Thesis

SimonHu

26th May 2004

Abstract

The aim of this project has two parts: a) is to use the method of Nonparametric Regression to estimate the variance of $logreturn$ at time t locally,by concerning on calculating the weighted average of r_t^2 in the two windows (whose length is h (scale factor) each) on both sides of the point at time t (expectation of $logreturn$) based on:

$$r_t^2 = \mu_t + \epsilon_t \tag{1}$$

and check the volatilities of market model and the sample autocorrelations of $\frac{r_t}{\sqrt{\mu(t)}}$ and $|\frac{r_t}{\sqrt{\mu(t)}}|$ (Here, i am going to use two different datasets:S&P 500 and Dow Jones, b) after that, it is to estimate the market model by using the joint bivariate normal distribution and linearity of the function:

$$E(\tilde{r}_{it}|r_{mt}) = \alpha_i + \beta_i r_{mt} \tag{2}$$

[1] (Here,i am going to use these two different datasets (S&P 500 and Dow Jones) and IBM stock price to calculate $\tilde{r}_{it}$ and r_{mt})

[1]We include a tilde()over the symbol to identify a random variable,when we refer to a specific value of the variable,the tilde is dropped.

Contents

Chapter 1

Introduction

The study of stock market has been popular till now. In studying this field, we gained a lot of knowledge regarding *Nonparametric Regression* method and $Non - Stationary\ Multivariate\ Model$ for *Financial Returns*. In the real world, actually we have a great amount of data, but we could not use the whole dataset to do the estimation because it is too expensive and time consuming, then the *Nonparametric regression* function emerged:

$$Y_i = m(X_i) + \epsilon_i \tag{1.1}$$

and the corresponding fucntions regarding kernel function and weight function and Leave-One-Out method, ct. Here, we need to know a few things about *Smoothing Techniques* such as (a) which kernel we use and (b) which kind of *Smoothing* we use.
Here we are going to use two different kernels: *Epanechnilov kernel* and *Normal Kernel* and the smoothing method: $k - NN$. And the next step is to construct the Leave-One-Out function to get two different h values, then construct the mean function of variances of daily logreturns (Actually the expectation of variance of daily logreturns) to check the estimation since we have to care about two most important factors: (*bias* and *variance*) and we have to balance between these two factors.
In estimating the market model and mastering the regulations and rules of the stock market, we have to know the behaviors of stock market, which means we have to know something about the stock market $returns(r_{it}, r_{mt})$ and use the $returns(r_{it})$ on one security and on the index (r_{mt}) to do the

multivariate analysis.
We will consider non-linear dynamics of modelling of financial returns here, usually, the sample autocorrelation functions (SACF) of such time series are of strong evidence of the non-linear nature. For example, the daily logreturns of the closing prices of the S&P 500 from January 1,1988 to November 21,2003 will result in time series of 4009 observations (the same dataset will be used to estimate the market model). It is obvious that returns show almost no autocorrelations at all lags, while the absolute returns have higher correlations over several hundred lags (which we call it long memory in volatility) (see Figure 1.1). From the figure, we see that the SACF of the absolute returns remains almost constant after declining quite fast for the first few lags, showing the evidence of long-range dependence in the time series of absolute returns.

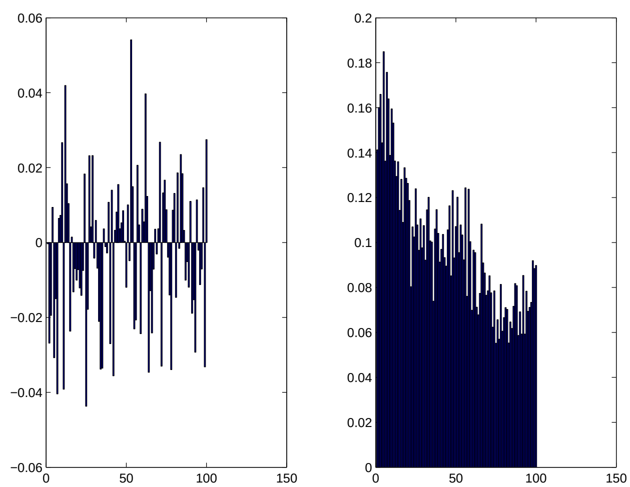

Figure 1.1: SACF of S&P 500 logreturns (Left) and absolute logreturns (Right). We did not show any confidence intervals for the correlation because of unknown dependency structure

But, before we start to estimate the market model, we are going to explain something regarding bivariate normality and linearity between individual security and index.
Define a new random variable,

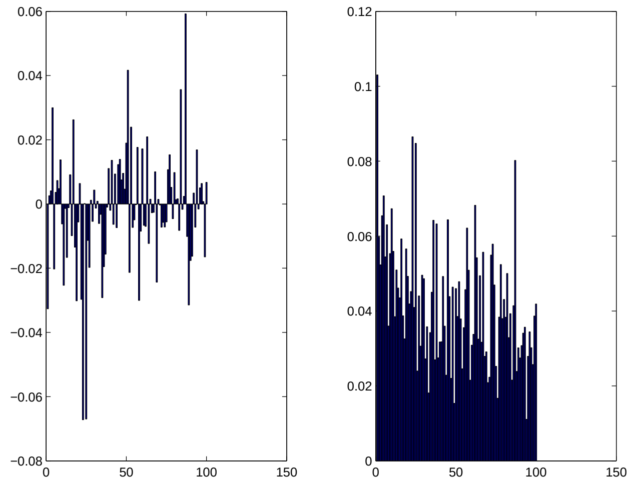

Figure 1.2: SACF of Dow Jones logreturns (Left) and absolute logreturns (Right). We did not show any confidence intervals for the correlation because of unknown dependency structure

$$\tilde{y} = \sum_{i=1}^{n} a_i \tilde{y}_i \tag{1.2}$$

which is a linear combination, that is the sum of weighted values of $\tilde{y}_1,...,\tilde{y}_n$. Assuming that each of linear combination of $\tilde{y}_i$ is normal (for any choice of weights $(a_1, ..., a_n)$, then the joint distribution of $\tilde{y}_1,...,\tilde{y}_n$ is multivariate normal. There is one property of multivariate normal distribution that we can use. If the joint distribution of $\tilde{R}_1,...,\tilde{R}_n$ is multivariate normal, then the joint distribution of any two different linear combination of $\tilde{R}_1,...,\tilde{R}_n$ is bivariate normal and this implies that the joint distribution of the return on one security and on one "market" is bivariate normal. Now, we are going to check the normality of R_{it}(the return of IBM stock) and R_{mt}(the return of two "market" indexes: S&P 500 and Dows), here we will use normality plot and QQ plot as two techniques. The following figures are plotted based on S&P 500 and IBM stock price and the last two figures are plotted based on Dow Jones and IBM stock price.

