ABSTRACT

Title of dissertation: Generalized Volatility Model And Calculating VaR Using A New Semiparametric Model

Haiming Guo, Doctor of Philosophy, 2005

Dissertation directed by: Professor Benjamin Kedem
Department of mathematics

The first part of the dissertation concerns financial volatility models. Financial volatility has some stylized facts, such as excess kurtosis, volatility clustering and leverage effects. A good volatility model should be able to capture all these stylized facts. Among the volatility models, ARCH, GARCH, EGARCH and stochastic volatility models are the most important. We propose a generalized volatility model or GVM in this part, which is a generalization of all the ARCH family and stochastic volatility models. The GVM adopts the structure of the generalized linear model (GLM). GLM was originally intended for independent data. However, using partial likelihood, GLM can be extended to time series, and can then be applied to predict financial volatility. Interestingly, the family of ARCH models are special cases of GVM. Also, any covariates can be added easily to a GVM model. As an example, we use GVM to predict the realized volatility. Because of the availability of high frequency data in today's market, we can calculate realized volatility directly. We compare the prediction results of GVM with that of other classical models. By the measure of mean square error, GVM is the best among these the models.

The second part of this dissertation is about value at risk (VaR). The most common methods to compute VaR are GARCH, historical simulation, and extreme value theory. A new semiparametric model based on density ratio is developed in Chapter three. By assuming that the density of the return series is an exponential function times the density of another reference return series, we can derive the density function of the portfolio's distribution. Then, we can compute the corresponding quantile or the VaR. We ran a monte carlo simulation to compare the semiparametric model and the traditional VaR models under many different scenarios. In several cases, the semiparametric model performs quite satisfactorily. Furthermore, when applied to real data, the semiparametric model performs best among all the considered models using the metric of failure rate.

Generalized Volatility Model And Calculating VaR Using A New Semiparametric Model

by

Haiming Guo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:

Professor Benjamin, Kedem, Chair/Advisor
Professor Frank, Alt
Professor Stephen, Mount
Professor Galit, Shumeli
Professor Wolfgang, Jank

This dissertation is dedicated to my parents and my beloved wife Yanwen!

ACKNOWLEDGMENTS

I would like to thank all the people whose help and support over my graduate studies has made this thesis possible and who have made my graduate experience unforgettable.

First and foremost, I would like to thank my thesis advisor, Dr. Benjamin Kedem, for his invaluable guidance over the past years. He has taught me a great deal of interesting mathematics especially time series. As an advisor, he gave me the opportunity on challenging research problems and he has been a constant source with brilliant ideas. I consider myself lucky to select him as my thesis advisor.

I am also deeply grateful to my beloved wife Yanwen. Without her consistent support, none of my work would have been possible. Thanks for her inspirations and invaluable encouragements over the past few years.

Finally, I owe my deepest thanks and gratitude to my parents, my sister and my grandmother; the gift of unbounded love and support has no equal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

An Overview of Volatility Models

1.1 Introduction

Volatility means variance conditional on underlying financial assets, such as exchange rates and stock returns. Volatility modeling has been a very active area of research in recent years when it became clear that volatility drives financial markets. A volatility model should be able to forecast volatility. The forecasted volatility of financial returns is routinely used as a simple measure of risk, and it enters directly into derivative pricing formulas such as the Black-Scholes formula. Furthermore, these forecasts are also used in risk management such as the calculation of value at risk, derivative pricing and hedging, portfolio selection and many other financial activities. There are a huge number of research papers on volatility models. In this chapter, we focus on univariate volatility models. In this section, we introduce model notation. In the second section, some stylized facts of financial volatility are described. A good volatility model should be able to capture all these characteristics of financial volatility. In section three, we review the family of ARCH models, and section four is an introduction to stochastic volatility models.

Next, we establish our notation. Let P_t be the asset price at time t. The

log-return is defined as:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln\left(1 + \frac{P_t - P_{t-1}}{P_{t-1}}\right) \tag{1.1}$$

The quantity $\frac{P_t - P_{t-1}}{P_{t-1}}$ is called *relative return*, if the relative return is small, by Taylor series expansion we can argue that:

$$r_t \approx \frac{P_t - P_{t-1}}{P_{t-1}}$$

Therefore, log-returns correspond approximately to percentage changes of a financial position, a fact convenient in data analysis. Log-returns are often preferred to relative returns because the multiperiod log-return is simply the sum of the involved one-period log-returns, a property relative returns do not share. In addition, log-returns have more attractive statistical properties. In the rest of the dissertation, when we talk about the return series we mean the log-return series. We define the conditional mean and conditional variance as:

$$m_t = E_{t-1}[r_t] \tag{1.2}$$

$$h_t = E_{t-1}[(r_t - m_t)^2] \tag{1.3}$$

where $E_{t-1}[u]$ is the conditional expectation of some variable u given past information $\mathcal{F}_{t-1}$. Without loss of generality this suggests that the return series r_t is generated according to the following process:

$$r_t = m_t + a_t \tag{1.4}$$

$$a_t = \sqrt{h_t}\epsilon_t \tag{1.5}$$

where $E_{t-1}[\epsilon_t] = 0$ and $V_{t-1}[\epsilon_t] = 1$, h_t and ϵ_t are independent. Hence, given $\mathcal{F}_{t-1}$, we can see that m_t is 'known' and the volatility of r_t is given in terms of the shock

term h_t. All the volatility models in this chapter are concerned with the modeling of a_t, or in other words, h_t and ϵ_t.

1.2 Stylized facts about volatility in financial time series

A special feature of financial volatility is that it is not directly observed but can be estimated. In this dissertation, we use high frequency intraday data to calculate approximate volatility. This enables us to measure the performance of different volatility models. We'll discuss the details in later chapters. Although volatility is not observed directly, some stylized facts about it are widely documented. These empirical findings stem from financial time series such as the exchange rate returns or stock returns. A reliable volatility model should be able to capture and reflect the following stylized facts.

i. Excess Kurtosis.

Kurtosis is defined as the standardized 4th moment of the distribution. When the unconditional distribution of financial time series is compared to the normal distribution, fatter tails are observed. The kurtosis of the standard normal distribution is 3, whereas for many financial time series, the typical kurtosis estimates range from 4 to 50, indicating extreme non-normality (Mandelbrot (1963) [34], Fama (1963,1965) [16] [17], Engle and Pattern (2001) [15]).

ii. Volatility Clustering.

Volatility clustering refers to the fact that large changes of financial returns

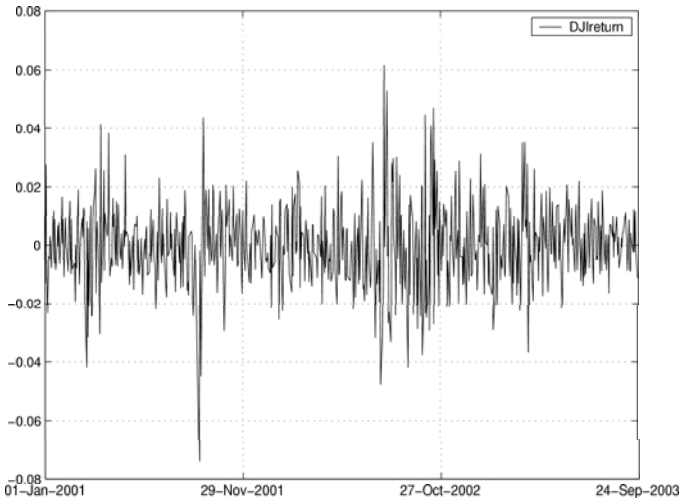

Figure 1.1: DJIA return

tend to follow large changes, whereas small changes tend to follow small changes. To illustrate this, figure 1.1 is a plot of Dow Jones Index log-returns from Jan 01, 2001 to September 2003. Volatility clustering is quite clear in the figure. For example, the changes of the returns of the few days around September 11, 2001 are much larger than the changes before that. Volatility clustering is an indication of persistence of shocks ϵ_t in (1.5).

iii. Leverage Effects.

Leverage effects refer to the phenomenon that the fluctuation of asset prices or returns tend to be negatively correlated with the changes of volatility. In other words, the volatility of the assets is affected asymmetrically by positive and negative innovations. The changes of volatility tend to be large if the changes of returns are negative. On the other hand, the changes of volatility tend to be small if the changes of return are positive. Many empirical works such as Black (1976) [2] , Nelson (1991) [37], Engle and Ng (1993) [14] all find evidence of leverage effects.

iv. Co-movements in volatility

When we look at financial time series across different markets, for example the exchange rate returns of different currencies or the stock returns from the same industry, we can often observe a joint variation or fluctuation of the time series. Such evidence has been found by Engle, Ng and Rothschild (1990) [13], Engle, Ito and Lin (1990) [10] among others. In addition to the co-movements in volatility of time series, some exogenous variables may also influence volatility. For example, Anderson and Bollerslev (1998) [4] find evidence that the volatility of the deutsch mark-dollar exchange rate increases when the US macroeconomic data are announced. Easley and O'Hara (1992) [8] find that the volatility of the US stock market is related to the trading volume, quote arrivals, forecastable events such as dividend announcements, and market closures.

1.3 ARCH and GARCH models

Following the structure of equation (1.5), we begin to talk about modeling the volatility term h_t. The seminal work of volatility modeling is the ARCH model of Engle (1982) [9]. ARCH means autoregressive conditional heteroscedasticity. The importance of the ARCH model is that it provides a systematic framework for volatility modeling.

The simplest ARCH(1) is as the following:

$$r_t = m_t + a_t \tag{1.6}$$

$$a_t = \sqrt{h_t}\epsilon_t \tag{1.7}$$

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 \tag{1.8}$$

where as mentioned in (1.3), r_t is the log-return series, $m_t = E_{t-1}[r_t]$ is the conditional expectation of r_t given past information, ϵ_t is assumed to follow the standard normal or standardized Student-t distribution with mean 0 and variance 1. As for m_t, we use a simple autoregressive model to model r_t in this dissertation. We can also include the volatility term h_t in the mean equation as the following:

$$r_t = \mu + \gamma h_{t-1} + a_t \tag{1.9}$$

$$m_t = \mu + \gamma h_{t-1} \tag{1.10}$$

This model is called garch-in-mean model developed by Engle, Lilien and Robins (1987) [11]. However, we will not talk about this model in this dissertation. Our focus in this chapter is about volatility only. Our goal in the dissertation is to model h_t. We have described the ARCH(1) model in (1.8). In the general linear ARCH(q) model, the conditional variance is assumed to be a function of the past q squared innovations,

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 \tag{1.11}$$

For this model to be well defined with positive conditional variance, the parameters must satisfy $\alpha_0 > 0$ and $\alpha_i \geq 0, i = 1, ...q$. Defining $v_t = a_t^2 - h_t$, the ARCH(q)

model may be re-written as

$$a_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + v_t$$

So the ARCH(q) model corresponds directly to an AR(q) model for the squared innovations a_t^2. The process is covariance stationary if and only if the sum of the positive autoregressive parameters is less than one, in which case the unconditional variance equals $Var(a_t) = \frac{\alpha_0}{1-\alpha_1-.....-\alpha_q}$. Therefore, we require

$$0 < 1 - \alpha_1 - ..... - \alpha_q < 1$$

The GARCH model introduced by Bollerslev (1986) [3] is a generalization of the ARCH model. The GARCH(p,q) model is defined by the equation:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{1.12}$$

where $p \geq 0, q > 0, \alpha_0 > 0, \alpha_i \geq 0, i = 1, ...q, \beta_i \geq 0, i = 1, ...p.$

In a GARCH model, an autoregressive term of the volatility is added. So, the GARCH formulation provides added flexibility over the linear ARCH models for parameterizing the conditional variance.

The advantages of ARCH and GARCH models are that they capture the features of fat tails and volatility clustering. The volatility clustering is due to the past a_t terms included in the model. As for the excess kurtosis, consider the ARCH(1) model as an example. It can be shown that the unconditional 4th moment $m_4 = E(a_t^4)$ of ARCH(1) is:

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

So, the unconditional kurtosis of a_t is:

$$\frac{E(a_t^4)}{[Var(a_t)]^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3$$

The weaknesses of ARCH and GARCH models is also clear:

1. They can not capture the leverage effects due to the positive and negative shocks that have the same effects on volatility.

2. They put restrictions on the parameters α_t and β_t.

3. They do not capture the co-movements of volatility time series.

To overcome the drawbacks of the ARCH model, Nelson (1991) [37] introduced the Exponential GARCH (EGARCH) model. The simplest EGARCH model can be written as:

$$\ln(h_t) = \alpha_0 + \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)] + \beta\ln(h_{t-1}) \tag{1.13}$$

A more general EGARCH(p,q) model can be written as:

$$\ln(h_t) = \alpha_0 + \frac{1+\sum_{j=1}^{q}\beta_j B^j}{1-\sum_{i=1}^{p}\alpha_i B^i} g(\epsilon_{t-1}) \tag{1.14}$$

where B is the lag operator such that $Bg(\epsilon_t) = g(\epsilon_{t-1})$, and $g(\epsilon_{t-1})$ allows for the asymmetric effects between positive and negative asset returns. The function $g(\epsilon_t)$ is defined by:

$$g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)], \tag{1.15}$$

where θ and γ are real constants. $E(|\epsilon_t|)$ is the expectation of $|\epsilon_t|$. When ϵ_t follows the standard normal distribution, $E(|\epsilon_t|) = \sqrt{2/\pi}$; when ϵ_t follows the standardized

Student-t distribution with n degrees of freedom, we have:

$$E(|\epsilon_t|) = \frac{2\sqrt{n-2}\Gamma((n+1)/2)}{(n-1)\Gamma(n/2)\sqrt{\pi}}$$

The advantages of the EGARCH model when compared with the GARCH model are:

1. The conditional variance of EGARCH is constrained to be non negative by the logarithmic assumption of the volatility term h_t without the restrictions on parameters in the GARCH model.

2. EGARCH allows for the leverage effects in the model. Positive and negative returns have asymmetric effect on the volatility term. If we rewrite $g(\epsilon_t)$ as:

$$g(\epsilon_t) = \begin{cases} (\theta+\gamma)\epsilon_t - \gamma E(|\epsilon_t|)] & : \quad \epsilon_t \geq 0 \\ (\theta-\gamma)\epsilon_t - \gamma E(|\epsilon_t|)] & : \quad \epsilon_t < 0 \end{cases}$$

we can see this property from the expression of $g(\epsilon_t)$.

In addition to GARCH and EGARCH models, there are many other ARCH formulations with similar properties. See Bera, Lee and Higgins (1990) [31], Zakoian (1994) [51] and Taylor (1986) [44]. In the next section, we discuss another extension of ARCH called stochastic volatility model.

1.4 Discrete time Stochastic Volatility Model

Stochastic volatility (SV) models treat the volatility term h_t as a stochastic process with an additional noise term other than ϵ_t. In this section we discuss only the discrete time formulation of stochastic volatility. The continuous time version is

described in Hull and White (1987) [23]. Continuous time stochastic volatility leads to generalizations of the well known Black-Scholes model.

The simplest discrete time stochastic volatility model [44] can be written as:

$$a_t = \sqrt{h_t}\epsilon_t \tag{1.16}$$

$$\ln(h_t) = \alpha + \beta \ln(h_{t-1}) + \eta_t \tag{1.17}$$

where ϵ_t, the same as in the ARCH model, is assumed to follow the standard normal or the standardized student-t distribution with mean 0 and variance 1, and η_t follows an assumed distribution with mean 0 and variance σ_η^2. ϵ_t and η_t are assumed to be independent. In other words, in this simplest SV model, $\ln(h_t)$ follows an AR(1) process. It's easy to generalize this model and $\ln(h_t)$ may follow many other stochastic processes.

We now describe some properties of the SV model. If η_t is Gaussian and $|\beta| < 1$, then $\ln(h_t)$ is a standard Gaussian autoregression with the mean and variance given by:

$$\mu_h = E(\ln(h_t)) = \frac{\alpha}{1-\beta} \tag{1.18}$$

$$\sigma_h^2 = Var(\ln(h_t)) = \frac{\sigma_\eta^2}{1-\beta^2} \tag{1.19}$$

If ϵ_t is Gaussian, the kurtosis of a_t is $3e^{\sigma_h^2}$, which is greater than 3. So the SV model also inherits the fat tail property of ARCH families. The asymmetric behavior in asset returns can also be captured in SV models by letting $Cov(\epsilon_t, \eta_t)$ be negative. By

adding some exogenous variables in equation (1.17), the co-movements of volatility can also be captured easily.

All these are advantages of SV models. In addition, SV models can capture all the stylized facts we mentioned earlier. The weakness of SV models is the difficulty we encounter in the estimation of parameters. Unlike ARCH or GARCH models, we cannot use the usual maximum likelihood estimation or ordinary least square methods in the estimation of the SV model parameters because of the innovation of η_t . Some feasible methods of estimation in SV models are Kalman Filtering (Nelson (1991) [37], Nelson and Foster (1994)) [38], Generalized Method of Moments (GMM), Quasi maximum Likelihood (QML), Markov Chain Monte Carlo (MCMC) (Shepard (1996) [43]), etc. But different algorithms may produce very different estimates and standard errors can be very large (Mahieu and Schotman (1998) [33]).

Chapter 2

Generalized Volatility Models (GVM)

2.1 Basic idea of generalized volatility model

The structure of this chapter is organized as follows. The first section is a brief introduction of the generalized volatility model. In the 2nd Section, we introduce the theoretical framework of the generalized volatility model (GVM). Section 3 provides two examples of applying this model to forecasting 20 day standard deviation. Section 4 applies GVM to forecasting realized volatility. The last section is a conclusion.

Before we talk about the generalized volatility model, let's go over the basic ARCH family models and stochastic volatility models. We use the same notation as in Chapter 1. Let r_t be a time series of interest; in this dissertation r_t is the log-return series of some financial asset. Let $\mathcal{F}_{t-1}$ be the information set at $t-1$. As in equation (1.5), $r_t = m_t + a_t = m_t + \sqrt{h_t}\epsilon_t$ in all the volatility models.

As we mentioned in equation (1.11), the ARCH(q) model is defined by:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 \tag{2.1}$$

where $\alpha_0 > 0, \alpha_i \geq 0, i = 1, ...q$.

A generalization of this is the GARCH(p,q) model in equation (1.12) where

h_t has the form:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{2.2}$$

where $p \geq 0, q > 0, \alpha_0 > 0, \alpha_i \geq 0, i = 1, ...q, \beta_i \geq 0, i = 1, ...p$

From Chapter 1, we know that a sensible volatility model should be able to capture the stylized facts of financial volatility described in Section 1.2. The advantages of ARCH and GARCH models are that they can capture excess kurtosis and volatility clustering. But they cannot capture the leverage effects and co-movements in volatility. And both ARCH and GARCH models have a lot of restrictions on the parameters α_i, β_i. To capture the leverage effects, Nelson (1991) [37] proposes an Exponential GARCH (EGARCH) model. As in Section 1.4, an EGARCH(p,q) model 1.14 can be written as:

$$\ln(h_t) = \alpha_0 + \frac{1 + \sum_{j=1}^{q} \beta_j B^j}{1 - \sum_{i=1}^{p} \alpha_i B^i} g(\epsilon_{t-1}) \tag{2.3}$$

where B is the lag operator such that $Bg(\epsilon_t) = g(\epsilon_{t-1})$. The function $g(\epsilon_t)$ allows for the asymmetric effects between positive and negative asset returns, and is defined by:

$$g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)], \tag{2.4}$$

where θ and γ are real constants. $E(|\epsilon_t|)$ is the expectation of $|\epsilon_t|$.

The EGARCH model overcomes some drawbacks of the GARCH model, but it does not consider the co-movements of financial volatility and the fact that some exogenous variable may have impacts on volatility.

An alternative volatility model is the stochastic volatility model. This model considers the volatility term h_t itself also as a stochastic process. The simplest discrete time stochastic volatility model 1.17 (Taylor (1986) [44]) can be written as:

$$\ln(h_t) = \alpha + \beta \ln(h_{t-1}) + \eta_t \tag{2.5}$$

Adding the innovation η_t increases the flexibility of the model. We can extend the model easily to capture all the stylized facts of volatility in section 1.2. And this model can be generalized to a continuous time model. But because this model uses two innovations ϵ_t and η_t, the estimation of the parameters of the model becomes quite difficult.

