Marmara Üniversitesi
İngilizceİktisatBölümü

Marmara University
Department of Economics

Exchange Rate Direction Forecasting and Usage of ANN

Ekrem KILIÇ

January 2008

Independent Research Paper No: IRP200801

Suggested Citation:
Kılıç, E., 2008, "Exchange Rate Direction Forecasting and Usage of ANN" Marmara University, Department of Economics, Göztepe Campus, Istanbul, Independent Research Paper No: IRP200801

Exchange Rate Direction Forecasting and Usage of ANN

Ekrem KILIÇ*

January 2008

Working Paper No: 200801

ABSTRACT

Forecasting exchange rates is one of the biggest deals that econometricians have been challenged for years. Thanks to Engle's Nobel winning work, we have managed to develop a way for forecasting the volatility of exchange rate returns. In addition to this, increasing availability of high frequency data emerged a great literature on realized volatility. Recent theoretical studies showed a direct relation between return volatility predictability and return sign predictability. This implies that return signs could be predicted. On the other hand, advanced econometric models such as artificial neural networks (ANN) provides highly flexible functional forms to cover nonlinear relations in the data. In this study, we forecasted daily return sign of USD/TRL exchange rate and showed that ANN is an appropriate tool for sign forecasting.

JEL Code: C45, C53, F31

Keywords: exchange rates, market timing, neural networks, volatility.

*) Marmara University, Department of Economics.

1. INTRODUCTION

Forecasting exchange rate returns has attracted the attentions of the researchers for many years. The efficient market hypothesis, however, proposes that all available information in the market is immediately constituted, hence no one can earn profits based on the publicly available data (Fama, Efficient Capital Markets: A Review of Theory and Empirical Work, 1970). Direct implication of this hypothesis is that it is impossible to forecast future returns, because the returns immediately reflect all information currently known. There are some empirical studies with contradictory evidence which claims markets are not fully efficient at least in very short horizons, therefore return forecasting is possible(Jegadeesh, 1990). Some other studies provided some evidence for conditional mean dependence of returns at long horizons ((Fama & French, 1988) and (Jegadeesh, 1990), and(Mark, 1995)). This implies long horizon returns can be forecasted. Although, there is a continuing debate on return forecasting and efficiency of markets, since this paper focuses on sign or direction forecast of the returns, we do not need level of the market returns to be forecastable.

Direction forecasting, on the other hand, is also as interesting as the level forecast is, because an accurate forecast of the direction basically means making money. A model which is able to forecast the sign of the tomorrow's return on some stock or exchange rate is good enough for market practitioners. Therefore, the sign forecasting may possibly be considered as a source of automated trading algorithms. Automated trading, or algorithmic trading, is roughly the use of computer programs for entering instant orders to the market which are based on a predefined algorithm. Today, there are many hedge funds, called quant funds, those are totally relies on this kind of algorithms.

In the literature, there are many studies on forecastability on sign directions as well. Breen, Glosten, & Jagannathan (1989) used the negative correlation between short-term interest rates and nominal excess returns on stocks and developed a model for sign forecasting. Pesaran & Timmerman(1995) examines mainly predictability of US stock returns. Their mainfinding is predictive power of various economic factors change over time,especially in accordance with volatility changes. They also consider signforecasting and its robustness. Gencay (1998a,1998b) explored technicaltrading rules and their predictive power. He used single layer feedforwardnetworks to forecast sign of Dow Jones Industrial Index and analyzed the profit for comparing models. Gencay (1999) investigated the predictive powerof simple technical trading rules on foreign exchange rate. Christoffersen& Diebold (2006) examined the sign dependence in detail and suggested theusage of a volatility link. They showed that volatility dependence producessign dependence, as long as expected returns are nonzero, one should expectsign dependence, given the overwhelming evidence of volatility dependence.They found that due to its special nonlinear nature, sign dependence is notlikely to be identified via analysis of sign autocorrelations, runs tests,or traditional market timing tests.

