Forecasting Volatility of Turkish Markets:

A Comparison of Thin and Thick Models

Ekrem Kilic

Marmara University

Faculty of Economics and Administrative Sciences

ekremk@finecus.com

-Abstract-

Volatility of financial markets is an important topic for academics, policy makers and market participants. In this study first I summarized several specifications for the conditional variance and also define some methods for combination of these specifications. Then assuming that the squared returns are the benchmark estimate for actual volatility of the day, I compare all of the models with respect to how much efficient they are to mimic the realized volatility. At the same time I used a VaR approach to compare these forecasts. With the help of these analyses I examine if combination of the forecast could outperform the single models.

Index of Contents

1 Introduction

Volatility of financial markets is an important topic for academics, policy makers and market participants. First, the volatility of an asset can be thought as a measurement of risk. After recognition of internal models to determine the capital charges of banks by Basel Committee, market participants deal with efficient computation of Value-at-Risk (VaR). Since volatility forecast plays a central role in VaR computation, forecasting asset volatility became more important for them. In the context of risk, forecasting volatility also received great concern from the policy makers. Especially in emerging markets, the financial liberalization process is not a smooth path. During last two decades, emerging markets experienced many financial crises caused or were leaded by huge capital inflows or outflows. While that is the situation, volatility is a good indicator for monitoring financial stability and understanding the mechanisms and exact relations behind those crises. Because stability of economy is much related with the stability of its financial market, volatility of financial markets have also many macroeconomic aspects. Another crucial role played by volatility is in the pricing of derivatives securities. Especially commonly used Black & Scholes formula is mainly based on the asset volatility. Although we don’t have organized derivative security markets in Turkey, the volume of trade in derivative market is growing both in Turkey and all over the world. From financial institutions’ point of view, problems like choosing the optimal portfolio or hedging portfolios are also mainly based on the volatility.

Parallel to its importance related with various aspects, there is huge amount of literature on forecasting volatility. We will use many models including very simple ones like moving average type models and also strictly complex ones like asymmetric GARCH specifications. But my main concern will be combine all these forecast using different methods and show the robustness of thick models in forecasting. The idea of combination is first expressed by J.M. Bates and C.W.J. Granger (1969). In a recent work by C.W.J. Granger and Y. Jeon (2001) they called this alternative way as “Thick Modeling”. The idea is quite simple, combination or thick modeling uses many alternative specifications to produce one robust result (Granger and Jeon, 2001). Diebold and Lopez (1995) express the idea as;

> “Regardless of whether one forecast is "best," however, the question arises as to whether competing forecasts may be fruitfully combined -- in similar fashion to the construction of an

> asset portfolio -- to produce a composite forecast superior to all the original forecasts. Thus, forecast combination, although obviously related to forecast accuracy comparison, is logically distinct and of independent interest."

As Diebold and Lopez (1995) mentioned, the combination of forecasts is also a distinct area of interest. The works on combination produced some methods for combining. I will use several methods such as equal weighting, trimming and weighting with loss functions.

Organization of this paper will be as follows. First I will do a basic setup for all models. Secondly, I will shortly introduce single specifications that I used. Thirdly the combining methods will be reviewed. Fourthly, the empirical result for the Turkish data will be examined by using some loss functions. Finally I will conclude.

2 General Assumptions and Data

The data set I use is the ISE100 index. ISE100 index covers 100 leading stocks from several sectors quoted on Istanbul Stock Exchange (ISE). ISE100 is a weighted index by market capitalizations. ISE100 data that I use starts from 2 January 1989 to 26 March 2004 and consists of 3999 daily observations. Therefore we have 3998 returns. I used logarithmic returns defined as follows;

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right)*100 \qquad (2.1)$$

where P_t is the level of ISE100 index at time t.

In this study we use realized volatility as benchmark to compare models. In general realized volatility can be defined as;

$$\sigma_t^{\,2} = \left(r_t - \mu\right)^2 \qquad (2.2)$$

where μ is the mean of return. And in this model if the mean of return is equal to zero, Eq.(2.2) can be re-written as;

$$\sigma_t^{\,2} = r_t^{\,2} \qquad (2.2a)$$

But before proceeding we should test whether these data series have zero median, by using Sign test which is a distribution free non-parametric test for zero median. Attractive feature of this test is that it imposes no distribution to the series. The sign test can be formulated as follows;

$$S_t = \sum_{t=1}^{n} I_t\left(r_t\right) \qquad (2.3)$$

where

$$I_t = \begin{cases} 1 & if \quad r_t > 0 \\ 0 & otherwise \end{cases}$$

Then test statistic is as follows (Diebold, 1995);

$$\frac{S - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \underset{asy}{\sim} N(0,1) \qquad (2.4)$$

By recursively applying the sign test for the sample, I observe that; if we look at the sub-sample covers 2000-2004, we can say there is zero median except February 2001 which was the time that market crash occurred and 11 September-15 October 2001 after terrorist attacks in the USA. But the period that includes 1994 to 2000 we reject the null that claims the median is zero.

By the sign test simply, we show that for the data set the number of loses can not be thought equal to number of gains. So for the daily volatility we need to use mean. Therefore we use Eq. 2.2 for defining realized volatility.

Figure 2.1 plots the series. As we can see in this graph, the ISE is highly volatile. It reaches a peak of 20000 at January 2000, and after crisis it declines to 7500. Especially after crisis we can see sharp increases and declines. However, this graph is not so useful to exhibit the volatility. Figure 2.2 shows the return of ISE100. By this graph we can identify the high volatility of late 1980’s and early 1990’s too. The squared returns which we use as benchmark are shown at Figure 2.3. In this graph we can clearly discern that there are more volatile periods and also relatively less volatile periods. Therefore even by the graph it is obvious that return is heteroscedastic. Now let us consider whether in presence of this heteroscedastic pattern, the return is forecastable or not, whether the history of the series could say anything about the tomorrow or not. In Figure 2.4 the correlogram of the return series are shown. The first order autocorrelation is quite low and the all higher order autocorrelations are lower than first order. This shows that the return series have very low level of memory. Because we used daily frequency, it is reasonable to have such a result. Further if we consider Figure 2.5 which exhibit correlogram of the squared daily returns, the autocorrelation rises noticeably. The first order autocorrelation is 0.3 and the higher order autocorrelations are not negligible. This is the evidence of volatility clustering and suggest that even return is not predictable, we can predict the volatility.

Table 2.1 shows the descriptive statistics of the return series. What we should focus with this table is normality test of Jarque-Bera. For the null hypothesis of normality, we reject very strongly. Kurtosis statistics also signals same problem. Figure 2.6 shows the estimated density against normal density. The main difference from normal distribution is original distribution has fat tails. QQ plot shows same. These are some evidences which suggest that we should consider some other distributions rather than normal. But in this study we concern with a comparison of thin and thick models, therefore we assume normality for Maximum Likelihood estimations.

To compare the forecast efficiencies of thin and thick models, we need at least two sub-samples for estimation and the forecast. But we divide data into three sub-samples. For the thin models, we use first two sub-samples to estimate parameters, and the last for forecasting. For thick models we use first sub-sample to estimate parameters of the models, second sub-sample to combine their forecasts and the last sample for forecasting. Finally for the combination by bootstrapping we use first two sub-samples to estimate, and the last for forecasting. Our forecasting horizon is one day. To compare models we use three different loss functions, which briefly described at Appendix 2.

3 Thin Models

In this section I will introduce the single models that I used.

3.1 Random Walk

If we assume that the volatility follows a random walk, the optimal forecast of future is today’s volatility. Therefore we can define the volatility forecast as;

$$f_{t+1} = \sigma_t^2 \qquad (3.1)$$

3.2 Historical Average

If we assume that conditional expectation of the volatility is constant than the out-sample forecast will be just an average;

$$f_{t+1} = \frac{1}{t}\sum_{i=1}^{t}\sigma_i^2 \qquad (3.2)$$

where t is the forecast date.

3.3 Heteroscedastic Models

3.3.1 Simple Moving Average (MA)

Simple moving average model might be considered as a modified version of the historical average model. In historical averaging we used an equally weighted average of the sample and we use all data from beginning to end for this average. In moving average, we again uses an equally weighted average, however, this time we use only a specific window for averaging.

Then we can define simple moving average as;

$$f_{t+1} = \frac{1}{W}\sum_{w=1}^{W}\sigma_{t+1-w}^2 \qquad (3.3)$$

where W is the width of the window. I use; MA-5, MA-10, MA-15, MA-20 and MA-60.

3.3.2 Exponential Smoothing (ES)

Exponential smoothing weights past observations with exponentially decreasing weights to forecast future values. Therefore this is again a modified version of historical averaging. Instead of equally weighting, in exponential smoothing weights differ. The forecast value defined as;

$$f_{t+1} = (1-\lambda)\sigma_t^2 + \lambda f_t \qquad where \;\; 0 < \lambda < 1 \qquad (3.4)$$

By repeated substitutions we can re-write the forecast as;

$$f_{t+1} = (1-\lambda)\sum_{i=1}^{t} \lambda^{i-1}\sigma_i^2 \qquad (3.4a)$$

equation (3.4a) shows the forecast is equal to a weighting average. The weights $(1-\lambda)\lambda^t$ decrease geometrically. The value of λ, decay factor, estimated simply by minimizing the one week forecast errors.

RiskMetrics which is a common risk management program, called this method as EWMA[1] and they proposed that the optimum value of decay factor is 0.94 for daily data. Because it is easy to apply, many market participants used this calculation as benchmark. We also use this particular value of decay factor in our comparison.

3.3.3 Exponentially Weighted Moving Average (EWMA)

EWMA is the mixture of the previous two models; moving average and exponentially smoothing. The forecast is defined as;

$$f_{t+1} = \lambda f_t + (1-\lambda)\frac{1}{W}\sum_{w=1}^{W} \sigma_{t+1-w}^2 \qquad where \;\; 0 < \lambda < 1 \qquad (3.5)$$

As we can see this definition covers the exponential smoothing too. If we choose the W as one what we have is the same with the previous one.

In this study, we use EWMA-5, EWMA-10, EWMA-15, EWMA-20 and EWMA-60.

[1] Even they call this method as EWMA, in this study EWMA is another method. In comparison of models we denote exponentially smoothing 0.94 as RiskMetrics to avoid such confusions.

3.4 Autoregressive Conditional Heteroscedastic Models

3.4.1 ARCH

The autoregressive conditional heteroscedasticity (ARCH) model which proposed by Engle (1982) can be defined by;

$$r_t = \mu + \sigma_t \varepsilon_t \qquad (3.6)$$

$$\sigma_t^{\,2} = \lambda + \sum_{i=1}^{q} \alpha_i \left(r_{t-i} - \mu \right)^2$$

where $\varepsilon_t \sim iid(0,1)$. Hence one-day forward volatility forecast can be represented as;

$$f_{t+1} = \lambda + \sum_{i=1}^{q} \alpha_i \sigma_{t+1-i}^{\;2} \qquad (3.7)$$

This shows the proper forecast of the tomorrow’s volatility is based on the q most recent volatilities. Because of this, if we would not select a high order, ARCH might catch only the short memory of the data. However the conditional variance dynamics need a long memory model.

In this study we will use AR(1), AR(2),AR(3) and AR(4).

3.4.2 GARCH

The Generalized ARCH model is a response to the short memory dynamic of the ARCH models. Bollerslev (1986) defined GARCH by;

$$r_t = \mu + \sigma_t \varepsilon_t \qquad (3.8)$$

$$\sigma_t^{\,2} = \lambda + \sum_{i=1}^{q} \alpha_i \left(r_{t-i} - \mu \right)^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^{\;2}$$

Therefore one day forward volatility forecast is shown as;

$$f_{t+1} = \lambda + \sum_{i=1}^{q} \alpha_i \sigma_{t+1-i}^2 + \sum_{j=1}^{p} \beta_j f_{t+1-j} \qquad (3.9)$$

GARCH imposes that the proper forecast for the tomorrow's volatility is based not only on the recent volatilities and also previous forecasts which include the previous volatilities. Then the GARCH model is a long memory model.

In this study we will use GARCH (1, 1). We choose this specification by the Schwarz information criteria.

3.4.3 Asymmetric ARCH Models

In the stock market it is common to observe that downward movements leads more volatile periods than upward movements, this called leverage effect. We will use two different asymmetric ARCH models; EGARCH and TARCH.

3.4.3.1 Exponential GARCH

The Exponential GARCH (EGARCH) model is proposed by Nelson (1991). The specification for the volatility is;

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \log \sigma_{t-i}^2 + \sum_{j=1}^{q} \left(\alpha_j \left| \frac{\varepsilon_{t-j}}{\sigma_{t-j}} \right| + \gamma_j \frac{\varepsilon_{t-j}}{\sigma_{t-j}} \right) \qquad (3.10)$$

In this model leverage effect is represented by γ. So if γ is statistically insignificant then we could say there is no leverage effect in our sample.

If we assume that the errors are normally distributed, Eq 3.10 can be re-written as;

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \log \sigma_{t-i}^2 + \sum_{j=1}^{q} \left(\alpha_j \left| \frac{\varepsilon_{t-j}}{\sigma_{t-j}} - \sqrt{\frac{2}{\pi}} \right| + \gamma_j \frac{\varepsilon_{t-j}}{\sigma_{t-j}} \right) \qquad (3.10a)$$

The one day volatility forecast can be shown as;

$$\log f_{t+1} = \omega + \sum_{i=1}^{p} \beta_i \log f_{t+1-i} + \sum_{j=1}^{q} \left(\alpha_j \left| \frac{(r_{t+1-j} - \mu)}{\sqrt{f_{t+1-j}}} - \sqrt{\frac{2}{\pi}} \right| + \gamma_j \frac{(r_{t+1-j} - \mu)}{\sqrt{f_{t+1-j}}} \right) \qquad (3.11)$$

3.4.3.2 Threshold ARCH

Threshold ARCH is introduced independently by Zakoïan (1994) and Glosten, Jaganathan, and Runkle (1993)[2]. The model for the volatility is defined as;

$$\sigma_t^2 = \omega + \sum_{i=1}^{q}\left(\alpha_i \varepsilon^2{}_{t-i} + \gamma_i \varepsilon^2{}_{t-i} d_{t-i}\right) + \sum_{j=1}^{p} \beta_j \sigma^2{}_{t-j} \qquad (3.12)$$

where

$$d_t = \begin{cases} 1 & if \quad \varepsilon_t < 0 \\ 0 & otherwise \end{cases}$$

In such a model leverage effect is represented by γ. The effect of the bad news $\gamma + \alpha$, the effect of the good news is α. So if γ is statistically insignificant then we could say there is no leverage effect in our sample.

The proper forecast of tomorrow's volatility is;

$$f_{t+1} = \omega + \sum_{i=1}^{q}\left(\alpha_i \varepsilon^2{}_{t+1-i} + \gamma_i \varepsilon^2{}_{t+1-i} d_{t+1-i}\right) + \sum_{j=1}^{p} \beta_j f_{t+1-j} \qquad (3.13)$$

[2] TARCH also called as GJR (**G**losten, **J**aganathan, and **R**unkle)

4 Thick Models

One can use many methods to combine forecasts of different models and to have one robust forecast of tomorrow. I will introduce the methods I used. Of course, it is possible to produce many new methods by changing and mixing these methods. Thick modeling provides us a space of possibilities. In this sense it is so flexible.

4.1 Equally Weighted Combination

This is the simplest strategy to combine out-sample forecasts. This combination might be optimal when there is no significantly best single model to forecast. In other words, there would be no robust model for all periods, and then this is an optimal. Many of the models that are used in combination might be best forecasts in particular periods. For instance one model might best in the times of high volatility, however for less volatile periods, the model might be useless. Or another might be best for more stabile periods of the market. Of course another possibility is structural change. Because we deal with an emerging market, market conditions change so quickly. The regulations might change, the financial institutions might change, and therefore every model might be optimal for a short period of time.

We can show an equally weighted combination as follows;

$$F_t = \frac{1}{N}\sum_{i=1}^{N} f_t^i \tag{4.1}$$

where f_t^i is the forecast of ith single model at time t. As we can see in the Eq. 4.1, all forecasts are equally weighted. Hence the equally weighted combination of the forecasts is just an average of thin model forecasts.

4.2 Trimming

Trimming introduced by Stock and Watson (1999). Although I used trimming as a modified version of the equally weighted combination, one might apply this method to all other combination techniques.

Basically trimming follows a procedure like; first, rank all forecasts, second remove α % largest and smallest and use simple average of remaining forecasts as combined forecast.

$$F_t = \frac{1}{N-T} \sum_{i=1}^{N-T} f_t^i \qquad (4.2)$$

where T is the number of trimmed forecasts.

In this study, I removed smallest and largest forecast of the day. In fact this is nearly equal to 5% trimming.

4.3 Combination with OLS

Another way to combine forecasts might be usage of OLS. Up to, we saw equally weighted forecasts. But if we can weight better forecasts high and worse forecasts low, it will improve our combinations' ability to catch tomorrow. Therefore, the choice of weight becomes important.

At this point, OLS can be thought as a good tool to combine several forecasts. But usage of OLS causes some disadvantages too. For example in our case, we generate all volatility forecasts with same and one data that is realized volatility of today. And we optimize each model's parameters -if it has- to mimic tomorrow's realized volatility as best as possible. Therefore, while each forecast is correlated with actual volatility, they are highly correlated with each others. Then possibly we would have high degrees of Multicollinearity if we try to regress all forecasts that we have, on actual volatility. Instead, we can combine some of them.

Then we can estimate the parameters of the model by the following regression;

$$(r_t - \mu)^2 = \sum_{i=1}^{N} \alpha_i f_t^i + u_t \qquad (4.3)$$

then the combined forecast is as follows;

$$F_t = \sum_{i=1}^{N} \hat{\alpha}_i f_t^i \qquad (4.4)$$

where $\hat{\alpha}$ is the estimated weight vector and N is the number of forecasts combined.

4.4 Combination with Loss Functions

As we mentioned earlier the notion behind weighting the models is simply weight high better models and weight low the worse models. To solve such a problem the OLS is good solution, but due to Multicollinearity to combine all forecasts with OLS is not possible –at least in our case. Then usage of some loss functions to combine all forecasts might be the solution.

In this study I used three loss functions that are introduced at appendix2. In this section I used Linex loss function which gave me the best results among other loss functions.

The weights are calculated as follows;

$$w_i = \left(1 - \frac{l(\boldsymbol{\sigma}^2, \mathbf{f}^i)}{\sum_{i=1}^{N} l(\boldsymbol{\sigma}^2, \mathbf{f}^i)} \right) \frac{1}{N} \qquad (4.5)$$

where l is the loss function, $\boldsymbol{\sigma}^2$ is the vector of realized volatility and $\mathbf{f}^i$ is the forecast vector of the ith model. Then the combined forecast is;

$$F_t = \sum_{i=1}^{N} w_i f_t^i \qquad (4.6)$$

where N is the number of the single models to combine.

4.5 Ranking

Another weighting method is ranking. Since our main concern is weighting most the best and weighting least the worst, the ranking should work. I used a ranking procedure as follows; first I ranked all forecasts with respect to their loss function values. Secondly I weighted them with;

$$w_i = \frac{\left(N + 1 - d_i\right)^2}{\sum_{i=1}^{N} \left(N + 1 - d_i\right)^2} \qquad (4.7)$$

where d_i is the rank of ith forecast. Hence the combined forecast is;

$$F_t = \sum_{i=1}^{N} w_i f_t^i \qquad (4.8)$$

where N is the number of forecasts.

4.6 Combination with Bootstrapping

One of the major problems with the combination is loss of data, because one needs two samples to constitute a combination (one sample to estimate the parameters of the single specifications and one sample to choose the weights). The idea of combination with bootstrapping might be a solution to this problem.

The notion is quite easy, use all data that you have to estimate proper parameters for the single models and combine those forecasts with bootstrapped data. This should work by two reasons; first the single models would be more efficient because the sample size increased, and second the weights will be more efficient again due to the same reason.

This method might be used with all previous combination techniques. In this study we will use only ranking method with bootstrapping.

5 Comparison of the Models

In this chapter we will compare our findings. First the results from the models including thin and thick will be compared with some loss functions. Second we will use a Value at Risk comparison. As we mentioned in introduction for financial institutions the calculation of VaR is very important. And the calculation of the VaR is mainly based on volatility of asset, and then we will test the robustness of combined forecasts of volatility to RiskMetrics which is commonly used by market participants.