The goal of the generalized volatility model is to find an underling model to generalize both the SV and ARCH models. Ludger Hentschel's work (JFE, 1995, [32]) was the first try to nest all the ARCH models in one model. He proposed the following parametric model:

$$\frac{h_t^{\lambda/2} - 1}{\lambda} = \omega + \alpha h_{t-1}^{\lambda/2} f^{\mu}(\epsilon_t) + \beta \frac{h_{t-1}^{\lambda/2} - 1}{\lambda} \tag{2.6}$$

where $f(\epsilon_t) = |\epsilon_t - b| - c(\epsilon_t - b)$. For example, when $\lambda = \mu = 2$, $b = c = 0$, it corresponds to the GARCH model, when $\lambda = 0$, $\mu = 1$, $b = 0$, it corresponds to the EGARCH model.

In this chapter, I propose a generalized volatility model, more general than Ludger's model. To define the generalized volatility model, we define the past information set $\mathcal{F}_{t-1}$ formally next. Let

$$\boldsymbol{Z}_t = (\boldsymbol{Z}_{t,1}, ..., \boldsymbol{Z}_{t,p})',$$

be the p-dimensional vector of past explanatory variables or covariates, $t = 1, ..., N$ corresponding to h_t. We denote by $\mathcal{F}_{t-1}$ the σ-field generated by

$h_{t-1}, h_{t-2}, ..., r_{t-1}, r_{t-2}, ..., \boldsymbol{Z}_{t-1}, \boldsymbol{Z}_{t-2}, ...,$

$$\mathcal{F}_{t-1} = \sigma\{h_{t-1}, h_{t-2}, ..., r_{t-1}, r_{t-2}, ..., \boldsymbol{Z}_{t-1}, \boldsymbol{Z}_{t-2}, ...\}.$$

Then we can define the generalized volatility model, GVM as

$$r_t = m_t + a_t \tag{2.7}$$

$$a_t = \sqrt{h_t}\epsilon_t \tag{2.8}$$

$$g(E_{t-1}[h_t]) = \mathbf{Z}_{\mathbf{t-1}}^{'}\boldsymbol{\beta} \tag{2.9}$$

Where $E_{t-1}[h_t] = E[h_t|\mathcal{F}_{t-1}]$, the lefthand side of equation (2.9) is a function of the conditional expectation of the volatility term h_t given the past information $\mathcal{F}_{t-1}$, the righthand side is a linear form of the past information, and $\boldsymbol{\beta}$ is a constant coefficient vector. Notice that the right hand of ARCH family models are determined if $\mathcal{F}_{t-1}$ is given. In other words, in ARCH family models $E_{t-1}[h_t] = h_t$. So, in this way ARCH-like models are a special case of our generalized volatility model. For example,

ARCH: g is the identity function, $\boldsymbol{Z}_t = (\mathbf{1}, \mathbf{a_{t-1}})$

GARCH: g is the identity function, $\boldsymbol{Z}_t = (\mathbf{1}, \mathbf{a_{t-1}}, \mathbf{h_{t-1}})$

EGARCH: g is the logarithmic function, $\boldsymbol{Z}_t = (\mathbf{1}, \epsilon_{\mathbf{t-1}}, |\epsilon_{\mathbf{t-1}}| - \mathbf{E}(|\epsilon_{\mathbf{t-1}}|), \ln(\mathbf{h_{t-1}}))$

Ludger's Model: g is the box-cox transformation function of $\sqrt{h_t}$,

$\boldsymbol{Z}_t = (\mathbf{1}, \mathbf{h}_{\mathbf{t-1}}^{\boldsymbol{\lambda/2}}\mathbf{f}^{\mu}(\epsilon_{\mathbf{t-1}}), \frac{\mathbf{h}_{\mathbf{t-1}}^{\boldsymbol{\lambda/2}}-\mathbf{1}}{\lambda})$

In addition, GVM has some of the advantages of the stochastic volatility models. We consider the volatility process itself to be a stochastic process determined

by related covariate information. However, to avoid the estimation difficulty encountered in stochastic volatility, we do not add another innovation term η_t in the model. All the covariates $\mathbf{Z}'_{\mathbf{t-1}}$ in GVM are determined when $\mathcal{F}_{t-1}$ is given. Then we can use partial likelihood maximization method to estimate the GVM parameters in (2.9) (Kedem and Fokianos (2002), [30]). Using this method we get an estimate of $\boldsymbol{\beta}$ in (2.9), and apply the theory of generalized linear models to volatility series. Therefore, we can get an estimate of the volatility $E_{t-1}h_t$.

2.2 Theoretical framework of generalized volatility models (GVM)

The theoretical framework of generalized linear models (GLM) was developed by Nelder and Wedderburn (1972) [36], and it provides under some conditions a unified regression theory suitable for continuous, categorical, and count data. The theory of GLM was originally intended for independent data, but Kedem and Fokianos (2002) [30] extended GLM to dependent data, specifically time series data. In this section we apply GLM to financial volatility time series. We call it generalized volatility model (GVM).

2.2.1 Notation of GVM and some properties of exponential family of distributions

Let us denote by

$$v_t = E[h_t|\mathcal{F}_{t-1}] = E_{t-1}[h_t]$$

the conditional expectation of the volatility given the past. An important property of v_t is that:

$$E[v_t] = E[E[h_t|\mathcal{F}_{t-1}]] = E[h_t]$$

Our idea is to estimate v_t, and use it as a prediction of h_t. The problem is how to relate v_t to the covariates. In the GVM, we assume that the conditional distribution of h_t given the past belongs to the exponential family of distributions in natural or canonical form. That is, for $t = 1, ..., N$,

$$f(h_t; \theta_t, \phi|\mathcal{F}_{t-1}) = \exp\{\frac{h_t\theta_t - b(\theta_t)}{\alpha_t(\phi)} + c(h_t; \phi)\} \tag{2.10}$$

The parametric function $\alpha_t(\phi)$ is of the form ϕ/ω_t, where ϕ is a dispersion parameter, and ω_t is a known parameter.

For $t = 1, ..., N$, we assume there is a monotone function $g(.)$ such that

$$g(v_t) = \eta_t = \sum_{j=1}^{p} \beta_j Z_{(t-1)j} = \mathbf{Z'_{t-1}}\boldsymbol{\beta} \tag{2.11}$$

From $g(v_t(\theta)) = \eta_t(\theta)$, we can get

$$\theta_t = v^{-1}(g^{-1}(\eta_t)),$$

We denote

$$u(.) \equiv v^{-1}(g^{-1}(.))$$

Many probability density functions belong to the exponential family. If the probability density function is from the normal distribution with parameters v_t and s^2, for $t = 1, ..., N$, then

$$f(h_t;\theta_t,\phi|\mathcal{F}_{t-1}) = \exp\{\frac{h_t v_t - v_t^2/2}{s^2} - (h_t^2/s^2 + \log 2\pi s^2)/2\}$$

From which we see that $v_t = \theta_t$, $b(\theta_t) = \theta_t^2/2$, and $\phi = s^2$. The canonical link is the identity function,

$$g(v_t) = v_t = \mathbf{Z}_{\mathbf{t-1}}^{'}\boldsymbol{\beta}$$

The distribution of an exponential family has some very good properties. Since for every $t = 1, ..., N$,

$$\int f(h;\theta_t,\phi|\mathcal{F}_{t-1})dh = 1$$

Differentiation of both sides of the above equation with respect to θ_t yields

$$\int (\frac{h - b'(\theta_t)}{\alpha_t(\phi)}) f(h;\theta_t,\phi|\mathcal{F}_{t-1})dh = 0, \tag{2.12}$$

which implies the conditional mean

$$v_t = E[h_t|\mathcal{F}_{t-1}] = b'(\theta_t)$$

By further differentiation of equation (2.12) we obtain the variance:

$$Var[h_t|\mathcal{F}_{t-1}] = \alpha_t(\phi)b''(\theta_t)$$

2.2.2 Partial Likelihood Inference of GVM

The likelihood is defined as the joint distribution of the data as a function of the unknown parameters. When the data are independent or the dependence in the data is limited, we can maximize the likelihood to infer the unknown parameters. But for time series, the data are dependent. The likelihood is much more complicated

than that of independent data. Consider a time series $Y_t, t = 1, ..., N$ with a joint density of $f_\theta(y_1, ..., y_N)$ parameterized by a vector θ. Then the likelihood of Y_t is

$$f_\theta(y_1, ..., y_N) = f_\theta(y_1) \prod_{t=2}^{N} f_\theta(y_t | y_1, ..., y_{t-1}) \tag{2.13}$$

If we have time series Y_t and covariate process $\mathbf{Z_t}$, the joint distribution of them would be very complicated. The partial likelihood idea of Cox (1975) [6] is that we take only part of the likelihood for inference about the unknown parameters. Suppose we have a time series Y_t and a covariate vector process Z_t, then the partial likelihood of the observed series (Kedem and Fokianos (2002) [30]) is

$$PL(\boldsymbol{\beta}) = \prod_{t=1}^{N} f(y; \theta_t, \phi | \mathcal{F}_{t-1})$$

We have two time series Y_t and $\mathbf{Z_t}$ (covariate process) but we only consider the conditional likelihood of one time series given the other. Under the GVM structure, the log-partial likelihood $l(\beta)$ is given by

$$l(\boldsymbol{\beta}) = \sum_{t=1}^{N} \{ \frac{h_t u(\mathbf{z_{t-1}'}\boldsymbol{\beta}) - b(u(\mathbf{z_{t-1}'}\boldsymbol{\beta}))}{\alpha_t(\phi)} + c(h_t, \phi) \} \tag{2.14}$$

We shall use the notation

$$\nabla \equiv (\frac{\partial}{\partial \beta_1}, \frac{\partial}{\partial \beta_2}, ..., \frac{\partial}{\partial \beta_p})'$$

When we take the first derivative of the log-partial likelihood function with respect to the vector of the unknown parameters $\boldsymbol{\beta}$, we get the partial score

$$\boldsymbol{S_N}(\boldsymbol{\beta}) \equiv \nabla l(\boldsymbol{\beta}) = \sum_{t=1}^{N} \boldsymbol{Z}_{t-1} \frac{\partial v_t}{\partial \eta_t} \frac{(h_t - v_t(\boldsymbol{\beta}))}{Var[h_t | \mathcal{F}_{t-1}]}$$

The solution of the score equation,

$$\boldsymbol{S_N}(\boldsymbol{\beta}) \equiv \nabla l(\beta) = 0$$

is denoted by $\hat{\boldsymbol{\beta}}$, and is referred to as the maximal partial likelihood estimator (MPLE) of $\boldsymbol{\beta}$. Some standard numerical algorithms such as fisher scoring method can be used to solve the above score equation. Under certain regularity assumption, $\hat{\boldsymbol{\beta}}$ has some asymptotical properties (Kedem and Fokianos (2002) [30]).

1. $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, and $\hat{\boldsymbol{\beta}}$ is asymptotically normally distributed. That means,

$$\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}$$

in probability, and

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N_p(0, \mathbf{G^{-1}}(\boldsymbol{\beta})),$$

in distribution, as $N \to \infty$. where $\mathbf{G}(\boldsymbol{\beta})$ is a certain positive definite matrix.

2. The following holds in probability, as $N \to \infty$:

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \frac{1}{\sqrt{N}}\mathbf{G^{-1}}(\boldsymbol{\beta})\boldsymbol{S_N}(\boldsymbol{\beta}) \to 0.$$

2.2.3 Deviance Analysis

The concept of deviance is used in the diagnostics process of the GVM. Let $f(h_t; \theta)$ (2.10) be the density function of the observation h_t given the parameter θ, v_t be the expectation of h_t, $v_t = E[h_t|\mathcal{F}_{t-1}]$, then the log partial likelihood (2.14) can be written as

$$l(\mathbf{v}; \mathbf{h}) = \sum_i \ln f(h_i; \theta)$$

where $\mathbf{v} = (v_1, v_2, ..., v_N)$. As a measure of goodness-of-fit criterion, we can define scaled deviance as:

$$D = 2l(\mathbf{h}; \mathbf{h}) - 2l(\mathbf{v}; \mathbf{h}) \tag{2.15}$$

A useful result of D is in the hypothesis testing of the goodness-of-fit of fitted model. Let D_0 be the scaled deviance corresponding to $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, and D_1 to $H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1$, where $\boldsymbol{\beta}_1$ is a p-dimensional vector and $\boldsymbol{\beta}_0$ is a subvector of $\boldsymbol{\beta}_1$with dimension q, $q < p$. Then for large N, if ϕ is known, under H_0,

$$D_0 - D_1 \sim \chi^2_{p-q}$$

If ϕ is unknown, we can use an alternative statistic for testing H_0, that is,

$$\frac{(D_0 - D_1)/(p-q)}{D_1/(N-p)} \sim F_{p-q,N-p} \tag{2.16}$$

2.2.4 Box-Cox Transformation and the Link Function in GVM

Let's review our generalized volatility model (2.9):

$$g(E_{t-1}[h_t]) = \mathbf{Z}'_{\mathbf{t-1}}\boldsymbol{\beta} \tag{2.17}$$

In GVM, we estimate $E_{t-1}[h_t]$, the conditional expectation of h_t. We assume that h_t are from an exponential family. In the empirical work of this dissertation, generally, the h_t are dependent and non-Gaussian. To simplify the computation procedure of parameter estimation, we apply a box-cox transformation to h_t and transform them to a normal random variable coordinatewise. And for the normally distributed h_t, we can use the identical link function in the parameter estimation process. Then we can get an estimate of $E_{t-1}[h_t]$. The box-cox transformation process is described in the following paragraphs.

Suppose $h_1, h_2, ..., h_n$ are positive historical volatility series. Assume for $\lambda \in (-3, 3)$, $i = 1, ..., n$,

$$\omega_i = \begin{cases} \frac{h_i^\lambda - 1}{\lambda} & : \quad \lambda \neq 0 \\ \ln h_i & : \quad \lambda = 0 \end{cases}$$

are normally distributed with mean μ and constant variance σ^2, $\omega_i \sim N(\mu, \sigma^2)$. So ω_i series are the 'new' observations, λ, μ, σ^2 are unknown parameters. Box and Cox (1964) estimate the parameters by the maximum likelihood method. Assume $g(h)$ is the distribution of h_i, $f_\omega(\omega)$ is the distribution of w_i. Then,

$$g(h) = f_\omega(\omega(h))\frac{d\omega}{dh} = f_\omega(\omega(h))h^{\lambda-1}$$

We can get the log likelihood function $l(\lambda)$ given by:

$$l(\lambda) = C - n/2 \ln \widehat{\sigma}^2(\lambda) + (\lambda - 1)\sum_{i=1}^{n} \ln(h_i)$$

where, $\widehat{\sigma}^2(\lambda)$ is given by:

$$\widehat{\sigma}^2(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(\omega_i(\lambda) - \frac{1}{n}\sum_{i=1}^{n}(\omega_i(\lambda)))^2$$

Therefore, we only need to maximize the log likelihood function $l(\lambda)$, we can get an estimate of the parameter λ, then we can apply the GVM to the 'new' transformed observation series ω. In our empirical work in this dissertation, under many cases, the value of λ is very close to zero.

2.3 Two examples of GVM to forecast sample standard deviation

2.3.1 GVM to predict the sample standard deviation for the Dow Jones Industrial Index returns

In this section, we emphasize our methodology and do not strive to obtain the best model. The in-sample data we use are the daily return series of Dow Jones Industrial Average (DJIA) and Nasdaq Index from January 1st, 2001 to September 23rd, 2003. the out-of-sample data are 10 days after the above period. The financial volatility series are unobservable. We simply use the 20-day sample standard deviation of the return of DJIA as a measure of its volatility $\sqrt{h_t}$ term to show how the GVM model can be applied. We want to predict the 20-day standard deviation of DJIA returns and use the Nasdaq returns data as the covariate data. In this section, the term 'volatility' means the 20-day standard deviation series $\sqrt{h_t}$. In the next section we will give a formal definition of 'realized volatility' and apply GVM to forecast realized volatility.

First, Let's have a look at the time series plot of the above two in-sample stock indices in Figure 2.1. From the graph, roughly, we can see that the Dow Jones and Nasdaq indices are related to each other as they co-move. For example, when the Dow Jones index is high, Nasdaq is also high, when the Dow is low, Nasdaq is also low. Our interest is to predict the volatility of the Dow Jones index. A natural thought is to include the data of Nasdaq as a covariate when we predict the volatility of Dow Jones index. Next, let's formalize our model. Let r_t be the log-return series

of DJIA, the plot of the return of the DJIA is Figure 2.2. We want to predict the volatility of DJIA returns. Our idea is to apply the generalized volatility model to the prediction. So the volatility or the standard deviation of DJIA return using the above definition is the response of the model. In the GVM, we should assume the distribution of the response process. A Gaussian assumption is a natural idea. So next let's check the normal probability plot of the volatility.

The graph in Figure 2.3 gives a normal probability plot for graphical normality testing. The plot has the sample data displayed with the plot symbol '+'. Superimposed on the plot is a line joining the first and third quartiles of each column of data (a robust linear fit of the sample order statistics). This line is extrapolated out to the ends of the sample to help evaluate the linearity of the data. If the volatility comes from the normal distribution, the plot should appear linear. But from this figure, it's far away from the straight line. We can't assume that our volatility data are normally distributed. The plot and histogram of volatility are attached as Figure 2.5, and Figure 2.6.

We now check the norm plot of the logarithm of volatility. The plot is shown in Figure 2.4. It's not exactly linear, but quite close to the straight line. So, in this example we shall assume that volatility is log normally distributed. In fact, we can apply the Box-Cox transformation method to the DJIA volatility series to validate this assumption. The graph of the log-likelihood function (2.2.4) is in Figure 2.7. The best λ is -0.45, which is not far from zero. We may assume the log-data come from the normal distribution. The histogram of log(volatility) is attached as Figure

2.8, and it appears reasonably symmetric.

In this section, we consider the following GVM of the data. Assume y_t is the logarithm of volatility of DJIA at time t, then $y_t = ln(\sqrt{h_t})$, z_t is the logarithm of the volatility of the Nasdaq index at time t.

$$v_t = E[\ln(\sqrt{h_t})|\mathcal{F}_{t-1}] = C + \alpha y_{t-1} + \theta r_{t-1}^2 + \gamma z_{t-1}$$

After we estimate the parameters of the model using partial likelihood maximization, the model is given by:

$$v_t = -0.078 + 0.99 \times y_{t-1} + 2.54 \times r_{t-1}^2 - 0.01 \times z_{t-1}$$

Finally, we use this model to predict 13 days out-of-sample volatility and get Figure 2.9 of the prediction. In the figure, the dashed line represents the real data, the solid line represents the predicted value and the error bar is the 99% confidence interval. From the picture, we can see that 6 times out of 13 predictions are close to or are within the 99% confidence interval, 4 of them are close to the 99% confidence interval, 3 of them are far beyond the 99% confidence interval. Notice that the coefficient of the covariate z_{t-1} is very small if we compare it with the other coefficients. We can do a deviance analysis about z_{t-1}. Consider the hypothesis H_0: the model with z_{t-1} as a covariate, and H_1: the model without z_{t-1} as a covariate. We can calculate the statistic $\frac{(D_0-D_1)/(p-q)}{D_1/(N-p)}$, by (2.16), the statistic follows $F_{p-q,N-p}$ distribution. The statistic for this example is 0.1941 with p-value 0.3403. So we reject the hypothesis H_0. The impact of z_t is very small in this model. Since λ in the box-cox transformation is close to zero, we also did a Box-Cox transformation for the same DJIA

data to predict the standard deviation. The two plots of the prediction are attached as Figure 2.10. The results of the prediction are very similar. It is reasonable to do just a log-transform for the data.