The artificial neural networks are also widely used for forecasting foreignexchange rates and their returns. Kuan& Liu (1995) investigated theout-of-sample forecasting ability of feedforward and recurrent neuralnetworks on foreign exchange rates. They were able to create some networkmodels which have significant market timing ability. Zhang & Hu (1998)examined the effects of the number of input and hidden layers, and at thesame time the size of the training sample on the in-sample and out-of-sampleperformance. Using empirical analysis, they found that neural networksoutperform linear models, particularly when the forecast horizon is short,and the number of input layers has a greater impact on performance than thenumber of hidden layers.

In this paper, we compared the forecasting strength of the models that uses the relation proposed by Christoffersen& Diebold (2006) versus ANN models. we considered both parametric and nonparametric distributions to extend the setup of Christoffersen& Diebold (2006). In addition, we used different volatility models to examine the effect of the choice of volatility model. In the nonparametric connection part, we considered both Kernel density estimator and a time-dependent density estimator. Using these nonparametric models, we investigated whether distribution assumption increase or decrease the performance. Finally as an alternative method, we employed ANN for sign forecasting.

In the following section, we will summarize the ideas of sign dependence and explain how we can extend it by using nonparametric methods. After that we will describe the forecasting models and parameter assumptions. In the empirical analysis part, first we will give the general information of the data, secondly, we will discuss the model evaluation methods and thirdly report the results. Finally we will conclude.

2. On Sign Dependence

Christoffersen& Diebold (2006) showed that sign of return can be forecastable via volatility dependence. The main idea behind this approach is forecasting the conditional probability density function of tomorrow's return and calculating the probability of having a positive return.

To show this connection, let us start with some definitions. Let $\{r_i\}_{i=0}^{t}$be the series of returns on exchange rate and $\Im_t$be the publicly available information set at time t. Then, the volatilityforecast for tomorrow or conditional variance is $\sigma_{t+1|t}^2 = Var(r_{t+1}|\Im_t)$. The mean dependence can alsobe defined in a same manner; $\mu_{t+1|t} = E(r_{t+1}|\Im_t)$.

Further if we assume that $r_{t+1}|\Im_t \sim F(\mu_{t+1|t}, \sigma_{t+1|t}^2)$ where $F(.)$ is any distribution that is dependent only on its mean and variance. However it is mentioned before, we do not need a mean dependence assumption. Therefore if there is no mean dependence the conditional distribution turns $r_{t+1}|\Im_t \sim F(\mu, \sigma_{t+1|t}^2)$. At this point, we have a conditional distribution which is based on the conditional variance. Then we can find the probability of having a positive return as;

$$P(r_{t+1} > 0|\Im_t) = 1 - P(r_{t+1} \leq 0|\Im_t) \quad (1)$$

$$= 1 - P(\frac{r_{t+1} - \mu}{\sigma_{t+1|t}^2} \leq -\frac{\mu}{\sigma_{t+1|t}^2}) \quad (2)$$

$$= 1 - \tilde{F}(-\frac{\mu}{\sigma_{t+1|t}^2}) \quad (3)$$

$\tilde{F}(.)$is the distribution of standardized return. If there is mean dependence,3 becomes $\tilde{F}(-\frac{\mu_{t+1|t}}{\sigma^2_{t+1|t}})$.