5.1 Comparison with Loss Functions

In this study, I used Root Mean Squared Errors (RMSE), Theil-U and Linex loss functions. For all loss functions the results are very similar.

First of all, TARCH (2, 2) is the best model among the all thin and thick models. This model outperforms others with respect to all loss functions. But the equally weighted combination is slightly worse than the TARCH (2, 2) again with respect to all loss functions.

The combined models generally work well. If we consider the best five models, the four of them are always combined models. Even TARCH (2, 2) is better than thick models the difference is very small and negligible. Moreover when we replicate the forecast procedure for some other sub-samples of the data the combined forecasts always outperforms others. The different nature of this period is strongly related with the deep crisis that we have in this period. Due to the crisis the leverage effect which is well represented by the TARCH model dominates the period. Hence if we consider all models we can say that the thick models are possibly better than thin models for forecasting one day forward volatility.

Among the combined forecasts the equally weighted combination performs well. According to all loss functions it is the second leaded by the TARCH (2, 2). But for all loss functions the difference is very small. Again if we consider different sub-samples generally ranking with bootstrapping combination is better than equally weighted combination. However the difference between these two is generally very small too. Then we can say that even equally weighting works well enough to combine forecasts.

One can also realize that the bootstrapped combination is clearly better than the other. Therefore this might be evidence for that one can handle with the loss of data problem with bootstrapped combination. It is also important to notice that the sample size of equally weighted combination and other combined models are not same, because the weighted combination models need to out of sample forecasts for combinations. However the equally weighted combination does not need to out of sample forecasts for estimation of the weights than its sample size twice of other combination models. Therefore this is another evidence for the importance of the sample size, because it performs better than other combinations.

It is interesting to see that the RiskMetrics model is one of the worst. First, all of the combinations are better than the RiskMetrics. And many of the single specifications are also better than the RiskMetrics. But I should mention that for other sub-samples the RiskMetrics is generally second best after combined forecasts. In fact if we consider several sub-samples one can say the RiskMetrics model is one of the best among the thin models. But definitely the combinations are the best in this sense. The reason for the worse performance in this period is again related with the crisis. Because the RiskMetrics can not cover the leverage effect, it is worse in this period.

For the asymmetric models, we can easily say that the TARCH model is better than the EGARCH model. Exponential behavior of the model might be problematic in many cases. It generally over estimates the volatility. On the other hand both TARCH model is good with respect to all loss functions. But again if we consider sub samples it is not that much easy to compare the EGARCH and TARCH. Some time TARCH does well and some times EGARCH. In general TARCH is slightly better, because it has a more stabile nature as a model. This also shows that we need to combine forecasts. Because the best single model definitely changes, using all of them is a better way.

5.2 VaR Comparison

Value at Risk is simply the worst possible return (which is possibly a loss) of an asset or portfolio with a given confidence level. Therefore VaR is nothing but a quantile analysis. For example if the confidence level is α, then VaR gives us $1-\alpha$ lower tail of the return distribution.

There some methods to calculate VaR. In general, we can there are two types of methods to calculate VaR; first is the non-parametric methods and the second is parametric methods. For parametric methods the estimation of the volatility is vital. The volatility is on the centre of these methods.

In this study we will calculate VaR for ISE100. It is similar like one would have a portfolio which consists of the stocks in ISE100 and weights of each stock in the portfolio are same with ISE100 too. Hence the formula for the VaR is quite simple one;

$$VaR_t^i = N(\alpha).f_t^i + \mu \qquad (5.1)$$

where VaR_t^i is the VaR at time t, calculated by the ith volatility model, N(.) is a function gives the cumulative standard normal distribution, α is the confidence level, μ is the average return and f_t^i is again volatility at time t, forecasted by the ith model.

After a quick introduction for the VaR concept, let us return back to the VaR comparison of our forecasts. The Basel Committee announced some rules for backtesting of the VaR models. In our further analysis, we will use a similar method with this framework. They want financial institutions to calculate VaR with 99% of confidence level and record their exceptions of last year. And according to the number of these exceptions they define some zones (see Table 5.5).

Since we use 99% confidence interval, we can use the 2.5 and 10 as cut-off value for exception respectively for 252-day period and 1000-day. In Table 5.6 we can see the number of exceptions for each model. The first 15 models are lower than our cut off value while all symmetric ARCH family and TARCH (1, 1) made higher mistake than we could tolerate. An important point is that there is no “Thick Model” that has higher exceptions than our cut-off value. In this sense, if one has a combination for volatility, he or she possibly had no problem. For one who uses thin model, however, he must choose a correct model, other wise he/she would make mistake.

If we consider the zones that defined by Basel Committee, the first 19 is in the Green Zone. This means all combinations are at the Green Zone. On the other hand GARCH (1, 1), TARCH (1, 1) and Exponentially Smoothing are at the Yellow Zone. And all lags of the ARCH and the Random Walk are at the Red Zone. Even tough the Basel’s criterion is more tolerable than our first one, the result did not change. Again all combinations performed well, they are all in green zone.

Another important point, one should realize is that, VaR performance is meaningless if we do not consider their performance with loss functions. On the other hand one model might have no exception, but this does not mean it is a good model, because our objective is not only have less exception but also mimic the realized volatility as good as possible. For example in VaR comparison the Historical Average seems like the best model, because it has no exception. However, it highly overestimates the volatility, because the ISE100 index was historically more volatile respectively its recent volatility. Figure 5.2 shows the graphs of the VaR calculated by each model against the realized volatility. It is clear by the graphs; some of the models like Historical Average, RiskMetrics and MA (60) consistently overestimate the volatility. Therefore we should look at the Table 5-1 again; as we can see these three models were not as good as the combinations. Result is the same if we consider other loss functions.

Hence if we consider both comparison with loss functions and VaR comparison, it becomes clearer that all combinations are better than single specifications. And also particularly the ranking with the bootstrapped combination is the best. We can see that for both comparisons combination with bootstrapping improved the efficiency of the combined forecast.

6 Conclusion

The main purpose of this study was comparison of "Thin" and "Thick" models with Turkish data. All our finding shows that the combination of the single specifications is possibly better than a single specification. Because a model might better than a combination in a specific period there is no model that is better than a combined forecast in every period. The behavior of the market changes over time. And also if you deal with an emerging market, it changes very quickly. Then one of the major reasons for why is combination better is these quick changes in market structure. One can understand it as institutional change or other might think of a deep change in the market participants' expectations or a deep external change, like entering EU … one can count numerous reasons to change.

Another important point is the nonlinear nature of these combinations. In this study we used linear models; however this is a very restrictive assumption on the behavior of the series. The combination provides an enormous flexibility. Even one imposes some restriction on the data in single models, when they are combined what it becomes do not have any restrictions. In this sense they are extremely flexible.

The normality assumption that we did in models that are estimated by Maximum Likelihood provides another room for better estimations. In fact, with Maximum Likelihood estimation we always impose a distribution to data, it can be normal or student or some other distribution. At the end of the day, even we use a very good distribution that well fits the real distribution; this is a big restriction on the data. In this point of view, the combination is again very flexible method. Because one can combine some models that are estimated by assuming normal distribution and some other models that are estimated by assuming any other distributions. Then these models are combined, it is neither normal nor any other distribution. For example, if one knows the real model but not the real distribution, by estimating model with several distributions, it is possible, at least theoretically, to have real distribution of the data by weighted combinations. This is an advantage of "Thick Models".

To improve the combinations efficiency, we used a bootstrapped sample to combine the forecasts. Our results show that it works. To express why it does works, first we should consider an efficient weighting of the models is always better than the equally weighting, because with equally weighting, we has no punishment for the bad performers and no reward for the good performers, like a benevolent father. And we combine them equally. But if we would combine them in most efficient way, it would be better for sure. In this study we use a primitive weighted combination model, ranking as an example. Both comparisons showed that the ranking which is combined with bootstrapping is better. The reason is obvious I think. In an emerging market, one must use the recent data for estimation of the model parameters, because every thing can change over a period like one-year. But to weight them we use recent data for efficient combination. With bootstrapping we lost no data. We used all data to estimation of the parameters and combined those reshuffled or bootstrapped data. For example without bootstrapped combination if economy had a recent crisis, there would be no change in the parameters of the models. With bootstrapping, since one can use all data for estimation of model parameters, this is not the case.

To sum up, the Turkish data shows the robustness of the “Thick Models” over thin models. To impose one specification on a series seems misleading. The possibility of different distributions and nonlinearity might create room for better, more flexible models. And also usage of bootstrapping to combine forecasts can improve the results.

References

Balaban, E. (1999), Comparative forecasting performance of symmetric and asymmetric conditional volatility models of an exchange rate, Discussion Paper 9905, Central Bank of the Republic of Turkey

Bates, J.M. and C.W.J. Granger (1969) “The Combination of Forecasts.” *Operational Research Quarterly* 20, 451-468.

Diebold, F. X. and Jose A. Lopez (1996) Forecast Evaluation and Combination, in *Handbook of Statistics* G.S. Maddala and C.R. Rao eds., North-Holland, Amsterdam.

Diebold, F.X. and P. Pauly, (1987), Structural change and the combination of forecasts, *Journal of Forecasting* 6, 21-40.

Diebold, F.X., (1988), Serial correlation and the combination of forecasts, *Journal of Business and Economic Statistics* 6, 105-111.

Gonzalez-Rivera, G., T. Lee, and S. Mishra (2004), Forecasting Volatility: A Reality Check Based on Option Pricing, Utility Function, Value-at-Risk, and Predictive Likelihood, To Appear in *International Journal of Forecasting*

Granger, C. W. J. and Yongil Jeon (2001), Thick Modeling

Granger, C.W.J. (2000): “An Overview of Nonlinear Macroeconomic Empirical Models.” *Journal of Dynamic Macroeconomics*

J.P. Morgan (1996), *RiskMetrics Technical Document*, 4th Edition

Jorion, P. (2000), *Value at Risk,* 2nd ed. McGraw Hill, New York

Merton, R.C., (1981), On market timing and investment performance I: An equilibrium theory of value for market forecasts, *Journal of Business* 54, 513-533.

Stock, J.H. and M.W. Watson (1999): “Forecasting Inflation.” *Journal of Monetary Economics* 44, 293-335.

Stock, J.H. and M.W. Watson (2003): “Combination Forecasts Of Output Growth in A Seven-Country Data Set” Harvard University, Cambridge

Theil, H., (1961), Economic Forecasts and Policy North-Holland, Amsterdam

Yu, Jun (1999), Forecasting volatility in the New Zealand stock market Department of Economics, University of Auckland, Auckland, New Zealand

Appendix1-Graphs

Figure 2.1 shows the ISE100 index along the period.

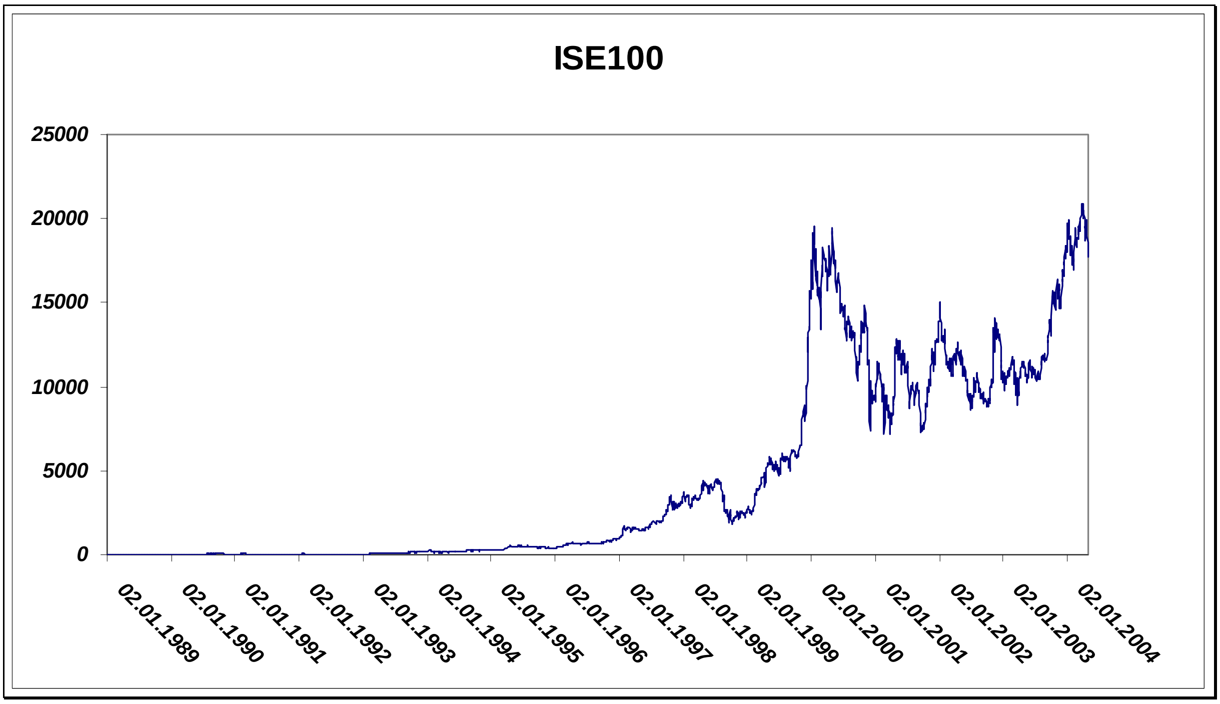


Figure 2.2 shows the return of ISE100 over period.

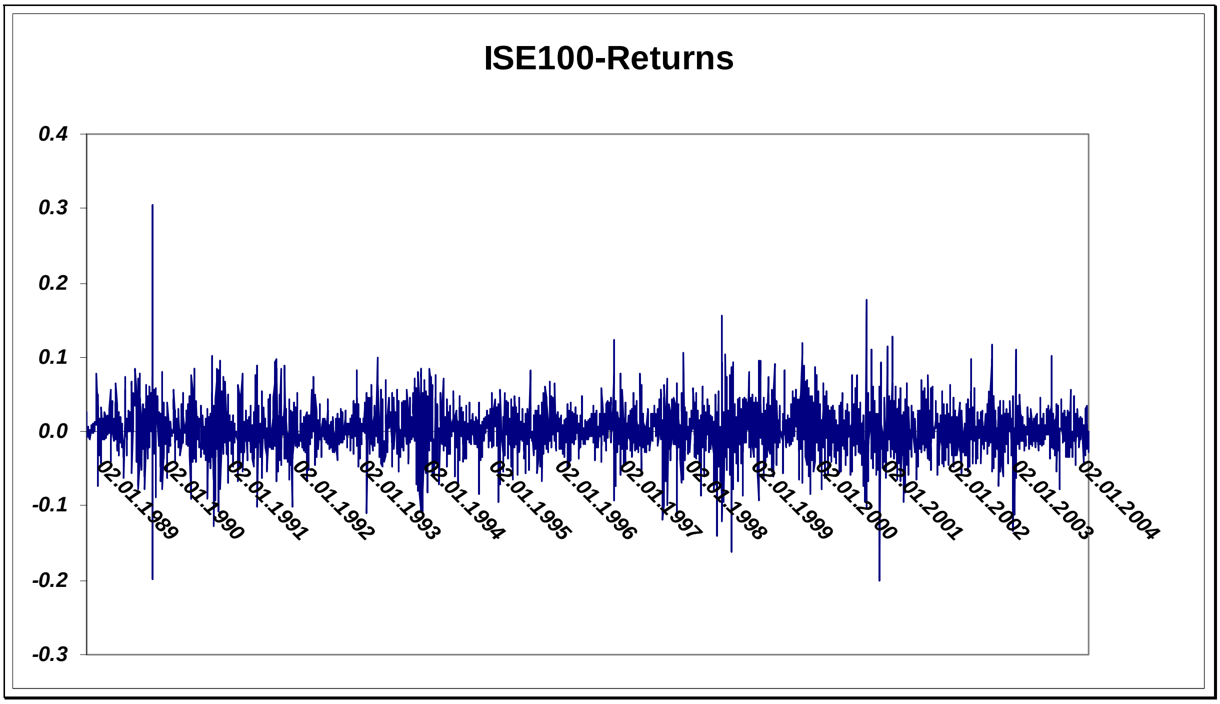


Figure 2.3 plots squared return of ISE100.