Figure 2.1: Dow Jones (upper curve) and Nasdaq (lower curve) Index

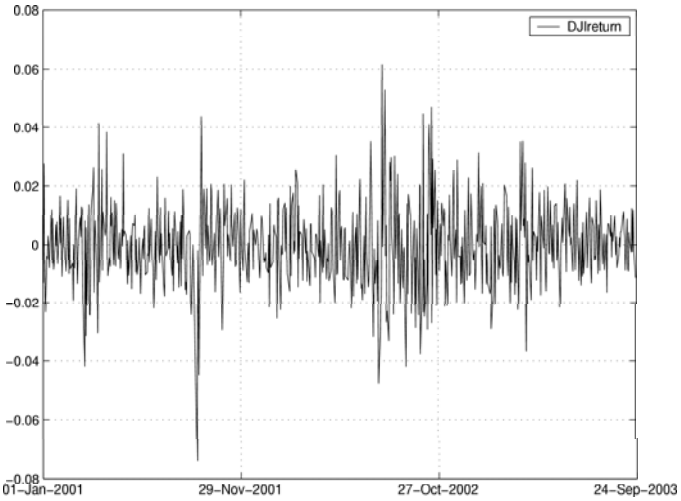

Figure 2.2: DJIA return r_t

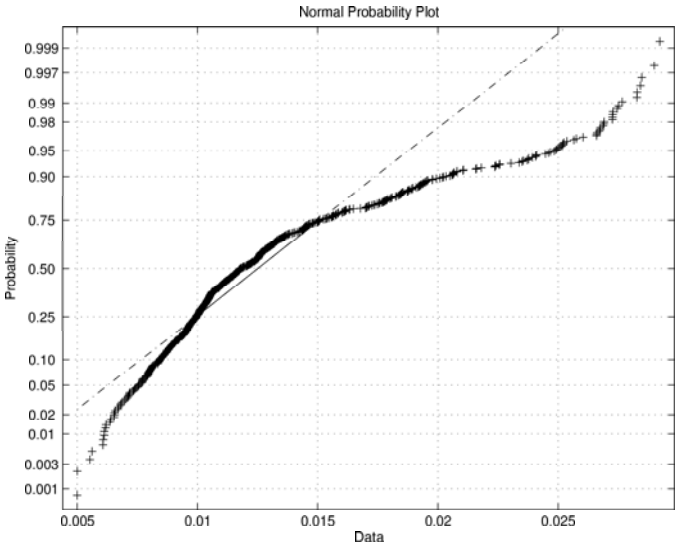

Figure 2.3: Normplot of the volatility data $\sqrt{h_t}$

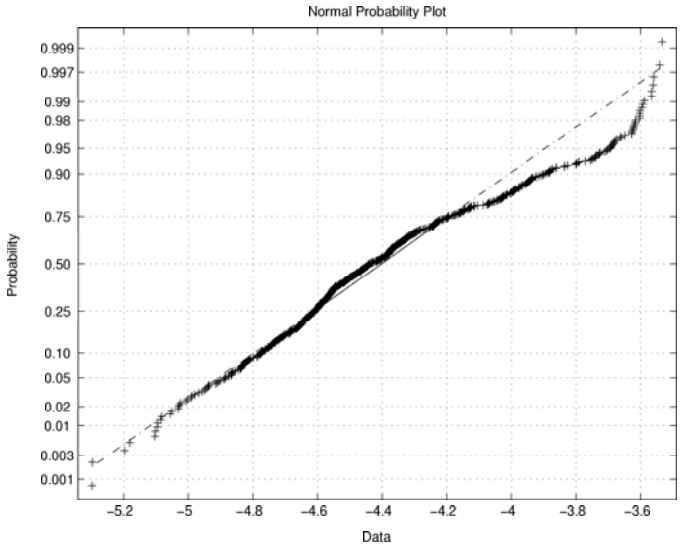

Figure 2.4: Normplot of the Ln(volatility), $ln(\sqrt{h_t})$

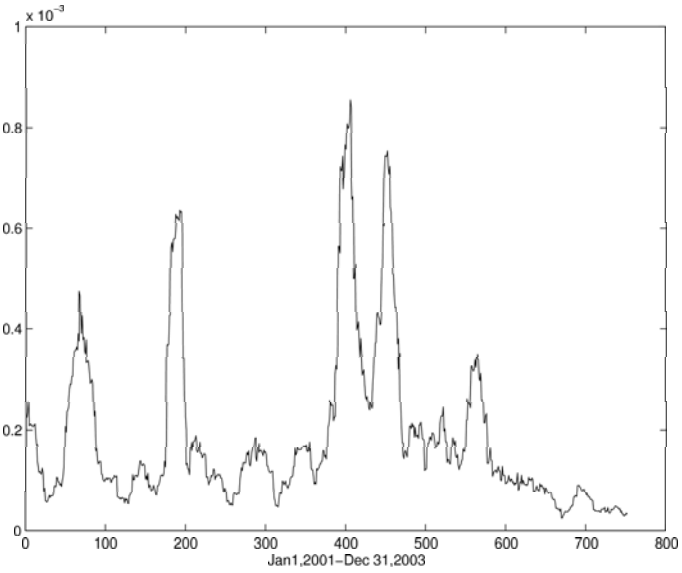

Figure 2.5: 20-day standard deviation of DJIA return $\sqrt{h_t}$

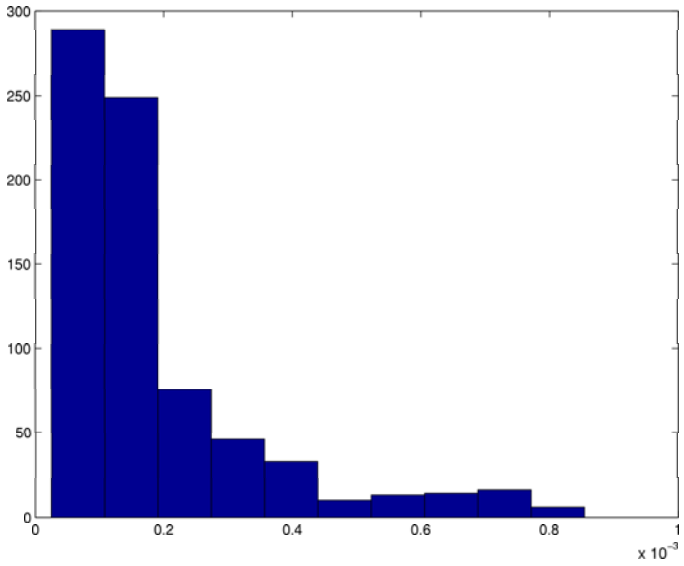

Figure 2.6: Histogram of DJIA volatility $\sqrt{h_t}$

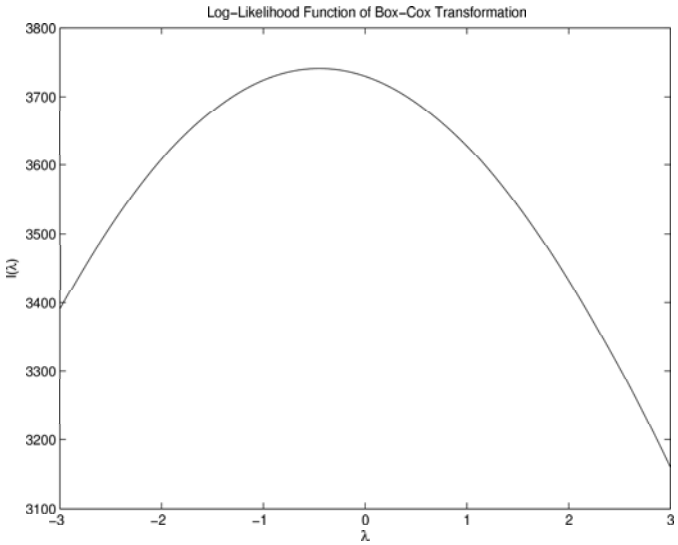

Figure 2.7: Log-likelihood Function of DJIA volatility $\sqrt{h_t}$

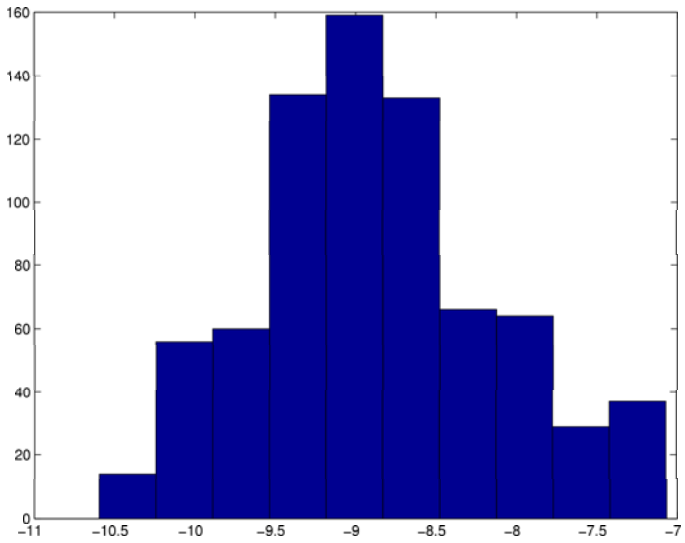

Figure 2.8: Histogram of Ln(volatility), $ln(\sqrt{h_t})$

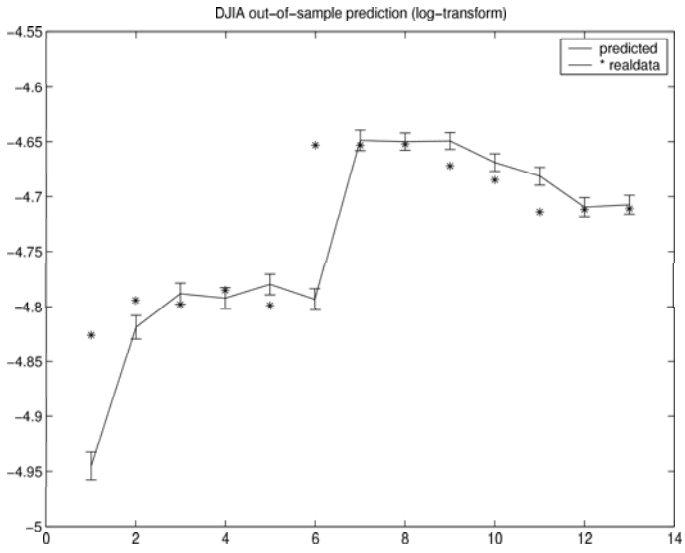

Figure 2.9: Out of Sample Prediction for DJIA $\sqrt{h_t}$, both predicted data and real volatility data $\sqrt{h_t}$ are after log-transform

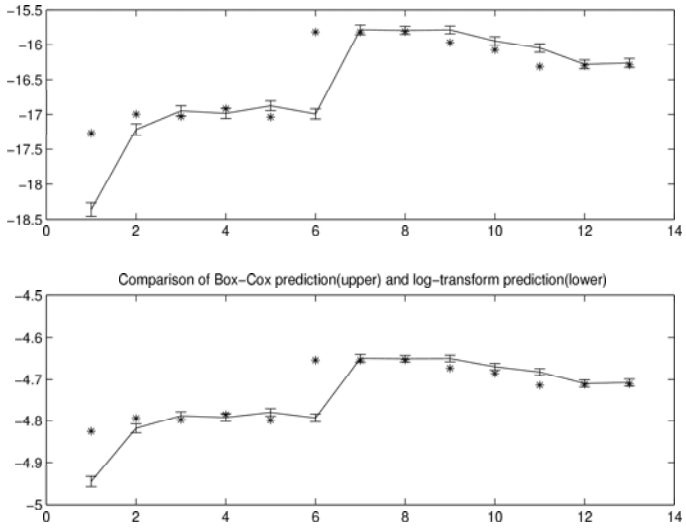

Figure 2.10: Comparison of the Out of Sample Prediction for DJIA $\sqrt{h_t}$, upper curve for Box-Cox transformation, lower curve for log-transformation

2.3.2 GVM to predict IBM stock return standard deviation

We follow the same procedure to predict IBM volatility (20-day standard deviation) using Microsoft stock price return as a covariate. Figure 2.14 gives the IBM sample standard deviation from February, 2000 to September, 2003. Figure 2.15 shows the Microsoft volatility during the same period. Economically we know that both IBM and Microsoft are IT giants. From the two plots, we can also see that the two volatility series have something in common. Generally, during the same time period, when the IBM volatility is high, that of Microsoft is also high, and vice versa. That's why we use Microsoft as a covariate to predict the IBM volatility.

Assume as above. log-IBM volatility is normally distributed. The histogram and qqplot of log(IBMvolatility) are attached. We can also apply the Box-Cox transformation method to the IBM volatility series. The value of λ is 0.089, also close to 0.

The graph of the log-likelihood function (2.2.4) is in Figure 2.11.

So these two examples show that it is reasonable to log-transform the data. Assume y_t is the logarithm of volatility of IBM at time t, then $y_t = log(h_t)$ and r_t is the return of IBM at time t, z_t is the logarithm of volatility of Microsoft at time t, m_t is the return of Microsoft at time t.

In this section, we consider the following model:

$$v_t = E[\ln(h_t)|\mathcal{F}_{t-1}] = \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \alpha_3 y_{t-1} + \alpha_4 m_{t-1} + \alpha_5 m_{t-1}^2 + \alpha_6 z_{t-1}$$

Using partial likelihood maximization, we get the estimates of the parameters of this model:

$$v_t = -0.13 - 0.08 r_{t-1} + 24.11 r_{t-1}^2 + 0.95 y_{t-1} - 0.03 m_{t-1} + 1.71 m_{t-1}^2 + 0.02 z_{t-1}$$

As in Section 3.1, we use this model to predict 10 days out-of-sample volatilities and get figure 2.16 about the prediction. In this figure, the stars represent the real data, the solid line represents the predicted value and the error bar is the 99% confidence bounds. For these data, 9 out of 10 of the real data are within the 99% prediction errorbars, quite satisfactory.

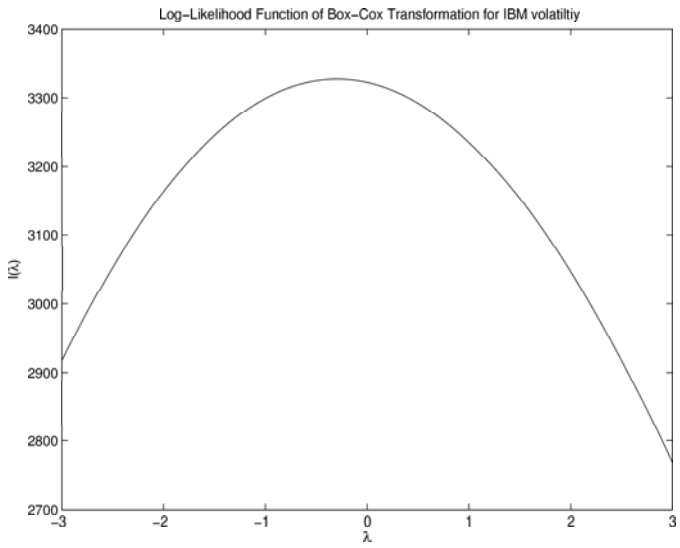

Figure 2.11: Log-likelihood Function of IBM volatility

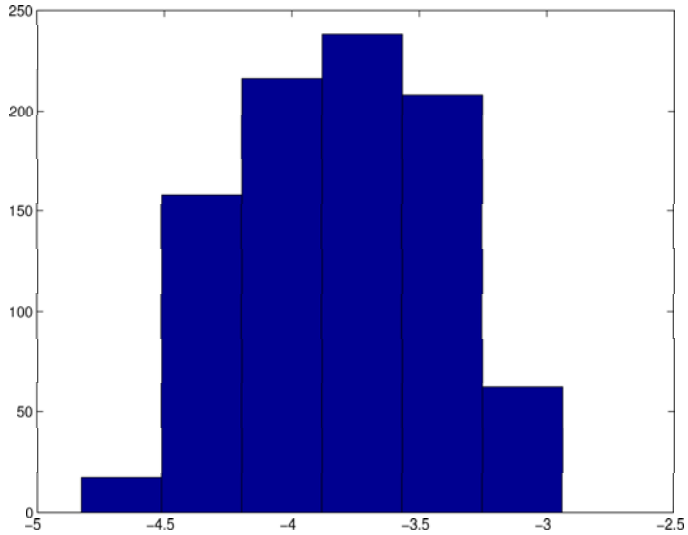

Figure 2.12: Histogram of log-IBM volatility $ln(\sqrt{h_t})$

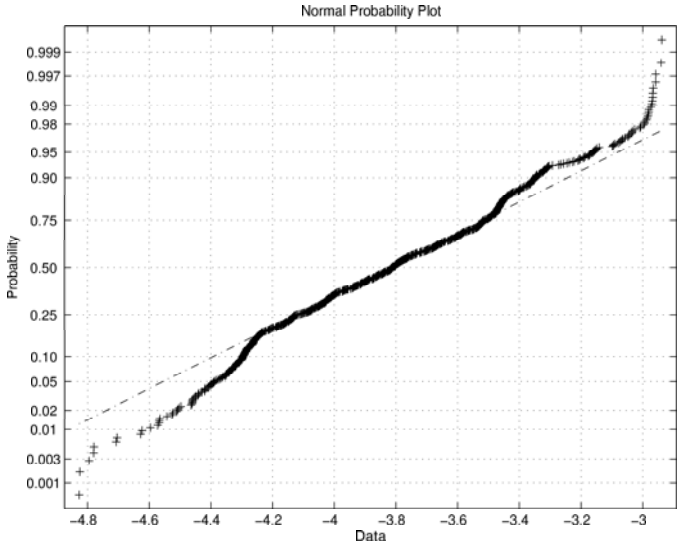

Figure 2.13: Normplot of log-IBM volatility $ln(\sqrt{h_t})$

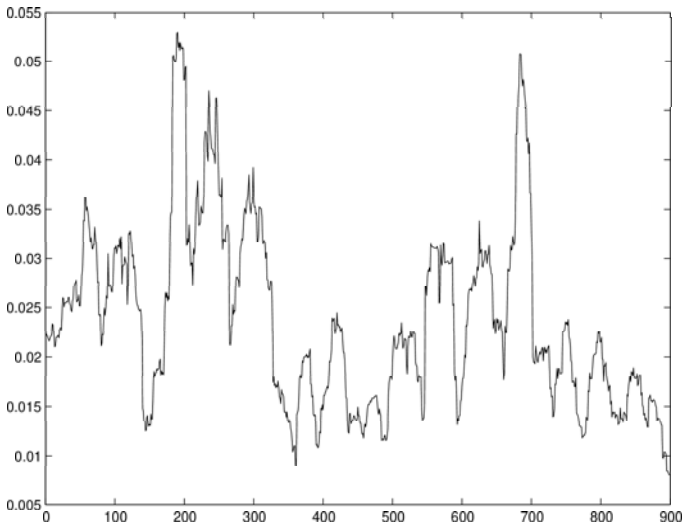

Figure 2.14: IBM volatility (Feb 2000- September 2003) $\sqrt{h_t}$

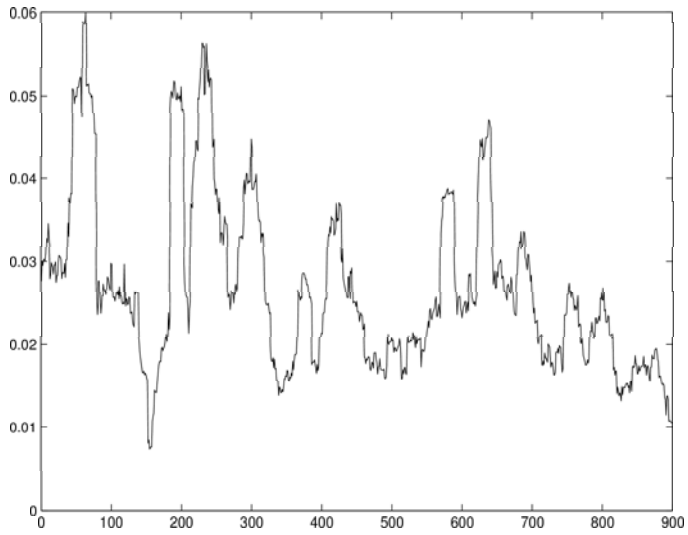

Figure 2.15: Microsoft Volatilities (The covariate, Feb 2000-Sep 2003)

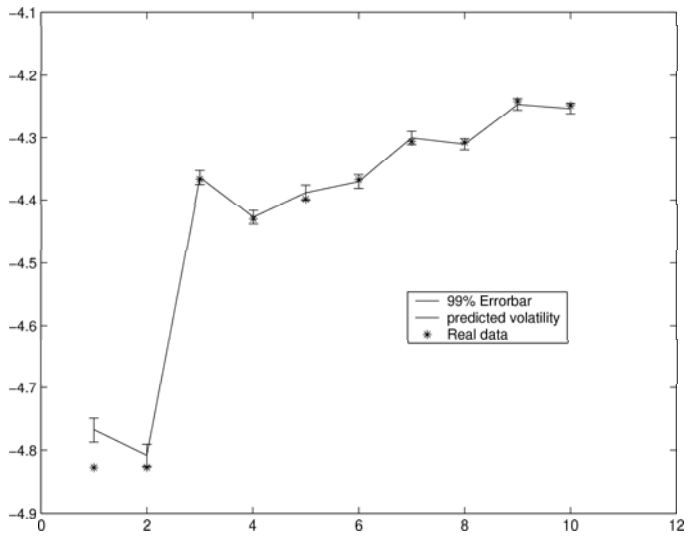

Figure 2.16: IBM Prediction

2.4 GVM to forecast realized volatility

2.4.1 Realized Volatility

In this section, we apply GVM to forecasting realized volatility. Volatility can not be observed directly. In the above section, we use 20-day sample standard deviation of returns as a measure of volatility. But how to choose the observation frequency is a big problem. When the sample standard deviation is calculated using a large number of observations, some stylized facts of volatility such as volatility clustering and leverage effects tend to disappear. When the sample standard deviation is calculated using a small number of observations, it's subject to large error.