The approach that we quickly summarized is based on a parametric distribution. Now, let us extend this model by considering a nonparametric distribution of standardized returns. A simple nonparametric density estimator is histograms. A simple histogram can be formally described as follows;

$$\tilde{f}(x) = \frac{1}{nh}\sum_{i=1}^{N} 1(x_i \in [x_0, x_0 + h]) \tag{4}$$

where$x \in [x_0, x_0 + h]$,$h$is bandwidth,$1(.)$is the indicator function, and $n$is the number of bins. After defining this histogram, we can calculate the cumulative distribution function by integrating the density function. If we use histograms for standardized returns, the alternative of 3 is simply the following;

$$\tilde{\Phi}(x) = \sum_{j=1}^{n} \tilde{f}(x_j > x) \tag{5}$$

$$\tilde{\Phi}(x > 0) = 1 - \sum_{j=1}^{n} \tilde{f}(x_j \leq \frac{\mu}{\sigma^2_{t+1|t}}) \tag{6}$$

A bit more sophisticated way of estimating nonparametric density functions is kernel density estimation. The histogram itself is also kernel density estimator. However a more convenient way is to employ Gaussian (normal) kernels. The Gaussian kernel density estimator can be formalized as;

$$\tilde{f}(x) = \frac{1}{nh}\sum_{i=1}^{N} \phi\left(\frac{x_0 - x_i}{h}\right) \tag{7}$$

where$\phi(.)$ is standard normal distribution. Again by numerical integrating 7, we may have a cumulative distribution function.

Hence an alternative to 3 is the usage of non-parametric density estimators such as Gaussian kernel density estimator. One can easily construct a density estimator of standardized returns with this kernel and may substitute 3.

The main distinction between histogram and Gaussian kernel is their weightings for the data. From this perspective, although Gaussian kernel density estimators are good alternatives for parametric models, they do not deal with the time dependence in the data. In other words, the weighting of each observation in a Gaussian kernel is independent of the occurrence time of concerning data. Therefore a time dependent weighting may be generate better results. A weighting scheme which assigns highest weight to the most recent data and the lowest weight to the oldest data, is a basic but good way to cover this time dependency. However since we are assuming a parametric structure for the weighting scheme, the resulting model should be classified as a semi-parametric method. A possible weighting scheme is the following;

$$w_{t-i} = \lambda^i \left(\frac{1-\lambda}{1-\lambda^\tau}\right), \qquad 0 < \lambda < 1 \tag{8}$$

whereτ is the time number of observations and λ is decaying factor.

3. Models for Sign Forecasting

3.1 Distribution Connection

As we mentioned in the previous section, Christoffersen& Diebold (2006) proposed an approach that is basically uses the volatility forecasting for sign forecast and they called this method as "volatility connection". However, their explanation and description of the methodology is applicable if we model the distribution without using volatility. Hence, we prefer to call the approach as "distribution connection". Regarding to modeling approach for the return distribution, the models we will use in this context can be classified broadly into two classes. First approach uses parametric distribution and assumes that the market returns follow a specific distribution, commonly normal and t distributions are employed. Second approach, on the other hand, assumes no-distribution for the returns instead it models the empirical cumulative density of the data by using Kernels or time dependent weighting schemes.

In this paper, we will use both of these approaches. The following sections describe different distribution connection implementations that we used in this study.

3.1.1 Distribution Connection with Parametric Distributions

The models that we use in this part are based on a distribution assumption. In this paper, we will consider two distributions; normal and t. Both distribution assumptions require volatility estimation, for this purpose; GARCH(1,1) model , Riskmetrics model and realized volatility are employed.

Now let us start with volatility models. There is a huge literature on volatility modeling and a variety of models are proposed. Especially the models of conditional heteroscedasticity are proven as robust and quite popular. Although it had many extensions, Bollerslev (1986) provides a common form of conditional heteroscedasticity models. This form can be shown as follows;

$$\sigma_t^2 = \omega + \sum_{i=0}^{q} \alpha_i r_{t-i} + \sum_{j=0}^{p} \beta_j \sigma_{t-j}^2 \tag{9}$$

wherer_{t-i} is the excess return on mean; $r_t = \mu_t + \sigma_t \epsilon_t$. Under unconditionality of the mean, μ_t can be set to a constant such as μ. In this study, however, we used 6-month moving average as conditional mean. GARCH model can be estimated by using maximum likelihood approach, therefore GARCH model itself requires a distribution assumption. Again, we used both normal and t distributions. When normal distribution is assumed for sign direction, it is used in the estimation of the GARCH model as well.

