introduce the definition of B-spline space. Let $S_n^{(d)}$ be the space of B-splines on $[a, b]$ of order $d \ge 1$. The space $S_n^{(d)}$ consists of all the functions s satisfying

- s is a polynomial of degree $d-1$ which we fix on each of the subintervals $[a_j, a_{j+1})$, where $k = 0, ..., N-1$.
- s is $d-2$ continuously differentiable on the interval $[a, b]$, for $d \ge 1$.

Chapter 3

Nonparametric Methods For Financial Volatility

3.1 Functional Gradient Descent algorithm

Audrino and Bühlmann (2009) proposed a flexible model that is based on a high dimensional parameterization from a B-spline basis expansion. Below is a brief introduction of the model and the algorithm.

3.1.1 Model

Let us denote returns in its logarithm form:

$$Y_t = \log(P_t) - \log(P_{t-1}) \approx (P_t - P_{t-1})/P_{t-1}, \tag{3.1}$$

where the P_t is the price. In the following, we assume:

$$\begin{aligned} Y_t &= \mu_t + \sigma_t \epsilon_t, \\ \sigma_t^2 &= f(X_{t-1}, \sigma_{t-1}^2). \end{aligned}$$

and we also assume that

$$\mu_t = \alpha_0 + \alpha_1 Y_{t-1}$$

follows a simple auto-regressive AR(1) model. We consider the squared volatility, i.e. variance, as a function of Y_{t-1}, and σ_{t-1}^2. Thus, we construct a nonparametric GARCH(1,1) model.

$$\sigma_t^2 = f(Y_{t-1}, \sigma_{t-1}^2),$$

while the unknown function $f(\cdot, \cdot)$ can be non-linear or even not smooth. Estimating the unknown function $f(\cdot, \cdot)$ is our goal. The common non-parametric techniques have the advantages that include generality which is often discounted by decreased or non-improved average

prediction performance. However, non-parametric methods show poor performance at edges that are of major interest in practical applications. Strong sensitivity of choosing smoothing parameters is another difficulty.

Audrino and Bühlmann (2009) studied volatility σ_t^2 as a series of additive components of simple bivariate B-spline basis functions on a predictor space $\Re \times \Re^+$ arising from the lagged values $(Y_{t-1}, \sigma_{t-1}^2)$. The reason that we use log-transform on volatility is that we can avoid positivity restrictions and then freely adopt a convex loss function. In detail, Audrino and Bühlmann (2009) model:

$$\begin{aligned} \log\{\sigma_t^2(\theta)\} &= \log\{f_\theta(Y_{t-1}, \sigma_{t-1}^2(\theta))\} \\ &= g_{\theta_0}\{Y_{t-1}, \sigma_{t-1}^2(\theta)\} + \sum_{j_1=1}^{k_1}\sum_{j_2=1}^{k_2} \beta_{j_1,j_2}\{Y_{t-1}, \sigma_{t-1}^2(\theta)\} \end{aligned}$$

where $g_{\theta_0}(\cdot,\cdot)$ is the initial function as a starting point, and is proposed as a logarithm of a parametric GARCH(1,1) process. The parameter vector θ is composed by $\{\theta_0, \beta_{j_1,j_2}, j_1 = 1, ..., k_1, j_2 = 1, ..., k_2\}$ and β_{j_1,j_2} will be discussed in short. Our main idea is to estimate the second term $\sum_{j_1=1}^{k_1}\sum_{j2=1}^{k_2}\beta_{j_1,j_2}\{Y_{t-1}, \sigma_{t-1}^2(\theta)\}$ by constructing the bivariate B-spline basis functions $\beta_{j_1,j_2}(\cdot,\cdot)$ in order to improve $g_{\theta_0}\{Y_{t-1}, \sigma_{t-1}^2(\theta)\}$, which is our initial point. We can state the multivariate B-splines as products of univariate B-splines and estimate them in an easy way, which means:

$$B_{j_1,j_2}\{Y_{t-1}, \sigma_{t-1}^2(\theta)\} = B_{j_1}(Y_{t-1})B_{j_2}\{\sigma_{t-1}^2(\theta)\}.$$

where B_{j_1} is the $j_1 th$ B-spline function of value Y_{t-1} and B_{j_2} is the $j_2 th$ B-spline function of value $\sigma_{t-1}^2(\theta)$. Both of them represent piecewise polynomial functions and can be approximated a general continuous, non-parametric conditional variance function. Different degrees of B-spline can lead to flexible shape of the conditional variance function. In our simulation, we choose the degree of the $B_{j_1}(X_{t-1})$ as 3 and the degree of the $B_{j_2}(\sigma_{t-1}^2)$ as 2, following the suggestion of Audrino and Bühlmann (2009). For the number of breaks, we choose the empirical α-quantiles of the corresponding predictor variables with $\alpha = i/\text{mesh}$, $i = 1, ...,$ mesh $- 1$, and mesh $\in N$.