Realized volatility is another method to measure volatility. It can be constructed by summing up intraday squared returns. Assuming that a day is divided into N equidistant periods, let $p_{i,t}$ denote the logarithmic price at time i at day t, if $r_{i,t}$ denotes the intradaily return of the i^{th} interval of day t, then $r_{i,t}$ is defined by:

$$r_{i,t} = p_{i,t} - p_{i-1,t} = \log(price_{i,t}) - \log(price_{i-1,t})$$

The daily volatility for day t can be written as:

$$\left[\sum_{i=1}^{N} r_{i,t}\right]^2 = \sum_{i=1}^{N} r_{i,t}^2 + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} r_{j,t} r_{j-i,t} \tag{2.18}$$

If we assume the $r_{i,t}$ have mean zero and are uncorrelated, $\sum_{i=1}^{N} r_{i,t}^2$ is a consistent and unbiased estimator of the daily variance h_t (ABDL, 1999, [1]). $\sum_{i=1}^{N} r_{i,t}^2$ is consistent because of the following argument. We assume the instantaneous returns are generated by

$$dp_t = \sigma_t dW_{p,t} \tag{2.19}$$

where p_t is the logarithmic price at time t, $dW_{p,t}$ is a standard wiener process. Since one period return (here is one day) $r_{t+1} = p_{t+1} - p_t$, from 2.19, the conditional variance of r_t would be $\int_0^1 \sigma^2_{t+\tau} d\tau$. And by the theory of quadratic variation (Karatzas and Shreve (1998) [27]),

$$p \lim_{N\to\infty} (\int_0^1 \sigma^2_{t+\tau} d\tau - \sum_{i=1}^{N} r^2_{i,t}) = 0 \tag{2.20}$$

When the intraday data are available, $r_{i,t}$ are observed. So $\sum_{i=1}^{N} r^2_{i,t}$ is called realized volatility.

If we sum sufficiently many $r^2_{i,t}$, we can get an error free measure of the daily volatility. However, if we choose a very high sampling frequency shorter than 5 minutes, this measure is plagued by a bias caused by various market microstructure effects which include bid-ask bounce, price discreteness, or nonsynchronous trading. [1] propose the use of 5-minute returns to calculate realized volatility. In this chapter, we also use this sampling frequency. Now we have the observed realized volatility as a measure of the latent volatility. We can use traditional time series models such as ARIMA models to forecast the realized volatility. But those models don't consider the stylized facts of the financial volatility. And we can also use the ARCH family models to forecast it. But the conventional ARCH models mainly use the information from the daily return series. They don't make full use of the availability of the high frequency data and the past realized volatility series. The GVM models can fill these gaps very well. They give the flexibility of adding covariates in the model and at the same time they keep the main structure of ARCH family models so that they can capture the stylized facts of financial volatility. In the next section,

we apply AR(3), GARCH(1,1) and EGARCH models to forecast IBM stock return realized volatility. In the third section, we apply GVM to predict IBM realized volatility and compare it with the results of the previous section.

2.4.2 Traditional time series models and GARCH models to forecast realized volatility

The data we use in this chapter are daily IBM return series and IBM realized volatility series from January 2002 to December 2003. We separate the data into two parts. The data from January 2002 to September 2003 are the in-sample data, from October 2003 to December 2003 are the out-of-sample data. We use the in-sample data to fit the models, predict the realized volatility during the out-of-sample period and compare it with the out-of-sample realized volatility. The graphs of the data are in the Figures 2.17 and 2.18 and some statistics of the in-sample data are in Table 2.1

IBM	mean	variance	skewness	kurtosis
returns	-7.26×10^{-4}	5.23×10^{-4}	0.23	6.26
realized volatility	2.36×10^{-4}	7.41×10^{-9}	1.58	8.85

Table 2.1: IBM statistics

From both data sets, the kurtosis are much great than 3, the kurtosis of the standard normal distribution. This means the distribution of IBM return has fatter tail than that of normal. The unconditional mean of the return series is -7.26×10^{-4}. To fit a GARCH or EGARCH model, we need to get an error series with mean 0.

We fit the return series with an AR(2) model empirically,

$$r_t = -0.00079 - 0.0024 r_{t-1} + 0.022 r_{t-2} + a_t$$

Then we fit the GARCH(1,1) and EGARCH(1,0) models to the a_t series. The GARCH(1,1) model is given by:

$$a_t = \sqrt{h_t}\epsilon_t \tag{2.21}$$

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta h_{t-1} \tag{2.22}$$

With the different assumptions of ϵ_t, the GARCH(1,1) model parameters are given in Table 2.2.

The GARCH model cannot capture the leverage effect. So we also fit the data to the simplest EGARCH(1,0) model. EGARCH(1,0) model has the form:

$$\ln h_t = \alpha_0 + \alpha_1 \epsilon_{t-1} + \alpha_1 |\epsilon_{t-1}| + \beta \ln h_{t-1} \tag{2.23}$$

After we fit the in-sample data, we get the estimated EGARCH model:

$$\ln h_t = -7.57 - 0.17\epsilon_{t-1} - 0.17 |\epsilon_{t-1}| + 0.75 \ln h_{t-1} \tag{2.24}$$

The family of ARCH models was built without the concept of realized volatility. So when we fit the GARCH(1,1) and EGARCH models, we only need the daily return series to fit the models. The first few volatility data we simply use the squared return data as a substitute. The ARCH family models don't make use of the information in realized volatility. Since we want to predict the out-of-sample realized volatility, it is natural to compare these models with the traditional ARMA

model of realized volatility series. By checking the PACF of IBM realized volatility series, we fit the IBM realized volatility series with the AR(3) model.

$$RV_t = 6.3 \times 10^{-5} + 0.36RV_{t-1} + 0.14RV_{t-2} + 0.23RV_{t-3} + \eta_t \tag{2.25}$$

where RV represents IBM realized volatility.

When we get the parameters of the corresponding GARCH(1,1), EGARCH and AR(3) models, we can apply them to forecast the out-of-sample realized volatility using the in-sample data. The mean square errors of predicted volatility from the above models and the true realized volatility are listed in table 2.3. The MSE is listed in 1,5,10,15,20 days period.

From the Table 2.3, we can see that for the IBM data, the traditional AR(3) is better than the GARCH and EGARCH models from the point of view of MSE. That's because the GARCH and EGARCH don't make use of the historical realized volatility series. For the GARCH model, if we assume the error terms have a student t distribution, the MSE is a little smaller than that of normal errors.

ϵ_t	α_0	α_1	β
Standard Normal	8.14×10^{-5}	0.20	0.65
Student T(5.34)	7.99×10^{-5}	0.19	0.67

Table 2.2: IBM GARCH(1,1)

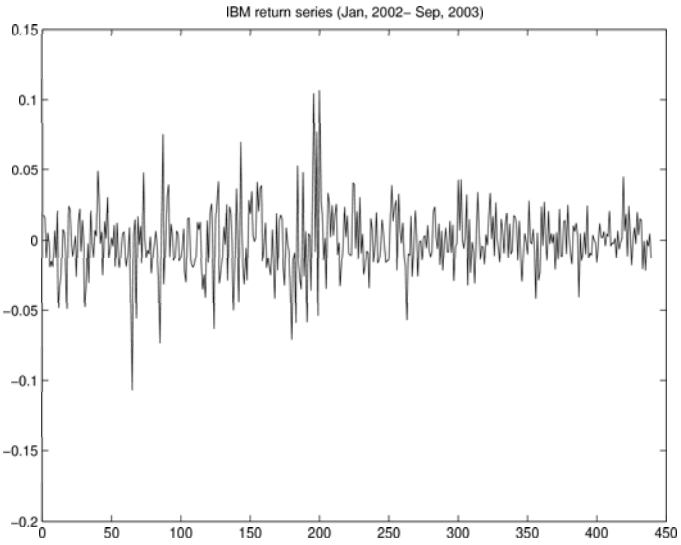

Figure 2.17: IBM return series

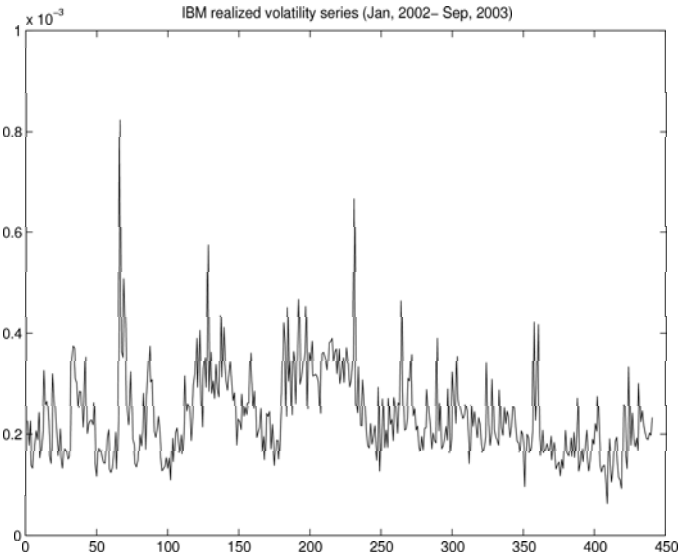

Figure 2.18: IBM realized volatility

MSE(in days)	1	5	10	15	20
GARCH1,$\times 10^{-7}$ Normal errors	0.6012	0.1686	0.1090	0.1300	0.1083
GARCH2,$\times 10^{-7}$ Student t(5.3) errors	0.5817	0.1679	0.1089	0.1263	0.1053
EGARCH, $\times 10^{-7}$	0.2882	0.1698	0.1436	0.1498	0.1272
AR(3), $\times 10^{-7}$	0.1535	0.0711	0.0537	0.0546	0.0456

Table 2.3: MSE comparison with true realized volatility

2.4.3 GVM to forecast realized volatility

In this section, we use the same IBM data as in previous and we apply GVM to predict the out-of-sample realized volatility. To apply GVM, we need to assume the in-sample volatility series are from a certain exponential family. We can have a look at the histogram of logarithm of IBM realized volatility series in Figure 2.19 and its qqplot in Figure 2.20. From the two figures, we can see that the distribution of logarithmic volatility series can be assumed normal. We can also use the box-cox transformation method to the IBM volatility series. The value of λ is -0.1246 also close to 0. The graph of the log-likelihood function 2.2.4 is in Figure 2.23. So, we assume the IBM volatility series follow the lognormal distribution. In the application, we apply GVM to the logarithmic volatility series and use the canonical link of normal distribution (identity link) to estimate the parameters in the model. We take a similar structure as in EGARCH model and add a covariate term. To find a good covariate series, we check the graph of trading volume of IBM stock

return and its realized volatility from January 2002 to September 2003. The graph is in Figure 2.21. They are very closely related to each other. We can compute the correlation coefficient, it is 0.71. So for IBM data, its trading volume seems a very good covariate. The GVM can be written as:

$$E[\ln(h_t)|\mathcal{F}_{t-1}] = \alpha_0 + \alpha_1\epsilon_{t-1} + \alpha_2\left|\epsilon_{t-1}\right| + \beta \ln h_{t-1} + \gamma VOLUME_{t-1} \qquad (2.26)$$

The GVM we fit is the following:

$$\ln h_t = -3.77 - 0.0145\epsilon_{t-1} + 0.0274\left|\epsilon_{t-1}\right| + 0.5606 \ln h_{t-1} - 0.5387 VOL_{t-1}$$

where VOL means the trading volume of IBM stock, in unit of one million shares. The standard error of the GVM parameters are in the table 2.4. The effect of the covariate the trading volume is obvious. After we estimate the parameters of GVM, we can also apply it to predict out-of-sample volatility. The graph of the predicted volatility and true volatility is in Figure 2.22. We also compute the MSE of the predicted series and true series in 1,5,10,15,20 days. The results are in the Table 2.5

Comparing this table with Table 2.3, we can see that the MSE of GVM are much smaller than that of all the other models. GVM improves greatly the prediction of realized volatilities as judged by the mean square error measure.

We follow the same procedure and comparison for the EXXON and COKE stock returns series. First we get a return series and use an AR(2) model to get its errors. For the errors series, we use GARCH(1,1) models with normal noise and student t noise, egarch to predict the volatility. For the realized volatility series, we use AR(2) and our GVM models to predict the volatility. We can also use the

box-cox transformation method to the EXXON and COKE volatility series. The value of λ for EXXON is 0.55, that of COCA COLA is 0.87, in between 0 and 1. The graphs of the log-likelihood function 2.2.4 are in Figure 2.24 and Figure 2.25.

We compare the mean square errors of all these volatility forecasting models for 1,5,10,15,20 days using the out-of-sample data. The in-sample and out-of-sample periods are the same as that of IBM data. In-sample period is from January 2002 to September 2003. Out-of-sample period is from October 2003 to December 2003. The models and MSE are in Table 2.6, 2.8, 2.9, 2.10. For the EXXON stock return series, we fit an AR(2) model by:

$$r_t = -0.00022 - 0.11r_{t-1} - 0.025r_{t-2} + a_t$$

Then we fit the in-sample error series a_t and realized volatility series as in Table 2.6. The mean square errors are in Table 2.8

For the Coca cola stock return series, we fit an AR(2) model by:

$$r_t = -0.00017 - 0.052r_{t-1} - 0.082r_{t-2} + a_t$$

Then we fit the in-sample error series a_t and realized volatility series as in Table 2.9. The mean square errors are in the table 2.10. From both MSE tables of EXXON and COKE models, the MSE of GVM are obviously smaller than that of the other models. We also attach the figures of the predicted volatility using GVM and the realized volatility in Figure 2.26 and 2.27.

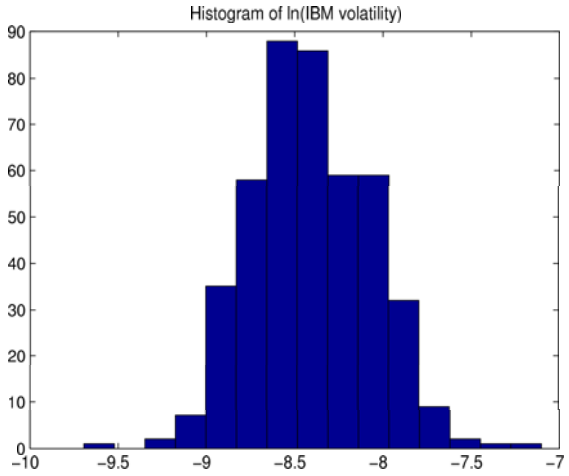

Figure 2.19: Histogram of logarithm of IBM realized volatility series

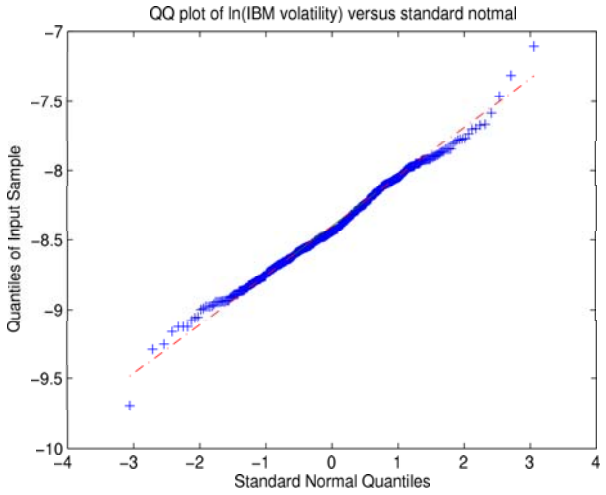

Figure 2.20: QQplot of logarithm of IBM realized volatility series

GVM Parameters	-3.7702	-0.0145	0.0274	0.5606	0.5387
Standard Errors for GVM Parameters	0.4820	0.0097	0.0175	0.0533	0.5596