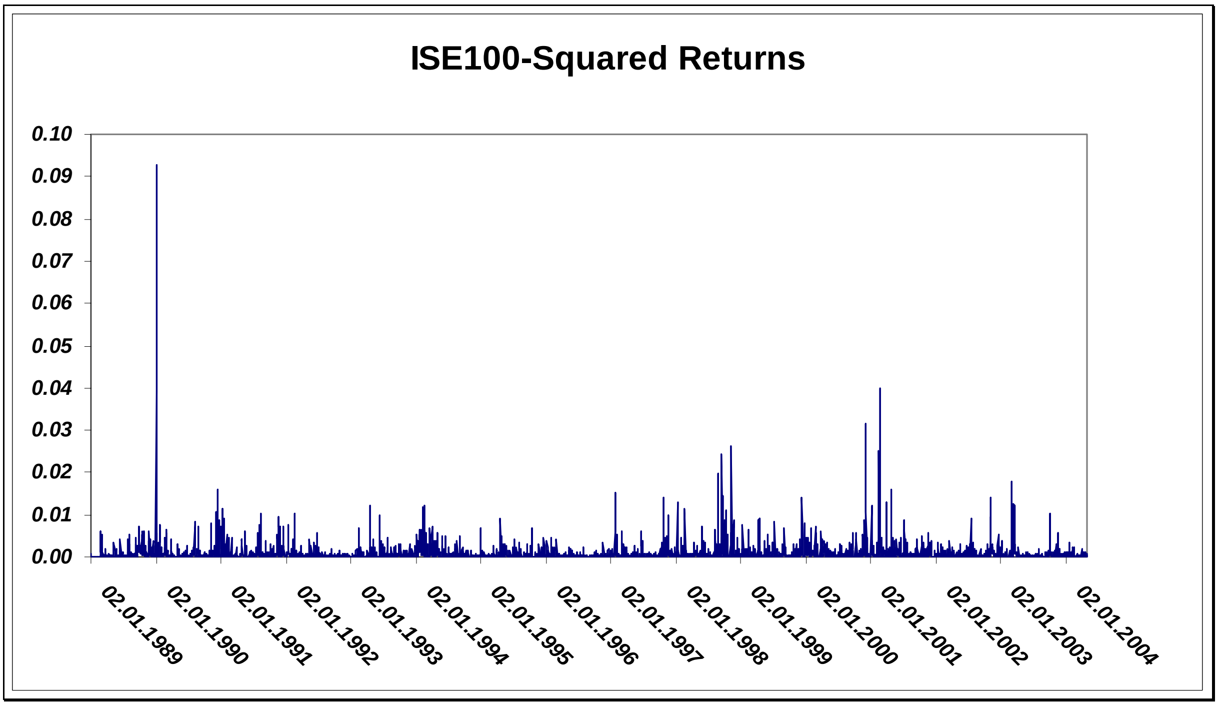


Figure2.4 shows correlogram of the returns

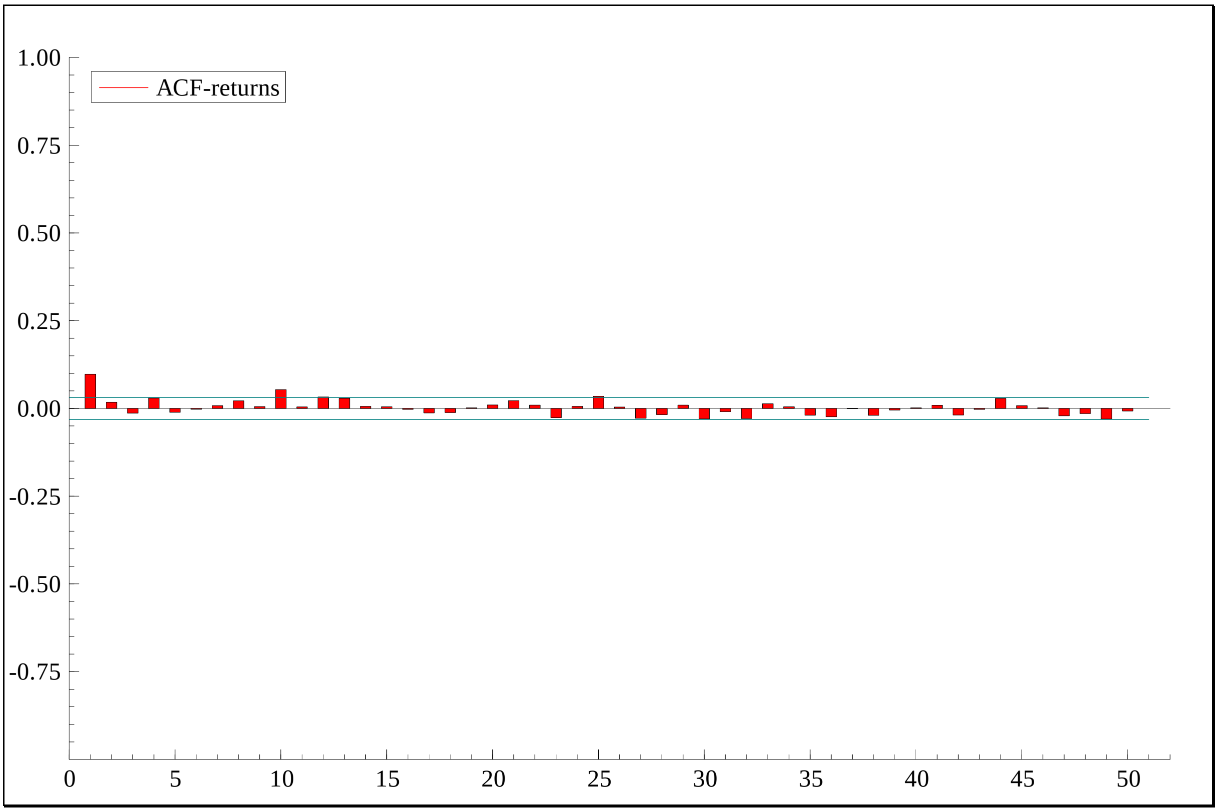


Figure 2.5 shows the correlogram of the squared returns

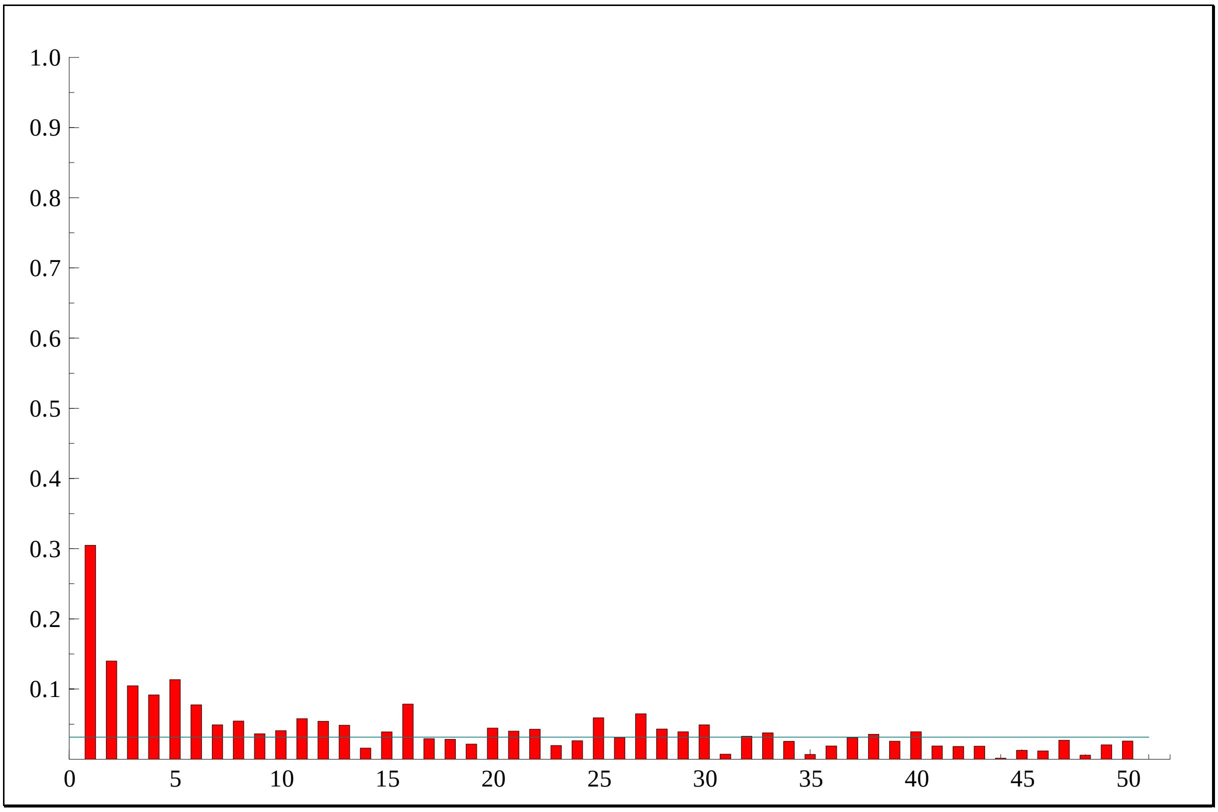


Figure 2.6 estimated density versus normal density graph and Quantile-Quantile graph (against normal).

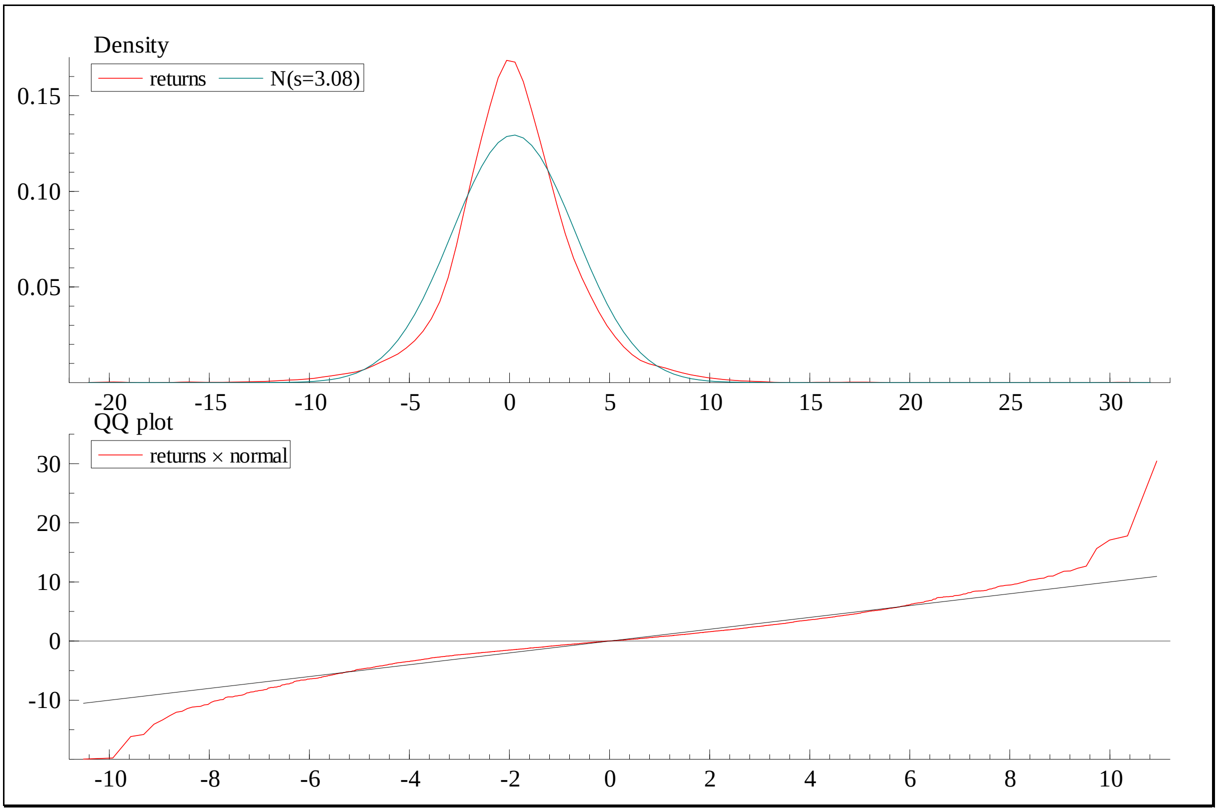


Table 2-1 shows the descriptive statistics

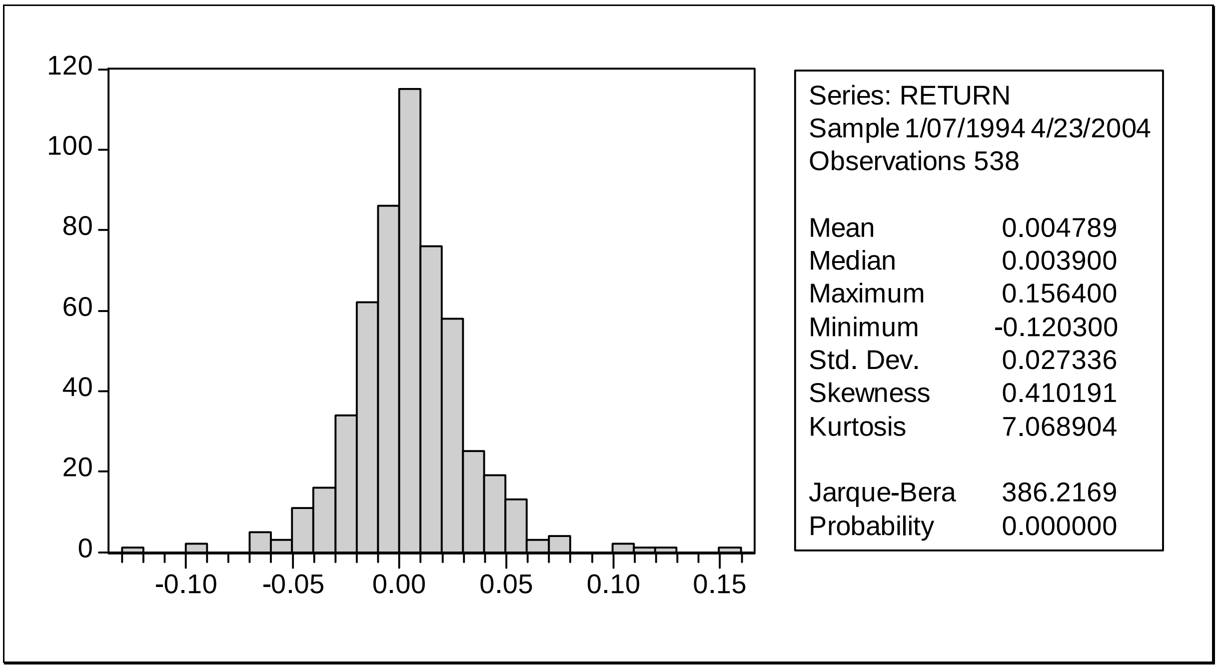
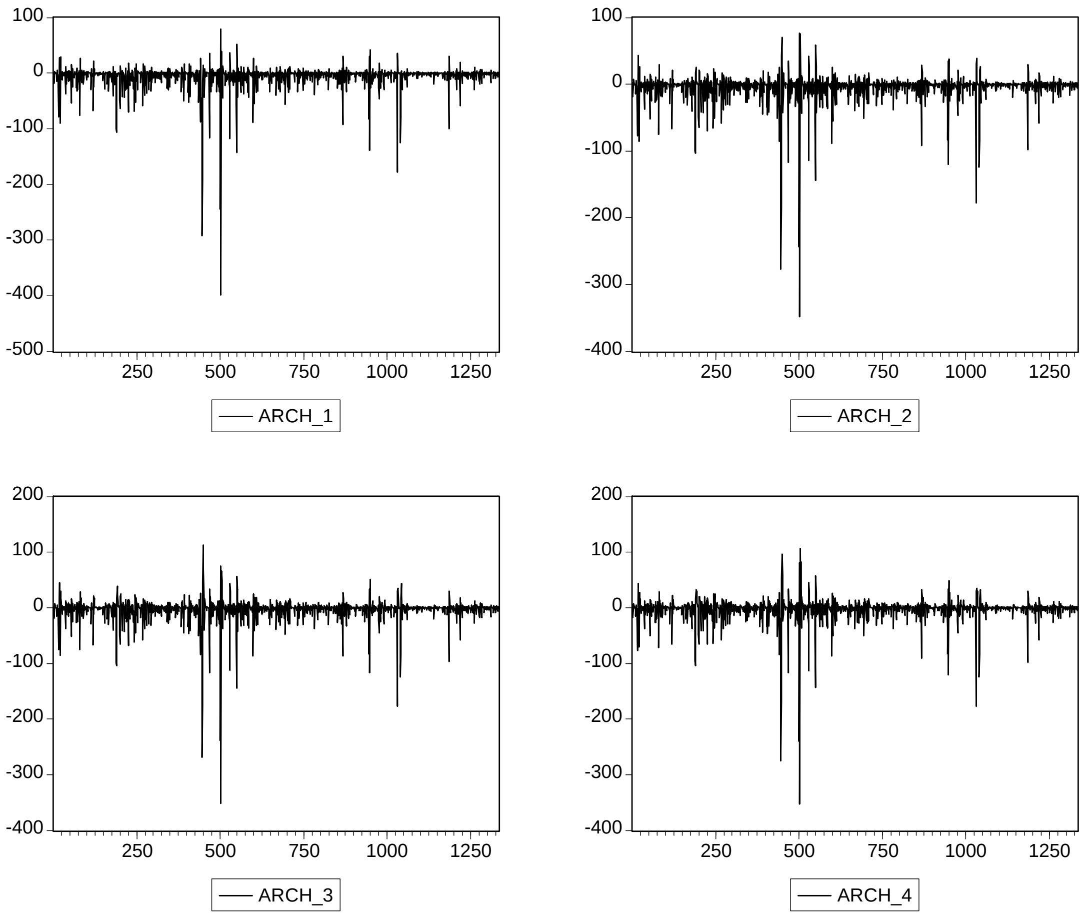
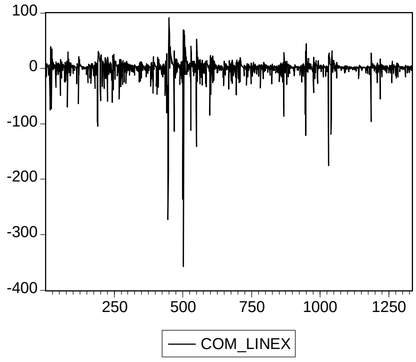
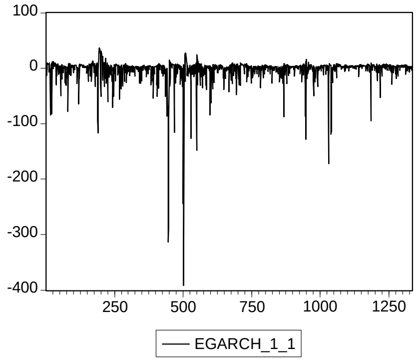


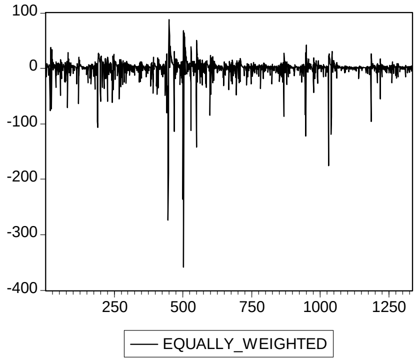
Figure 5.1 differences between forecasted and realized volatilities