Suppose ϵ_t following the standard normally distribution, then the negative log-likelihood in the model is shown as:

$$
\begin{aligned}
-\log\{L(\alpha,\theta;Y_2^T)\} &= \sum_{t=1}^{T}\frac{1}{2}\left[\log(2\pi)+\log\{\sigma_t^2(\theta)\}+\frac{\{Y_t-\mu_t(\alpha)\}^2}{\sigma_t^2(\theta)}\right] \\
&= \sum_{t=1}^{T}\frac{1}{2}\left\{\log(2\pi)+g_\theta\{Y_{t-1},\sigma_{t-1}^2(\theta)\}+\frac{\{Y_t-\mu_t(\alpha)\}^2}{\exp\left[g_\theta\{Y_{t-1},\sigma_{t-1}^2(\theta)\}\right]}\right\},
\end{aligned}
$$

where $g_\theta\{Y_{t-1},\sigma_{t-1}^2(\theta)\}=\log\{\sigma_t^2(\theta)\}$ as our initial starting point. We estimate the model by using the *functional gradient descent* (FGD for short) algorithm, which includes the following three components:

- loss function
- its partial derivative
- an initial starting estimate

The loss function can be described as:

$$
\lambda(y,g)=\frac{1}{2}\left\{\log(2\pi)+g+\frac{y^2}{\exp(g)}\right\},
$$

where $y=Y-\mu$, the centered mean. Partial derivative is needed with respect to the g in order to maximize the negative log-likelihood function:

$$
\frac{\partial\lambda(y,g)}{\partial g}=\frac{1}{2}\left\{1-\frac{y^2}{\exp(g)}\right\}.
$$

The use of a componentwise least squares method is proposed to fit one B-spline basis function at a time and get the maximum log-likelihood estimation from our initial starting point, the GARCH(1,1) model.

We describe the *co-ordinatewise gradient descent* algorithm in Audrino and Bühlmann (2009) in details:

- Step 1 (*initialization*): As stated before, we choose GARCH(1,1) model as $g_{\theta_0}\{Y_{t-1}, \sigma^2_{t-1}(\theta)\}$ to be the starting point. And $\hat{\alpha}$ and $\hat{\theta}$ are estimated in Equation 3.2, 3.3

$$\begin{aligned} \hat{\mu}(t) &= \hat{\alpha}_1 + \hat{\alpha}_2 Y_{t-1}, & (3.2) \\ \exp\{\hat{g}_0(t)\} &= \hat{\theta}_{0,1} + \hat{\theta}_{0,2} Y^2_{t-1} + \hat{\theta}_{0,3} \exp\{\hat{g}_0(t-1)\}. & (3.3) \end{aligned}$$

- Step 2 (*projection of the gradient to the B-splines*): The negative gradient vector is:

$$U_t = -\frac{1}{2}\left[1 - \frac{(Y_t - \hat{\mu}_t)^2}{\exp\{\hat{g}_{m-1}(t)\}}\right], \quad t = 2, ..., T.$$

  where $\exp\{\hat{g}_{m-1}(t)\}$ is the updated part of logarithm volatility. We use the B-spline method to get the bivariate basis functions and regress U_t onto the spline basis function $B_d[X_{t-1}, \exp\{\hat{g}_{m-1}(t-1)\}], (t = 2, ...T)$ in order to get the least squares among the basis functions.

$$\hat{S}_m = \text{argmin}_{1 \le d \le k} \sum_{t=2}^{T} \left\{ U_t - \hat{\beta}_d B_d[Y_{t-1}, \exp\{\hat{g}_{m-1}(t-1)\}] \right\}^2.$$

  where $d = (d_1, d_2)$ as the bivariate basis index which has the least square among k_1, k_2. $\hat{\beta}_d$ is the least squares estimated coefficient. We also use $k = (k_1, k_2)$ as the bivariate order of the B-splines, while k_1 is the number of univariate B-spline basis functions for X_{t-1} and k_2 for σ^2_{t-1}. As mentioned before, $k_1 = (mesh-1)+3$ and $k_2 = (mesh-1)+2$.