Table 2.4: IBM models, Standard Errors for GVM Parameters

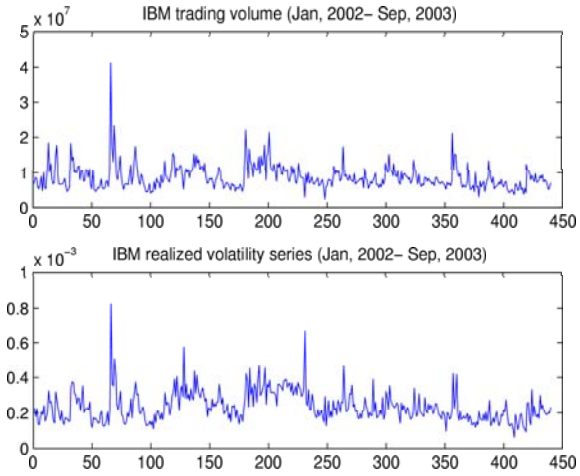

Figure 2.21: IBM realized volatility versus its trading volume

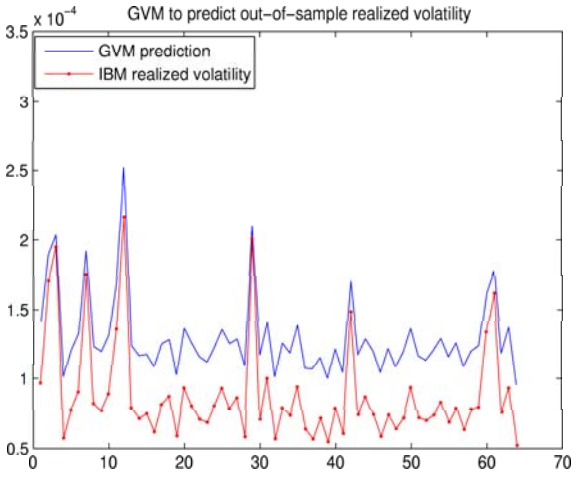

Figure 2.22: IBM realized volatility versus predicted volatility

Model	1	5	10	15	20
GVM $\times 10^{-8}$	0.1752	0.1100	0.1220	0.1235	0.1345

Table 2.5: GVM MSE comparison with true realized volatility

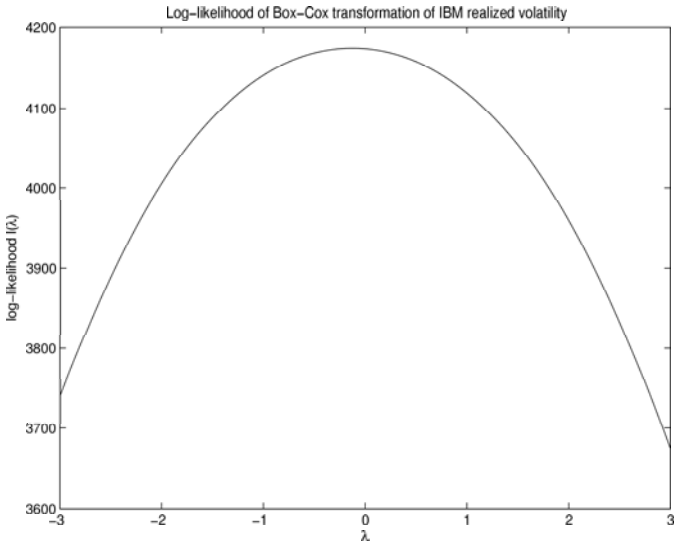

Figure 2.23: Log-likelihood Function of IBM realized volatility

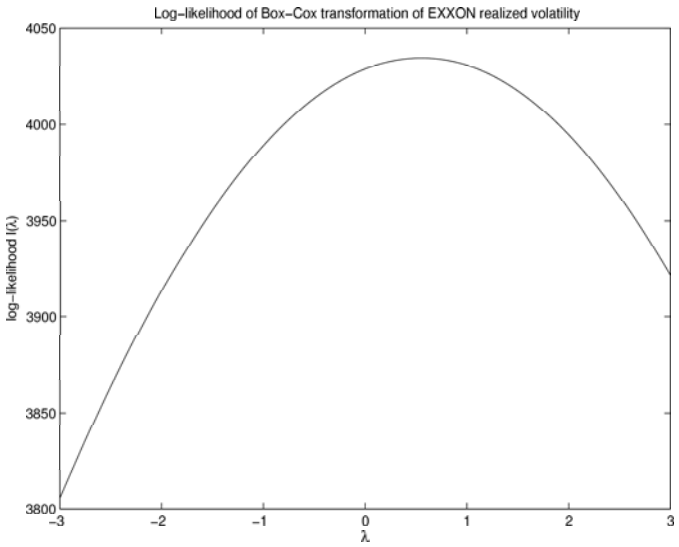

Figure 2.24: Log-likelihood Function of EXXON realized volatility

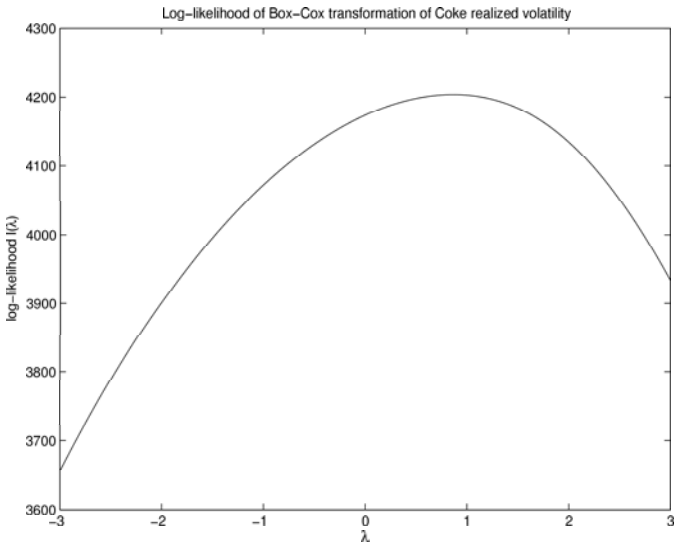

Figure 2.25: Log-likelihood Function of Coca Cola realized volatility

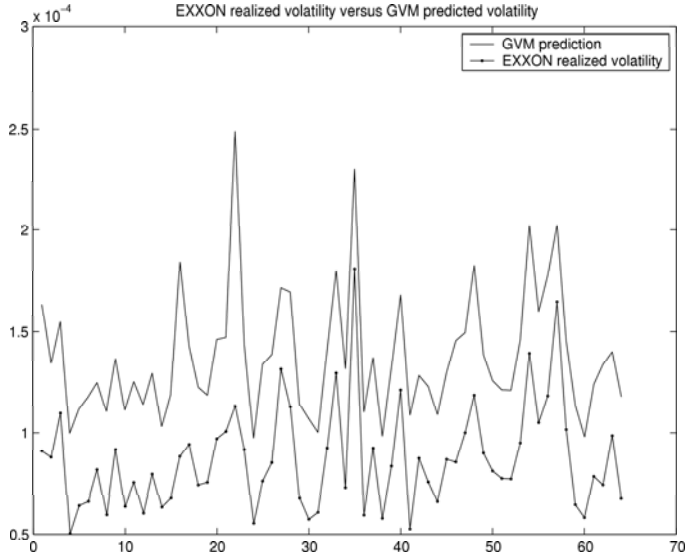

Figure 2.26: EXXON realized volatility versus predicted volatility

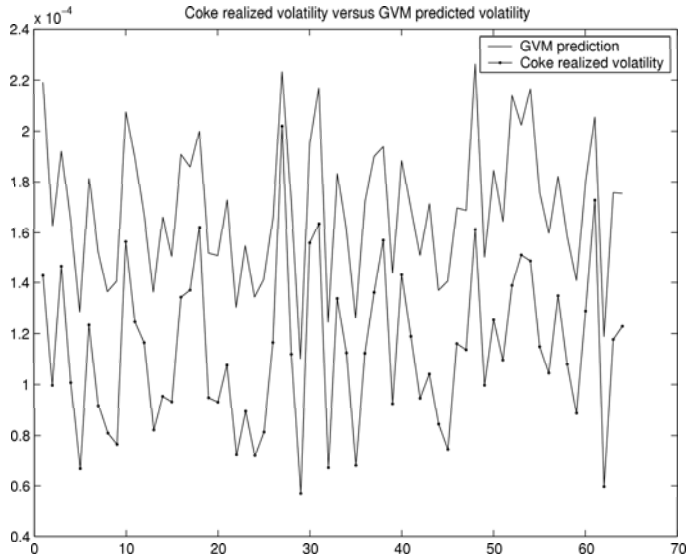

Figure 2.27: Coca Cola realized volatility versus predicted volatility

Models	Formula
GARCH1	$h_t = 3.44 \times 10^{-6} + 0.093 a_{t-1}^2 + 0.89 h_{t-1}$
GARCH2	$h_t = 2.09 \times 10^{-5} + 0.22 a_{t-1}^2 + 0.71 h_{t-1}$
EGARCH	$\ln h_t = -8.15 - 0.065 \epsilon_{t-1} - 0.065 \lvert\epsilon_{t-1}\rvert + 0.75 \ln h_{t-1}$
AR(3)	$RV_t = 7.63 \times 10^{-5} + 0.41 RV_{t-1} + 0.21 RV_{t-2} + 0.16 RV_{t-3} + \eta_t$
GVM	$\ln h_t = -2.83 - 0.0090 \epsilon_{t-1} + 0.12 \lvert\epsilon_{t-1}\rvert + 0.66 \ln h_{t-1} + 0.20 VOL_{t-1}$

Table 2.6: EXXON models, GARCH1 is with normal noise, GARCH2 is with student-t distributed noise, VOL is the trading volume, in the unit of one million shares.

GVM Parameters	-2.8316	-0.0090	0.1154	0.6599	0.2012
Standard Errors for GVM Parameters	0.3482	0.0125	0.0189	0.0406	0.3636

Table 2.7: EXXON models, Standard Errors for GVM Parameters

MSE(in days)	1	5	10	15	20
GARCH1,$\times 10^{-7}$ Normal errors	0.7260	0.1548	0.0792	0.0533	0.0420
GARCH2,$\times 10^{-7}$ Student t(10.2) errors	0.5377	0.1292	0.0660	0.0448	0.0439
EGARCH, $\times 10^{-7}$	0.0811	0.0687	0.0612	0.0569	0.0612
AR(3), $\times 10^{-7}$	0.6217	0.3739	0.2060	0.1506	0.1200
GVM,$\times 10^{-7}$	0.0526	0.0284	0.0256	0.0252	0.0280

Table 2.8: MSE comparison with true realized volatility for EXXON return series

Models	Formula
GARCH1	$h_t = 7.49 \times 10^{-6} + 0.064 a_{t-1}^2 + 0.91 h_{t-1}$
GARCH2	$h_t = 1.48 \times 10^{-4} + 0.17 a_{t-1}^2 + 0.27 h_{t-1}$
EGARCH	$\ln h_t = -7.81 - 0.10 \epsilon_{t-1} - 0.10 \lvert\epsilon_{t-1}\rvert + 0.75 \ln h_{t-1}$
AR(3)	$RV_t = 1.11 \times 10^{-4} + 0.32 RV_{t-1} + 0.16 RV_{t-2} + 0.14 RV_{t-3} + \eta_t$
GVM	$\ln h_t = -3.69 - 0.0082 \epsilon_{t-1} + 0.092 \lvert\epsilon_{t-1}\rvert + 0.56 \ln h_{t-1} - 0.013 VOL_{t-1}$

Table 2.9: Coca Cola realized volatility models, GARCH1 is with normal noise, GARCH2 is with student-t distributed noise, VOL is the trading volume, in the unit of one million shares.

MSE(in days)	1	5	10	15	20
GARCH1,$\times 10^{-7}$ Normal errors	0.3874	0.0966	0.0582	0.0404	0.0334
GARCH2,$\times 10^{-7}$ Student t(4.95) errors	0.1125	0.1249	0.0950	0.0908	0.0798
EGARCH, $\times 10^{-7}$	0.2009	0.1313	0.1257	0.1192	0.1295
AR(3), $\times 10^{-7}$	0.4067	0.2210	0.1382	0.1149	0.0984
GVM,$\times 10^{-7}$	0.0584	0.0394	0.0365	0.0363	0.0340

Table 2.10: MSE comparison with true realized volatility for Coca Cola return series

GVM Parameters	-3.69	-0.0082	0.092	0.56	-0.013
Standard Errors for GVM Parameters	0.4428	0.0165	0.0174	0.0386	0.5326

Table 2.11: Coca Cola stock return realized volatility models, Standard Errors for GVM Parameters

2.5 Conclusion

We suggest a new model using generalized linear models to predict financial volatilities in this paper. What's the advantage of this volatility model? We know that a number of stylized facts about the volatility of financial asset prices emerged over the years. These facts include: volatility clustering, mean-reverting, asymmetric innovation impact, exogenous impact, heavy tail, forecast evaluation. As mentioned in Engle and Pattern (2000) [15], a good volatility model must be able to capture and reflect these stylized facts. Since we can include any covariates in our model, we can reflect almost all the above stylized facts. For example, for volatility clustering, we can use historical data as covariates similar to GARCH model. For asymmetric innovation, we can do the same way as in regime-switching model, separate them into several parts so that they would have different weights as covariates. Heavy tails can be reflected from the assumption. And the other facts are quite obvious in our model. Furthermore, because of the flexibility of GVM, we can also try GARMA and binary models in the future research.

From the application part we can see that the prediction results are quite acceptable. Compare the mean square error of GVM and the other traditional models, the mean square error of GVM is the smallest. It is interesting to notice that when we use GVM to predict realized volatility there is a gap between the predicted volatility and true volatility. The predicted realized volatility are always higher the realized volatility. A possible reason is that the realized volatility in this

dissertation is computed by the summation of 5-minute return variance. The exact realized volatility should be the sum of instantaneous variance. Therefore, a gap can be expected. The exact realized volatility should be a little larger than 5-minute sum. And the predicted realized volatility are closer to the exact realized volatility.

There are many open questions in this field. For example, in the GVM model, we need to assume the time series is from some specific distribution and then choose a link function, is the assumption of the distribution good? Can we try other links? In fact, using the data of this paper we also try the link function of gamma distribution and inverse Gaussian distribution. Unfortunately, the residuals' square errors are extremely large. How to choose the link function, how to choose the appropriate covariates and how to do the goodness of fit of the models are all very interesting and important problems. And for financial time series, the macroeconomic news announcements have a big impact on volatility. Can we put this kind of news in our model and how? In this chapter, we do not aim to obtain the best model. How to obtain the best model from the measure of mean square error? And another future direction is about the option pricing using the GVM. All these problems need deeper research.

Chapter 3

Value at Risk

3.1 Introduction

There are many sources of risk in financial markets including credit risk, operational risk, liquidity risk, and market risk. These and other downturns may lead to the further risk of panic reaction which can be harmful to both domestic and international financial markets. Among these risks, market risk is tantamount to the uncertainty of future earnings due to changes in market conditions such as changes in the rates of interest, unemployment, or inflation. Value at Risk (VaR) is a probabilistic measure of financial risk often applied to assessing market risk, conditional on portfolio holdings and historical market data. More precisely, VaR can be defined as the maximal potential loss of a portfolio of financial instruments during a given time period for a given probability. The review article by Duffie and Pan (1997) [7] and the book by Jorion (2000) [25] provide a good introduction to this topic.

Let V_t be the value of some financial portfolio at time t, and let $\Delta V_t(l)$ be the difference between the values at time t and $t+l$. Then the VaR of a *long position* over the time horizon l with probability p is given by

$$p = Pr[\Delta V_t(l) \leq VaR] = F_l(VaR) \tag{3.1}$$

where $F_l(x)$ is the cumulative distribution function of $\Delta V_t(l)$. Long position means

owning some financial assets. Since the holder of a long position suffers a loss when $\Delta V_t(l) < 0$, the VaR in (3.1) is negative when p is small. On the contrary, *Short position* is a speculative action. The holder of a short position borrows a certain asset and sells them in market in advance and will buy the asset back and return them at a future time. So the holder of a *short position* suffers a loss when $\Delta V_t(l) > 0$. We can define VaR of a short position by

$$p = Pr[\Delta V_t(l) \geq VaR] \tag{3.2}$$

and here VaR is positive when p is small.

The great popularity of VaR among practitioners is mainly due to its simplicity. Using a single number we can assess a future financial loss with probability p over a given time horizon. VaR is used pervasively by banks, including the World Bank and the Bank for International Settlements, to make recommendation for capital adequacy requirement based on VaR estimates [25]. That makes the estimation of VaR very important to financial institutions.

The estimation of VaR is a challenging statistical problem which involves the estimation of the quantiles of the distribution of a financial asset return. The estimation methods of VaR can be classified into three broad categories: parametric, nonparametric, and semiparametric. The parametric method generally consists of two steps. First, the volatility of a portfolio is estimated using GARCH or IGARCH models. Then VaR is estimated from the estimated volatility and an assumed distribution of the return process [3],[7],[9], [35]. The typical nonparametric method entails historical simulation. We'll discuss those most used VaR methods in more

details in the next section.

Another useful quantity is *log-return* defined as

$$r_t = \ln(\frac{P_t}{P_{t-1}}) \tag{3.3}$$

where P_t and P_{t-1} are the values of the assets at times t and $t-1$. There are several reasons for the use of log-returns. First, we see from (3.1) and (3.2) that value changes of assets can be written in terms of the returns of the assets. Second, log-returns correspond approximately to percentage changes of a financial position, a fact convenient in data analysis. Third, the return of an asset is a complete and scale-free summary of an investment. Fourth, return series are easier to handle than price series because the former has more attractive statistical properties. Finally, there is an obvious connection between returns and VaR which itself may be viewed as a percentage, a perspective taken in what follows. Thus, the dollar amount of VaR is the cash value of the financial position times the VaR of the log-returns.

In this chapter we develop a new semiparametric method, based on a density ratio model, to estimate VaR. No distributional assumptions are made. Yet, under a certain "tilt" assumption we are able to determine an optimal distribution of the return process directly and from this estimate the corresponding VaR. The semiparametric VaR estimate is then compared via a simulation with traditional VaR estimates obtained from historical records and from a GARCH model and other methods.

The rest of the paper is organized as follows. Section 3.2 describes the VaR methodologies. We classify the existing VaR models into three categories: paramet-

ric, nonparametric and semiparametric. And we briefly introduce most commonly used models in these categories. Section 3.3 describes briefly our semiparametric methodology, and in Section 3.4 we discuss its implementation and comparison with other methods using a simulation. In the last section, we briefly summarize the main results of the paper.

3.2 VaR Methodologies

For the existing models to calculate VaR, they all have a general framework. First step is to estimate the distribution of portfolio returns. Then we can get the quantile of the distribution and therefore get the VaR of the portfolio. The difference between these models is how to estimate the distribution. We mentioned in the first section, we can classify the existing models into three categories: parametric, nonparametric and semiparametric models. We'll briefly talk about the most commonly used models in this section. Among these models: parametric models include Riskmetrics, GARCH and GVM; nonparametric models include Historical simulation and extensions of it; semiparametric models include Extreme Value Theory, Quantile Regression.

3.2.1 Parametric Models

The models in this section such as RiskMetrics (1996), GARCH and GVM propose a specific parameterization for the volatility process and return series. Then we can calculate value at risk from the model. The structure is as the following, which is

same as in (1.5):

$$r_t = m_t + a_t \tag{3.4}$$

$$a_t = \sqrt{h_t}\epsilon_t \tag{3.5}$$

Notice that m_t is the conditional mean of r_t. We assume that ϵ_t follows a certain standard distribution such as normal or student-t distribution and we want to estimate the quantile, or the VaR of r_t. If we have estimates of m_t and h_t as $\hat{m}_t$ and $\hat{h}_t$, and if we assume q is the quantile of the distribution of ϵ_t, then the estimate of VaR can be obtained from the following formula:

$$VaR = \hat{m}_t + \sqrt{\hat{h}_t} \times q \tag{3.6}$$

We have described the GARCH and GVM in the first two chapters. The simplest GARCH(1,1) model is given by 2.22:

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1} \tag{3.7}$$

where r_t is log-return series, m_t is the conditional expectation of r_t, a_t is the noise term, h_t is the volatility equation, ϵ_t is standard residuals, we assume ϵ_t is identically independent distributed with mean 0 and variance 1. This model has three crucial elements:the specification of volatility equation h_t, the mean equation m_t which can be specified by the general autoregressive model, and the standard residuals ϵ_t. With the assumption ϵ_t from certain distribution we can use maximum likelihood to estimate the parameters in the volatility equation h_t. The most common assumption is ϵ_t is from standard normal or standard student-t distribution. Once we estimate all the parameters in the equation, we can calculate it quantile or value at risk of

the return series r_t from the distribution. For example, if ϵ_t is from standard normal distribution and we want to compute 5% quantile, it equals to m_t plus -1.645 (the 5% quantile of standard normal distribution) times $\sqrt{h_t}$. For the GVM method, as long as we can estimate the volatility process we can compute VaR by the same method as in the GARCH method.

For the Riskmetrics, its origin is also from GARCH. In Riskmetrics, we assume m_t equals to zero and the volatility model follows an IGARCH model as the following:

$$r_t = \sqrt{h_t}\epsilon_t \tag{3.8}$$

$$h_t = \alpha r_{t-1}^2 + (1-\alpha)h_{t-1} \tag{3.9}$$

where α usually set to be 0.94 or 0.97 empirically. The advantage of Riskmetrics is that it can compute multiperiod value at risk easily. Suppose $r_t[k]$ is the log return from time $t+1$ to $t+k$, then $r_t[k] = r_{t+1} + ... + r_{t+k-1} + r_{t+k}$. And $h_t[k] = Variance(r_t[k]|F_t)$ is the k period volatility.

Notice that from (3.9), we can get

$$h_t = h_{t-1} + (1-\alpha)h_{t-1}(\epsilon_{t-1}^2 - 1)$$

for all t, so $E_{t-1}(h_t) = E_{t-1}(h_{t-1}) = h_{t-1}$ and by induction we can get $E_t(h_{t+i}) = h_{t+1}$ for $i = 2, 3, ..., k$. We can derive that $h_t[k] = Variance(r_t[k]|F_t) = kh_{t+1}$. The conditional variance of $r_t[k]$ is proportional to the time horizon k. Correspondingly, if VaR is the daily value at risk, then the k-day value at risk would be $\sqrt{k} \times VaR$. Riskmetrics is simple and easy to apply. But it is based on the assumption that the

mean equation is $m_t = 0$. If this assumption is not true, then the multiperiod value at risk formula is also wrong. And all these parametric models are subject to the misspecification of the variance equation and the chosen distribution. How relevant are these misspecification in the estimation of value at risk is mainly an empirical issue.

3.2.2 Nonparametric Models

The most common nonparametric model to calculate VaR is Historical Simulation. To apply the Historical Simulation, one needs to select a window of observations, generally half year to two years. Then we sort the return series within this window from lowest to highest. Then the θ-th quantile is the θth return of the observation to its left. For example, if we 1000 data ordered from smallest to largest, then the 5% quantile is the 50th smallest data. But this method has an implied assumption. This method assumes implicitly that the distribution within the window and beyond the window are the same. Logically we can say all the return series are from the same distribution. This implicit assumption is not reasonable. The advantage of Historical Simulation is its simplicity and it has no distribution assumption in the model.

3.2.3 Semiparametric Model: Extreme Value Theory

Using extreme value theory to calculate VaR is classified as a semiparametric VaR model. Extreme value theory is a classical topic in the statistical literature. Consider

a return series $\{r_1, ..., r_n\}$, let $M_n = max\{r_1, ..., r_n\}$, $m_n = min\{r_1, ..., r_n\}$. Notice that $M_n = -min\{-r_1, ..., -r_n\}$, so we only consider the distribution of the minimum m_n in this section. Assume that the return series r_t are from independent identical distribution with a CDF $F(x)$. If we denote the CDF of m_n to be $\widetilde{F}(x)$, then

$$\widetilde{F}(x) = 1 - [1 - F(x)]^n \tag{3.10}$$

As $n \to \infty$, we can see that $\widetilde{F}(x)$ becomes degenerate. It's either 1 or 0. Therefore, in EVT, we study the distribution of $(m_n - a_n)/b_n$ instead of CDF of m_n. Here $\{a_n\}$ and $\{b_n\}$ are two sequences with respect to n and $a_n > 0$. a_n is called location parameter, b_n is called scale parameter. By the theorem of Fisher and Tippett, if there exist such a_n, b_n and some degenerate distribution H such that $(m_n - a_n)/b_n \to H$ in distribution. Then,

$$H \equiv H_k(x) = \begin{cases} 1 - \exp[-(1+kx)^{1/k}] & : \quad k \neq 0 \\ 1 - \exp[-\exp(x)] & : \quad k = 0 \end{cases} \tag{3.11}$$

where $1 + kx > 0$. When $k = 0$, the special case $H_0(x) = \lim_{x \to 0} H_k(x)$. H is called the generalized extreme value (GEV) distribution. It describes the limit distribution of normalized minimum $(m_n - a_n)/b_n$. k is the crucial parameter in GEV. It determines the shape of the GEV distribution, especially the tail behavior of the limiting distribution. So k is referred to as the shape parameter. Genedenko (1943, [45]) showed that the GEV consists of three types of limiting distribution.