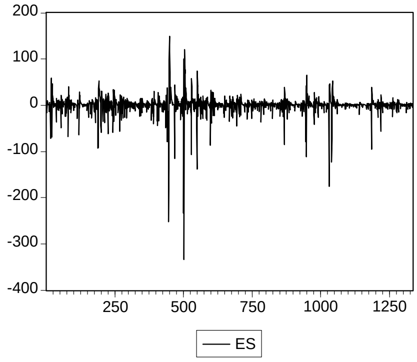


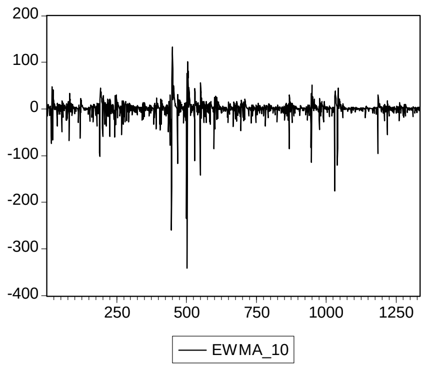
continued

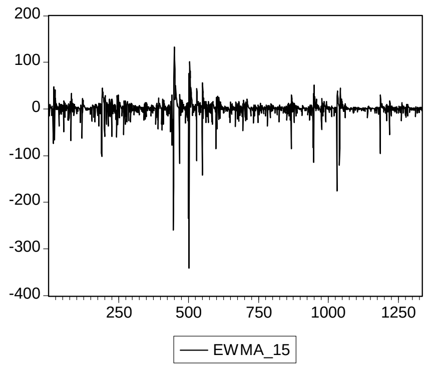

100
0
-100
-200
-300
-400
250 500 750 1000 1250
COM_LINEX

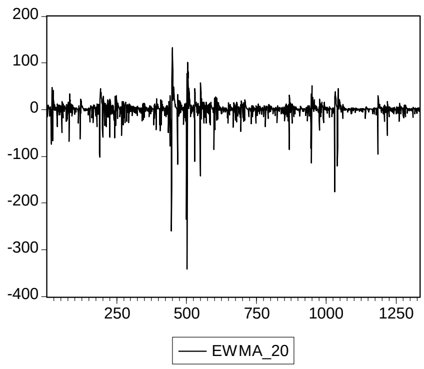

100
0
-100
-200
-300
-400
250 500 750 1000 1250
EGARCH_1_1

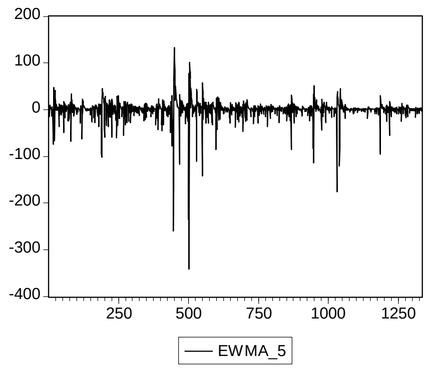

100
0
-100
-200
-300
-400
250 500 750 1000 1250
EQUALLY_WEIGHTED


200
100
0
-100
-200
-300
-400
250 500 750 1000 1250
ES

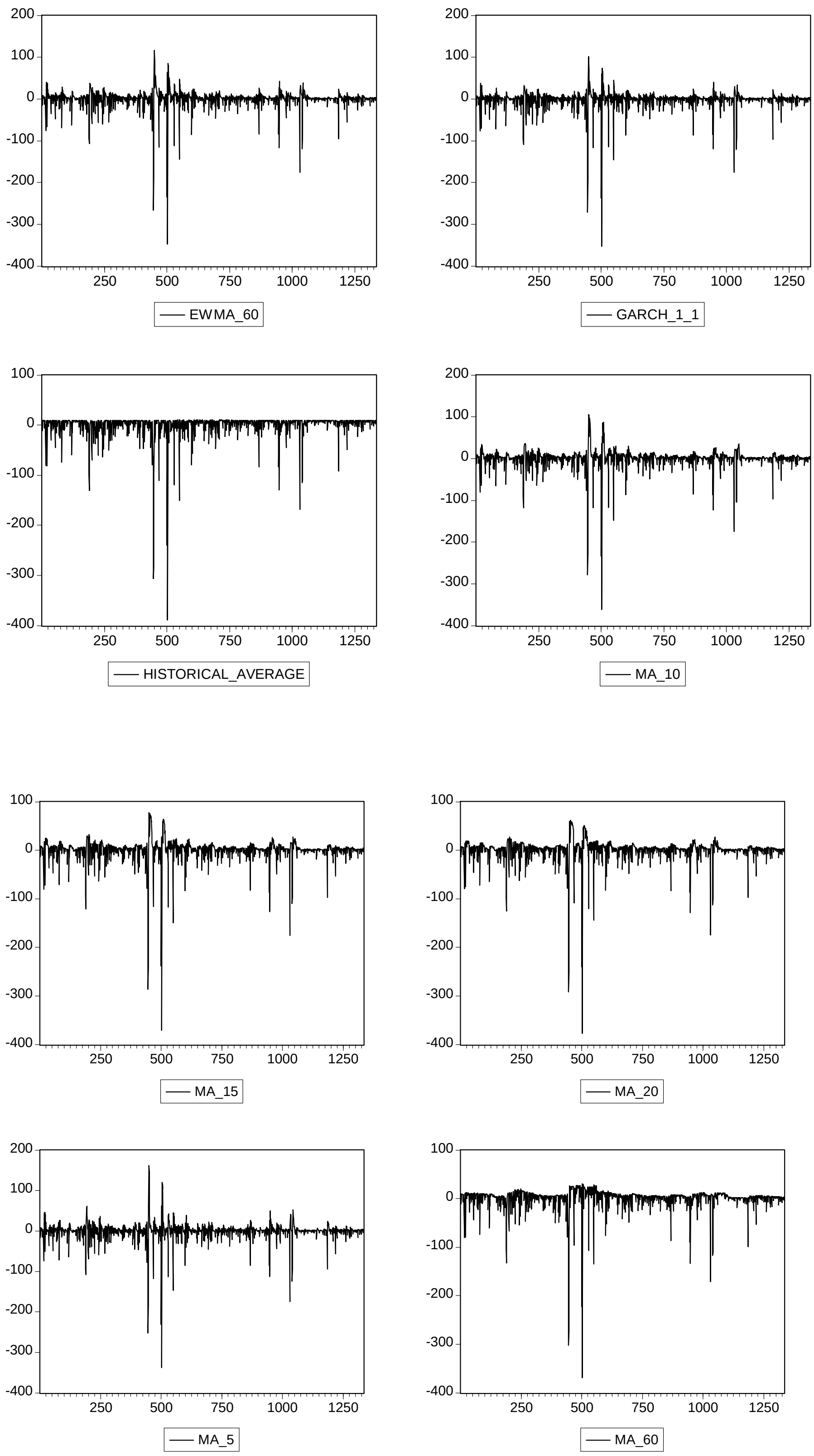

200
100
0
-100
-200
-300
-400
250 500 750 1000 1250
EWMA_10


200
100
0
-100
-200
-300
-400
250 500 750 1000 1250
EWMA_15

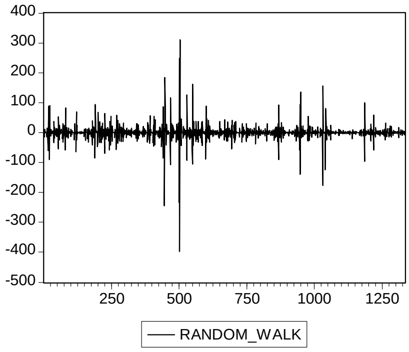

200
100
0
-100
-200
-300
-400
250 500 750 1000 1250
EWMA_20

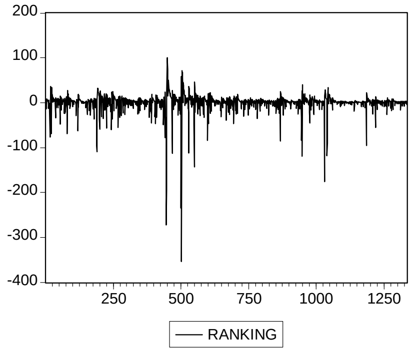

200
100
0
-100
-200
-300
-400
250 500 750 1000 1250
EWMA_5

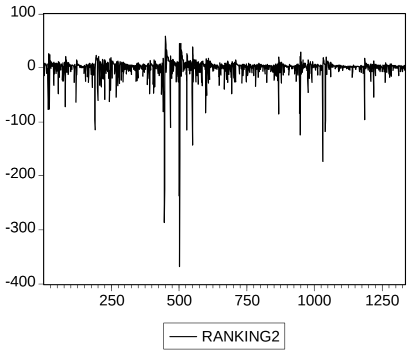
continued

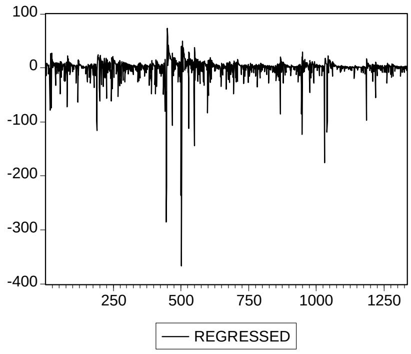


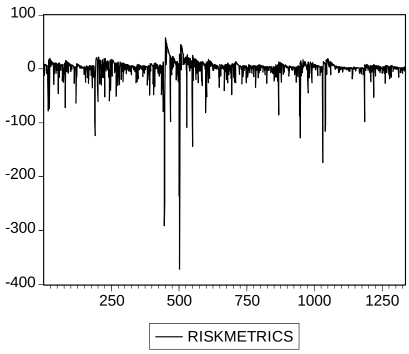
continued

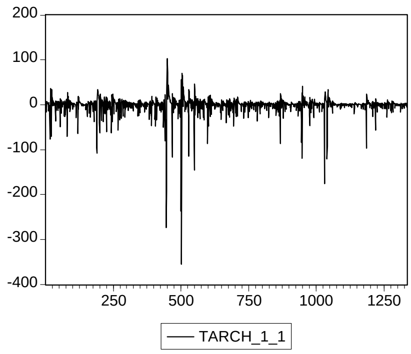

400
300
200
100
0
-100
-200
-300
-400
-500
250
500
750
1000
1250
RANDOM_WALK

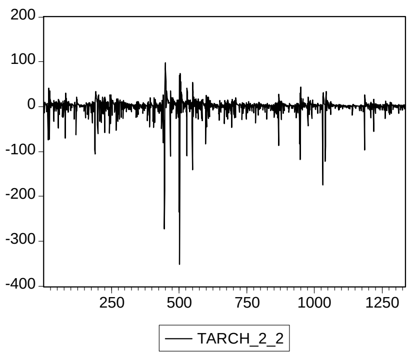

200
100
0
-100
-200
-300
-400
250
500
750
1000
1250
RANKING


100
0
-100
-200
-300
-400
250
500
750
1000
1250
RANKING2


100
0
-100
-200
-300
-400
250
500
750
1000
1250
REGRESSED


100
0
-100
-200
-300
-400
250
500
750
1000
1250
RISKMETRICS


200
100
0
-100
-200
-300
-400
250
500
750
1000
1250
TARCH_1_1


200
100
0
-100
-200
-300
-400
250
500
750
1000
1250
TARCH_2_2

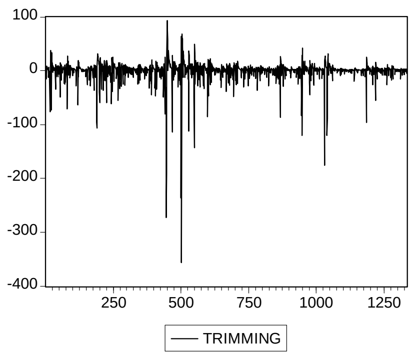

100
0
-100
-200
-300
-400
250
500
750
1000
1250
TRIMMING

Table 5-1 shows the results from RMSE

	Model	RMSE
1	TARCH(2,2)	22.91991365
2	Equally Weighted	22.96373095
3	Trimming	22.98106452
4	Cobined with linex	22.99324737
5	Ranking2	23.02576792
6	ARCH(2)	23.09702707
7	ARCH(3)	23.18128627
8	GARCH(1,1)	23.18874241
9	Regressed	23.21372854
10	Ranking	23.24955042
11	TARCH(1,1)	23.25378047
12	ARCH(4)	23.34511884
13	EWMA(60)	23.52851636
14	EWMA(10)	23.55580143
15	EWMA(15)	23.55580163
16	EWMA(5)	23.55699927
17	EWMA(20)	23.55813809
18	RiskMetrics	23.72942839
19	ES	23.73136505
20	MA(60)	23.83023953
21	ARCH(1)	23.9846088
22	Historical Average	24.22853009
23	MA(20)	24.57949551
24	MA(10)	24.5818688
25	EGARCH(1,1)	24.58408306
26	MA(15)	24.59981647
27	MA(5)	24.63746929
28	Random Walk	28.49191514

Table 5-2 shows the results from Theil-U

	Model	Theil-U
1	TARCH(2,2)	0.647061094
2	Equally Weighted	0.649558148
3	Trimming	0.650533049
4	Cobined with linex	0.651225098
5	Ranking2	0.653059291
6	ARCH(2)	0.657152274
7	ARCH(3)	0.661952362
8	GARCH(1,1)	0.662354354
9	Regressed	0.663763791
10	Ranking	0.665814414
11	TARCH(1,1)	0.666072584
12	ARCH(4)	0.671327272
13	EWMA(60)	0.681888747
14	EWMA(10)	0.683476151
15	EWMA(15)	0.683476162
16	EWMA(5)	0.683545778
17	EWMA(20)	0.683611976
18	RiskMetrics	0.693563032
19	ES	0.693712363
20	MA(60)	0.699465194
21	ARCH(1)	0.708634014
22	Historical Average	0.723038689
23	MA(20)	0.744170584
24	MA(10)	0.744296172
25	EGARCH(1,1)	0.744428401
26	MA(15)	0.745401846
27	MA(5)	0.747662773
28	Random Walk	1

Table 5-3 shows the results from Linex with a=10

	Model	Linex(a=10)
1	TARCH(2,2)	0.000276616
2	Equally Weighted	0.0002783
3	Trimming	0.000278589
4	Cobined with linex	0.000278773
5	ARCH(2)	0.000281647
6	Ranking2	0.000281688
7	ARCH(3)	0.000282833
8	GARCH(1,1)	0.000283265
9	Ranking	0.000284423
10	TARCH(1,1)	0.000284917
11	Regressed	0.000285734
12	ARCH(4)	0.00028662
13	EWMA(10)	0.000289555
14	EWMA(15)	0.000289555
15	EWMA(5)	0.000289562
16	EWMA(20)	0.00028957
17	EWMA(60)	0.000289988
18	ES	0.000292248
19	RiskMetrics	0.000299041
20	MA(60)	0.000302235
21	ARCH(1)	0.000306601
22	MA(5)	0.000314221
23	Historical Average	0.000314237
24	MA(10)	0.000317014
25	MA(15)	0.000319233
26	MA(20)	0.000319899
27	EGARCH(1,1)	0.000323979
28	Random Walk	0.000411936

Table 5-4 shows the results from Linex with a=20

	Model	Linex(a=20)
1	TARCH(2,2)	0.001170723
2	Equally Weighted	0.001180686
3	Trimming	0.001181287
4	Equally Weighted	0.001181564
5	Cobined with linex	0.001181652
6	ARCH(2)	0.001194832
7	ARCH(3)	0.00119634
8	GARCH(1,1)	0.001199385
9	Ranking	0.00120297
10	Ranking2	0.0012034
11	TARCH(1,1)	0.001206718
12	ARCH(4)	0.001211639
13	EWMA(20)	0.001214202
14	EWMA(5)	0.001214259
15	EWMA(15)	0.001214324
16	EWMA(10)	0.001214324
17	Regressed	0.001218222
18	ES	0.001218699
19	EWMA(60)	0.001220922
20	RiskMetrics	0.00127728
21	MA(60)	0.001293685
22	MA(5)	0.001307458
23	ARCH(1)	0.001314972
24	MA(10)	0.001337386
25	Historical Average	0.001353581
26	MA(15)	0.001354432
27	MA(20)	0.001362565
28	EGARCH(1,1)	0.001397605
29	Random Walk	0.001686253

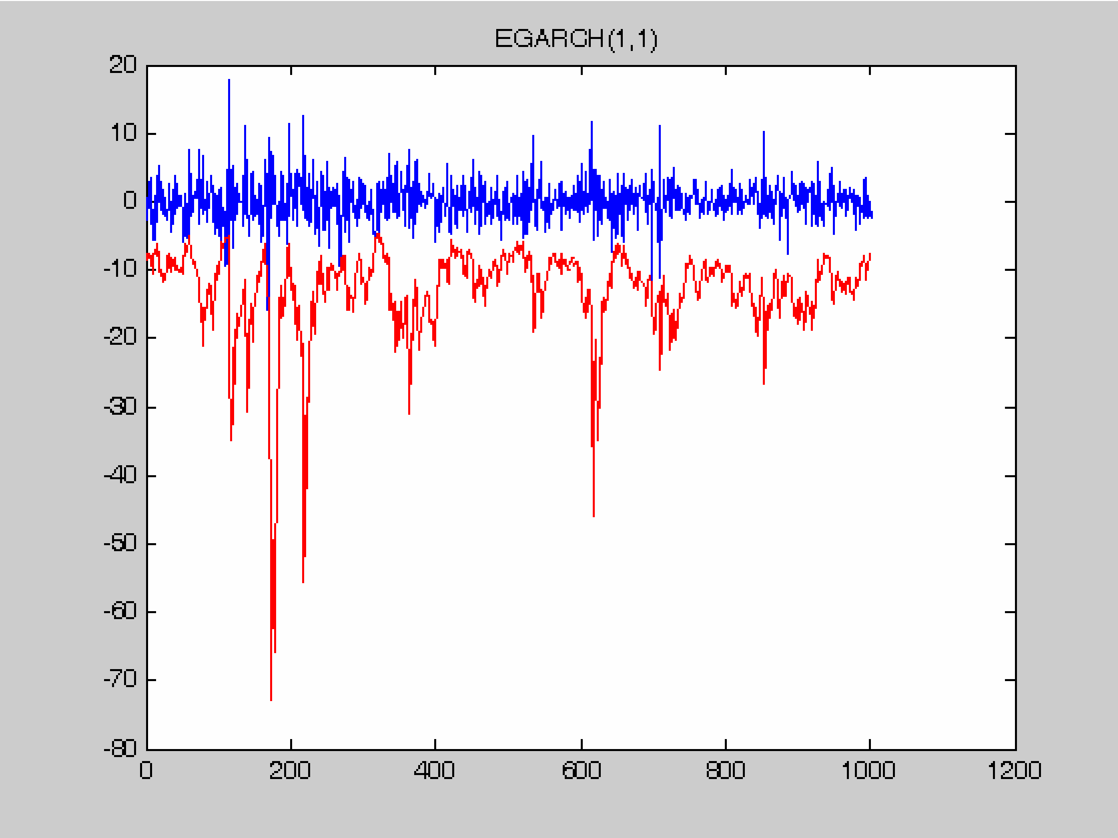
Table 5-5 the zones defined by Basel Committee

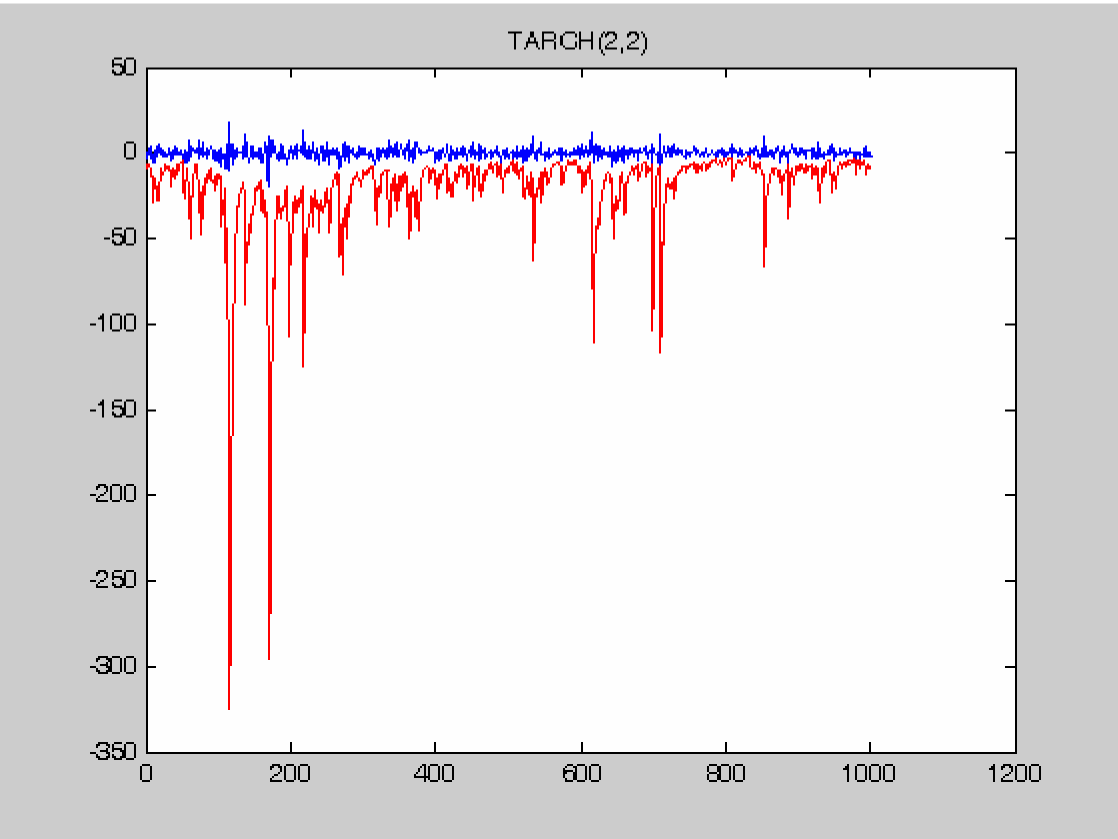
	Number of Exceptions	Increase in k
Green	0-4	0.00
Yellow	5	0.40
	6	0.50
	7	0.65
	8	0.75
	9	0.85
Red	10+	1.00

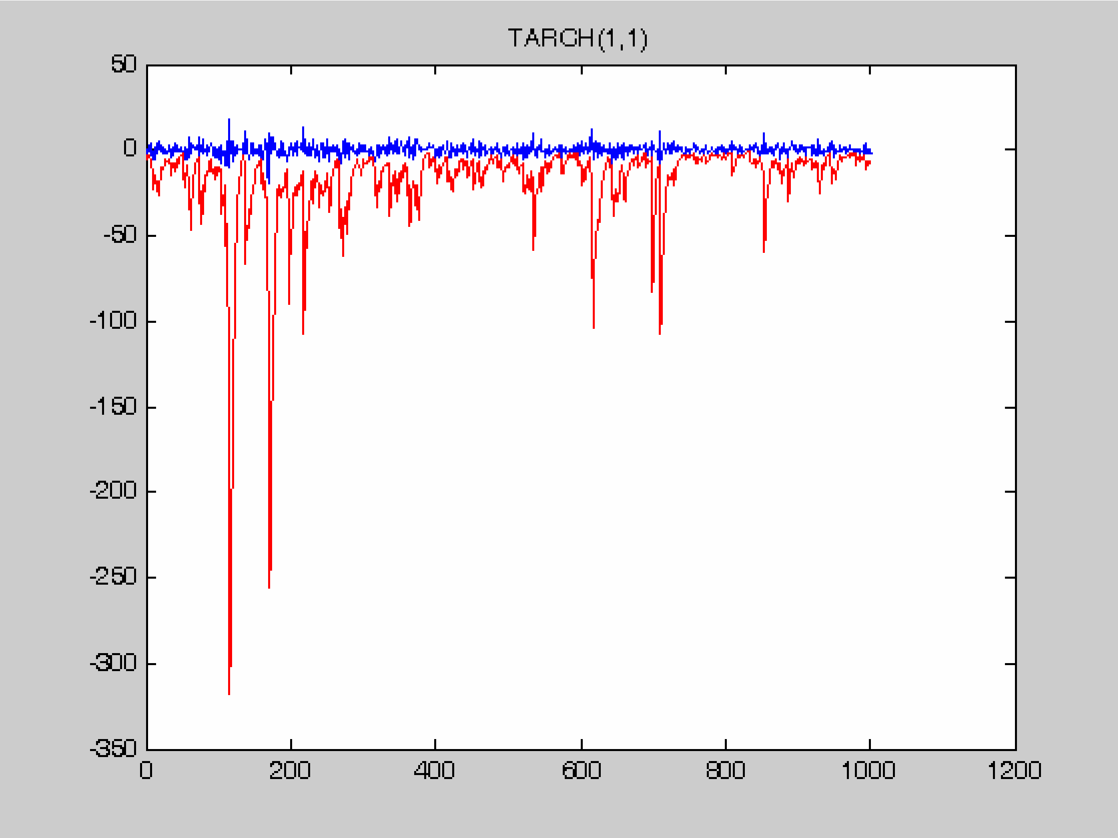
Table 5-6 shows the number of exceptions with 99% confidence level.