- Step 3 (*line search*): Every time $\hat{g}_{m-1}$ is updated, we need to seek an optimization for the step length and the criteria is:

$$\hat{\beta}_{\hat{S}_m} = \text{argmin}_\omega \sum_{t=2}^{T} \lambda(Y_t - \hat{\mu}_t, \hat{g}_{m-1}(t) + \omega B_{\hat{S}_m}[Y_{t-1}, \exp\{\hat{g}_{m-1}(t-1)\}]).$$

  Hence, we update our $\hat{g}$ function every time by doing:

$$\hat{g}_m(t) = \hat{g}_0(t) + \sum_{m=1}^{M} \hat{\beta}_{\hat{S}_m} B_{\hat{S}_m}[Y_{t-1}, \exp\{\hat{g}_{m-1}(t-1)\}].$$

- Step 4 (*iteration and stopping*): We do loops for steps 2 and 3. We adopt the cross-validation method and the stopping point (when $m = M$) is chosen in the training set

when the testing set gets optimal. Thus, overfitting is avoided. In the end, we get our final estimation of the $\hat{g}$ function:

$$\hat{g}_M(t) = \hat{g}_0(t) + \sum_{m=1}^{M} \hat{\beta}_{\hat{S}_m}[Y_{t-1}, \exp\{\hat{g}_{m-1}(t-1)\}].$$

Shrinkage is also introduced in step 3 in order to reduce the variance of the estimated B-spline components. We update $\hat{\beta}_{\hat{S}_m} B_{\hat{S}_m}$ by:

$$\kappa \hat{\beta}_{\hat{S}_m} B_{\hat{S}_m}, \quad \text{with } 0 < \kappa \leq 1.$$

Empirically, the value is determined within $\{0.1, 0.2\}$.

3.2 Semi-parametric aGARCH modeling

As mentioned in the last chapter, flexible non-parametric GARCH models very often outperform the standard GARCH models when applied to real data with many lagged variables. But it also presents great challenges in smoothing high dimensional and strongly correlated time series data.

Therefore, additive models are proposed to overcome these difficulties. Carroll (2002) and Yang (2002) introduced a truncated version of the nonparametric GARCH model with a finite number of lags J:

$$\sigma_t^2 = \sum_{j=1}^{J} \beta_0^{j-1} m_0(Y_{t-j}), \quad \beta \in [\beta_1, \beta_2]$$

There are various ways are used to estimate m_0 and β_0, and existing methods include marginal integration kernel smoothing, backfitting algorithm, etc. Wang, *et al.* (2011) developed a semi-parametric method with a well-justified theory and a fast algorithm to implement the method. They used the polynomial splines to estimate the additive model.

Consider a stationary time series $\{Y_t\}_{t=1}^{T}$, with

$$Y_t = \sigma_t \epsilon_t, \quad t = 1, 2, ..., T,$$

which satisfies:

$$\begin{aligned} Y_t^2 &= c + \sum_{j=1}^{J} m_j(Y_{t-j}) + \varepsilon_t, \\ \varepsilon_t &= \sigma_t^2(\epsilon_t^2 - 1), \end{aligned}$$

where $m_j(y) = \beta_0^{j-1} m_1(y)$.

Wang, *et al.* (2011) introduce a least square risk function, denote it as $R(\beta)$ as:

$$\begin{aligned} R(\beta) &= E[\sum_{j=1}^{j} \{m_j(Y_t) - \beta^{j-1} m_1(Y_t)\}^2] \\ &= \sum_{j=1}^{J} \{(\beta_0^{j-1} - \beta^{j-1})^2\} E\{m_1(Y_t)^2\}. \end{aligned}$$

Their goal is to estimate the m_1 function and the coefficient parameter β_0 by using the polynomial spline smoothing.