1. $k = 0$, the Gumbel family. The CDF is:

$$H_k(x) = 1 - \exp[-\exp(x)] \tag{3.12}$$

2.$k < 0$, the Fréchet family. The CDF is:

$$H_k(x) = \begin{cases} 1 - \exp[-(1+kx)^{1/k}] & : \quad x < -1/k \\ 1 & : \quad otherwise. \end{cases} \tag{3.13}$$

3. $k > 0$, the Weibull family. The CDF is:

$$H_k(x) = \begin{cases} 1 - \exp[-(1+kx)^{1/k}] & : \quad x > -1/k \\ 0 & : \quad otherwise. \end{cases} \tag{3.14}$$

Among these distributions, the Fréchet family include stable and Student-t distribution. Gumbel family consists of normal and log-normal distribution. The probability density function of the GEV distribution in (3.11) is given by:

$$h_k(x) = \begin{cases} (1+kx)^{1/k-1} \exp[-(1+kx)^{1/k}] & : \quad k \neq 0, 1+kx > 0 \\ \exp[x - \exp(x)] & : \quad k = 0 \end{cases} \tag{3.15}$$

Several methods are available to implement EVT. We apply a parametric maximum likelihood method in this dissertation. Our purpose is to estimate the location parameter a_n, scale parameter b_n and the shape parameter k. Since for each sample data series there is only one sample minimum. To implement EVT by MLE we need more sample minima. Therefore in the first step, we divide the whole data set into some non overlapping subsamples. Then on each subsample, there exist a sample minimum. For example, if there is a sample data series $\{r_1, ..., r_N\}$, for simplicity we can assume that $N = n * q$. Then we can divide the data into p subsamples and each subsample contains n data. Suppose $m_{n,i}$ is the subsample

minimum, $i = 1, ..., q$. If n is large enough, we assume EVT applies to each subsample. From equation 3.15, we can get the distribution of $x_{n,i}$. That is,

$$h(m_{n,i}) = \begin{cases} \frac{1}{b_n}[(1+kx_i)^{1/k-1}]\exp[-(1+kx_i)^{1/k}] & : \quad k \neq 0, 1+kx_i > 0 \\ \frac{1}{b_n}\exp[x_i - \exp(x_i)] & : \quad k = 0 \end{cases} \tag{3.16}$$

where $x_i = (m_{n,i} - a_n)/b_n$. In this way, we can construct the likelihood function for each subsample minimum $m_{n,i}$.

$$L(m_{n,1}, ..., m_{n,p}|a_n, b_n, k) = \prod_{i=1}^{q} h(m_{n,i})$$

And by the maximization of L, we can obtain the estimate of a_n, b_n, k.

In the second step, through the generalized extreme value distribution, we can get the corresponding value at risk for the subsample minimum. Let p^* be a small probability, r^* is the corresponding p-th quantile of the subsample minimum. Then by the CDF of the generalized extreme value distribution, we have:

$$p^* = \begin{cases} 1 - \exp[-(1+k(r^*-a_n)/b_n)^{1/k}] & : \quad k \neq 0 \\ 1 - \exp[-\exp(r^*-a_n)/b_n] & : \quad k = 0 \end{cases} \tag{3.17}$$

We can obtain the p-th quantile of the subsample minimum by the above equation. That is:

$$r^* = \begin{cases} a_n - \frac{b_n}{a_n}\{1 - [-\ln(1-p^*)]^k\} & : \quad k \neq 0 \\ a_n + b_n \ln[-\ln(1-p^*)] & : \quad k = 0 \end{cases} \tag{3.18}$$

In the third step, from a given small probability p^* and quantile r^* for the subsample minimum, we can get the VaR for the whole data series. For a data

series $\{r_t\}$, using the relationship in (3.10), we can get:

$$p^* = 1 - [1 - P(r_t \leq r_*)]^n$$

If we set $p = P(r_t \leq r_*)$, then $p^* = 1 - (1 - p)^n$, by the equation 3.18, we can get the VaR by the following expression:

$$VaR = \begin{cases} a_n - \frac{b_n}{a_n}\{1 - [-n\ln(1-p)]^k\} & : \quad k \neq 0 \\ a_n + b_n \ln[-n\ln(1-p)] & : \quad k = 0 \end{cases}$$

Similarly, for a short position with a small probability p, by the above parametric maximum likelihood method to implement EVT we can get VaR by:

$$VaR = \begin{cases} a_n + \frac{b_n}{a_n}\{1 - [-n\ln(1-p)]^k\} & : \quad k \neq 0 \\ a_n + b_n \ln[-n\ln(1-p)] & : \quad k = 0 \end{cases} \tag{3.19}$$

3.3 A Density Ratio Model

The rest of this chapter are mainly from the working paper "Time Series Prediction via Density Ratio Modeling" (with Kedem B. and Gagnon R., [28]).

Our basic idea is that of a "reference" probability density and its "distortions". It is motivated from a well known fact regarding exponential families of probability distributions.

For $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$, let $g(x, \boldsymbol{\theta})$ be the probability density of a k-parameter exponential family,

$$g(x, \boldsymbol{\theta}) = d(\boldsymbol{\theta}) S(x) \exp\left\{\sum_{i=1}^{k} c_i(\boldsymbol{\theta}) T_i(x)\right\}$$

Then, with $\alpha = \log[d(\boldsymbol{\theta}_1)/d(\boldsymbol{\theta}_2)]$, $\boldsymbol{\beta} = (c_1(\boldsymbol{\theta}_1) - c_1(\boldsymbol{\theta}_2), ..., c_k(\boldsymbol{\theta}_1) - c_k(\boldsymbol{\theta}_2))'$, and $\boldsymbol{h} = (T_1(x), ..., T_k(x))'$, we obtain the ratio

$$\frac{g_1(x)}{g_2(x)} \equiv \frac{g(x, \boldsymbol{\theta}_1)}{g(x, \boldsymbol{\theta}_2)} = \exp\{\alpha + \boldsymbol{\beta}'\boldsymbol{h}(x)\} \tag{3.20}$$

or

$$g_1(x) = \exp\{\alpha + \boldsymbol{\beta}'\boldsymbol{h}(x)\} g_2(x) \tag{3.21}$$

Numerous well known parametric families including the normal, gamma, beta, and Rayleigh, satisfy (3.20) [26]. In the normal case with mean μ and variance σ^2, $\boldsymbol{\theta} = (\mu, \sigma^2)$, the density ratio (3.20) is obtained with

$$\begin{aligned} \alpha &= \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\mu_2^2}{2\sigma_2^2} - \frac{\mu_1^2}{2\sigma_1^2} \\ \boldsymbol{\beta} &= \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}, \frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right)' \\ \boldsymbol{h}(x) &= (x, x^2)' \end{aligned}$$

In the gamma case with shape r and scale λ, $\boldsymbol{\theta} = (r, \lambda)$, the density ratio (3.20) is obtained with

$$\begin{aligned} \alpha &= \log\frac{\lambda_1^{r_1}\Gamma(r_2)}{\lambda_2^{r_2}\Gamma(r_1)} \\ \boldsymbol{\beta} &= (\lambda_2 - \lambda_1, r_1 - r_2)' \\ \boldsymbol{h}(x) &= (x, \log x)' \end{aligned}$$

And in the Rayleigh case with scalar parameter θ, β and $h(x)$ in (3.20) are scalars,

$$\alpha = \log\frac{\theta_2^2}{\theta_1^2}$$

$$\begin{aligned} \beta &= \frac{1}{2\theta_2^2} - \frac{1}{2\theta_1^2} \\ h(x) &= x^2 \end{aligned}$$

The tilt relationship (3.21) holds true for exponential families, but in general it suggests a way to relate or regress probability densities. In applications we view (3.21) as a plausible and sensible model.

3.3.1 A Density Ratio Model for Time Series

Consider now the following $m = q + 1$ time series,

$$\begin{aligned} x_{1t} &= f_1(\boldsymbol{z}_{1,t-1}) + \epsilon_{1t}, \quad t = 1, ..., n_1 \\ &\cdot \\ &\cdot \\ &\cdot \\ x_{qt} &= f_q(\boldsymbol{z}_{q,t-1}) + \epsilon_{qt}, \quad t = 1, ..., n_q \\ x_{mt} &= f_m(\boldsymbol{z}_{m,t-1}) + \epsilon_{mt}, \quad t = 1, ..., n_m \end{aligned} \tag{3.22}$$

where the vectors $\boldsymbol{z}_{k,t-1}$ contain past values of covariate time series possibly including even past values of $x_{1t}, ..., x_{qt}, x_{mt}$, and the ϵ_{kt} are independent noise components [30]. For the moment assume the f_i are known. In applications, as demonstrated in the last example, the f_i are estimated and the ϵ_{kt} are replaced by the residuals. An important special case is m-dimensional autoregressive process with additive independent noise and known coefficients. Suppose that for each t, ϵ_{jt} is distributed according to an unknown probability density $g_j(x)$, $j = 1, ..., q, m$. Define $g(x) =$

$g_m(x)$ as the *reference* density. Then, emboldened by the density ratio model (3.20), we shall assume that each g_j satisfies (3.20) relative to g, or equivalently, with scalars α_j, $p \times 1$ vectors $\boldsymbol{\beta}_j$, and a known $p \times 1$ vector of real-valued functions $\boldsymbol{h}(x)$,

$$g_j(x) = \exp\{\alpha_j + \boldsymbol{\beta}_j' \boldsymbol{h}(x)\} g(x), \quad j = 1, ..., q \tag{3.23}$$

The objective is to estimate all the $\alpha_j, \boldsymbol{\beta}_j$, the reference density g, and the corresponding cdf G, for the purpose of predicting the future reference value $x_{m,t+1}$, using the combined "data" from all the m "samples"

$$\boldsymbol{\tau} \equiv \Big\{(\epsilon_{1t}, ..., \epsilon_{1n_1}), ..., (\epsilon_{q1}, ..., \epsilon_{qn_q}), (\epsilon_{m1}, ..., \epsilon_{mn_m})\Big\} \tag{3.24}$$

of length $n = n_1 + ... + n_q + n_m$. This is the same as estimating the $\alpha_j, \boldsymbol{\beta}_j$ and g, conditional on some initial values if necessary, from the entire observed differenced data set of size $n = n_1 + \cdots + n_m$,

$$\begin{aligned}
x_{1t} &- f_1(\boldsymbol{z}_{1,t-1}), \quad t = 1, ..., n_1 \\
&\cdot \\
&\cdot \\
&\cdot \\
x_{qt} &- f_q(\boldsymbol{z}_{q,t-1}), \quad t = 1, ..., n_q \\
x_{mt} &- f_m(\boldsymbol{z}_{m,t-1}), \quad t = 1, ..., n_m
\end{aligned} \tag{3.25}$$

This in particular means that g is estimated from all the observations and covariate data and not just from "its own" time series $(x_{m1}, ..., x_{mn_m})$ of length $n_m < n$.

The exponential "tilt" relationships (3.23) relative to a reference or baseline density g enable a semiparametric inference about the $\alpha_j, \boldsymbol{\beta}_j$, and about g and the

corresponding distribution function G, based on the combined data set consisting of all the m time series and their covariate data. Using the estimator $\hat{G}$ of G we can estimate future probabilities of events formulated in terms of the "reference" $x_{m,t+1}$ conditional on $\boldsymbol{z}_{m,t}$. Model (3.23) is a special case of weighted distributions and biased sampling [39],[46],[47].

Example: Bivariate AR. The linear system

$$x_t = a_1 x_{t-1} + a_2 y_{t-1} + \epsilon_t$$

$$y_t = b_1 x_{t-1} + b_2 y_{t-1} + \eta_t \tag{3.26}$$

$t = 1, ..., N$, with independent Gaussian noise components $\epsilon_t \sim \mathrm{N}(0, \sigma_1^2)$ and $\eta_t \sim \mathrm{N}(0, \sigma_2^2)$, satisfies the density ratio model. We have

$$\frac{g_\epsilon(x)}{g_\eta(x)} = \exp\left\{\log(\sigma_2/\sigma_1) + (1/2\sigma_2^2 - 1/2\sigma_1^2)x^2\right\} \equiv e^{\alpha+\beta x^2} \tag{3.27}$$

or $m = 2, q = 1$, $g_1 = g_\epsilon$, $g = g_\eta$, and (3.23) reduces to

$$g_\epsilon(x) = e^{\alpha+\beta x^2} g_\eta(x) \tag{3.28}$$

3.3.2 Semiparametric Estimation

The exponential "tilt" relationships (3.23) relative to a reference or baseline density g enable a semiparametric inference about the $\alpha_j, \boldsymbol{\beta}_j$, and about g and the corresponding distribution function G, based on the combined dataset $\boldsymbol{\tau}$ consisting of all the m datasets.

A maximum likelihood estimator of $G(x)$ can be obtained by maximizing the likelihood over the class of step cdf's with jumps at the values $\tau_1, ..., \tau_n$ [20],[22],[41],[42].

Let $w_j(\tau) = \exp\left\{\alpha_j + \boldsymbol{\beta}_j' \boldsymbol{h}(\tau)\right\}$, $j = 1, ..., q$, and $p_i = dG(\tau_i)$, $i = 1, .., n$. Then the likelihood becomes,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_q, G) = \\ \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} \exp(\alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{h}(\epsilon_{1j})) \cdots \prod_{j=1}^{n_q} \exp(\alpha_q + \boldsymbol{\beta}_q' \boldsymbol{h}(\epsilon_{qj})) \\ = \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} w_1(\epsilon_{1j}) \cdots \prod_{j=1}^{n_q} w_q(\epsilon_{qj}) \end{aligned} \tag{3.29}$$

When the data are dependent, (3.29) may be viewed as a partial likelihood discussed in terms of time series in [30]. Fix $\boldsymbol{\alpha}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_q$. Then maximizing (3.29) with respect to the p_i subject to the constraints

$$\sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i[w_1(\tau_i) - 1] = 0, ..., \ \sum_{i=1}^{n} p_i[w_q(\tau_i) - 1] = 0$$

we obtain [20],

$$p_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 w_1(\tau_i) + \cdots + \rho_q w_q(\tau_i)} \tag{3.30}$$

where $\rho_j = n_j / n_m$, $j = 1, ..., q$, are the relative series sizes. Substituting the p_i in (3.30) back into (3.29) gives the log-likelihood as a function of $\boldsymbol{\alpha}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_q$ only,

$$\begin{aligned} l = -n \log n_m \ &- \ \sum_{i=1}^{n} \log[1 + \rho_1 w_1(\tau_i) + \cdots + \rho_q w_q(\tau_i)] \\ &+ \ \sum_{j=1}^{n_1} [\alpha_1 + \boldsymbol{\beta}_1 h(\epsilon_{1j})] + \cdots + \sum_{j=1}^{n_q} [\alpha_q + \boldsymbol{\beta}_q h(\epsilon_{qj})] \end{aligned} \tag{3.31}$$

By maximizing (3.31) with respect to $\boldsymbol{\alpha}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_q$ we obtain the estimators $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_q$, and by substitution,

$$\hat{p}_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 \exp(\hat{\alpha_1} + \hat{\boldsymbol{\beta}_1}' \boldsymbol{h}(\tau_i)) + \cdots + \rho_q \exp(\hat{\alpha_q} + \hat{\boldsymbol{\beta}_q}' \boldsymbol{h}(\tau_i))} \tag{3.32}$$

Therefore, with $I(B)$ the indicator of the event B,

$$\begin{aligned}\hat{G}(\tau) &= \sum_{i=1}^{n} \hat{p}_i I(\tau_i \le \tau) \\ &= \frac{1}{n_m} \cdot \sum_{i=1}^{n} \frac{I(\tau_i \le \tau)}{1 + \rho_1 \exp(\hat{\alpha_1} + \hat{\boldsymbol{\beta}}_1{}' \boldsymbol{h}(\tau_i)) + \cdots + \rho_q \exp(\hat{\alpha_q} + \hat{\boldsymbol{\beta}}_q{}' \boldsymbol{h}(\tau_i))}\end{aligned} \tag{3.33}$$

Asymptotic properties of $\hat{G}$, and its optimality over the empirical distribution function obtained only from the reference sample $\epsilon_{m1}, ..., \epsilon_{mn_m}$ ignoring all the other samples, are discussed in a sequence of papers by Zhang [48],[49],[50]. In particular, it is shown in [50] that $\sqrt{n}(\hat{G} - G)$ converges to a Gaussian process with mean zero and a rather complex covariance structure under some assumptions. In addition, it can be shown [20],[42],[48],

$$\sqrt{n}\left[(\hat{\boldsymbol{\alpha}}', \hat{\boldsymbol{\beta}}_1{}', ..., \hat{\boldsymbol{\beta}}_q{}') - (\boldsymbol{\alpha}_0', \boldsymbol{\beta}_{10}', ..., \boldsymbol{\beta}_{q0}')\right]' \Rightarrow \mathrm{N}(\mathbf{0}, \boldsymbol{S^{-1}} \boldsymbol{V} \boldsymbol{S^{-1}}) \tag{3.34}$$

where $\boldsymbol{\alpha}_0, \boldsymbol{\beta}_{10}, ..., \boldsymbol{\beta}_{q0}$ are the true values of the parameters and $\boldsymbol{S}, \boldsymbol{V}$ are given in the appendix. Here $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_q)'$.

Now, from $\hat{G}$ we can compute probabilities and quantiles as we demonstrate next in relation to VaR estimation.

3.3.3 Prediction

Since $x_{m,t+1} = f_m(\boldsymbol{z}_{m,t}) + \epsilon_{m,t+1}$ and $\epsilon_{m,t+1} \sim G$, we have the useful probability approximation at $t+1$ conditional on $\boldsymbol{z}_{m,t}$,

$$\begin{aligned}P(x_{m,t+1} \le x \mid \boldsymbol{z}_{m,t}) &= G(x - f_m(\boldsymbol{z}_{m,t})) \\ &\approx \hat{G}(x - f_m(\boldsymbol{z}_{m,t})) \\ &= \sum_{i=1}^{n} \hat{p}_i I(\tau_i \le x - f_m(\boldsymbol{z}_{m,t}))\end{aligned} \tag{3.35}$$

where we now replace the noise τ_i by actual observations. Thus, for

$$1 \le r \le m, \quad 1 \le t \le n_r, \quad n_0 \equiv 0, \quad i = n_1 + \cdots + n_{r-1} + t$$

we have $\tau_i = \epsilon_{rt} = x_{rt} - f_r(\boldsymbol{z}_{r,t-1})$, and

$$\hat{p}_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 \hat{w}_1 (x_{rt} - f_r(\boldsymbol{z}_{r,t-1})) + \cdots + \rho_q \hat{w}_q (x_{rt} - f_r(\boldsymbol{z}_{r,t-1}))} \tag{3.36}$$

From (3.35) we can get the predicted values $E(x_{m,t+1} \mid \boldsymbol{z}_{m,t})$.

3.4 Simulations and Applications of Density Ratio Model to Value at Risk

3.4.1 Simulations

To apply our method to the calculation of VaR, we need two or more datasets of financial returns, considering one of the datasets as a "reference". That is, it corresponds to the reference probability density.

Consider now two financial series P_t and $\widetilde{P}_t$. Then the corresponding returns r_t and $\widetilde{r}_t$ are relative changes defined as,

$$\ln(P_t) = \ln(P_{t-1}) + r_t \tag{3.37}$$

$$\ln(\widetilde{P}_t) = \ln(\widetilde{P}_{t-1}) + \widetilde{r}_t \tag{3.38}$$

When the temporal changes in the financial series are not large, a Taylor series expansion to one term shows that the returns are really percentage changes. We may view r_t and $\widetilde{r}_t$ as the residuals (ϵ's) of two regression models. In accordance

with the above discussion, if $\widetilde{r}_t$ has a reference probability density $\widetilde{g}(\cdot)$, and r_t has density $g(\cdot)$, then we model the relationship between the densities by

$$g(x) = \exp(\alpha + \beta x^2)\widetilde{g}(x) \tag{3.39}$$

Note the slight change in notation; here $\widetilde{g}(x)$ is the reference probability density. The choice of $h(x) = x^2$ is justified in light of the previous discussion, and by the comparison results in the tables below.

Given data from r_t and $\widetilde{r}_t$, an application of the semiparametric method yields an estimate for $\widetilde{g}(x)$, from which we get the VaR (i.e. quantile) estimates of $\widetilde{r}_t$ corresponding to fixed probabilities. From this, under additional assumptions, we can also estimate the conditional predictive distribution of $\widetilde{P}_t$. But unlike before, the main interest here is in VaR estimates for $\widetilde{r}_t$.