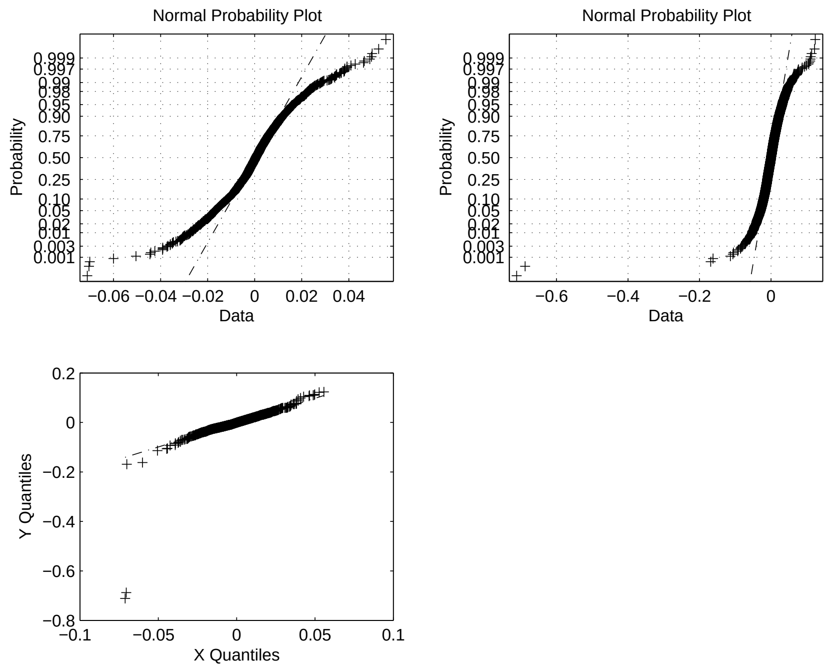

Figure 1.3: Normality plots of S&P 500 and IBM stock (top) and QQ plot of S&P 500 and IBM stock

The rest of the thesis is organized as:
In chapter 2, we present a few equations which are indispensable for the construction of the *Nonparametric Regression* functions like *weight* function and *density* function, etc.
In chapter 3, we present and introduce the regression model for multivariate returns.
In chapter 4, we check the fit of the market model by plotting different aspects of outcomes.
In chapter 5, we make a conclusion regarding what we do and employment of the methodology and acknowledge some things

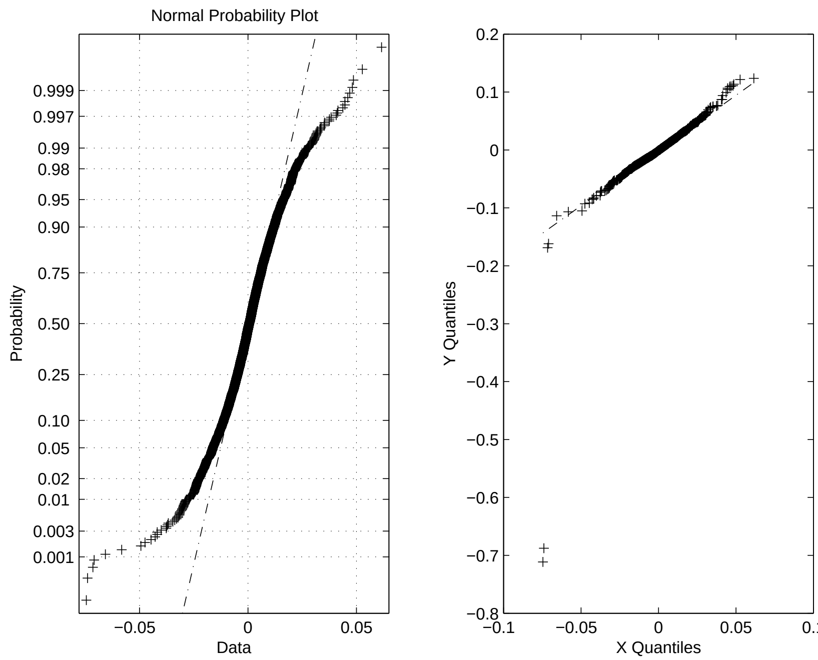

Figure 1.4: Normality plot of Dows 500 (left) and QQ plot of Dow Jones and IBM (right)

Chapter 2

Nonparametric Regression Equations

Here, the major regression smoothing method is to be given in the following and we are going to use a few corresponding equations.
First of all, we use *Epanechnikov kernel* which is:

$$K(u) = 0.75(1 - u^2)I(|u| \leq 1) \tag{2.1}$$

The corresponding weight function and density function are the following:

$$K_{h_n}(u) = {h_n}^{-1}K(u/h_n) \tag{2.2}$$

$$\hat{f}_{h_n}(x) = n^{-1}\sum_{i=1}^{n} K_{h_n}(x - X_i) \tag{2.3}$$

$$W_{ni}(x) = K_{h_n}(x - X_i)/\hat{f}_{h_n}(x) \tag{2.4}$$

Since we know the kernel function and the corresponding weight function and the density function, we are going to construct the *Leave-One-Out* expression based on those and use this expression to construct the *cross-validation*

function with the modified smoothers.
The jth observation in the whole dataset is left out:

$$\hat{m}_{h,j}(X_j) = n^{-1} \sum_{i \neq j} W_{hi}(X_j) Y_i \tag{2.5}$$

The cross-validation function is:

$$CV(h) = n^{-1} \sum_{j=1}^{n} [Y_j - \hat{m}_{h,j}(X_j)]^2 W(X_j) \tag{2.6}$$

We minimize the above $CV(h)$ to get the optimal value of h in order to construct $Nadaraya - Watson\ estimator$, finally the regression function is:

$$\hat{m}_h = \frac{n^{-1} \sum_{i=1}^{n} K_h(x - X_i) Y_i}{n^{-1} \sum_{i=1}^{n} K_h(x - X_i)} \tag{2.7}$$