We consider the estimation of m_j based on all bounded measurable function on compact interval $[a, b]$, where a, b are fixed constants. We define $h \equiv (b-a)/(N+1)$ as the spacing between knots next to each other, and space $S_n^{(2)} = S_n^{(2)}[a, b]$ as the linear space spanned by $\{1, b_{j,k}, j = 1, ..., J, k = 1-p, ..., N\}$. We still use the least square method to find out the coefficients in front of each splines in order that:

$$(\hat{\lambda}_0', \hat{\lambda}_{1,1-p}', ..., \hat{\lambda}_{J,N}') = \text{argmin}_{R^{1+J(N+p)}} \sum_{t=J+1}^{T} \{Y_t^2 - \lambda_0 - \sum_{j=1}^{J} \sum_{k=1-p}^{N} \lambda_{j,k} b_{j,k}(Y_{t-j})\}^2.$$

Let $\hat{c} = n^{n-1} \sum_{t=J+1}^{T} Y_t^2$, in which $n = T - J$, be a $\sqrt{n}-$ consistent estimator of c by the Central Limit Theorem. Hence, the centered spline estimator of each component function is written as:

$$\hat{m}_j(y) = \sum_{k=1-p}^{N} \hat{\lambda}_{j,k} b_{j,k}(y) - \frac{1}{n} \sum_{t=J+1}^{T} \sum_{k=1-p}^{N} \hat{\lambda}_{j-k} b_{j,k}(Y_{t-j})$$

for $1 \leq j \leq J$. Next, in order to get the estimated coefficient β_0, we simply regress $\{\hat{m}_2(Y_t)\}_{t=J+1}^{T}$ on $\{\hat{m}_1(Y_t)\}_{t=J+1}^{T}$ and try to minimize the term:

$$\sum_{t=J+1}^{T} \{\hat{m}_2(Y_t) - \beta \hat{m}_1(Y_t)\}^2.$$

Overall, we can improve the performance by averaging over all the components by doing:

$$\hat{R}(\beta) = \frac{1}{n} \sum_{t=J+1}^{T} \sum_{j=1}^{J} \{\hat{m}_j(Y_t) - \beta^{j-1} \hat{m}_1(Y_t)\}^2$$

where $\hat{\beta} = \text{argmin}_{\beta \in [\beta_1, \beta_2]} \hat{R}(\beta)$.

Chapter 4

Numerical Examples

In this chapter, we provide some numerical examples to illustrate the behavior of the nonparametric methods, FGD and aGARCH, described in Chapter 3. We use R to realize all of the results and compare performance from both methods with a simple parametric GARCH(1,1) fit.

In all simulations below, we generate 3000 observations for each process. We discard the first 1000 observations which can be treated as warm-up period and use the second 1000 simulated data as in-sample period to estimate the model and the third 1000 data as out-sample period to test the model. In order to get fair results, we repeat simulations for 100 independent runs and compare their error means and standard deviations.

We quantify the goodness of fit as in-sample error and out-sample error, which are illustrated as:

$$\begin{aligned} \text{IS-L}_p &= \frac{1}{T}\sum_{t=1}^{T} |\sigma_t^2 - \hat{\sigma}_t^2|^p, \quad \text{for } p = 1, 2 \text{ (in-sample error)}; \\ \text{OS-L}_p &= \frac{1}{T}\sum_{t=T+1}^{2T} |\sigma_t^2 - \hat{\sigma}_t^2(X_{T+1}^{2T})|^p, \quad \text{for } p = 1, 2 \text{ (out-sample error)}. \end{aligned}$$

When $p = 1$, it is known as the mean absolute errors, and $p = 2$ is the squared volatility error.

In Sections 4.1, 4.2 and 4.3, we compare the nonparametric GARCH models, with standard GARCH(1,1) models using four different sets of simulations. The first and second datasets are generated using the standard GARCH(1,1) process. The third is generated using an asymmetric GARCH model, called the GJR model (Glosten, Jaganathan, and Runkle,

1993), and the fourth one is generated with more complicated models. We will continue to apply three models in the real financial data to see their performance in different situations.

4.1 SIMPLE GARCH(1,1) DATA EXAMPLE

We generate a time series as the following:

$$\begin{aligned} Y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2 \end{aligned} \tag{4.1}$$

where ϵ_t follows the standard normal distribution, and $\alpha_0 = 0.05$, $\alpha_1 = 0.20$ and $\beta_0 = 0.75$ for dataset A and $\alpha_0 = 0.20$, $\alpha_1 = 0.30$ and $\beta_0 = 0.45$ for model B. The news impact curve in the above model, $m(y) = \alpha_0 + \alpha_1 y^2$, is symmetric.

We generate the above two processes and compare three models: GARCH(1,1), aGARCH and FGD. The results are provided in Tables 4.1 and 4.1, respectively. The first value in each cell is the mean of the errors based on 100 runs and the second value is the corresponding standard deviation. The last column provides the computing time for each method.