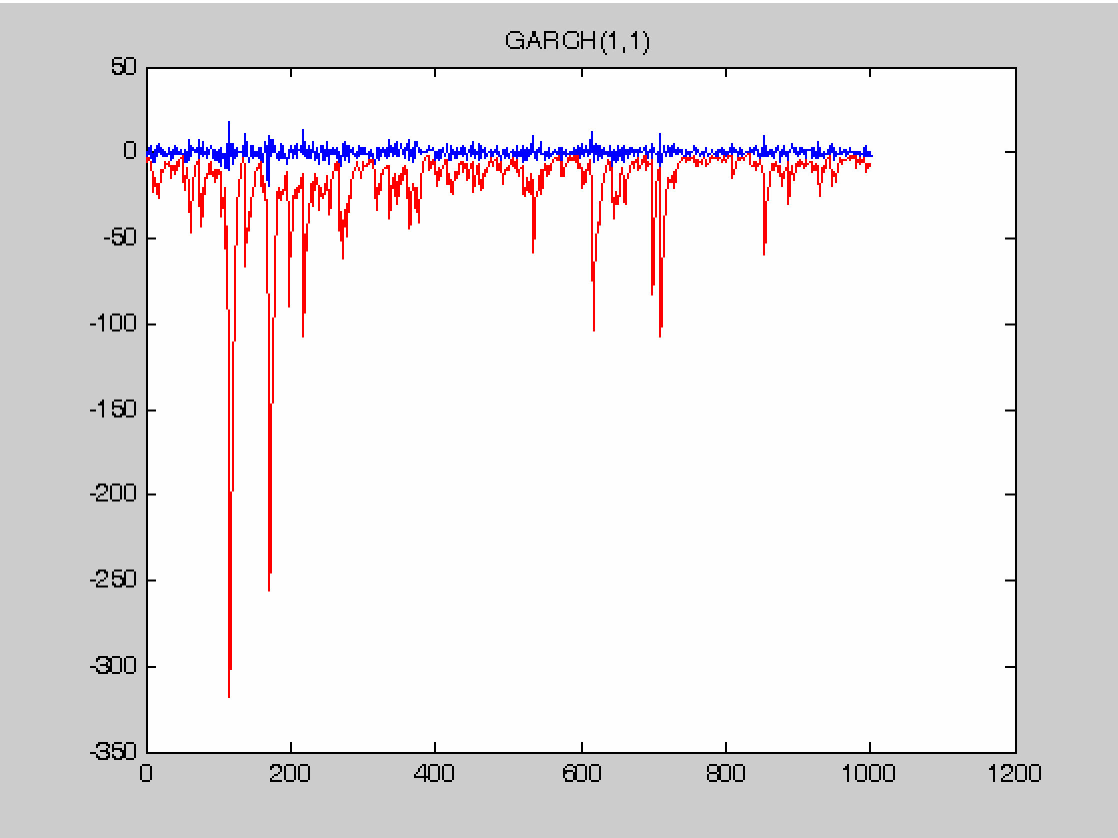
		Number of Exceptions		Percentage	
		last 252	long period (1000)	last 252	long period (1000)
1	Historical Average	0	0	0.00%	0.00%
2	MA(60)	0	0	0.00%	0.00%
3	Ranking2	0	1	0.00%	0.08%
4	RiskMetrics	0	1	0.00%	0.08%
5	EGARCH(1,1)	0	8	0.00%	0.60%
6	MA(15)	1	2	0.40%	0.15%
7	Cobined with linex	1	3	0.40%	0.23%
8	MA(20)	1	3	0.40%	0.23%
9	Ranking	1	3	0.40%	0.23%
10	Regressed	1	3	0.40%	0.23%
11	Equally Weighted2	2	4	0.79%	0.30%
12	EWMA(60)	2	5	0.79%	0.38%
13	TARCH(2,2)	2	5	0.79%	0.38%
14	MA(10)	2	6	0.79%	0.45%
15	Trimming	2	6	0.79%	0.45%
16	EWMA(10)	3	9	1.19%	0.68%
17	EWMA(15)	3	9	1.19%	0.68%
18	EWMA(20)	3	9	1.19%	0.68%
19	EWMA(5)	3	9	1.19%	0.68%
20	GARCH(1,1)	5	13	1.98%	0.98%
21	TARCH(1,1)	5	13	1.98%	0.98%
22	ES	6	15	2.38%	1.13%
23	MA(5)	7	23	2.78%	1.73%
24	ARCH(3)	18	72	7.14%	5.40%
25	ARCH(4)	19	51	7.54%	3.83%
26	ARCH(2)	35	123	13.89%	9.23%
27	Random Walk	38	143	15.08%	10.73%
28	ARCH(1)	62	237	24.60%	17.78%

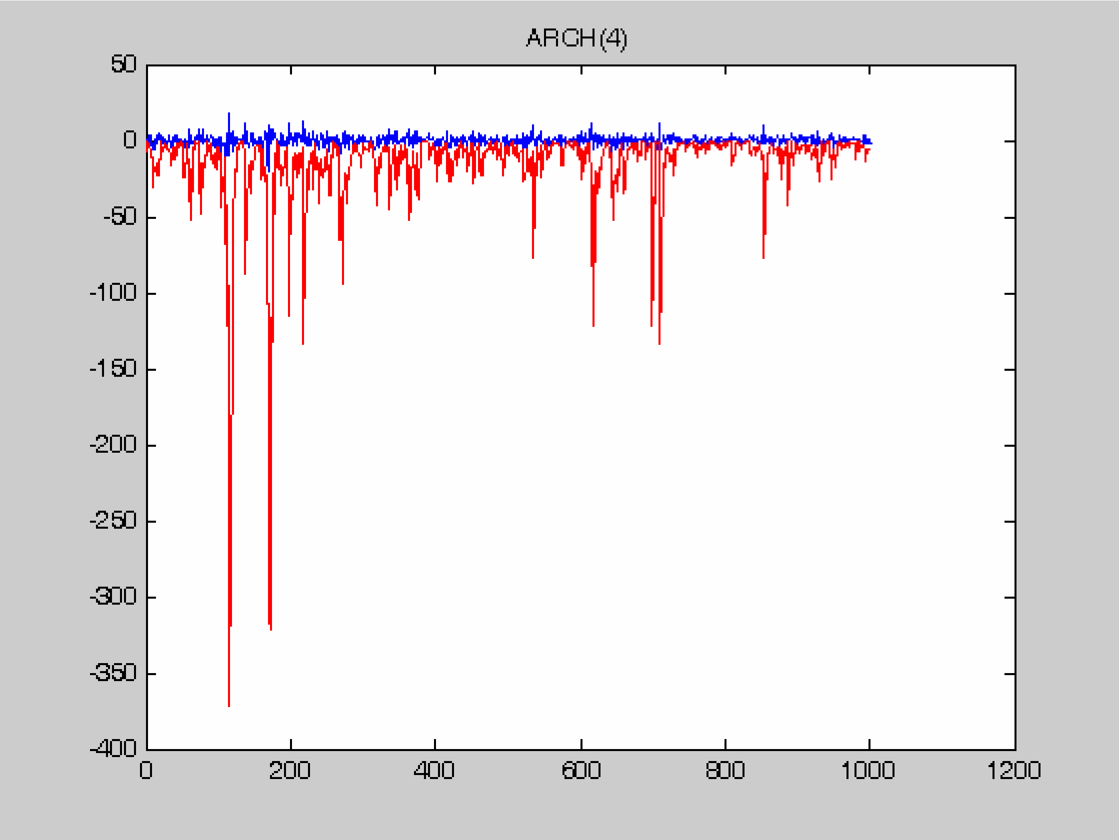
Figure 5.1 VaR calculated with forecasted volatility versus daily returns