The *shape* of the kernel weights is determined by K, whereas the *size* of the weights is parameterized by h.
The regression relationship can be modeled as the following:

$$Y_i = m(X_i) + \epsilon_i, i = 1, ..., n, \tag{2.8}$$

In equation 2.8, Y_i represents *square* of *logreturn* and $m(X_i)$ represents the expectation of *square* of *logreturn*.
Then we use the *Normal Kernel*:

$$K(u) = (2\pi)^{-0.5} \exp(-u^2/2) \tag{2.9}$$

For the weight function and density function of *Normal Kernel*, we have the similar forms as the ones displayed above, we will not display them here once more,please refer to the previous equations.

Chapter 3

Regression Model For Multivariate Returns

As we assumed above, we could use the bivariate normal distribution of $\tilde{R}_{it}$ and $\tilde{R}_{mt}$ and the linear relationship between $\tilde{R}_{it}$ and $\tilde{R}_{mt}$ to construct the matrix form of the multivariate analysis.
Denote R_t the $m \times 1$ dimensional vector of returns at time t, and we concentrate on the models with constant mean μ:

$$(R_t - \mu) - \phi(R_{t-1} - \mu) = U_t \tag{3.1}$$

Where ϕ is a diagonal matrix of auto-regressive parameters and (U_t) is the sequence of innovations.
We assume (U_t) to be a $non-stationary$ sequence of *independent* random vectors. The distribution of U_t is characterized by a changing covariance structure which is embodiment of complex market conditions. Here, we choose to model the covariance as a deterministic, smooth function of time and the approach leads to the following model:

$$U_t = S(t)\epsilon_t, t = 1, 2, ...n, where \tag{3.2}$$

$S(t)$ is an invertible, lower-triangular matrix and a smooth, deterministic function of time,

ϵ_t is iid sequence of random vectors with mutually independent coordinates, such that $E(\epsilon_t) = 0$, $Var(\epsilon_t) = I_m$
We should emphasize that random effects play role in volatility dynamics and this modelling approach only reflects both recent past and close future returns are embodiment of the same unspecified, exogenous economic factors. In our model, the economic factors are of the unconditional variance and our methodology quantifies the expression of these economic factors in the recent past (due to moment t) prices.
In order to fit this regression model to a time series of returns, we obtain the estimated innovations:

$$\hat{U}_t = (R_t - \bar{R}) - \hat{\phi}(R_{t-1} - \bar{R}), t = 1, 2, ..., n \tag{3.3}$$

Where $\bar{r}_i = n^-1 \sum_{i=1}^{n} r_{i,t}$ is the natural estimator for the mean μ and $\hat{\phi}$ is the $m \times m$ diagonal matrix whose non-zero entries are the auto-regression coefficients estimated-wise. Note that the estimated innovations $\hat{U}_t$ are supposed to be independent with covariance matrix $S(t)S'(t)$, a smooth function of t and the function S(t)S'(t) can be estimated by the standard non-parametric regression method by using the series $\hat{U}_t\hat{U}_t', t = 1, 2, ...n$.
The last step is to model the distribution of the estimated standardized innovations which are defined as:

$$\hat{\epsilon}_t = S^{-1}(t)\hat{U}_t, t = 1, 2, ..., n(3.4)$$

In our case we have the following model:

$$R = \Sigma\epsilon \tag{3.5}$$

$$R_{mt} = \sigma_m(t)\epsilon_m(t) \tag{3.6}$$

$$R_{st} = \beta(t)\sigma_m(t)\epsilon_m(t) + \sigma_s(t)\epsilon_s(t) \tag{3.7}$$

R_{mt} and R_{st} here represent the standardized returns on market and on one security at time t, and $\beta(t)$ is the correlation coefficient between R_{mt} and R_{st} at time $t, t = 1, 2, ..., n$,where

$$\beta(t) = Cov(R_{mt}, R_{st})/\sigma_m^2(t) \tag{3.8}$$

Because:

$$\begin{aligned} Cov(r_{mt}, r_{st}) = Cov(\sigma_m(t)\epsilon_{mt}, \beta(t)\sigma_m(t)\epsilon_{mt} + \sigma_s(t)\epsilon_{st}) \\ = Cov(\sigma_m(t)\epsilon_{mt}, \beta(t)\sigma_m(t)\epsilon_m(t) + Cov(\sigma_m(t)\epsilon_{mt}, \sigma_s(t)\epsilon_{st}) \\ = \sigma_m^2(t)\beta(t) + \sigma_m(t)\sigma_s(t) \\ = \sigma_m^2(t)\beta(t) \end{aligned} \tag{3.9}$$

So,we can estimate expectation of RR^T since we already know that ϵ is iid,

$$E(RR^T) = E(\Sigma\epsilon\epsilon^T\Sigma^T) = \Sigma E(\epsilon\epsilon^T)\Sigma^T = \Sigma\Sigma^T \tag{3.10}$$

Through matrix calculations it is interesting to see that:

$$\Sigma\Sigma^T = \left(\begin{array}{c|c} \sigma_m(t)^2 & \rho(t)\sigma_m(t)^2 \\ \beta(t)\sigma_m(t)^2 & \beta(t)^2\sigma_m(t)^2 + \sigma_s(t)^2 \end{array} \right)$$

As we found in (3.8), the relationship between $R_m(t)$ and $R_s(t)$ is nicely explained, the top right element divided by the first entry on the diagonal is just the $\beta_s(t)$ which can be interpreted as the risk of security s measured relative to the risk of the index m at time $t, t = 1, 2, ..., n$. We will estimate $\beta_s(t)$ and analyse $\beta_s(t)$ in the next chapter and it is also our point.