We compare our method with four other methods. The first is *historical simulation*, one of the most commonly used methods for VaR estimation. Suppose θ is the given probability. We choose a window (or stretch) of observations, sort the returns data from small to large, then the corresponding sample quantile is the VaR estimate.

The second method estimates VaR using a GARCH(1,1) model (e.g. see [30, p. 200] and references therein). Here it is assumed that $\widetilde{r}_t|F_{t-1} \sim \mathrm{N}(0, \sigma_t^2)$, where σ_t^2 is the conditional variance of r_t satisfying,

$$\sigma_t^2 = \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3.40}$$

This assumption provides a method to determine both the estimated volatility series and the quantile (or VaR) for a given probability. That is, the estimated VaR is the

quantile of the standard normal distribution times the estimated standard deviation at time t calculated from (3.40).

The third method is Riskmetrics model, the structure of Riskmetrics is the same as the above GARCH method. The only difference is the volatility equation of h_t. In Riskmetrics, h_t is given by:

$$h_t = \alpha r_{t-1}^2 + (1-\alpha)h_{t-1} \tag{3.41}$$

In practice, we treat $\alpha = 0.94$. Then, the estimated VaR is the quantile of the standard normal distribution times the estimated standard deviation $\sqrt{h_t}$.

The last method is to use extreme value theory to predict VaR. We talk about this method in detail in section 3.2.3.

In our simulation, we generate six different datasets. Each time, we apply the five methods to calculate VaR for one set of data. The first three datasets are generated from the same GARCH(1,1) model with the different innovation term ϵ_t. The dataset $\tilde{r}_t = \sigma_t \epsilon_t$, σ_t is in (3.40) with coefficients $(\alpha_0, \alpha_1, \beta_1) = (0.005, 0.15, 0.80)$. Then three $\tilde{r}_t = \sigma_t \epsilon_t$ *reference* series were generated similarly such that: a. ϵ_t follows the standard normal distribution and $(\alpha_0, \alpha_1, \beta_1) = (0.005, 0.15, 0.80)$. b. ϵ_t follows the standardized student-t with 3 degrees of freedom, mean 0 and variance 1, and $(\alpha_0, \alpha_1, \beta_1) = (0.005, 0.15, 0.80)$. c. ϵ_t follows the standardized Gamma(4,2) distribution with mean 0 and variance 1, and $(\alpha_0, \alpha_1, \beta_1) = (0.005, 0.15, 0.80)$. The standardization is in accordance with the requirement that the error term of GARCH models should have mean 0 and variance 1.

The fourth data set is generated from the above GARCH Gamma and GARCH

T data sets with the same coefficients. We generate the data from GARCH Gamma and GARCH T alternatively. In this way, we generate a non iid data set.

The fifth data set is generated from a TGARCH process and a standard normal distribution.TGARCH means threshold GARCH. In our TGARCH, σ_t^2 changes according to a clipped ϵ as in (3.43) below.

TGARCH can be defined as follows:

$$\widetilde{r}_t = \sigma_t \epsilon_t \tag{3.42}$$

$$\sigma_t^2 = \begin{cases} \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & : \quad \epsilon_t > 0 \\ \alpha_0 + \alpha_2 \widetilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & : \quad \epsilon_t \leq 0 \end{cases} \tag{3.43}$$

In empirical work, threshold models correspond to 'leverage effect'. In our simulation, $\alpha_0 = 0.005$, $\beta_1 = 0.80$, $\alpha_1 = 0.05$, $\alpha_2 = 0.15$. ϵ_t follows the standard Normal distribution with mean 0 and variance 1.

Then the true quantile (VaR) from the $\widetilde{r}_t$ series is the standard deviation at time t obtained from (3.40) times the quantile of the corresponding error terms ϵ_t with the indicated probabilities.

The last data set is generated similarly to the fifth. It is generated from a TGARCH process and a standardized Gamma(4,2) distribution with mean 0 and variance 1 as in 3.43. The parameters in TGARCH process is the same as the fifth data set. $\alpha_0 = 0.005$, $\beta_1 = 0.80$, $\alpha_1 = 0.05$, $\alpha_2 = 0.15$. ϵ_t follows the standardized Gamma(4,2) distribution with mean 0 and variance 1.

The plots of the simulated data and their histograms are attached in figure 3.1 to figure 3.6. In each of the six cases the corresponding GARCH models were used in generating 100 independent time series each of length 1001. The true VaR

is obtained for $t = 1001$. To apply the historical simulation method, we choose a sliding window of 300. For the GARCH and semiparametric method, we use the first 1000 data points. We would like to predict the VaR at day 1001 and compare it with true VaR on day 1001.

Recall, in our semiparametric model, $\widetilde{r}_t$ is the reference return. We make the "distorted" series r_t a simulation of stock index, or in other words, a combination of $\widetilde{r}_t$ and some other securities. To generate the "distorted" series r_t, we first generate data y by a GARCH(1,1) process with parameters $(\alpha_0, \alpha_1, \beta_1) = (0.005, 0.05, 0.90)$ times a process ϵ_t from the standard normal distribution. Then r_t is generated by $r_t = w_1 * \widetilde{r}_t + w_2 * y$, where w_1, w_2 are the weights of $\widetilde{r}_t$ and y. We assume $w_1 + w_2 = 1$. This is a combination or a portfolio of two assets $\widetilde{r}_t$ and y. If w_1 is large, it guarantees that $\widetilde{r}_t$ and r_t are highly correlated. If w_1 is small, that means $\widetilde{r}_t$ and r_t are weakly correlated. However, our model doesn't require that $\widetilde{r}_t$ and r_t are dependent. So we generate four cases of r_t with different weights: (a) $w_1 = 0.8$, $w_2 = 0.2$; (b) $w_1 = 0.2$, $w_2 = 0.8$; (c) $w_1 = 0.03$, $w_2 = 0.97$; (d) $w_1 = 0$, $w_2 = 1$;

We use two measures: mean square error (MSE), and failure rate to compare the VaR estimates obtained from the three methods: semiparametric, historical simulation, and GARCH. In the present simulation there are 100 runs and the MSE is defined as

$$\text{MSE} = (\sum_{\text{i=1}}^{100} (\text{EstimatedVaR} - \text{TrueVaR})^2)/100 \tag{3.44}$$

The MSE measure the deviation of the estimated VaR from the true VaR in the simulation. The other measure is called failure rate or exceedance ratio (ER).

It is a measure to check the adequacy of the models. Suppose p is the given small probability, n is the length of the time series, m is the length of the data with magnitude less than the VaR (if for the long position VaR is negative), if for the short position, VaR is positive, m is the length of the data with magnitude great than the VaR then ER is defined by

$$ER = m/n \tag{3.45}$$

If the sample is large enough, ER is an unbiased estimate of p, and we expect ER to be close to the given probability p. The closer to p, the better the model is. Otherwise, the model is inadequate. In our simulation tables, we write 95%, 99% to represent the short position. Under those two cases, ER should be close to small probability 5% and 1%.

The results of MSE and ER under different scenarios are given in Tables 3.1 to 3.12. In all of these simulation tables, we write 95%, 99% to represent the short position with small probability 5%, 1%. We use the notation DR to represent our semiparametric method based on density ratio, HS to represent historical simulation, GN to represent GARCH normal method, RM to represent Riskmetrics, EVT to represent the extreme value theory method.

From the tables we can make the following conclusions. For the data generated from the GARCH normal, table 3.1 and table 3.2 we see that the GARCH model and RiskMetrics are the best two models in terms of MSE. But the density ratio (DR) model is best in terms of exceedance ratio. For the data from TGARCH normal, we have similar results. The GARCH model and RiskMetrics are the best

two models in terms of MSE. But the density ratio (DR) model and historical simulation model are the best two best in terms of exceedance ratio. For the data generated from the t-GARCH, $t(3)$, table 3.3 to table 3.4, GARCH is best by MSE, whereas DR is best in terms of exceedance ratio. For the data generated from the Gamma-GARCH, gamma(4,2), table 3.5 to table 3.6, these models perform quite similar by all the measures. For the non iid data, from exceedance ratio, DR is best, from MSE, GARCH is the best. For the data generated from the threshold GARCH with $\epsilon_t \sim$ gamma$(4, 2)$, DR performs better than the other four models by the measures exceedance ratio. From the MSE measure, all the models perform quite similar. From all these simulation scenarios, the semiparametric density ratio model performs quite well according to the measure of the exceedance ratio. Notice that exceedance ratio checks the adequacy of the models. If that measure of the model deviates the given probability too much, the model is not sensible. We can say that our density ratio model is the most robust model. And this model preforms much better in nonormal cases than in normal cases. Especially in the TGARCH-Gamma case, it performs best by both the measures of bias and exceedance ratio. Considering that TGARCH-Gamma data has the features of heavy tail and leverage effects, which are some stylized facts of financial volatility we mentioned in Chapter 1, we can expect the DR model also works well in the real financial world. We will give an example of application of DR model in the next section.

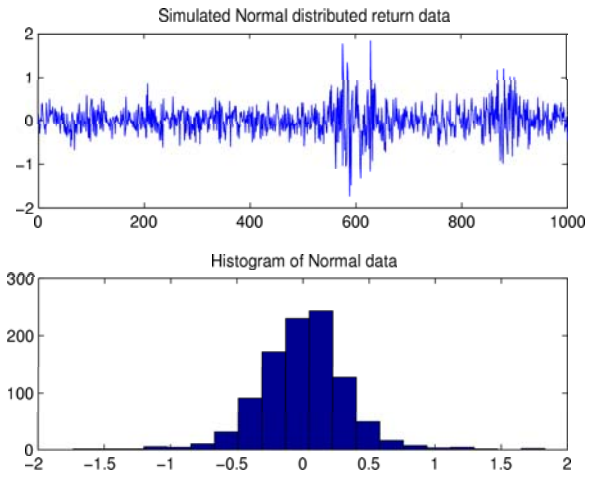

Figure 3.1: Simulated GARCH Normal returns

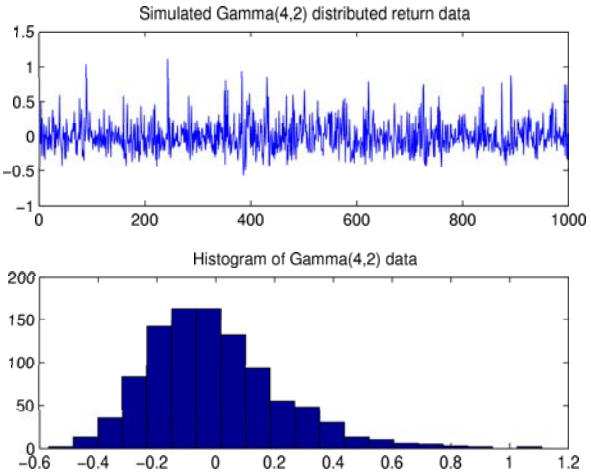

Figure 3.2: Simulated GARCH Gamma(4,2) returns

3.4.2 Applications

The previous section describes the simulation results. As an application, we apply the semiparametric method to the IBM, EXXON and COCA stock price returns individually, which are the reference series, and use the same Dow Jones Industrial

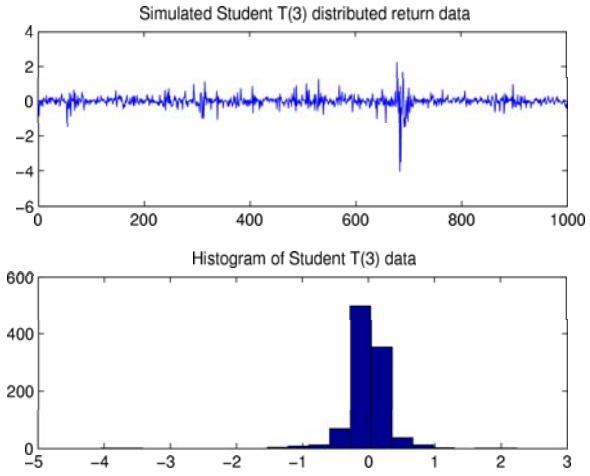

Figure 3.3: Simulated GARCH Student-t(3) returns

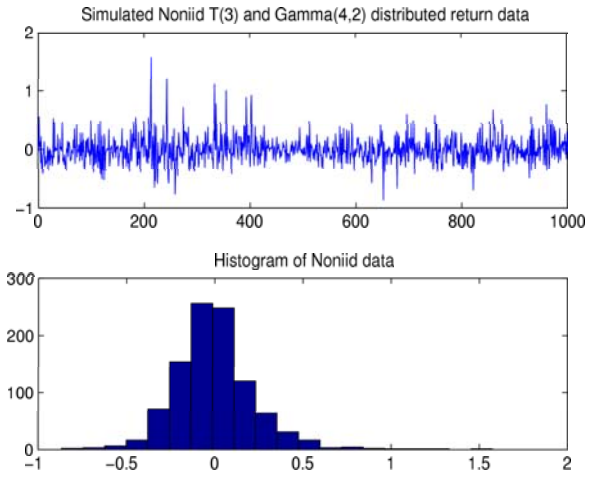

Figure 3.4: Simulated Non iid returns

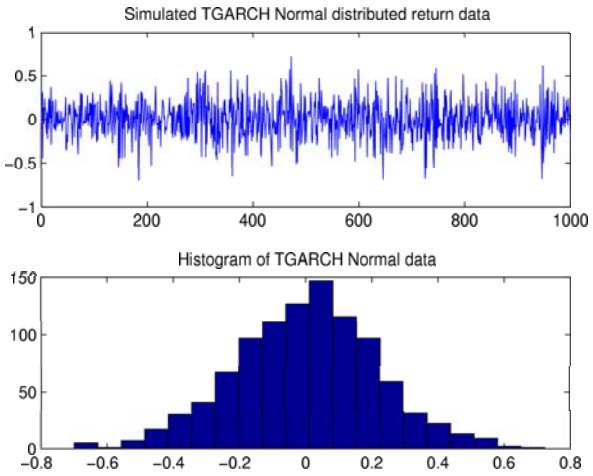

Figure 3.5: Simulated TGARCH Normal returns

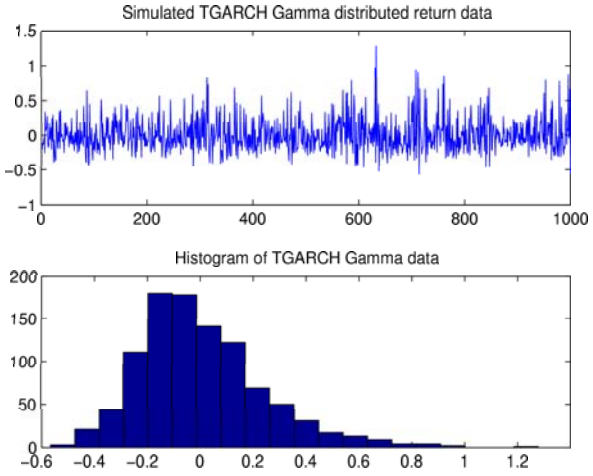

Figure 3.6: Simulated TGARCH Gamma returns

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	MSE	0.1376	0.1216	0.1070	0.1063	0.1443	0.0526	0.0659	0.0690
5%	MSE	0.0296	0.0286	0.0291	0.0293	0.0330	0.0144	0.0219	0.0219
95%	MSE	0.0270	0.0268	0.0264	0.0265	0.0296	0.0073	0.0130	0.0442
99%	MSE	0.0790	0.0808	0.0818	0.0819	0.0970	0.0467	0.0660	0.0917

Table 3.1: Normal case. MSE (3.44) measuring the distance between the true and estimated VaR.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	ER	0.0095	0.0106	0.0111	0.0111	0.0139	0.0219	0.0241	0.0152
5%	ER	0.0501	0.0507	0.0499	0.0496	0.0524	0.0598	0.0617	0.0752
95%	ER	0.0503	0.0504	0.0494	0.0492	0.0537	0.0669	0.0682	0.0750
99%	ER	0.0095	0.0106	0.0112	0.0113	0.0137	0.0278	0.0309	0.0153

Table 3.2: Normal case. ER(3.45) checks the adequacy of the models.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	MSE	0.3714	0.3471	0.3499	0.3549	0.7319	0.0641	0.1228	0.3409
5%	MSE	0.0534	0.0521	0.0545	0.0551	0.0654	0.0296	0.0325	0.0575
95%	MSE	0.0795	0.0637	0.0592	0.0587	0.0590	0.0111	0.0499	0.1009
99%	MSE	0.1203	0.1341	0.1276	0.1270	0.2222	0.0592	0.0768	0.1306

Table 3.3: t-case. MSE (3.44) measuring the distance between the true and estimated VaR.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	ER	0.0105	0.0133	0.0129	0.0128	0.0106	0.0221	0.0295	0.0135
5%	ER	0.0537	0.0390	0.0347	0.0342	0.0529	0.0463	0.0554	0.0666
95%	ER	0.0544	0.0398	0.0358	0.0352	0.0530	0.0441	0.0525	0.0684
99%	ER	0.0103	0.0135	0.0132	0.0131	0.0116	0.0216	0.0283	0.0143

Table 3.4: t-case. ER(3.45) checks the adequacy of the models.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	MSE	0.0534	0.0618	0.0688	0.0716	0.0785	0.0678	0.0552	0.0292
5%	MSE	0.0127	0.0162	0.0198	0.0208	0.0186	0.0105	0.0166	0.0115
95%	MSE	0.0256	0.0248	0.0245	0.0247	0.0279	0.0085	0.0142	0.0318
99%	MSE	0.1436	0.1293	0.1390	0.1382	0.1296	0.0669	0.0906	0.1130

Table 3.5: Gamma case. MSE (3.44) measuring the distance between the true and estimated VaR.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	ER	0.0116	0.0072	0.0058	0.0055	0.0141	0.0098	0.0134	0.0182
5%	ER	0.0581	0.0385	0.0321	0.0312	0.0535	0.0433	0.0460	0.0891
95%	ER	0.0480	0.0558	0.0571	0.0571	0.0526	0.0782	0.0802	0.0736
99%	ER	0.0087	0.0125	0.0160	0.0166	0.0139	0.0358	0.0406	0.0154

Table 3.6: Gamma case. ER (3.45) checks the adequacy of the models.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	MSE	0.2435	0.2618	0.2172	0.2177	1.1790	0.0760	0.0901	0.0614
5%	MSE	0.0310	0.0305	0.0326	0.0331	0.0305	0.0165	0.0156	0.0335
95%	MSE	0.0443	0.0424	0.0414	0.0416	0.0683	0.0080	0.0172	0.0538
99%	MSE	0.1347	0.1339	0.1303	0.1297	0.4364	0.0730	0.0726	0.0866

Table 3.7: Non iid case. MSE (3.44) measuring the distance between the true and estimated VaR.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	ER	0.0102	0.0105	0.0097	0.0096	0.0149	0.0170	0.0209	0.0158
5%	ER	0.0551	0.0387	0.0337	0.0331	0.0562	0.0432	0.0527	0.0777
95%	ER	0.0520	0.0501	0.0468	0.0464	0.0520	0.0612	0.0701	0.0669
99%	ER	0.0097	0.0142	0.0159	0.0160	0.0123	0.0318	0.0368	0.0139

Table 3.8: Non iid case. ER (3.45) checks the adequacy of the models.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	MSE	0.0445	0.0365	0.0347	0.0341	0.0404	0.0309	0.0363	0.0337
5%	MSE	0.0094	0.0088	0.0090	0.0090	0.0091	0.0072	0.0087	0.0057
95%	MSE	0.0115	0.0104	0.0101	0.0101	0.0119	0.0041	0.0057	0.0140
99%	MSE	0.0316	0.0306	0.0302	0.0302	0.0362	0.0243	0.0294	0.0351

Table 3.9: TGARCH Normal case. MSE (3.44) measuring the distance between the true and estimated VaR.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	ER	0.0097	0.0115	0.0119	0.0122	0.0127	0.0169	0.0175	0.0125
5%	ER	0.0495	0.0511	0.0508	0.0507	0.0525	0.0546	0.0593	0.0620
95%	ER	0.0505	0.0461	0.0452	0.0452	0.0508	0.0490	0.0534	0.0581
99%	ER	0.0101	0.0097	0.0096	0.0096	0.0122	0.0132	0.0153	0.0115

Table 3.10: TGARCH Normal case. ER (3.45) checks the adequacy of the models.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	MSE	0.0068	0.0258	0.0312	0.0312	0.0101	0.0319	0.0326	0.0047
5%	MSE	0.0028	0.0062	0.0073	0.0074	0.0030	0.0063	0.0080	0.0028
95%	MSE	0.0068	0.0077	0.0079	0.0080	0.0082	0.0046	0.0056	0.0083
99%	MSE	0.0337	0.0463	0.0507	0.0510	0.0418	0.0343	0.0368	0.0334

Table 3.11: TGARCH Gamma case. MSE (3.44) measuring the distance between the true and estimated VaR.

p		DR(a)	DR(b)	DR(c)	DR(d)	HS	GN	RM	EVT
1%	ER	0.0115	0.0027	0.0020	0.0020	0.0134	0.0033	0.0057	0.0141
5%	ER	0.0568	0.0270	0.0238	0.0237	0.0520	0.0283	0.0378	0.0713
95%	ER	0.0472	0.0586	0.0592	0.0593	0.0507	0.0693	0.0693	0.0596
99%	ER	0.0090	0.0181	0.0201	0.0203	0.0121	0.0289	0.0314	0.0123

Table 3.12: TGARCH Gamma case. ER (3.45) checks the adequacy of the models.