Table 4.1: Performance results for dataset A

	IS-1	IS-2	OS-1	OS-2	CPU time
GARCH(1,1)	0.3226±0.1681	0.7749±1.1507	0.3487±0.3357	0.7749±1.1508	<5min
aGARCH	0.3493±0.130	1.0647±1.3990	0.5491±0.3540	1.0681±1.1919	<5min
FGD	0.3934±0.2010	0.8972±0.7645	0.4755±0.4139	0.9294±0.8768	almost 2 hours

Table 4.2: Performance results for dataset B

	IS-1	IS-2	OS-1	OS-2	CPU TIME
GARCH(1,1)	0.204±0.0279	0.329±0.309	0.2050±0.046	0.3049±0.4146	<5min
aGARCH	0.2157±0.026	0.337±0.3229	0.3229±0.0414	0.4021±0.1351	<5min
FGD	0.2274±0.034	0.4021±0.2908	0.2425±0.073	0.5219±0.2376	almost 2 hours

Since the true underlying model is the standard GARCH(1,1) model, clearly, simple GARCH(1,1) will be good enough to fit these data. Tables 4.1 and 4.2 show that GARCH(1,1) outperforms the two non-parametric models in terms of both volatility estimation and prediction. We can see both in-sample L_1 and L_2 error of GARCH(1,1) is smaller than those of FGD and aGARCH, hence non-parametric models may not be the perfect choice. But the non-parametric methods aGARCH and FGD perform quite close to GARCH(1,1) model, which shows the validity of all three models. Out-of sample error measures the predictability of model. Similarly, aGARCH and FGD model present similar results as classical GARCH(1,1) model does.

4.2 GJR data example

In this section we generate dataset C based on the following GJR model:

$$\begin{aligned} Y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \delta_0 Y_{t-1}^2 I(Y_{t-1} < 0) + \beta_0 \sigma_{t-1}^2 \end{aligned} \tag{4.2}$$

where ϵ_t is standard normal, with $\alpha_0 = 0.2$, $\alpha_1 = 0.06$, $\delta_0 = 0.03$, and $\beta_0 = 0.90$. The news impact curves, $m(y) = \alpha_0 + \alpha_1 y^2 + \delta_0 y^2 I(y < 0)$, is asymmetry in terms of y, and we regard it as an GJR model. Hence, the standard GARCH(1,1) model can not catch the character of data.

Table 4.3: Performance results for dataset C

	IS-1	IS-2	OS-1	OS-2	CPU time
GARCH(1,1)	0.6573±0.1132	0.7526±0.2204	0.7351 ± 0.1590	0.7834 ± 0.6381	<5min
aGARCH	0.6766±0.1056	0.7722±0.2197	0.6540±0.2183	0.6284± 0.4696	< 5min
FGD	0.6858±0.2036	0.8115±0.2197	0.7754±0.3163	0.7974± 0.4026	almost 2 hours

Table 4.3 shows the performance of the three models. Not surprisingly, since the underlying news impact curve is asymmetric in the above model, GARCH(1,1) no longer shows

its advantages. FGD model also fails to do well compared to GARCH(1,1) indicating that non-parametric method may not appropriate in the simple data structure. aGARCH is the best of the three, due to the reason that it has the nonparametric part for Y_t while the parametric for σ_t, which is cater to the data above.

4.3 More complicated data example

We generate the processes using a more complicated non-parametric GARCH model and consider the following squared volatility function

$$\sigma_t^2 = f(Y_{t-1}, \sigma_{t-1}^2),$$

where

$$f(y, \sigma^2) = \begin{cases} 0.12 + 0.3\sigma^2 + \{1 - 0.3L - (1 - 10^{-6}L)(1 - L)^d\}y^2, & \text{if } y \leq 0; \\ (0.4 + 0.28Y_{t-1}^3)\exp(-0.15Y_{t-2}^2), & \text{if } y > 0. \end{cases}$$

In the above equations, L stands for the backshift operator and the expression $(1-L)^d$ is the binomial expansion and usually it is stated as the hypergeometric function; see Baillie, *et al.* (1996). In this simulation, we fix $d = 0.4$. Thus, the whole dataset will be a non-parametric GARCH process with long memory and asymmetric effects in volatility. We show our results in Table 4.4.