Chapter 4

Modelling Returns of S&P 500 and Dow Jones

4.1 Modelling Returns of S&P 500

In this chapter, we will apply the method we introduced in the previous chapter to the time series of daily returns of S&P 500 and Dow Jones from January 1,1988 to November 21,2003 (4009 observations). See figure 4.1 and 4.2.

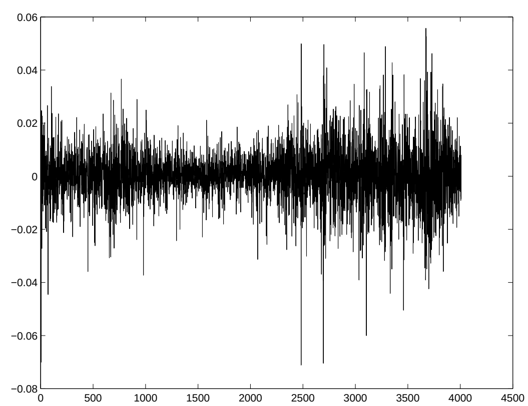

Figure 4.1: Log Returns of S&P 500 from 01/01/1998 to 11/21/2003

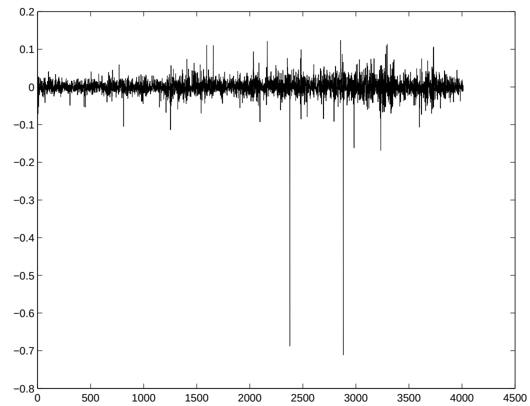

Figure 4.2: Log Returns of Dow Jones from 01/01/1998 to 11/21/2003

Here, we only use the one-sided evaluation weighted (Nadaraya-Watson) estimator to display the time-varying standard deviations:

$$\hat{\sigma}_1^2(t) = \frac{\sum_{1 \leq i \leq t} K_h(i-t) R_i^2}{\sum_{1 \leq i \leq t} K_h(i-t)} \tag{4.1}$$

4.1.1 Estimation Based on Epanechnikov Kernel

Due to equation(4.1), $K_h(\cdot) = h^-1K(\cdot/h)$ and K is the *Epanechnikov Kernel* with band width parameter $h = 20$, where $\tilde{R}_t$ is defined as:

$$\tilde{R}_t = X_t - \bar{X}_{t-1} \tag{4.2}$$

with $\bar{X}_t = n^-1 \sum_{i=1}^t X_i$ being the natural estimator for the mean μ based on returns up to day t and $\hat{\sigma}_1^2(\cdot)$ being an estimator of the unconditional variance $\sigma^2(\cdot)$ based only on the past information.

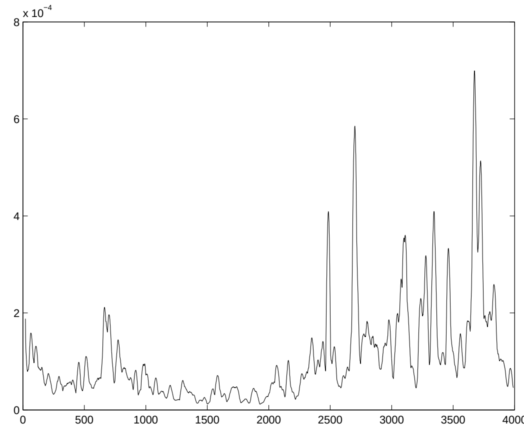

Figure 4.3: Estimated volatilities of index S&P 500 returns using one-sided smoothing based on Epanechnikov Kernel

The volatilities depicted in figure 4.3 reflect consecutive fluctations of being high, low and high which are less obvious from figure 4.1.
In the following picture 4.3, there displays $\beta(t)$ at time $t, t = 1, 2, ...n$ which interprets the risk of security s measured relative to the risk of index m or is interpreted as the market sensivity of the return on security s, we will see that the estimated values of $\beta(t)$ differ mostly between 0 and 2, with few exceptions which are quite large.
Let us have a look at the market model as follows:

$$E(\tilde{R}_{it}|R_{mt}) = \alpha_i + \beta_i R_{mt} \tag{4.3}$$

Again, we are going to show the estimated volatilities of security s in figure 4.5
Compared to figure 4.3, it is quite clear to note that the estimated volatilities of IBM stock are quite smaller than those of index S&P 500 at the most of the time range, only with a few exceptions in several short periods, from this, we get the feeling like IBM stock is safer than the whole index.
Based on our previous assumption that the innovations ϵ are of standard normal distribution with $E(\epsilon) = 0, Var(\epsilon) = I_m$, we need to have a look at SACFs of innovations to see the correlation relationship between them, since

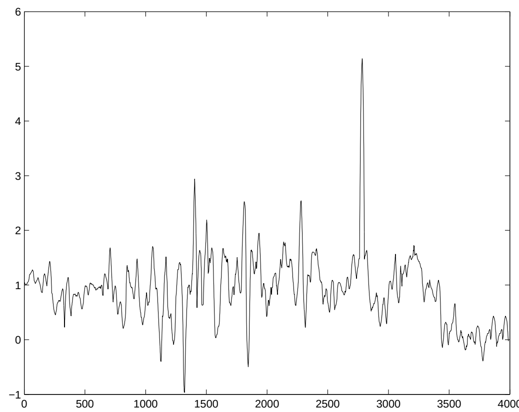

Figure 4.4: Estimated risk of security IBM measured relative to the risk of index S&P 500 based on Epanechnikov Kernel

we already know the innovations are independent theoretically.

In these two figures 4.6 and 4.7, we have a confidence level at 5% to measure the feasibility of the market model. Noting that SACFs of the market model vary mostly in the confidence interval which is from -0.0311 to 0.0311 (the boundary values of the interval are the same because ϵ belongs to standard normal distribution with mean zero), we might say the innovations are almost uncorrelated with this confidence test at 5%,let us review the market model which is:

$$r = \Sigma\epsilon \tag{4.4}$$

The market model suits the data fairy well.