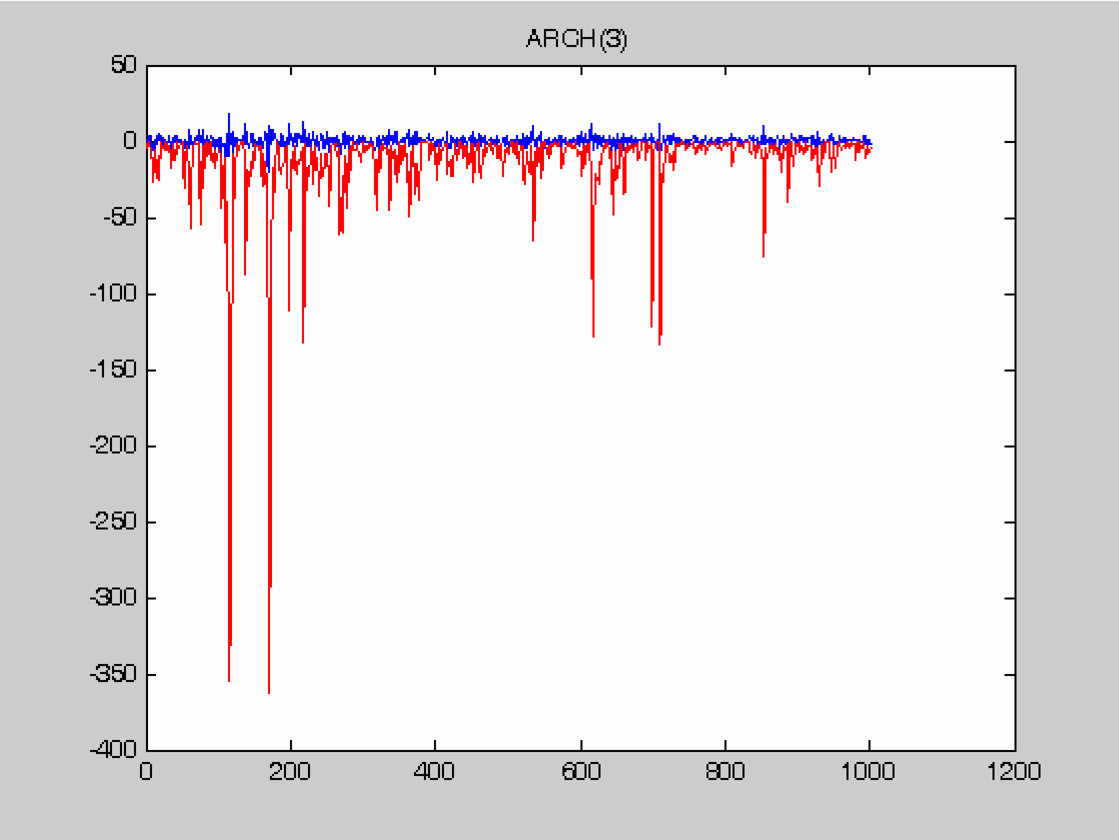


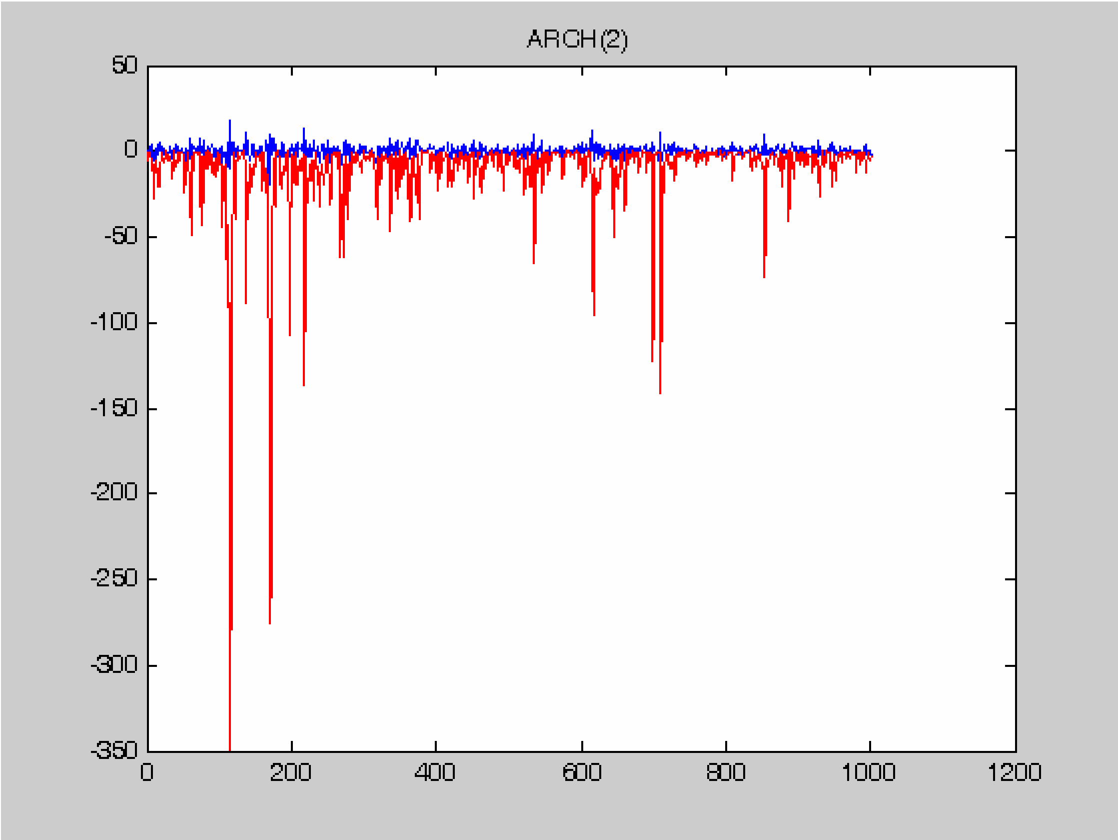


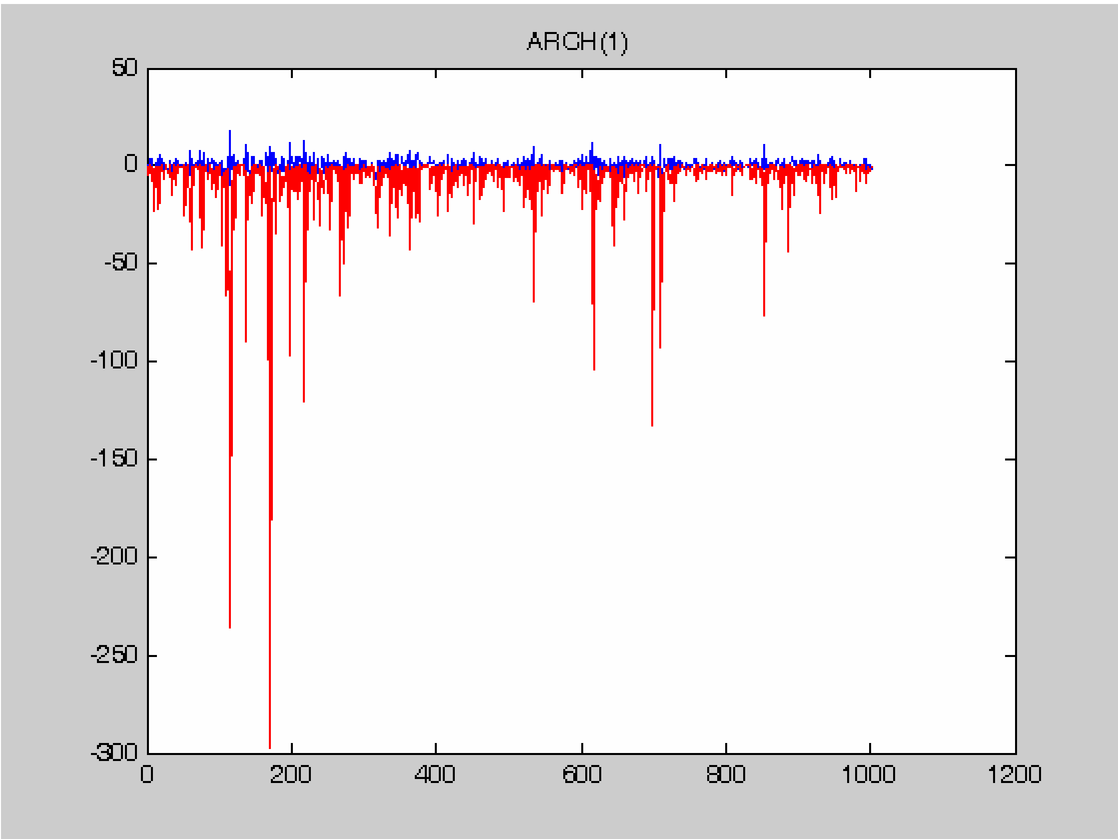

TARCH(1,1)
50
0
-50
-100
-150
-200
-250
-300
-350
0
200
400
600
800
1000
1200

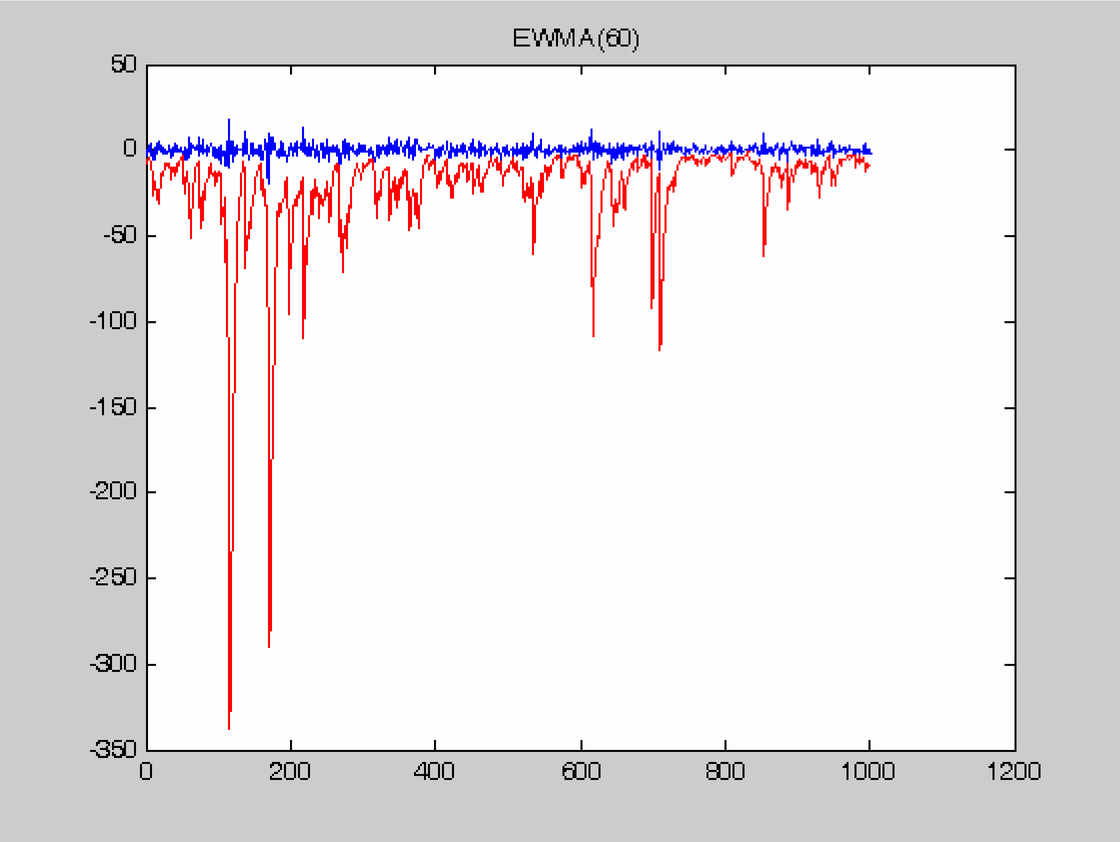

GARCH(1,1)
50
0
-50
-100
-150
-200
-250
-300
-350
0
200
400
600
800
1000
1200

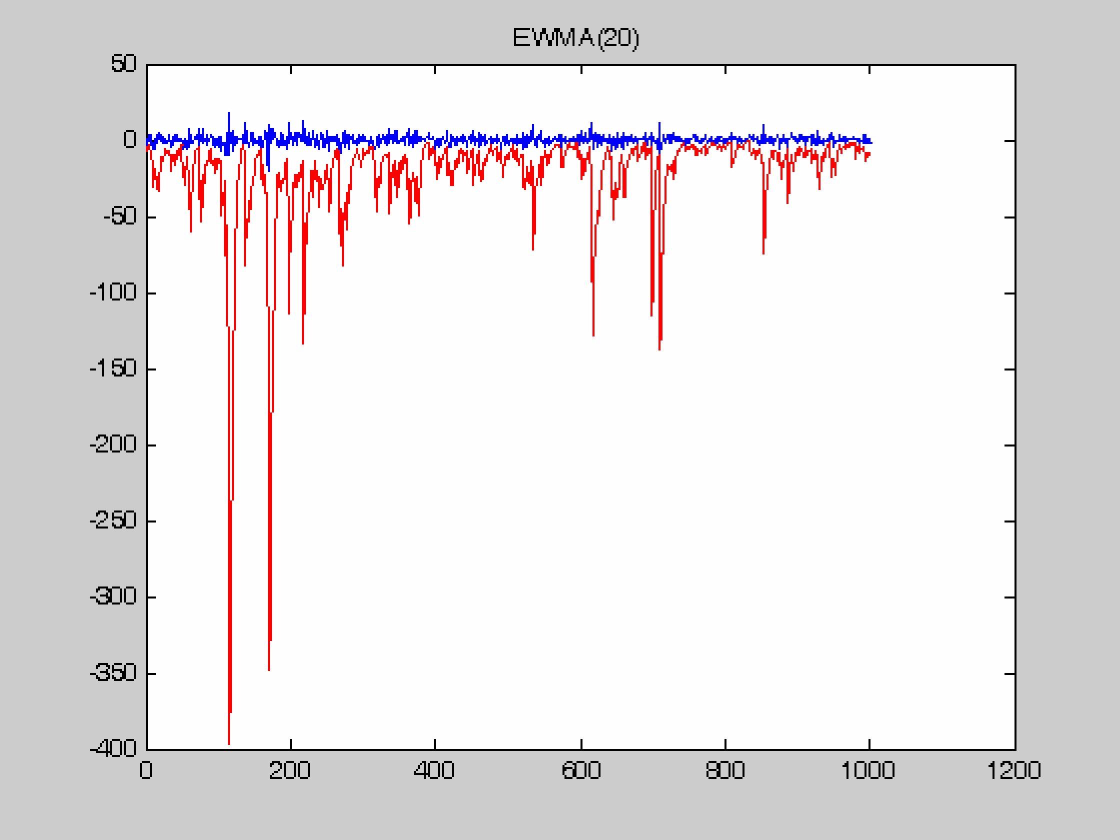

ARCH(4)
50
0
-50
-100
-150
-200
-250
-300
-350
-400
0
200
400
600
800
1000
1200

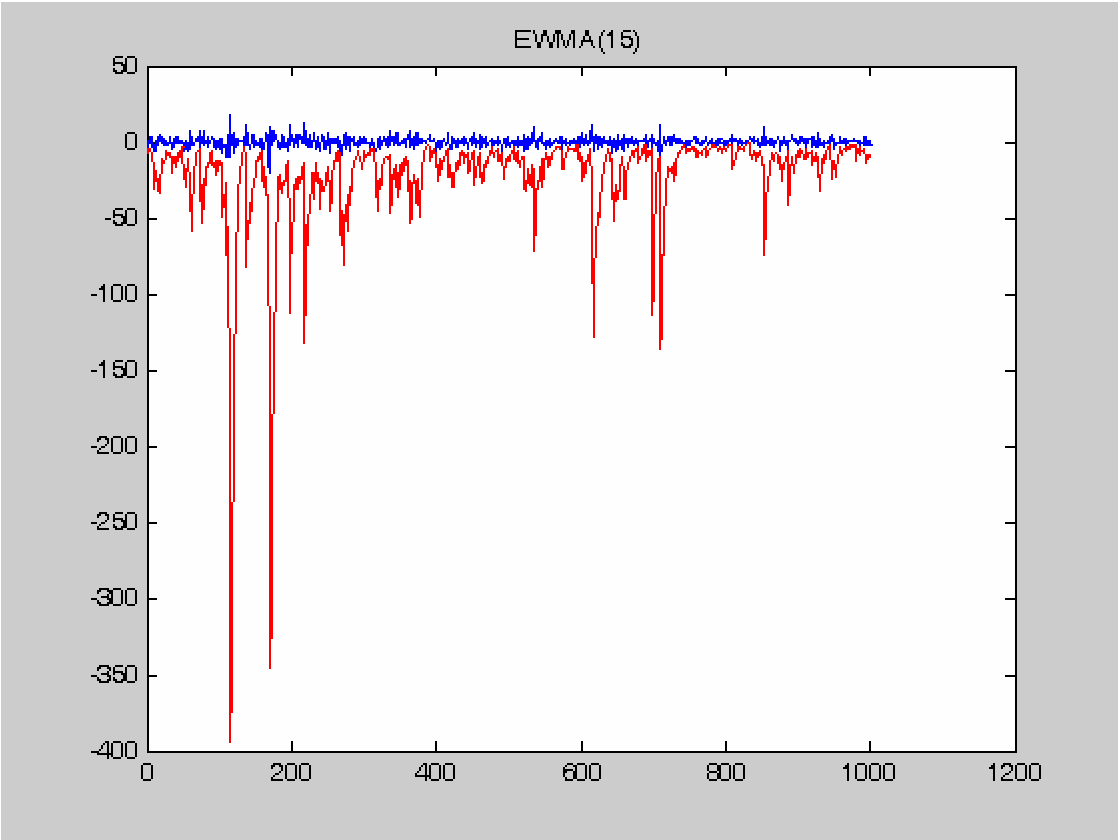

ARCH(3)
50
0
-50
-100
-150
-200
-250
-300
-350
-400
0
200
400
600
800
1000
1200

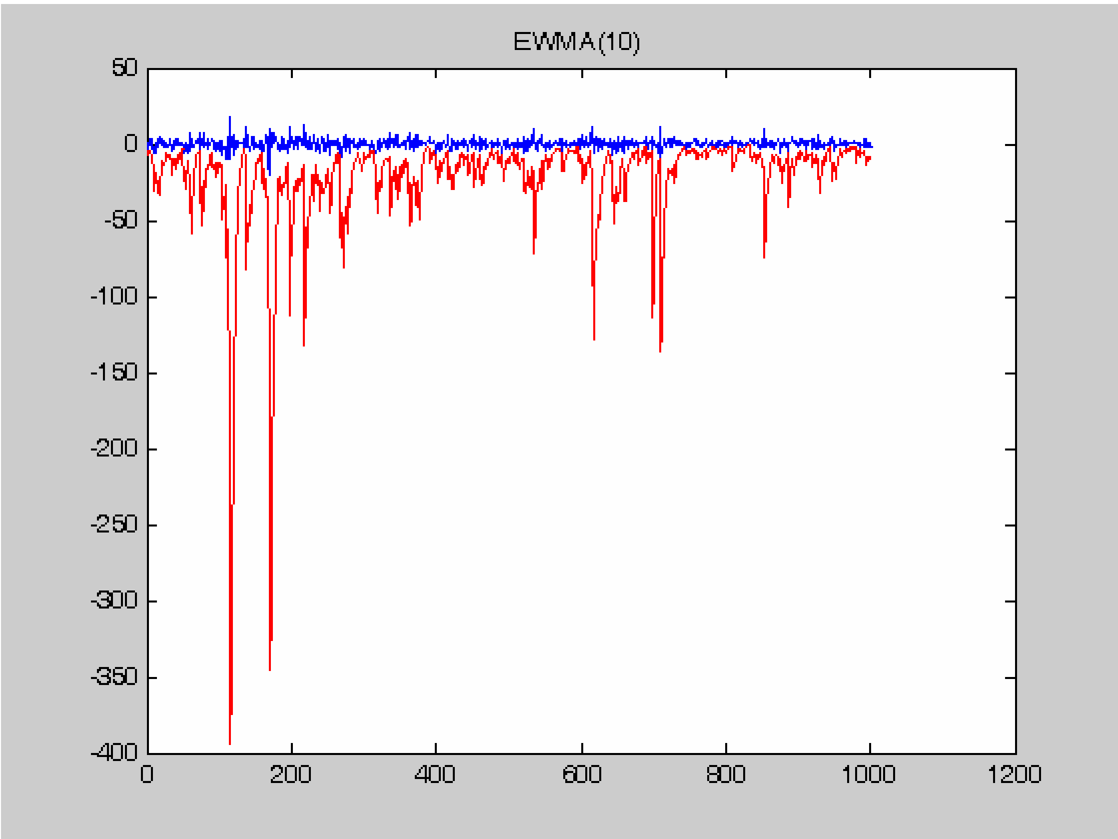

ARCH(2)
50
0
-50
-100
-150
-200
-250
-300
-350
0
200
400
600
800
1000
1200

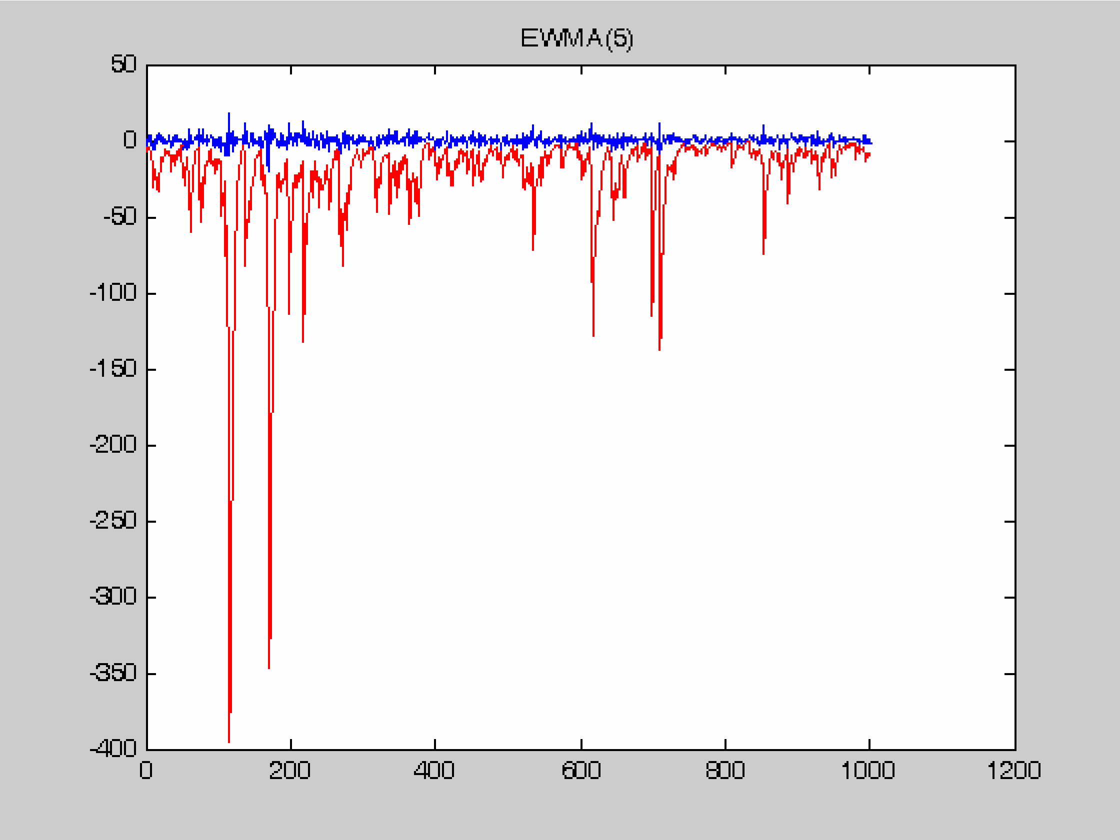

ARCH(1)
50
0
-50
-100
-150
-200
-250
-300
0
200
400
600
800
1000
1200

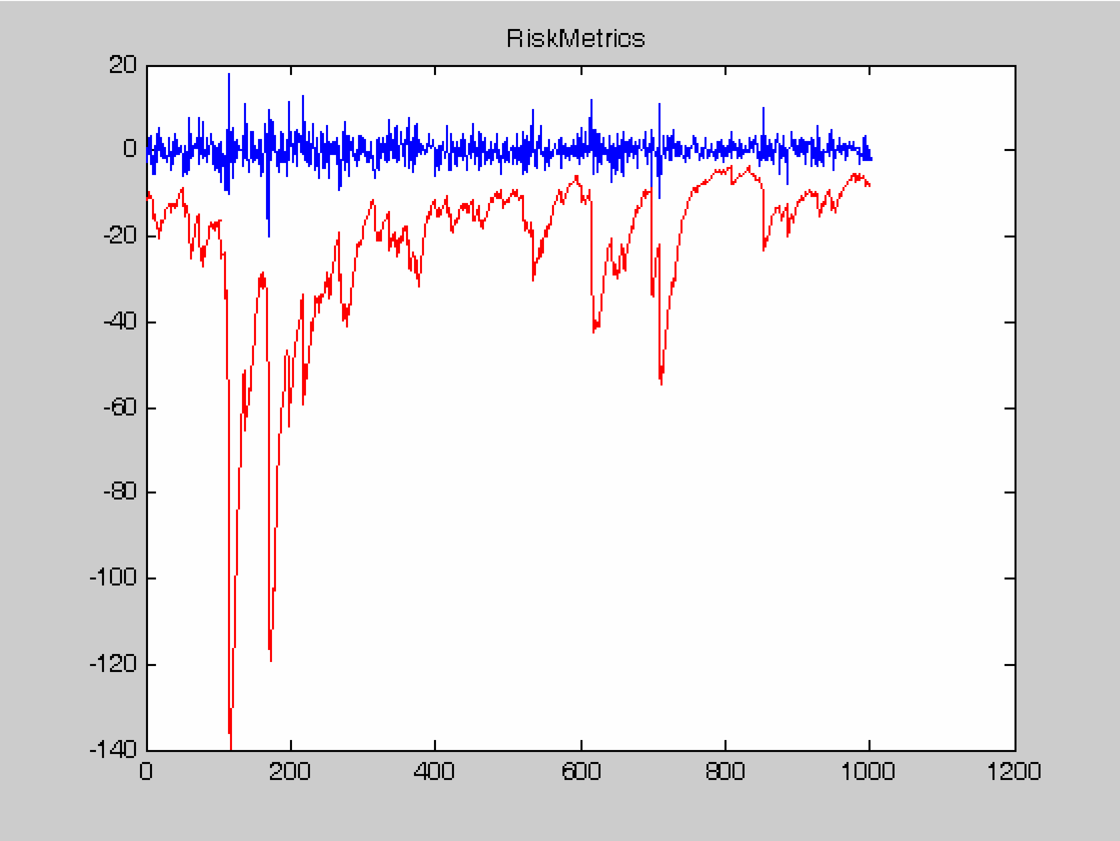

EWMA(60)
50
0
-50
-100
-150
-200
-250
-300
-350
0
200
400
600
800
1000
1200

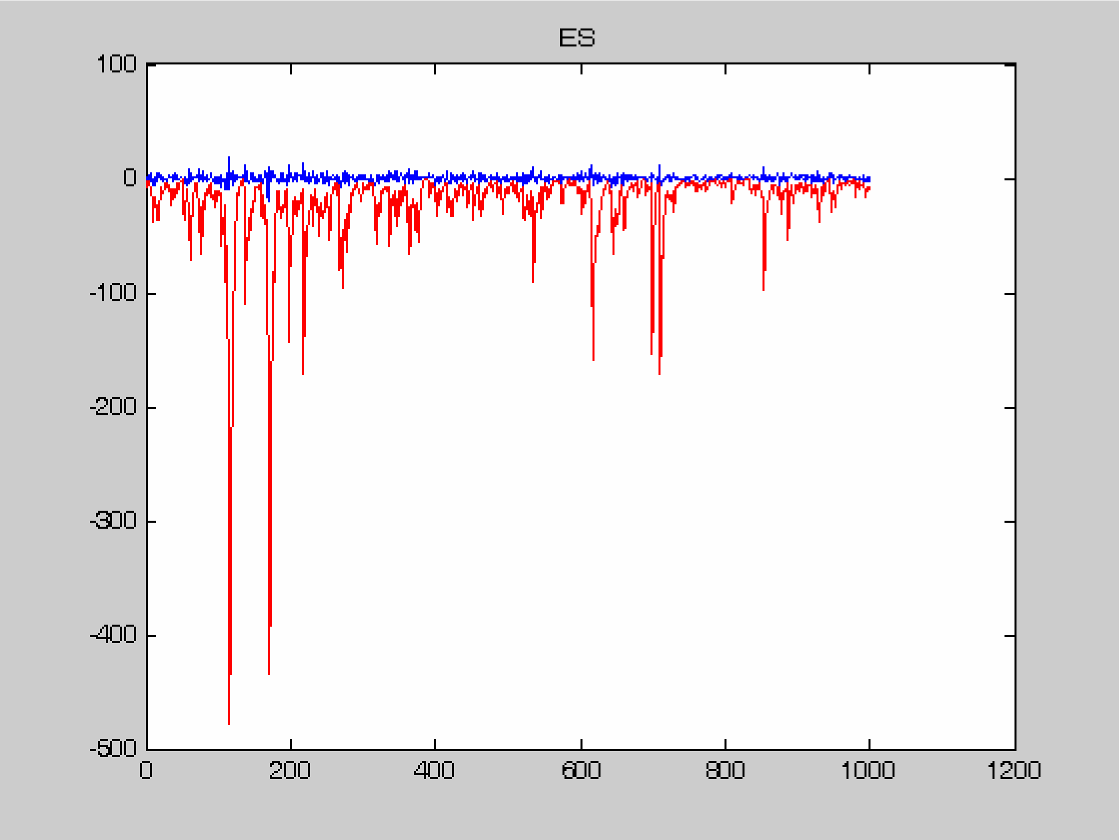

EWMA(20)
50
0
-50
-100
-150
-200
-250
-300
-350
-400
0
200
400
600
800
1000
1200

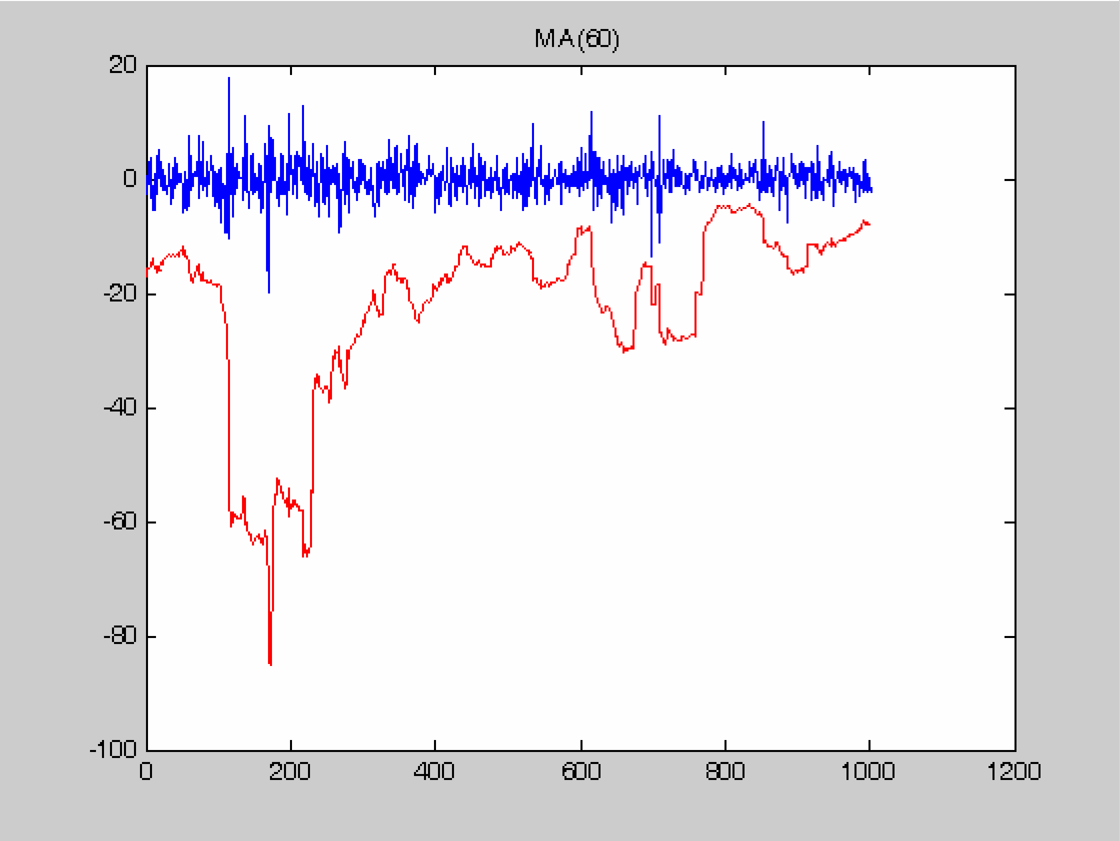

EWMA(15)
50
0
-50
-100
-150
-200
-250
-300
-350
-400
0
200
400
600
800
1000
1200

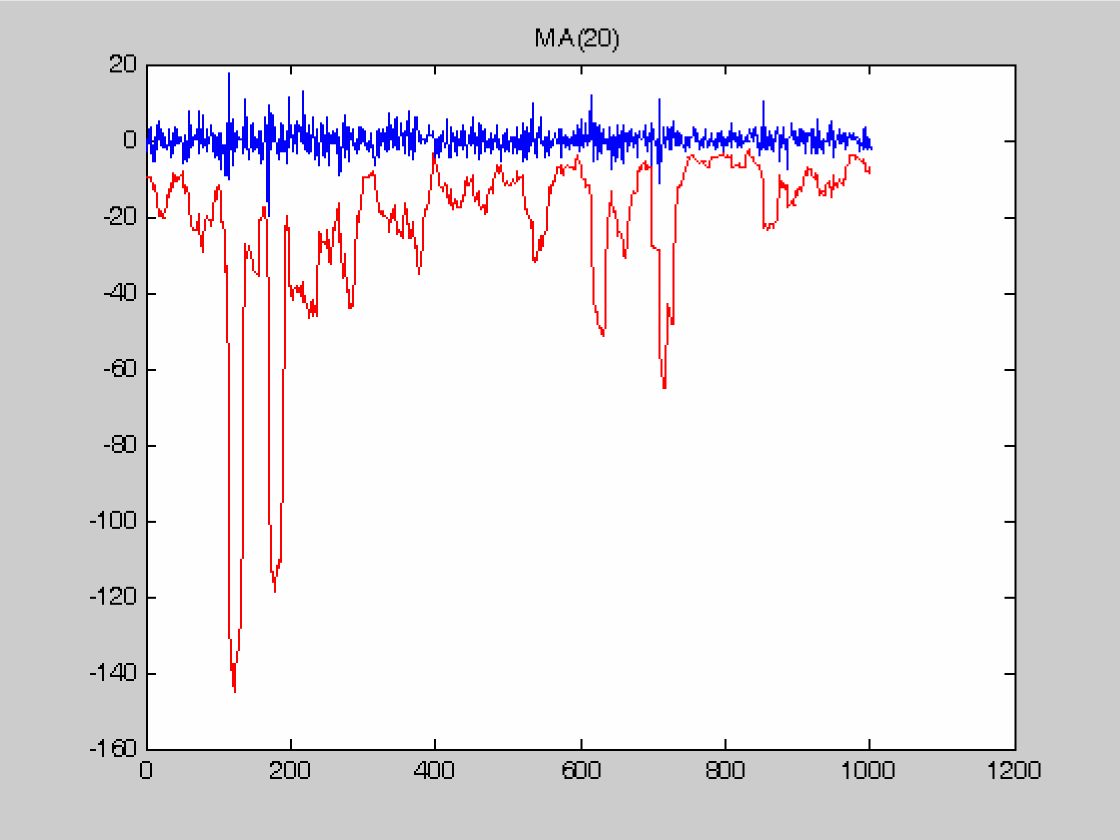

EWMA(10)
50
0
-50
-100
-150
-200
-250
-300
-350
-400
0
200
400
600
800
1000
1200

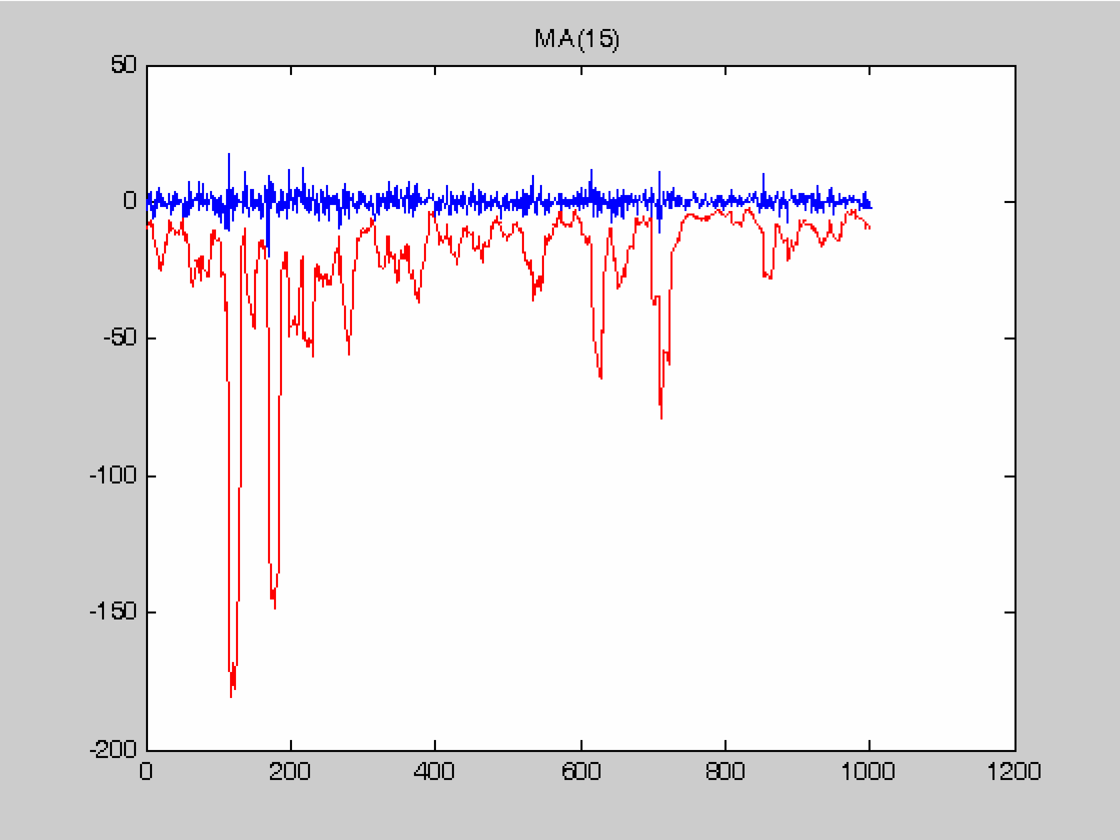

EWMA(5)
50
0
-50
-100
-150
-200
-250
-300
-350
-400
0
200
400
600
800
1000
1200

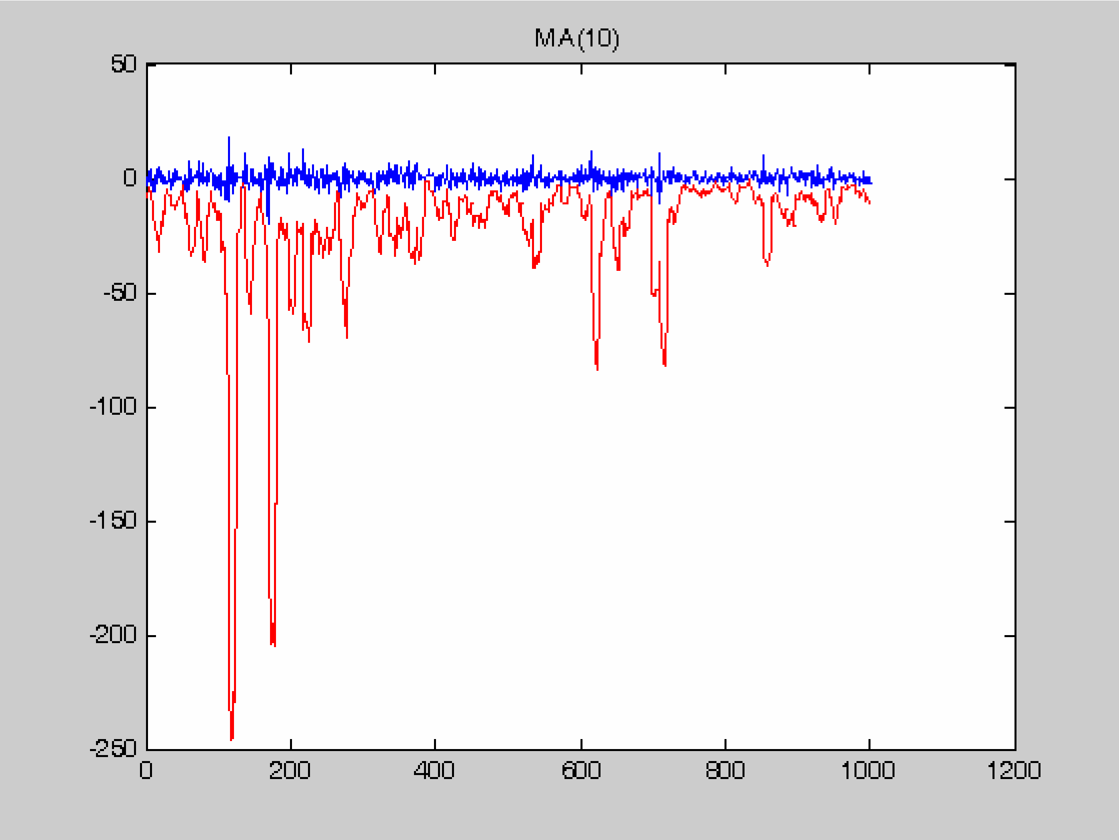

RiskMetrics
20
0
-20
-40
-60
-80
-100
-120
-140
0
200
400
600
800
1000
1200

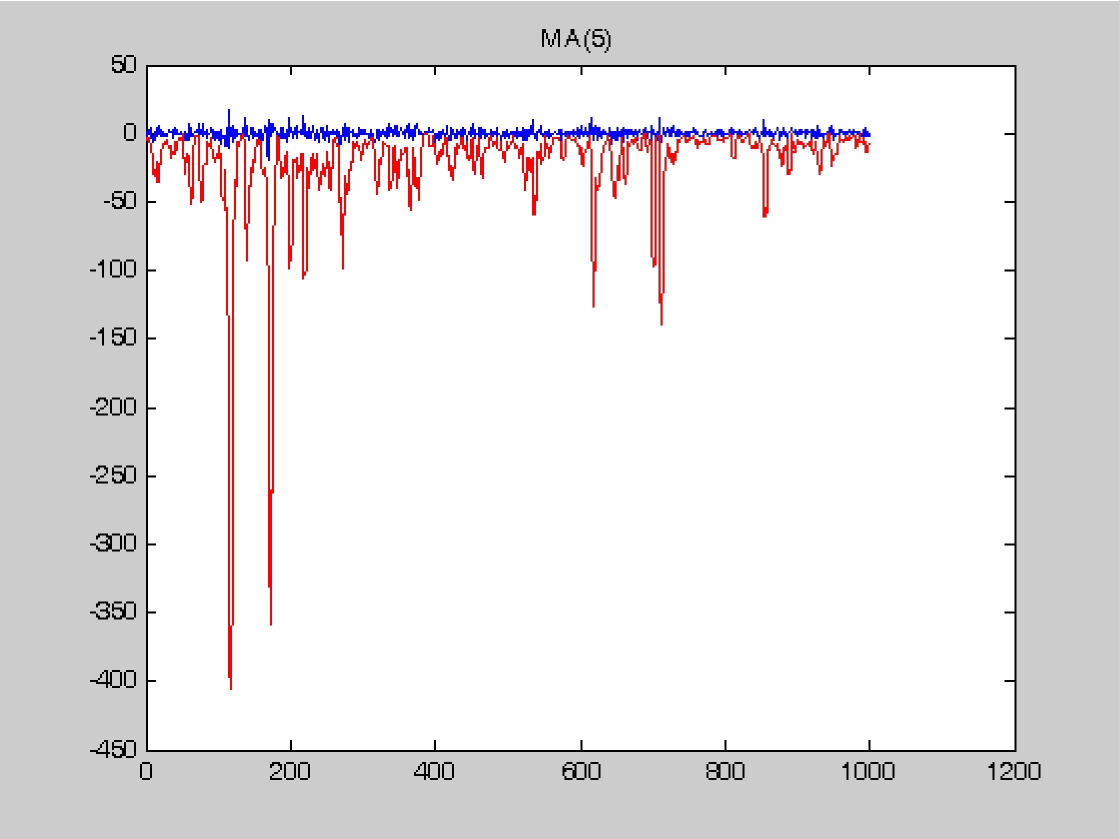

ES
100
0
-100
-200
-300
-400
-500
0
200
400
600
800
1000
1200

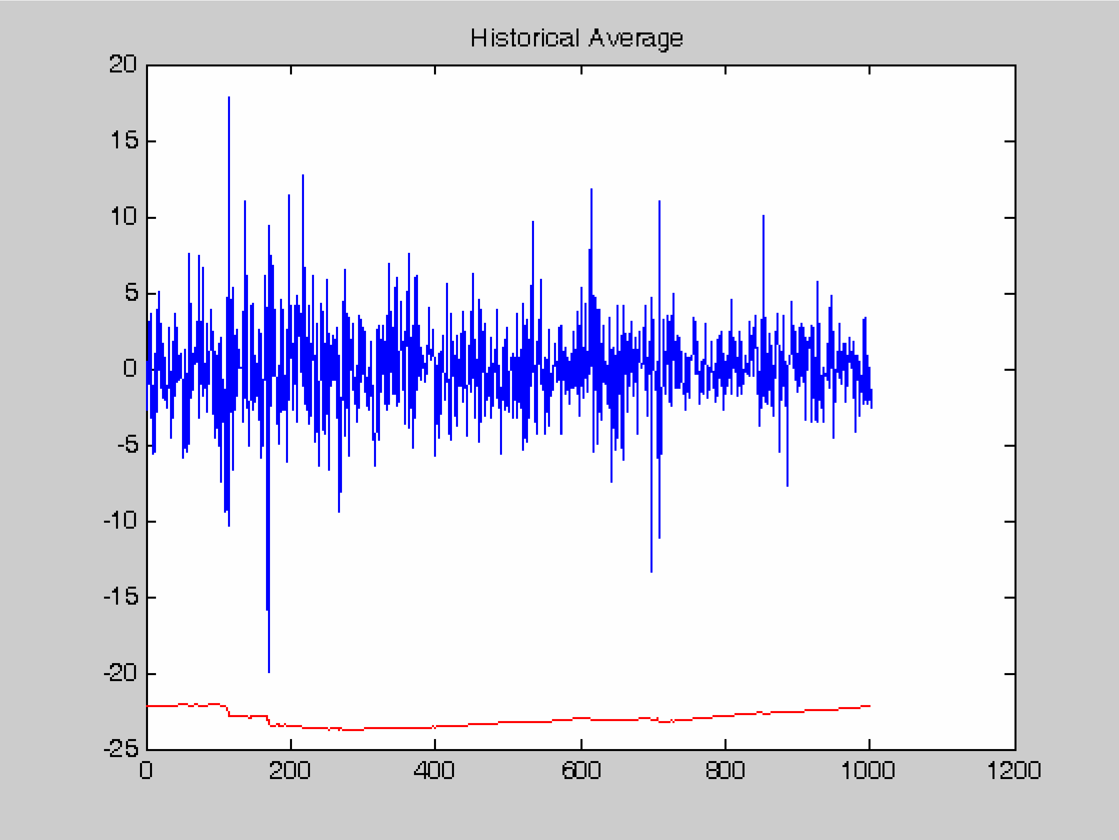

MA(60)
20
0
-20
-40
-60
-80
-100
0
200
400
600
800
1000
1200

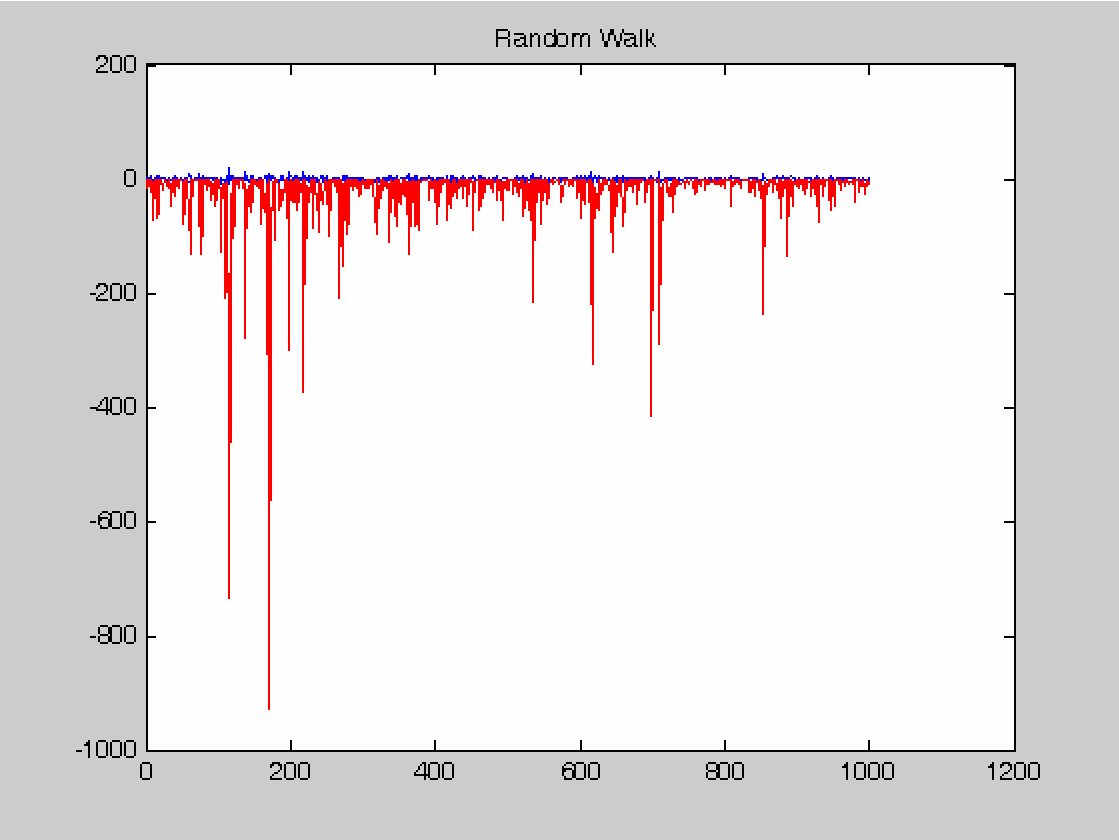

MA(20)
20
0
-20
-40
-60
-80
-100
-120
-140
-160
0
200
400
600
800
1000