Table 4.4: Performance results for dataset D

	IS-1	IS-2	OS-1	OS-2	CPU time
GARCH(1,1)	0.2255±0.0561	0.3665±0.358	0.2197±0.0751	0.3041±0.3601	1.5 hours
aGARCH	0.2059±0.049	0.3442±0.2443	02121±0.092	0.2492±0.301	1.5 hours
FGD	0.1802±0.042	0.1912±0.208	0.1846±0.039	0.1704±0.176	5∼6 hours

This time, both aGARCH and FGD perform better than GARCH(1,1) model since the original simulated data is more complicated and can not be simply explained by GARCH(1,1)

model. FGD shows better results than aGARCH, based on the fact that the latter is semi-parametric GARCH model and might not do well on fitting the non-parametric part on σ^2.

4.4 Real Data

4.4.1 S&P 500 weekly return

Standard & Poor 500, referred as S&P 500 is the free-float capitalization-weighted index focusing on 500 U.S. companies, which have large-cap common stocks. It is viewed as the important signal of U.S. economy, thus, studying its volatility is many investors' interest.

We collect 1000 weekly S&P 500 index from February 20, 1973 up to May 18, 1992 and consider their log-returns in percentages. We truncate Y_t by 0.005 and 0.995 quantiles during the fitting. Table 4.5 shows the plot of both S&P 500 index and S&P 500 return. Table 4.6 is the data description of S&P 500 dataset. We define the definition of return as in Equation 3.1:

Table 4.5: Data plot about the weekly S&P500 index and S&P500 return

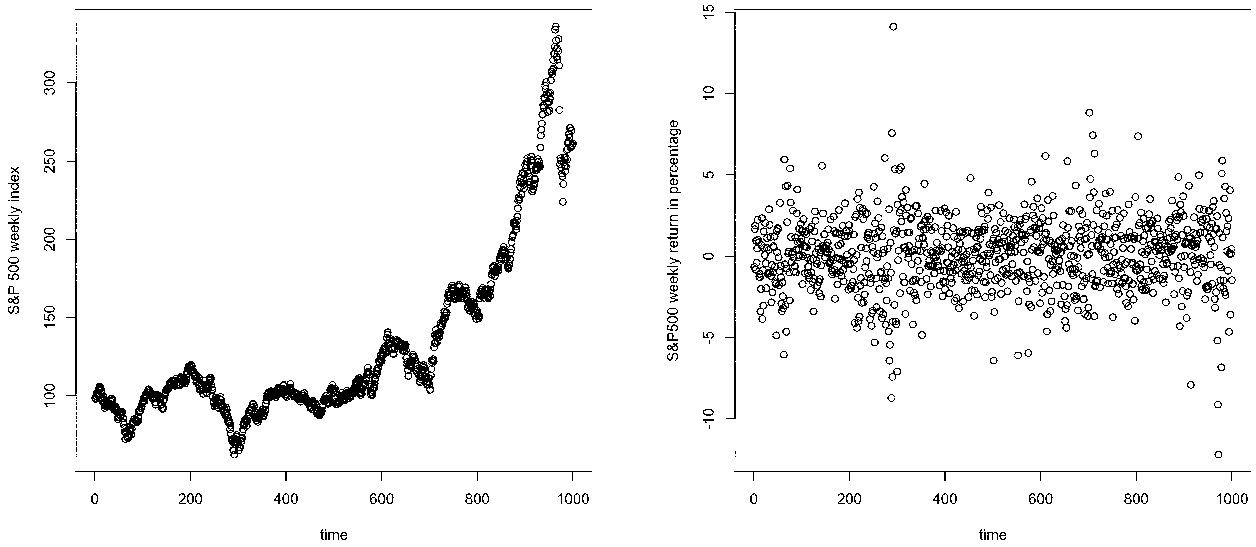

Table 4.6: Data summary of S&P 500 weekly log-return in percentage

	Mean	Standard deviation	Maximum	Minimum
S&P500 weekly return	0.121	2.234	14.116	-12.197

Note that we cannot calculate the true volatility based on the real dataset, hence we choose these two criteria to judge our results:

$$\text{Prediction Error} = \frac{1}{n}\sum_{t=J+1}^{T}(\hat{\sigma}_t^2 - Y_t^2)^2 \tag{4.3}$$

$$-\text{Log-likelihood} = -\sum_{t=J+1}^{T} log\left\{\hat{\sigma}_t\phi(\frac{Y_t}{\hat{\sigma}_t})\right\} \tag{4.4}$$

where J is the optimal lag chosen from the aGARCH model.