4.1.2 Estimation Based on Normal Kernel

In this section, we are going to use another kind of *kernel* which has different kernel weight function and corresponding density function, here, we are going to plot the volatilities of IBM stock and index S&P 500 again to see whether the model still suits the data by using another kernel, namely, whether we still have nice plots of SACFs of ϵ. As we did before, we still use the multivariate analysis based on the normal relationship and linearity between the stock

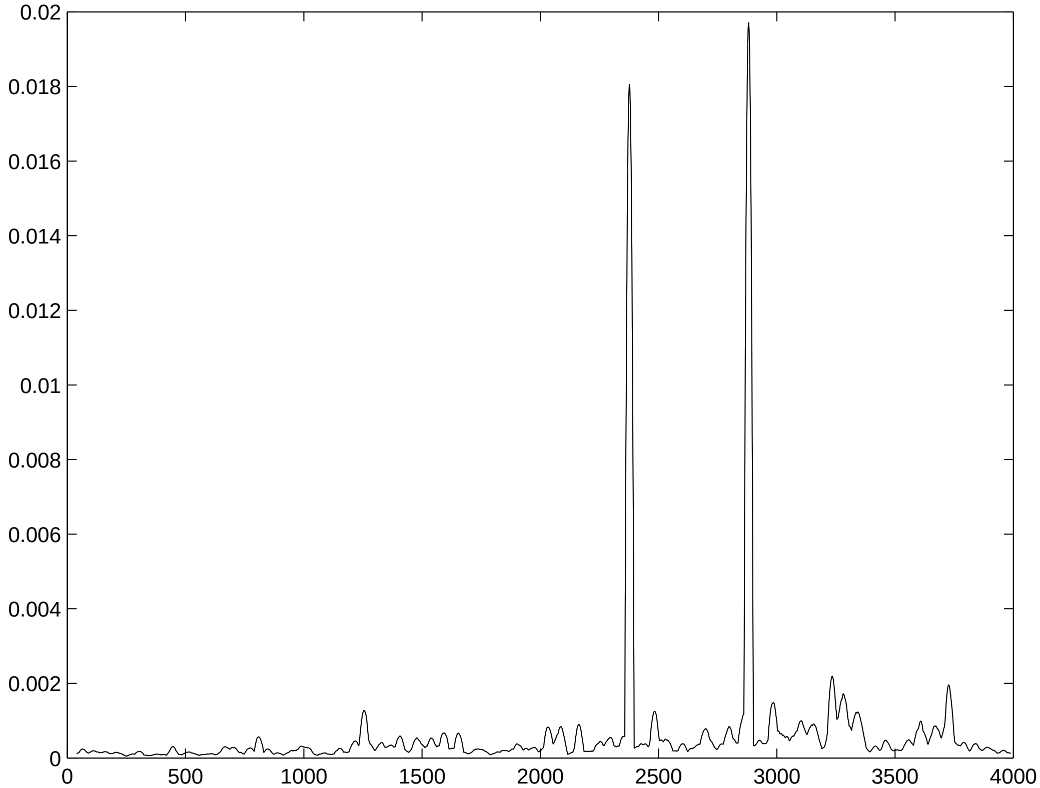

Figure 4.5: Estimated volatilities of IBM returns using one-sided smoothing based on Epanechnikov Kernel

and the index.

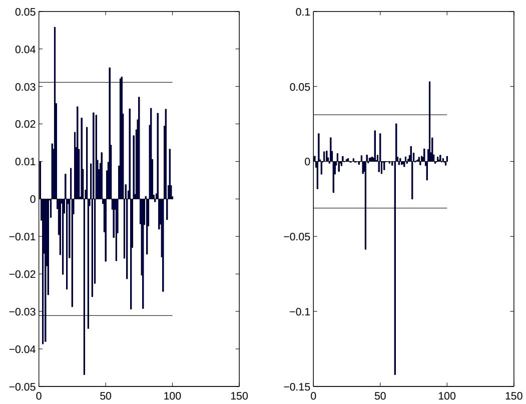

Figure 4.6: SACFs of innovations:the left one is of IBM stock, while the right one of index S&P 500 based on Epanechnikov Kernel

Where bandwidth $h = 24$ and kernel $K = 2\pi^{-0.5}exp(-u^2/2)$. Now, we have checked the market model by using two different kernels based on index S&P 500, through the figures from 4.3 to 4.12 which describe several features of the market model, we surprisingly can find that there is no big difference between features of market model based on two different kernels, and regarding separate feature of market model, we have small values of volatilities and small SACFs, which means we have a good estimate and the market model suits both indexes.

4.2 Modelling Returns of Dows

We have the estimates of market model based on dataset S&P 500 in the previous section, again, we are going to test the fit of the market model by using another dataset which is Dow Jones by using the same period of time period (the same number of observations) as we did before. Similarly, we will present plots of volatilities and β, SACFs of innovations as well.

4.2.1 Estimation Based on Epanechnikov Kernel

We have seen the estimated volatilites regarding Epanechnikov Kernel by using S&P 500, this time, as well,we are going to use the same Kernel to plot

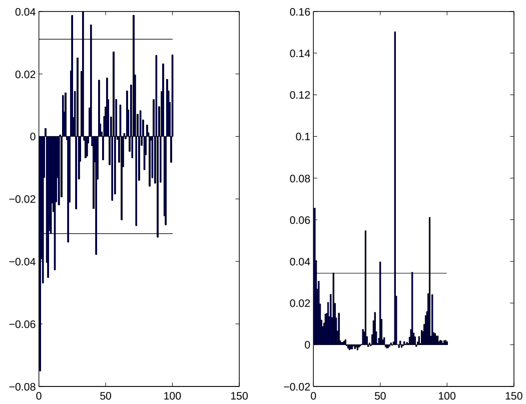

Figure 4.7: SACFs of absolute innovations:the left one is of IBM stock, while the right one of absolute innovations of index S&P 500 based on Epanechnikov Kernel

what we want to see, especially the dependency structure of innovations. Because of different dataset and kernel, we will use $h = 22$ here which is calculated based on Nonparametric Regression function.