A specific form of GARCH model is called Riskmetrics and especially it is a common tool for risk management practices. In its original form Riskmetrics has only a single parameter and requires no parameter estimation because it is predefined. However, since it is an integrated-GARCH(1,1) model the parameter can be estimated via maximum likelihood methods. The functional form of the Riskmetrics model is as follows;

$$\sigma_t^2 = (1-\lambda) r_{t-1} + \lambda \sigma_t^2 \tag{10}$$

whereλ is decaying factor and it is set to 0.94 as it is proposed by Riskmetrics. Equation 9 will be equal to 10, if the following conditions are satisfied;

$$\alpha + \beta = 1 \tag{11}$$
$$\omega = 0 \tag{12}$$
Since, Riskmetrics constrained the parameter values to satisfy $\alpha + \beta = 1$ condition, this model is an integrated-GARCH model.

Another alternative for volatility forecasting is to use realized volatility. Especially, as the data storage capacity of computers increased and trading orders are handled by computers in all over the world, intradaily data becomes accessible at least in more developed markets. Thus, a literature on this kind of frequent data set rapidly increases. One of the most popular analysis, in this context, is modeling realized volatility. Realized volatility is derived from the notion of integrated volatility which can be defined as follows;

$$IV_t^d = \sqrt{\int_{t-d}^{t} \sigma^2(\epsilon) d\epsilon} \tag{13}$$

whereϵ is the instant changes in the data. Integrated volatility provides an approach for calculating the real volatility occurred. Realized volatility is a discrete approximation of integrated volatility and it is defined as follows;

$$RV_t^\Delta = \sqrt{\sum_{j=1}^{\Delta} r_{t-j\Delta}^2} \tag{14}$$

whereΔ is the number of intradaily observations. As the number of intradaily observations, Δ, goes to infinity, realized volatility, RV_t^Δ , tends to integrated volatility, IV_t^d.

After describing the volatility models, now let us start to investigate how the "volatility connection" is implemented in this paper. In volatility connection models, I followed the similar approach to Christoffersen&Diebold(2006) by using different volatility models and assuming different return distribution. Christoffersen& Diebold (2006) describes conditional probability of a positive return as;

$$P(r_{t+1} > 0|\Im_t) = 1 - P\left(\frac{r_{t+1} - \mu}{\sigma_{t+1}} < \frac{\mu}{\sigma_{t+1}} |\Im_t\right) = \Phi\left(\frac{\mu}{\sigma_{t+1}}\right) \tag{15}$$

As it mentioned before, apart from them, we used 6-month moving average returns as conditional mean.

3.1.2 Distribution Connection with Non-parametric Distributions

An alternative to parametric distribution connection for sign forecast is to use non-parametric distribution forecasting. In this study, a Gaussian Kernel density estimator and a time-dependent weighting in the form of equation 9 is applied. For each density forecast last 252 observations are used recursively. Under both models, after constructing the distribution forecast, the conditional probability of a positive return is calculated as;

$$P(r_{t+1} > 0|\Im_t) = 1 - \tilde{F}\left(-\frac{\mu}{\sigma_{t+1|t}}\right) = \tilde{F}\left(\frac{\mu}{\sigma_{t+1|t}}\right) \tag{16}$$

3.2 Artificial Neural Networks

In this section artificial neural network (ANN) model for sign forecasting will be examined, however before that let us take look at the basic concepts in neural networks. ANNs are first developed to mimic neurons in the human brain. Therefore ANNs are information processing structures, containing some interconnected elements. These elements are called neurons. An

individual neuron receives input from direct data sources or other neurons and produces an output.
The input of a neuron is determined by the synapse. A full synapse, the most used type of synapses, provides a linear combination of inputs; the combination is defined by the weighting matrix of the synapse. The input for a single neuron is than;