Table 4.7: Fitting S&P 500 weekly returns

	−Log-likelihood	Volatility prediction error
GARCH(1,1)	2139.646	44.7886
aGARCH	2120.896	42.7434
FGD	2085.704	42.0431

4.4.2 Apple stock returns

We also investigate on some famous US stocks to compare three models. We select Apple company's daily stock from August, 1991 to August, 1995, which are 1000 observations in all. Again, we plot the Apple daily stock index and its return in Table 4.8 and present the data summary of the Apple stock in Table 4.9:

Table 4.8: Data plot about the weekly S&P500 index and S&P500 return

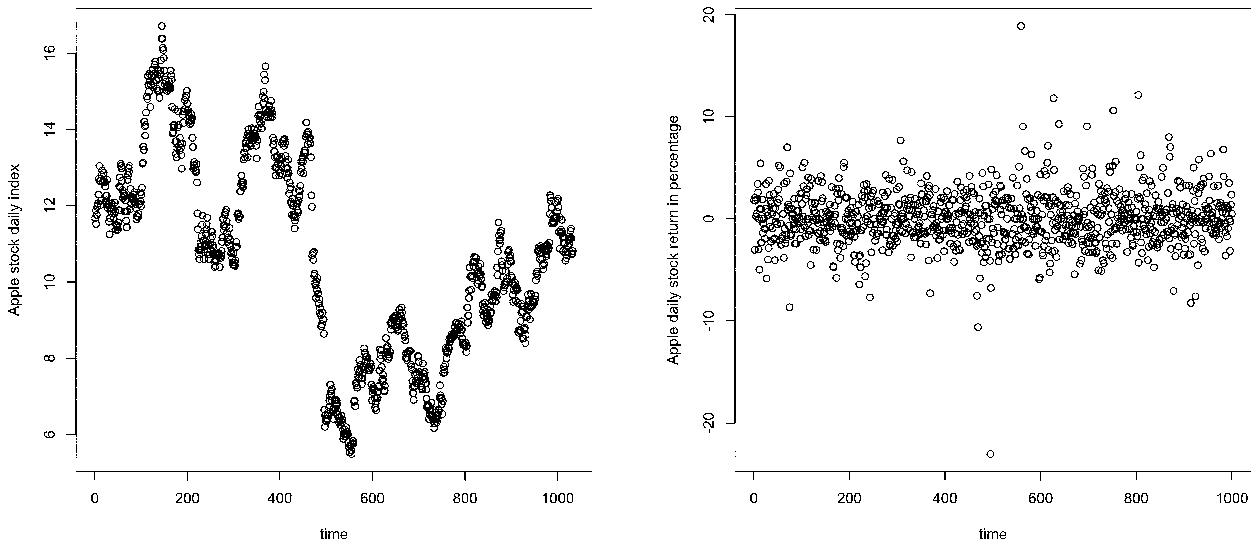

Table 4.9: Data summary of Apple daily log-return in percentage

	Mean	Standard deviation	Maximum	Minimum
Apple Daily Return	0.041	2.696	18.891	-23.032

Table 4.10: Fitting Apple stock daily returns

	−Log-likelihood	Volatility prediction error
GARCH(1,1)	2365.124	118.9165
aGARCH	2302.773	115.2463
FGD	2145.569	107.8452

Table 4.10 provide the $-$log-likelihood and volatility prediction error from three models. Both aGARCH and FGD model show significant improvement compared to GARCH(1,1) model. Note that in the previous simulated data, nonparametric models may not outperform GARCH model sometimes if the processes are generated using the parametric volatility models. In real life, most of the time series data is non-parametric and our two methods then show up their advantages. FGD is even a little better than aGARCH in some situations which may be the reason that aGARCH is semi-parametric model while if coefficient in front of σ^2 is not constant, it will not capture its character as well as FGD model does.

4.5 Remarks on the non-parametric mothods

The above results demonstrate that both aGARCH and FGD model is a good choice when fitting the non-parametric financial volatility. FGD is even more flexible when it comes to estimating a high dimension of unknown parameters. However, it is also stated in the tables that the computing time for FGD method is extremely longer than both GARCH and aGARCH model due to the fact that it needs to find an optimal structure in a high dimension every time and update the loss function by a fix length. It needs to repeat it until find the maximum likelihood. Therefore, computing efficiency will be a big problem for FGD method and how to optimize the algorithm will be further study's interest.