4.2.2 Estimation Based on Normal Kernel

This section again, includes several figures based on *normal kernel*, and the corresponding value of bandwidth is $h = 26$ here.Still, we can have a look at the plots and check the dependency structure of the innovations and check the fit of the market model to the dataset.

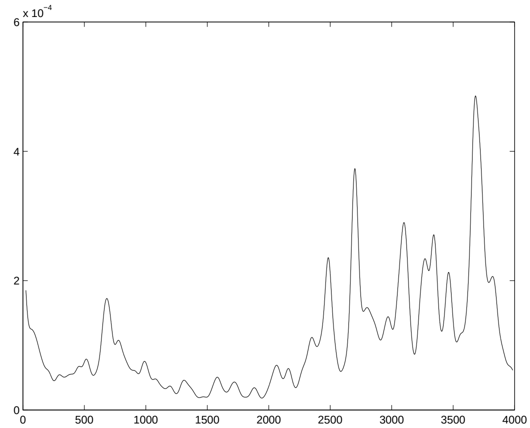

Figure 4.8: Estimated volatilities of index S&P 500 by using the one-sided smoothing based on Normal Kernel

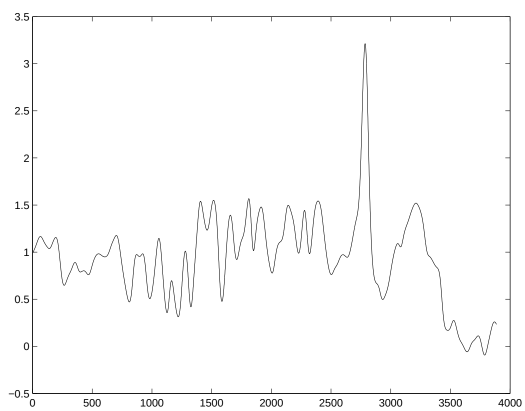

Figure 4.9: Estimated risk of IBM measured relative to index S&P 500 based on Normal Kernel

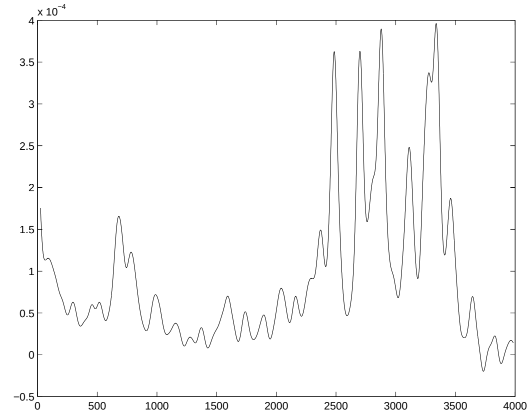

Figure 4.10: Estimated volatilities of IBM stock based on Normal Kernel

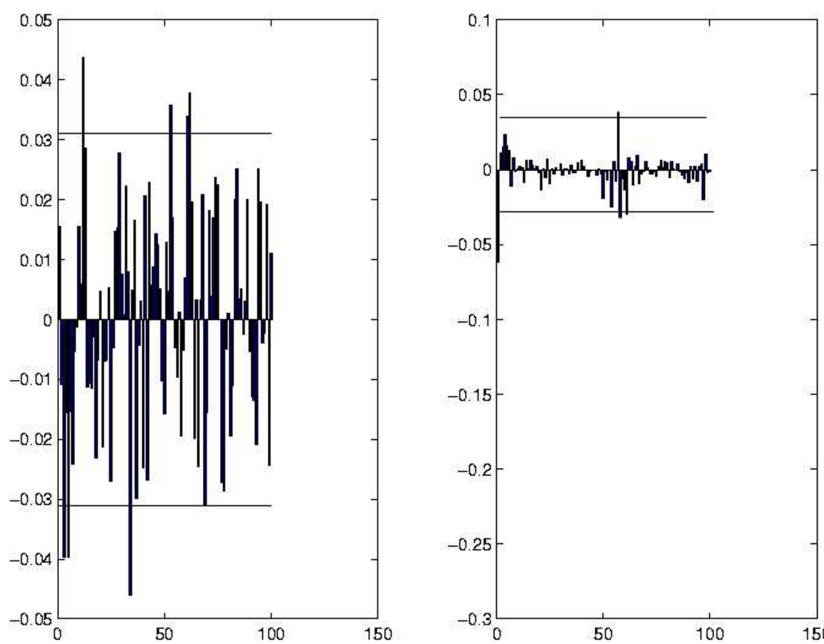

Figure 4.11: SACFs of innovations of index S&P 500(left) and IBM stock(right based on Normal Kernel

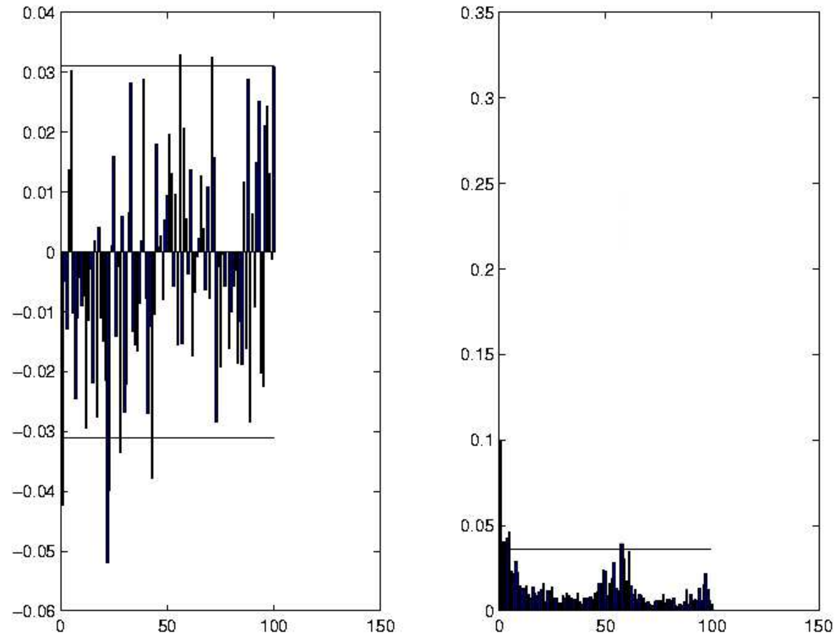

Figure 4.12: SACFs of absolute innovations of index S&P 500(left) and absolute innovatons of IBM stock(right) based on Normal Kernel.