1200

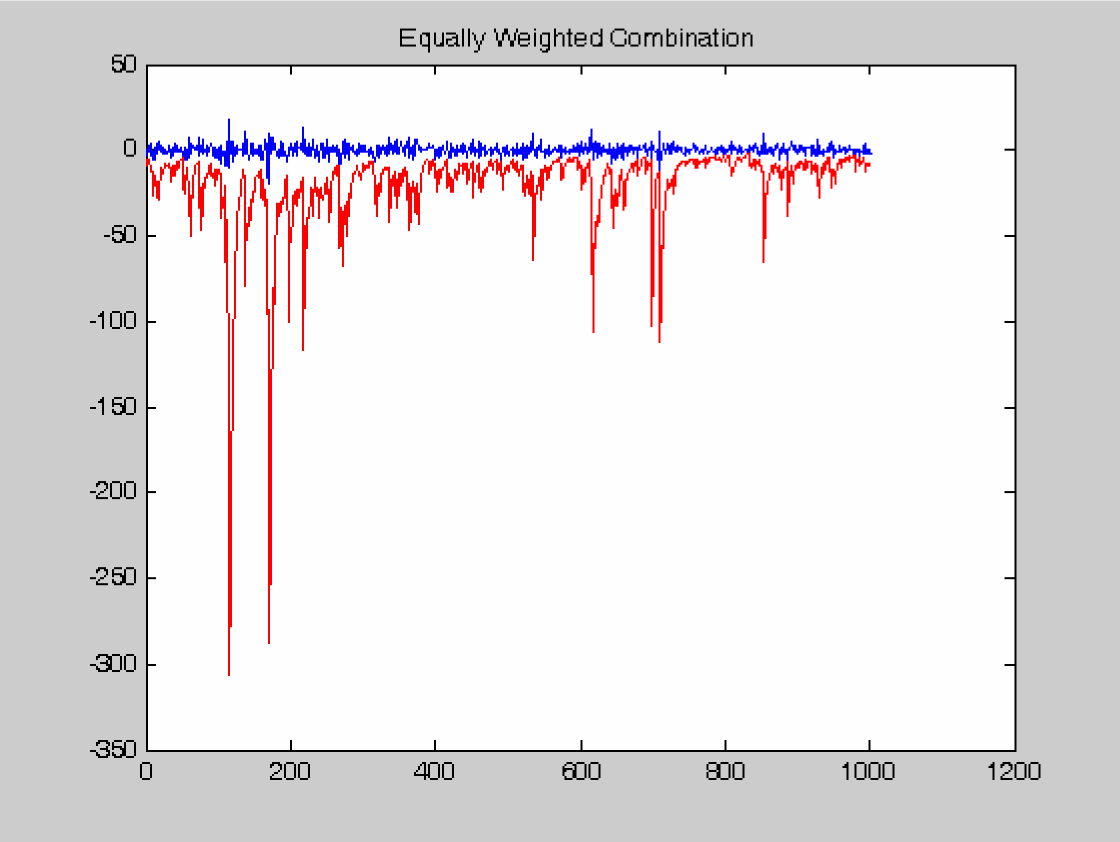

MA(15)
50
0
-50
-100
-150
-200
0
200
400
600
800
1000
1200

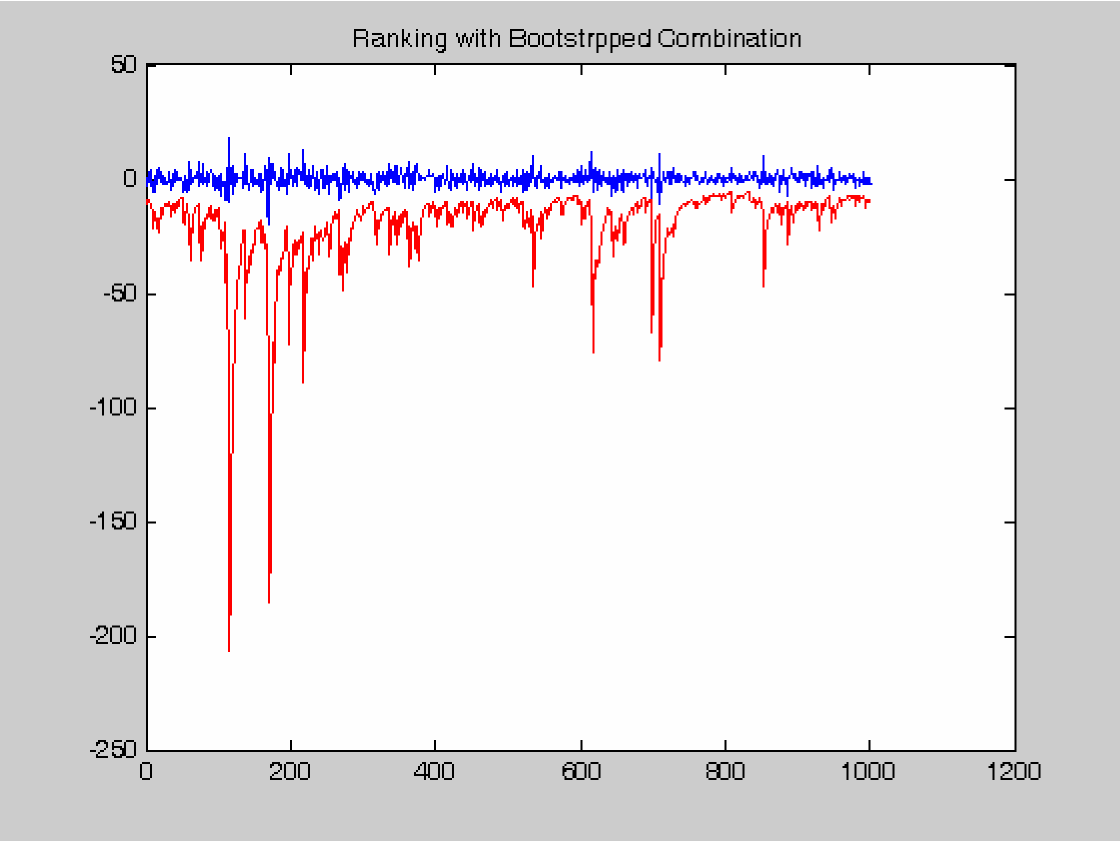

MA(10)
50
0
-50
-100
-150
-200
-250
0
200
400
600
800
1000
1200

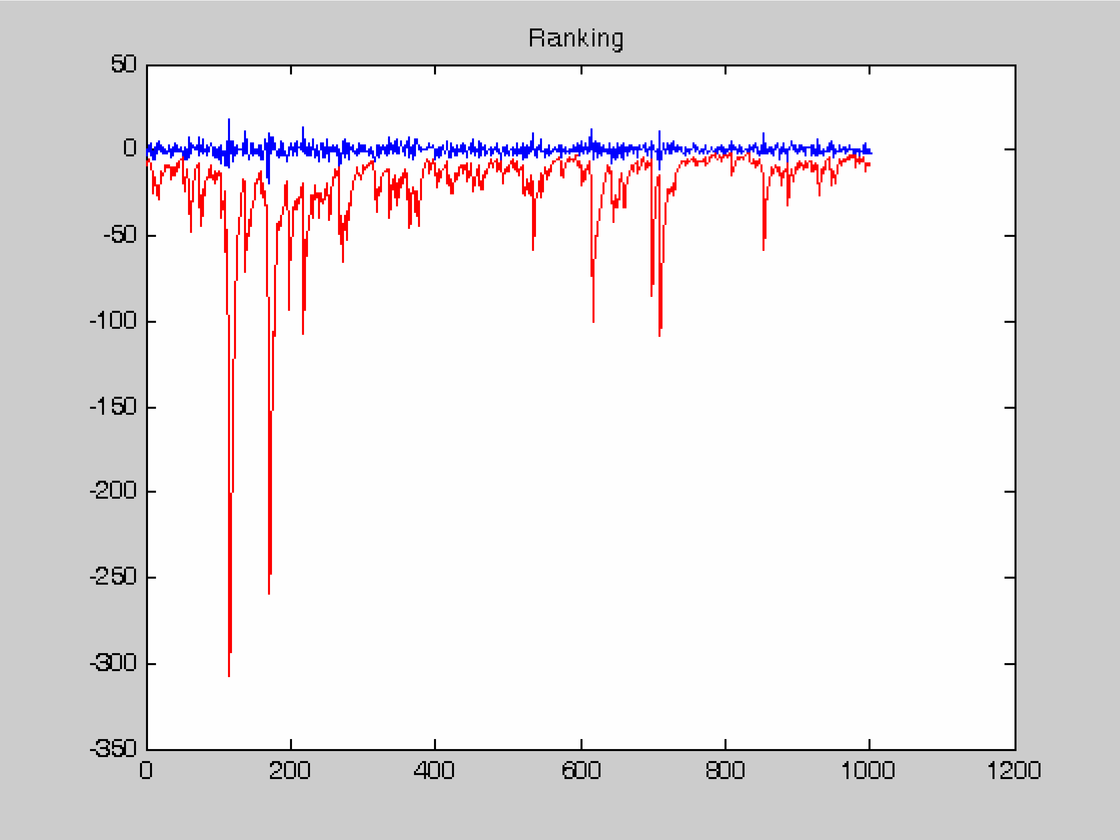

MA(5)
50
0
-50
-100
-150
-200
-250
-300
-350
-400
-450
0
200
400
600
800
1000
1200

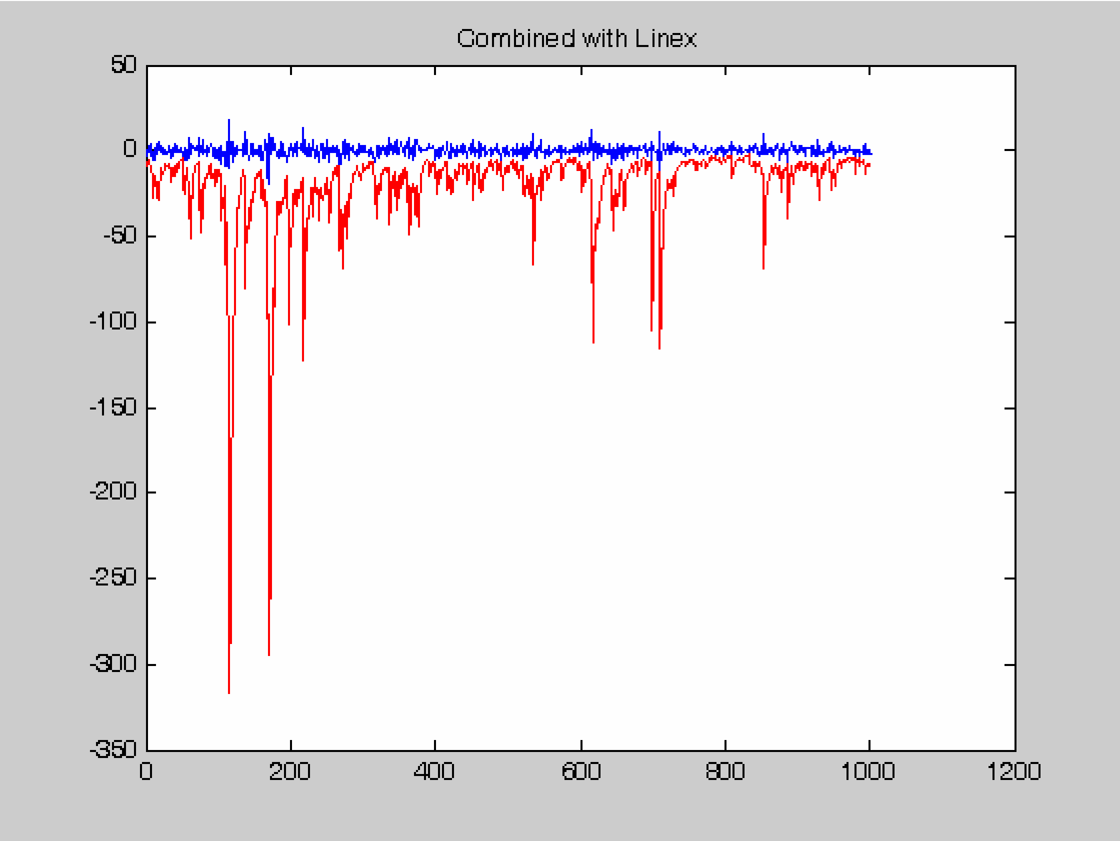

Historical Average
20
15
10
5
0
-5
-10
-15
-20
-25
0
200
400
600
800
1000
1200

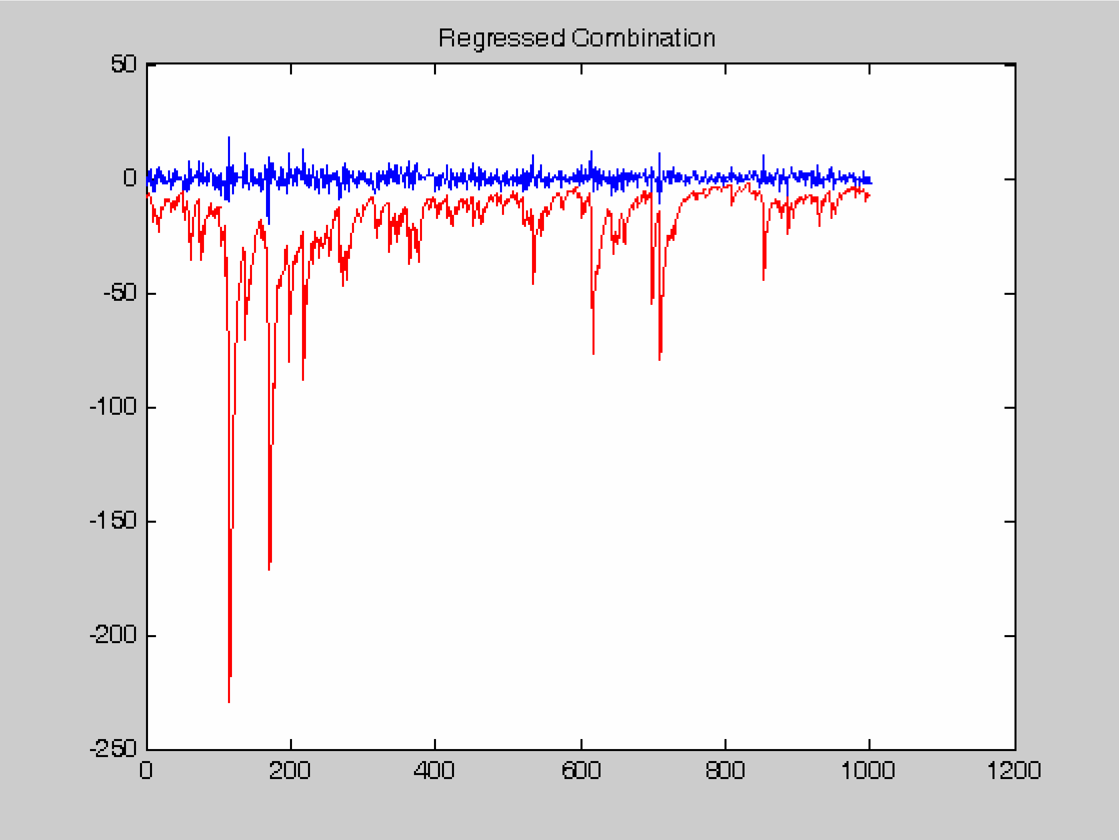

Random Walk
200
0
-200
-400
-600
-800
-1000
0
200
400
600
800
1000
1200

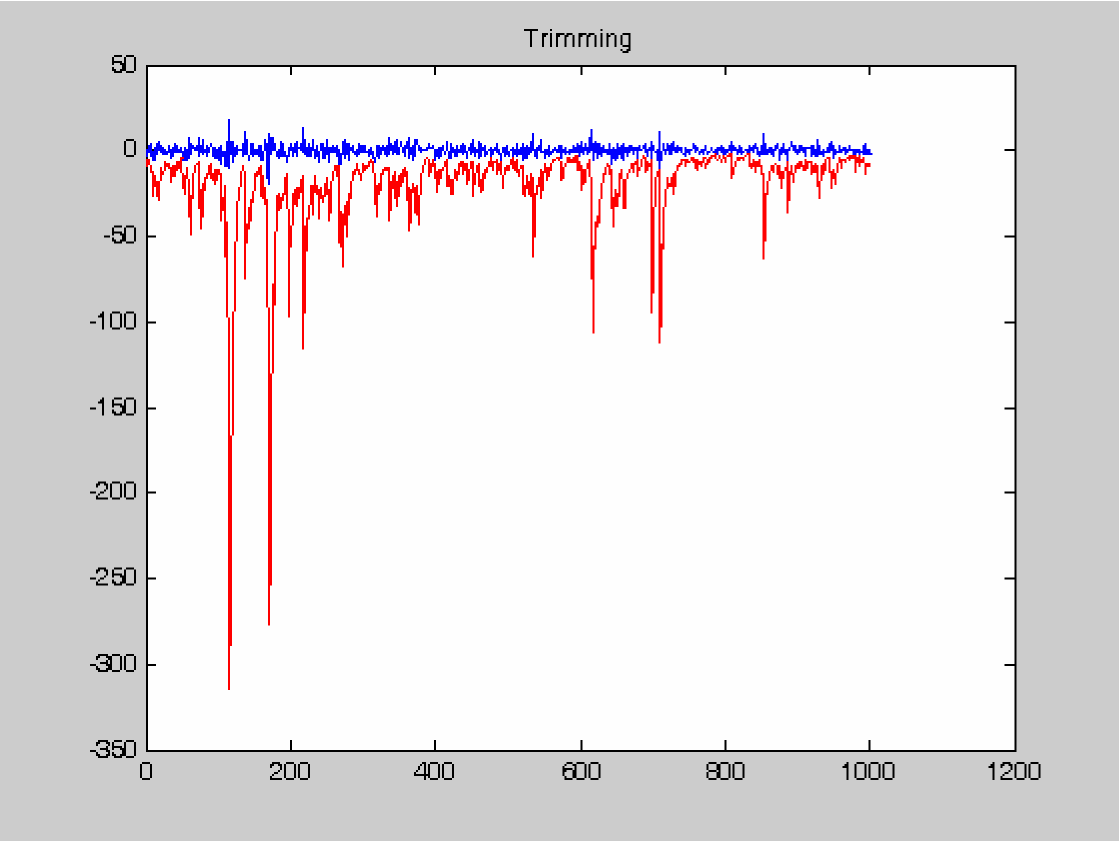

Equally Weighted Combination
50
0
-50
-100
-150
-200
-250
-300
-350
0
200
400
600
800
1000
1200


Ranking with Bootstrpped Combination
50
0
-50
-100
-150
-200
-250
0
200
400
600
800
1000
1200


Ranking
50
0
-50
-100
-150
-200
-250
-300
-350
0
200
400
600
800
1000
1200


Combined with Linex
50
0
-50
-100
-150
-200
-250
-300
-350
0
200
400
600
800
1000
1200


Regressed Combination
50
0
-50
-100
-150
-200
-250
0
200
400
600
800
1000
1200


Trimming
50
0
-50
-100
-150
-200
-250
-300
-350
0
200
400
600
800
1000
1200

Appendix2-Loss Functions

Root Mean Squared Error

Root Mean Squared Error (RMSE) can be defined as follows;

$$l_{RMSE} = \sqrt{\frac{1}{t}\sum_{i=1}^{t}\left(f_i - \sigma_i^2 \right)^2}$$

where t is the sample size, f is the forecast value, and σ^2 is the realized volatility. Although RMSE has a very simple functional form it is one of the common loss functions in the literature. It is symmetric, an overestimation and underestimation affect equally the loss value. However, for our case overestimating and underestimating should not be equally treated. For example, if we consider a financial institution, an overestimation means generally loosing some profit opportunities, but an underestimation means miscalculation of the risks and even default of the institution.