Chapter 5

Conclusion

In this thesis, we compare two non-parametric GARCH models with the simple GARCH(1,1) model for financial time series. The FGD method constructs an optimal tensor B-spline structure to get the maximum likelihood in a high dimension context. The aGARCH method uses polynomial spline smoothing on the news impact curve to obtain a nonparametric part in GARCH(1,1) model. Both non-parametric methods are more flexible and powerful for analyzing highly persistent financial time series data. We compare their performance with the GARCH(1,1) model by first using the GARCH(1,1) simulated data. Results show both non-parametric models do not outperfrom the simple GARCH(1,1) model, since the data are generated from the GARCH(1,1). Although the standard GARCH model is sufficient in that case, our results still present the validity of using both non-parametric models. We then applied three methods to datasets generated from a GJR model. Our results demonstrate the advantages of the non-parametric GARCH methods over the simple GARCH(1,1) model in these cases. aGARCH is a relatively simple semi-parametric GARCH model, and its in-sample error is close to that of the GARCH(1,1) but it predicts better in terms of out-sample errors. FGD method does not show any advantages among the three simply because the non-parametric method might not be the best choice for this dataset. Our last simulated data are generated from a complicated GARCH model. Both FGD and aGARCH improve the prediction performance compared to the simple GARCH(1,1) model. Later, we apply them into the real financial time series data. We choose to fit the S&P 500 weekly return and Apple stock daily return. Results indicate noticeable improvement compared to the GARCH(1,1) model.

Bibliography

[1] Audrino, F. and Bühlmann, P. (2009). Splines for financial volatility. *Journal of the Royal Statistical Society: Series B Statistical Methodology*, **71**, 655-670.

[2] Balbás, A. and Romera R. *et al* (2003). An overview of probabilistic and time series models in finance. *Statistics and Econometrics Series*, **05**. e login page

[3] Baillie, R.T., Bollerslev, T. *et al* (1996). fractionally integrated generalized autoregressive conditional heteroskedasticity. *J.Econometr.*, **74**, 3-30.

[4] Bera, A.K. and Higgins, M.L. (1993). A survery of ARCH models: properties, estimation and testing. *J. Econom. Surverys*, **7**, 305-3665.

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

[6] Breidt, F.J. and Davis, R.A. (1991). Time-reversibility, identifiability and independence of innovations for stationary time series. Preprint, Department of Statistics, Colorado State University, Fort Collins.

[7] Bühlmann, P. and McNeil, A.J. (2002). An algorithm for nonparametric GARCH modelling. *Comput. Statist. Data Anal.* **40**, 665-683.

[8] Carroll, R., Härdle, W., and Mammen, E. (2002). Estimation in ad additive model when the components are linked parametrically. *Econ. Theory* **18**, 886-912.

[9] Charfield, C. (2009). *The analysis of time series: an introduction*, CRC Press.

[10] Cressie, N. and Wikle, C. (2011). *Statistics for spatio-temporal data*, Wiley

[11] Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, **4**, 987-1007.

[12] Engle, R.(2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *The hournal of Economic Perspectives*, **15**, 157-168. e login page

[13] Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*, Springe.r

[14] Glosten, L.R., Jaganathan,R., and Runkle, D.E. (1993). Onthe relation between the expected value and the volatility of the nomial excess return on stocks. *Journal of Finance*, **48**, 1779-1801.

[15] Härdel, W. and Lütkepohl, H. (1997). A review of nonparametric time series analysis. *Statistics and Econometrics Series*, **05**.

[16] Lawrance, A.J. (1991). Directionality and reversibility in time seres. *Int. Statist. Rev.* **59**, 67-79.

[17] Linton, O. (2011). Semiparametric andSatheesh and Matthew (1988) nonparametric ARCH modelling, *Journal of Probability and Statistics*, 1-17.

[18] Nason, G. (2008). *Wavelet methods in statistics with R*. Springer.

[19] Poon, S. and Granger, C. (2003). Forecasting volatility in financial markets: a review, *Journal of Economic Literature*, **XLI**, 478-539.

[20] Tjostheim, D. (1994). Non-linear time series: a selective review. *Scandinavian Journal of Statistics*, **21**, 97-130.

[21] Wang, L., Feng, C. Song, X. and Yang, L. (2011). Efficient semiparametric GARCH modeling of financial volatility. *Statistica Sinica*. Forthcoming.

[22] Yang, L. (2002). direct estimation in an additive model when the components are proportional. *Statist. Sinica*. **12**, 801-821.