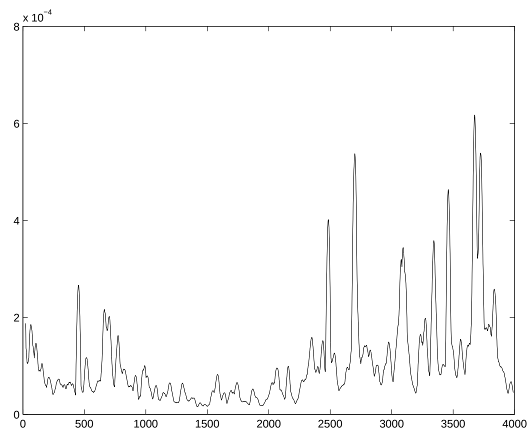

Figure 4.13: Estimated volatilities of index Dow Jones based on Epanechnikov Kernel

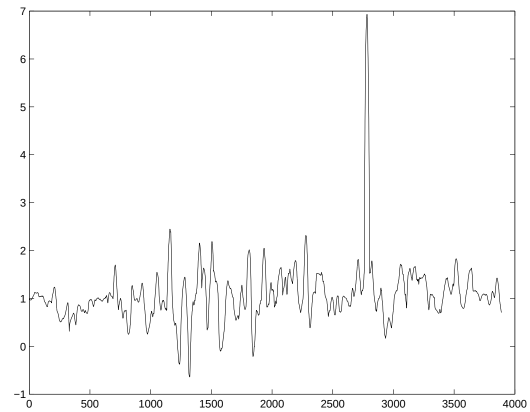

Figure 4.14: Estimated risk of IBM stock measured relative to index Dow Jones based on Epanechnikov Kernel

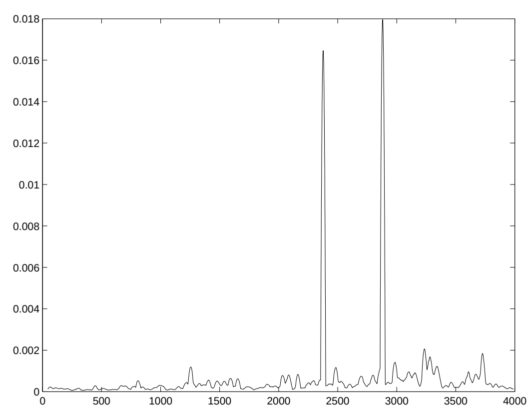

Figure 4.15: Estimated volatilities of IBM stock based on Epanechnikov Kernel

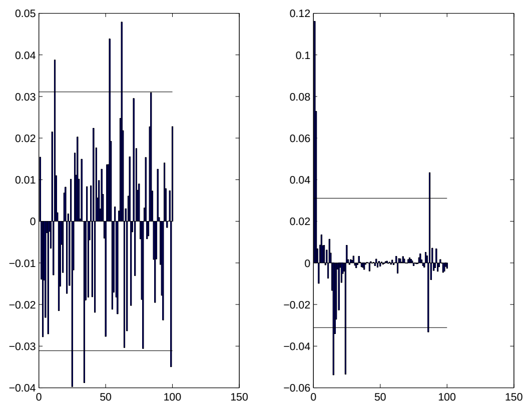

Figure 4.16: Estimated SACFs of innovations of index Dow Jones and IBM stock based on Epanechnikov Kernel

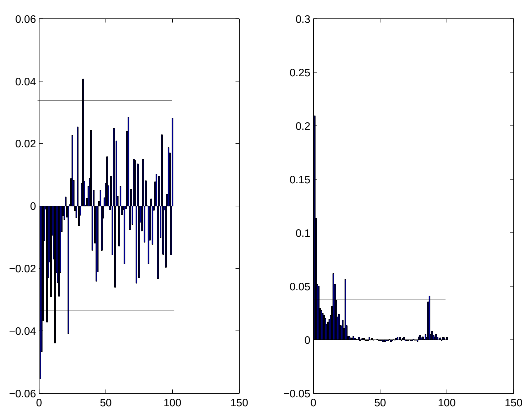

Figure 4.17: Estimated SACFs of absolute innovations of index and IBM stock based on Epanechnikov Kernel