Even tough it has a disadvantage expressed above; it is still a good criterion for testing the accuracy of our forecasts.

Theil-U

Theil-U loss function is a standardized version of the RMSE and can be defined as follows;

$$l_{theil-u} = \frac{\sum_{i=1}^{t-1}\left(f_{i+1} - \sigma_{i+1}^2 \right)^2}{\sum_{i=1}^{t-1}\left(\sigma_i^2 - \sigma_{i+1}^2 \right)^2}$$

where t is the sample size, f is the forecast value, and σ^2 is the realized volatility. As it seen this loss function uses random walk model as benchmark, then the value of random walk is just equal to 1. Because it the worst forecast, other models loss value is between one and zero. The benefit form usage of this loss function is a better comparison, because in standard RMSE, the difference between two loss values is meaningless. On the other hand with this standardized values, the difference between some models’ loss values becomes more meaningful.

Linex

The functional form of the Linex function is;

$$l_{linex} = \frac{1}{t}\sum_{i=1}^{t} e^{-a\left(f_i-\sigma_i^2\right)} + a\left(f_i-\sigma_i^2\right)-1$$

where t is the sample size, f is the forecast value, σ^2 is the realized volatility and a is the parameter which determines the degree of asymmetry in the function. For this loss function the value of “a” is very crucial. If the “a” is positive, then the loss function becomes exponential for underestimations and linear for overestimations. If the “a” is negative, then the loss function becomes exponential of for the overestimations and linear for the underestimations. In this study, we used two values of the “a”; 10 and 20. We did not use negative values of the “a” because it is not reasonable, for our study, to punish more the model that overestimates the volatility than another that underestimates it.