Average log-returns for the same period as the "distorted" series. The period is from January 2002 to September 2003. Since the true VaR is unknown, we only use one measure *exceedance ratio* (3.45) to test our model. We first test the adequacy of the models by ER measure. If the sample is large enough, ER is an unbiased estimate of p, and we expect ER to be close to the given probability p. The closer to p, the better the model is. Otherwise, the model is inadequate. In our result tables, we write 95%, 99% to represent the short position with probability 5%, 1%. Under those two cases, ER should be close to small probability 5% and 1%. We expect the ER to be not far from the given probability. If it's too far away, the model would be inadequate. From [25], the Basel Committee of International Settlement Bank has a rule for the ER measure. If ER is too far away, the volatility should be penalized by multiply by a factor coefficient to make the VaR larger than the estimated. The second measure is expected shortfalls. That measure is to test the conditional expected loss of the models.

We use all the five models in the application of real data. The results are in the tables 3.13 to 3.18. The notations are the same as that in the previous section. DR means the semiparametric density ratio model, HS means historical simulation model (we select a period length of 250 in this model) , GN means GARCH normal model, RM means RiskMetrics model, EVT means extreme value theory model. From the tables we can see that, for the measure of ER, DR model is almost always the best for all the three data sets. Comparatively, GN and RM are not adequate enough. The ER of them sometimes are too far away. For example, for IBM with probability 1%, the ER for GN is 0.04 and ER for RM is 0.07, but we expect it to

be around 0.01. In that case, ER for DR is 0.0091. As a consequence, GN and RM tend to underestimate the true VaR. Some other literature also verifies this. As for the other measure of expected shortfalls, although it's widely used, we can not see which model performs uniformly better from this measure. It's also interesting to see that EVT performs better with a small probability than with a larger one.

Given p		DR	HS	GN	RM	EVT
1%	VaR	-0.0629	-0.0537	-0.0395	-0.0308	-0.0582
5%	VaR	-0.0352	-0.0286	-0.0281	-0.0218	-0.0326
95%	VaR	0.0376	0.0334	0.0267	0.0218	0.0297
99%	VaR	0.0692	0.0768	0.0381	0.0308	0.0521

Table 3.13: VaR for IBM.

Given p		DR	HS	GN	RM	EVT
1%	ER	0.0091	0.0205	0.0410	0.0729	0.0137
5%	ER	0.0456	0.0866	0.0934	0.1298	0.0592
95%	ER	0.0456	0.0592	0.0888	0.1276	0.0706
99%	ER	0.0114	0.0068	0.0456	0.0661	0.0137

Table 3.14: ER of VaR for IBM. ER (3.45) checks the adequacy of the models.

Given p		DR	HS	GN	RM	EVT
1%	VaR	-0.0452	-0.0344	-0.0224	-0.0196	-0.0359
5%	VaR	-0.0266	-0.0194	-0.0158	-0.0139	-0.0194
95%	VaR	0.0268	0.0220	0.0162	0.0139	0.0190
99%	VaR	0.0476	0.0358	0.0229	0.0196	0.0384

Table 3.15: VaR for EXXON.

Given p		DR	HS	GN	RM	EVT
1%	ER	0.0091	0.0251	0.0729	0.0934	0.0205
5%	ER	0.0433	0.0957	0.1367	0.1708	0.0957
95%	ER	0.0410	0.0661	0.1162	0.1708	0.0934
99%	ER	0.0068	0.0205	0.0615	0.0888	0.0159

Table 3.16: ER of VaR for EXXON. ER (3.45) checks the adequacy of the models.

Given p		DR	HS	GN	RM	EVT
1%	VaR	-0.0411	-0.0339	-0.0255	-0.0203	-0.0380
5%	VaR	-0.0256	-0.0224	-0.0181	-0.0144	-0.0192
95%	VaR	0.0256	0.0228	0.0178	0.0144	0.0258
99%	VaR	0.0432	0.0389	0.0252	0.0203	0.0398

Table 3.17: VaR for Coca Cola.

Given p		DR	HS	GN	RM	EVT
1%	ER	0.0091	0.0182	0.0547	0.0934	0.0114
5%	ER	0.0547	0.0774	0.1071	0.1390	0.0957
95%	ER	0.0456	0.0592	0.1071	0.1526	0.0456
99%	ER	0.0046	0.0137	0.0501	0.0843	0.0114

Table 3.18: ER of VaR for Coca Cola. ER (3.45) checks the adequacy of the models.

3.5 Conclusion

We discussed the extension of a certain semiparametric method originally intended for independent random samples to time series prediction by assuming additive noise. Using the combined data from several time series and from accompanying known covariate data we have shown how to estimate the true probability distribution of a "reference" time series and use it in conditional prediction.

When the form of the regression functions f_i in (3.22) is known up to some parameters, the parameters can be estimated by least squares, by maximum likelihood, or by the method of moments as was indicated in the bivariate AR example. Another way to estimate the unknown parameters in f_i is to use the profile log-likelihood (3.31). When the f_i are estimated, the semiparametric procedure can be applied with the residual innovations replacing the unobserved noise as was done in the mortality example.

The noise sequences need not consist of independent data, and the likelihood can be replaced by partial likelihood, as argued in [30], as long as each noise sequence consists of identically distributed random variables. This leads to a similar analysis

as above.

The potential of the semiparametric approach in financial time series applications was demonstrated in VaR estimation where we combined or fused return information from two sources. It seems that the method is able to accommodate a wide spectrum of scenarios ranging from normal to highly non-normal processes.

It is important to note that we may combine the residuals of the main regression equation used in prediction with other data samples which are not necessarily residuals. This means that in (3.22) we may retain only the time series of interest

$$x_{mt} = f_m(\boldsymbol{z}_{m,t-1}) + \epsilon_{mt}, \quad t = 1, ..., n_m$$

and combine its residuals with data from other sources assuming the tilt model (3.23), for the purpose of estimating the distribution $G(x)$ of ϵ_{mt}.

The method depends on the choice of a distortion function $\boldsymbol{h}(x)$. Our experience with geophysical data indicates that for skewed data $h(x) = \log(x)$ is quite appropriate, while $h(x) = x$ is suitable for symmetrically distributed data with different means but close variances [29],[40]. In logistic discriminant analysis $h(x) = x$ or $\boldsymbol{h}(x) = (x, x^2)$ [48]. The problem of choosing $h(x)$ has been considered recently in [19] using the so called Box–Cox transformation. However the issue of estimating $h(x)$ in general is still open.

The question of goodness of fit can be tackled by measuring the discrepancy between $\hat{G}$ and the empirical distribution form the mth (reference) sample. This is studied in [42],[50].

Appendix A

Derivation of $\boldsymbol{S, V}$

The matrices $\boldsymbol{S, V}$ are derived by repeated differentiation of (3.31). Recall that the asymptotic covariance matrix of the estimates in (3.34) is given by the product

$$\boldsymbol{\Sigma} = \boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{S}^{-1} \tag{A.1}$$

First define

$$\nabla \equiv \left(\frac{\partial}{\partial\alpha_1}, ..., \frac{\partial}{\partial\alpha_q}, \frac{\partial}{\partial\boldsymbol{\beta}_1}..., \frac{\partial}{\partial\boldsymbol{\beta}_q}\right)' \tag{A.2}$$

Then $E[\nabla l(\alpha_1, ..., \alpha_q, \boldsymbol{\beta}_1, .., \boldsymbol{\beta}_q)] = \mathbf{0}$. To obtain the score second moments it is convenient to define $\rho_m \equiv 1,\ w_m(t) \equiv 1$,

$$E_j[\boldsymbol{h}(t)] \equiv \int \boldsymbol{h}(t) w_j(t) dG(t) \tag{A.3}$$

and,

$$A_0(j, j') \equiv \int \frac{w_j(t) w_{j'}(t) dG(t)}{1 + \sum_{k=1}^{q} \rho_k w_k(t)} \tag{A.4}$$

$$\boldsymbol{A}_1(j, j') \equiv \int \frac{\boldsymbol{h}(t) w_j(t) w_{j'}(t) dG(t)}{1 + \sum_{k=1}^{q} \rho_k w_k(t)} \tag{A.5}$$

$$\boldsymbol{A}_2(j, j') \equiv \int \frac{\boldsymbol{h}(t)\boldsymbol{h}'(t) w_j(t) w_{j'}(t) dG(t)}{1 + \sum_{k=1}^{q} \rho_k w_k(t)} \tag{A.6}$$

for $j, j' = 1, ..., q$. Then, the entries in

$$\boldsymbol{V} \equiv Var\left[\frac{1}{\sqrt{n}} \nabla l(\alpha_1, ..., \alpha_q, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_q)\right] \tag{A.7}$$

are,

$$\frac{1}{n}Var\left(\frac{\partial l}{\partial \alpha_j}\right) = \frac{\rho_j^2}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j) - \sum_{r=1}^m \rho_r A_0^2(j,r)\} \quad (A.8)$$

$$\frac{1}{n}Cov\left(\frac{\partial l}{\partial \alpha_j}, \frac{\partial l}{\partial \alpha_{j'}}\right) = \frac{\rho_j \rho_{j'}}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j') - \sum_{r=1}^m \rho_r A_0(j,r)A_0(j',r)\} \quad (A.9)$$

$$\frac{1}{n}Cov\left(\frac{\partial l}{\partial \alpha_j}, \frac{\partial l}{\partial \boldsymbol{\beta}_j}\right) = \frac{\rho_j^2}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j)E_j[\boldsymbol{h}'(t)] - \sum_{r=1}^m \rho_r A_0(j,r)\boldsymbol{A}_1'(j,r)\} \quad (A.10)$$

$$\frac{1}{n}Cov\left(\frac{\partial l}{\partial \alpha_j}, \frac{\partial l}{\partial \boldsymbol{\beta}_{j'}}\right) = \frac{\rho_j \rho_{j'}}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j')E_{j'}[\boldsymbol{h}'(t)] - \sum_{r=1}^m \rho_r A_0(j,r)\boldsymbol{A}_1'(j',r)\} \quad (A.11)$$

$$\begin{aligned}\frac{1}{n}Cov\left(\frac{\partial l}{\partial \boldsymbol{\beta}_j}, \frac{\partial l}{\partial \boldsymbol{\beta}_{j'}}\right) &= \frac{\rho_j \rho_{j'}}{1+\sum_{k=1}^q \rho_k}\{-\boldsymbol{A}_2(j,j') + E_j[\boldsymbol{h}(t)]\boldsymbol{A}_1'(j,j') \\ &+ \boldsymbol{A}_1(j,j')E_{j'}[\boldsymbol{h}'(t)] \\ &- \sum_{r=1}^m \rho_r \boldsymbol{A}_1(j,r)\boldsymbol{A}_1'(j',r)\} \\ &+ \frac{1}{n}\sum_{i=1}^{n_j}\sum_{k=1}^{n_{j'}} Cov[\boldsymbol{h}(\epsilon_{ji}), \boldsymbol{h}(\epsilon_{j'k})] \end{aligned} \quad (A.12)$$

The last term is 0 for $j \neq j'$ and $(n_j/n)Var[\boldsymbol{h}(\epsilon_{j1})]$ for $j = j'$.

Next, as $n \to \infty$,

$$-\frac{1}{n}\nabla\nabla' l(\alpha_1, ..., \alpha_q, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_q) \to \mathbf{S} \quad (A.13)$$

where $\mathbf{S}$ is a $q(1+p) \times q(1+p)$ matrix with entries corresponding to $j, j' = 1, ..., q$,

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j^2} \to \frac{\rho_j}{1+\sum_{k=1}^q \rho_k}\int \frac{[1+\sum_{k\neq j}^q \rho_k w_k(t)]w_j(t)}{1+\sum_{k=1}^q \rho_k w_k(t)} dG(t) \quad (A.14)$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j \alpha_{j'}} \to \frac{-\rho_j \rho_{j'}}{1+\sum_{k=1}^q \rho_k}\int \frac{w_j(t)w_{j'}(t)}{1+\sum_{k=1}^q \rho_k w_k(t)} dG(t) \quad (A.15)$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j \partial \boldsymbol{\beta}_j'} \to \frac{\rho_j}{1+\sum_{k=1}^q \rho_k}\int \frac{[1+\sum_{k\neq j}^q \rho_k w_k(t)]w_j(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^q \rho_k w_k(t)} dG(t) \quad (A.16)$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial\alpha_j\partial\boldsymbol{\beta}'_{j'}} \;\to\; \frac{-\rho_j\rho_{j'}}{1+\sum_{k=1}^{q}\rho_k}\int\frac{w_j(t)w_{j'}(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^{q}\rho_k w_k(t)}dG(t) \tag{A.17}$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}'_j} \;\to\; \frac{\rho_j}{1+\sum_{k=1}^{q}\rho_k}\int\frac{[1+\sum_{k\neq j}^{q}\rho_k w_k(t)]w_j(t)\boldsymbol{h}(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^{q}\rho_k w_k(t)}dG(t) \tag{A.18}$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}'_{j'}} \;\to\; \frac{-\rho_j\rho_{j'}}{1+\sum_{k=1}^{q}\rho_k}\int\frac{w_j(t)w_{j'}(t)\boldsymbol{h}(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^{q}\rho_k w_k(t)}dG(t) \tag{A.19}$$

BIBLIOGRAPHY

[1] Torben Anderson., Tim Bollerslev., Francis X. Diebold and Paul Labys, Modeling and Forecasting Realized Volatility, forthcoming Econometrica

[2] Black, F. (1976). The pricing of commodity contracts. *The Journal of financial economics*, 3, 167-179.

[3] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, 31, 307-327.

[4] Bollerslev, T. and Anderson, T. (1998) DM-Dollar Volatility: Intraday Activity Paterns, Macroeconomic Annoucements and Longer-Run Dependencies. *Journal of Finance*, 51(1), 219-265.

[5] Campbell, J.Y., Lo, A.W, and Mackinlay, A.C. (1997). *The Econometrics of Financial Markets*. Princeton University Press: Princeton.

[6] Cox, D.R. (1975). Partial Likelihood, *Biometrica*, 62, 269-276

[7] Duffie, D. and Pan J. (1997). An overview of value at risk. *The Journal of Derivatives*, Spring, 7-49.

[8] Easley, D. and O'Hara, M. (1992). Time and the process of security price adjustment. *Journal of Finance*, 47, 577-605.

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

[10] Engle, R., Ito, T. and Lin Wen-Ling. (1990) Meteor showers or heat waves? Heteroscedastic intra-daily volatility in the foreign exchange market. *Econometrica*, 58(3), 525-542.

[11] Engle, R., Lilien, D. and Robins, R (1987) Estimation of time varying risk premia in the term structure: ARCH-M Model. *Econometrica*, 55, 391-407.

[12] Engle, R. and Manganelli, S. (1999). CAViaR: Conditional Value at Risk by quantile regression. NBER, working paper 7342

[13] Engle, R., NG, V.K. and Rothschild,M. (1990). Asset pricing with a Factor-ARCH covariance structure. *Journal of Econometrics*, 45(2), 235-237.

[14] Engle, R., and NG, V.K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5), 1779-1801.

[15] Engle, R. and Pattern, A. (2001). What good is a volatility model? *Quantitative Finance*, 1(2), 237-245

[16] Fama, E.F. (1963) Mandelbrot and the Stable Paretian Distribution, *Journal of Business*, 36, 420-429.

[17] Fama,E.F. (1965) The behavior of stock market prices,*Journal of Business*, vol.38, 34-105.

[18] Fokianos, K.(2004). Merging information for semiparametric density estimation. *Journal of Royal Statistics Society*, **66**, 941-958.

[19] Fokianos, K. and Kaimi, I. (2004). On the effect of misspecifying the density of ratio model. Submitted.

[20] Fokianos, K., Kedem, B., Qin, J., and Short, D. A. (2001). A semiparametric approach to the one-way layout. *Technometrics*, **43**, 56-65.

[21] Gagnon, R., Kedem, B., and Qi, Y. (2004). On the efficiency of a semiparametric approach to the one-way layout. Submitted.

[22] Gilbert, P.B., Lele, S.R., Vardi, Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, 86, 27-43.

[23] Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42, 281-300.

[24] Jenkins, G.M. and Watts, D.G. (1968). *Spectral Analysis and its Application.* San Francisco: Holden-Day.

[25] Jorion, P. (2000). *Value at Risk: The New Benchmark for Managing FinancialRisk*

[26] Kay, R., and Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, **74**, 495-501.

[27] Karatzas,Ioannis and Shreve, Steven E. (1998). *Methods of Mathematical Finance*, Springer-Verlag, New York

[28] Kedem, B., Gagnon, R. and Guo, H. (2005). *Time Series Prediction via Density Ratio Modelling*, submitted.

[29] Kedem, B., Wolff, D.B., and Fokianos, K. (2004). Statistical Comparison of Algorithms. *IEEE Trans. Instrum. Meas.*, Vol. 53, 770-776.

[30] Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*, New York: Wiley.

[31] Bera, Anil K. and Higgins, Matthew L. and Lee, Sangkyu. (1992). Interaction between Autocorrelation and Conditional Heteroscedasticity: A Random-Coefficient Approach. *Journal of Business and Economic Statistics*, 10(2), 133-42.

[32] Ludger Hentschel, (1995), All in the family: Nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics*, 39, 71-104.

[33] Mahieu,R.J. and Schotman, P. (1998). An empirical application of stochastic volatility models, em Journal of Applied Econometrics), 13(4), 333-360

[34] Mandelbrot, B. (1963). The variation of certain speculative prices, *Journal of Business,36* ,394-416

[35] Morgan, J.P. (1996), Riskmetrics, Technical Document, 4th Edition, New York

[36] Nelder, JA. and Wedderburn, RWM. Generalized Linear Models, *Journal of the Royal Statistical Society* A 1972, 132, 107-120.

[37] Nelson, D. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.

[38] Nelson, D. and Foster, D. (1994). Assymptotical filtering theory for univariate ARCH models, *Econometrica*, 62, 1-41

[39] Patil, G.P., Rao, C.R., and Zelen, M. (1988). Weighted distributions. In *Encyclopedia of Statistical Sciences*, 9, 565-71, S. Kotz and N.L. Johnson Eds., New York: Wiley.

[40] Qi, Y. (2002). Classification of Microarray Data. MA Thesis, Department of Mathematics, University of Maryland, College Park.

[41] Qin, J., and J.F. Lawless (1994), "Empirical likelihood and general estimating equations," *The Annals of Statistics*, **22**, 300-325.

[42] Qin, J., and Zhang, B. (1997). A goodness of fit test for logistic regression models based on case-control data. in *Biometrika*, **84**, 609-618.

[43] Shephard, N.G. (1996) Statistical aspects of ARCH and stochastic volatility, in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielson (eds), *Time series Models*, Chapman and Hall, London.

[44] Taylor, S. *Modelling Financial Time Series*. Wiley: Chichester.

[45] Tsay, R. (2003). *Analysis of Financial Time Series*. Wiley: New York.

[46] Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *Annals of Statistics*, Vol. 10, 616-20.

[47] Vardi, Y. (1985). Empirical distribution in selection bias models. *Annals of Statistics*, Vol. 13, 178-203.

[48] Zhang, B. (2000a). M-estimation under a two sample semiparametric model. *Scand. Jour. of Statistics*, 27, 263-280.

[49] Zhang, B. (2000b). Quantile estimation under a two-sample semi-parametric model. *Bernoulli*, 6, 491-511.

[50] Zhang, B. (2000c). A goodness of fit test for multiplicative-intercept risk models based on case-control data. *Statistica Sinica*, 10, 839-865.

[51] Zakoian, Jean-Michel. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5), 931-955.