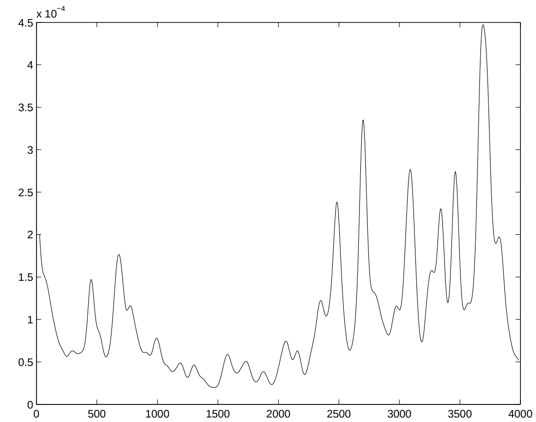

Figure 4.18: Estimated volatilites of index Dow Jones based on Normal Kernel

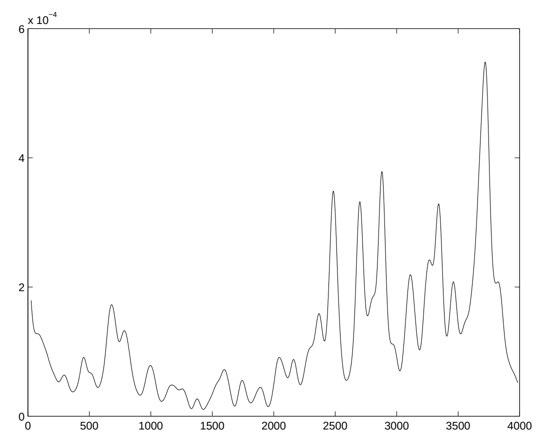

Figure 4.19: Estimated volatilites of IBM stock based on Normal Kernel

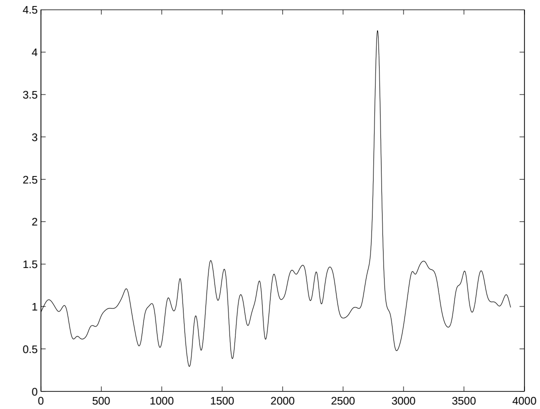

Figure 4.20: Estimated risk of IBM stock measured relative to index Dow Jones based on Normal Kernel

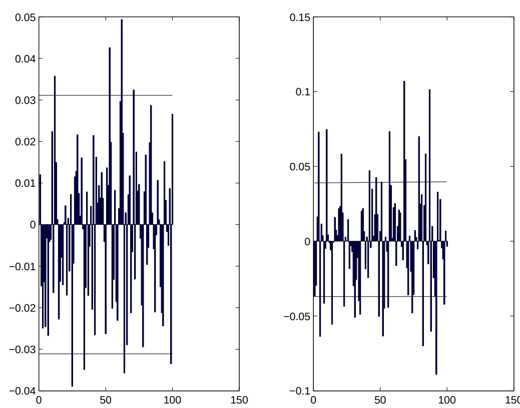

Figure 4.21: SACFs of innovations of IBM stock and index Dow Jones based on Normal Kernel

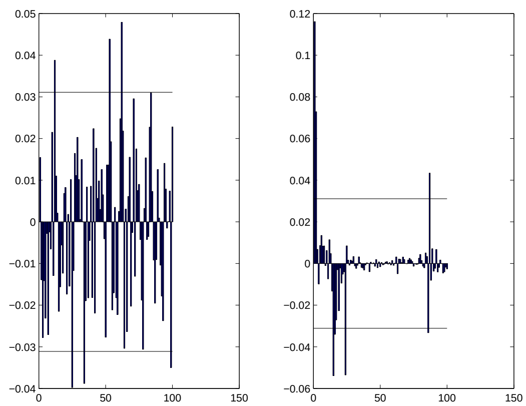

Figure 4.22: Estimated SACFs of absolute innovations of index Dow Jones and IBM stock based on Normal Kernel

Chapter 5

Conclusions and Acknowledgements

We have employed two different kernels: *Epanechnikov* and *Normal* and two different datasets: S&P 500 and Dows to test the market model by plotting several aspects of the outcomes, especially concerning the SACFs of innovations of IBM stock and these two datasets (Please check figures 4.5,4.9,4.13,4.17, we found that most estimates of SACFs are within the confidence interval [-0.0311,0.0311] which indicates the dependency structure of innovations and also matches the original assumption that innovations are independent. We also checked the plots of volatilities of the datasets and the security, we could see the fluctuation of volatilities due to the one-sided evaluation weighted estimator (see equation 2.7). In general, we could say the original assumptions of the market model are reasonable and the assumed market model fits the data quite well by checking the dependency structures of innovations.

Acknowledgement

Thanks God for giving me confidence and power to deal with and overcome difficulties. I am very grateful to my supervisor, Associate professor Catalin Starica for the instruction before the thesis and the guidance to me during the project especially when i was in dilemma. Thanks also are given to Ivar Gustafsson and Stig Larsson and Ann-Britt Karlsson for their support and help. Special thanks to my parents for their endless support and understanding especially when i was diffident and depressed.

Bibliography

[1] Wolfgang, Haerdle, *Applied Nonparametric Regression.* Version 27(1994)

[2] Eugene, Fama, *Foundation of Finance*,reprinted with permission,BAsic Book,1976.

[3] John, A.Rice, *Mathematical Statistics and data analysis*, second edition.

[4] J.D.Gibbons and S.Chakraborty, *Nonparametric Statistical Inference*,3rd edition,Marcel Dekker,New York 1992.

[5] P.Brockwell and R.Davis, *Introduction to Time Series and Forecasting*,(Springer,1996).

[6] Jay L.Devore, *Probability and Statistics for Engineering and Science*,3rd edition,Marcel Dekker,New York 1992.

[7] Gerald Keller,Brian Warrack, *Statistics for Management and Economics*,6th edition,Duxbury Press,1999.

[8] Laurence G.Grimm,Paul R.Yarnold, *Reading and Understanding Multivariate Statistics*,3rd edition,4th edition(January 1995).

[9] Barbara G.Tabachnick, *Using Multivariate Statistics*, 4th edition(August 9,2000).

[10] George E.P.Box, *Statistics for Experiments:An Introduction to Design,Data Analysis and Model Building*,1st edition(June 22, 1978).

[11] Richard A.Johnson,Dean W.Wichen, *Applied Multivariate Statistical Analysis*,3rd edition,1992.

[12] Richard J.Larsen,Morris L.Marx, *An introduction to Mathematical Statistics and Its Applications*,3rd edition(July 15,2000).

[13] David G.Kleinbaum, *Applied Regression Analysis and Multivariate Methods*,3rd edition,Brooks Cole(September 15,1997).

[14] David J.Bartholmew, *The Analysis and Interpretation of Multivariate Data for Social Scientists*,1st edition,CRC Press (February 26,2002).

[15] Catalin Starica,Stefano Herzel and Reha Tutuncu,*A Non-Stationary Multivariate Model for Financial Returns*,August(2002).

[16] Michael G.Schimek, *Smoothing and Regression:Approaches, Computation,and Application*,1st edition,Wiley-Interscience(August 15,2000).

[17] John Fox, *Multiple and Generalized Nonparametric Regression*,Sage Publications(May 1,2000).