$$\eta_j = \sum_{i=1}^{n} \omega_{ji} x_i + b_j = \omega_j \boldsymbol{x} + \boldsymbol{b_j} \qquad (17)$$

whereω_j is the *jth* vector of weighting matrix and b_j is the bias. Figure 1shows a single neuron.
The output of the neuron, on the other hand, is determined by the activation function. The computed values of inputs are transformed into the output value by employing an activation function. There are several alternatives for activation functions, but in general, functions those are able to map input values to the interval of [-1,1] or [0,1] are used. In this study we used logistic sigmoidal function that can be formalized as;

$$f(\eta_j) = \frac{1}{1 + \exp(-\eta_j)} \qquad (18)$$

whereη_j is the input. Logistic sigmoidal function maps inputs to [0,1] interval. Since our aim is to model a probability, this feature avoids the unpleasant outputs like negative probabilities.

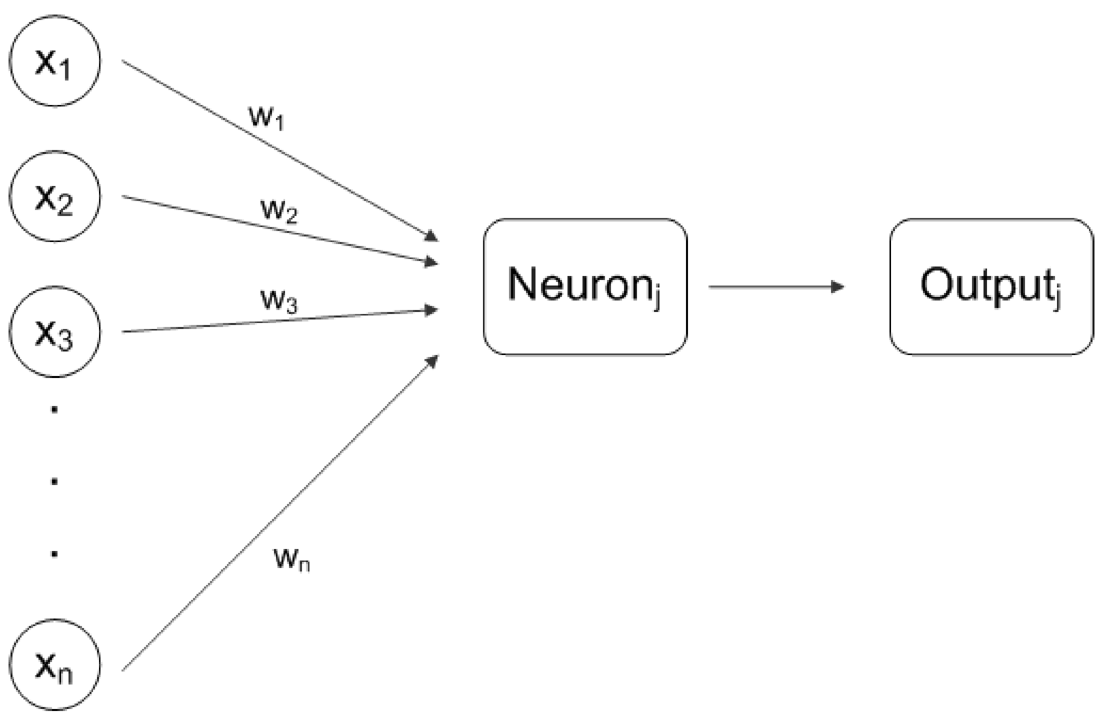

Figure 1 : Input / Output relations of a single neuron

Another important concept is layers. A layer is a group of neurons which are not connected with each other. Neurons within the same layer receive inputs from the same group of neurons and provide outputs for the same group of neurons. The first layer of the network is often called as input layer, last one as output layer, and the layers between them as hidden layer(s). Figure 2 shows a simple network with two hidden layers.
Parameter fitting procedure of synapses is called learning algorithm. Among several methods I preferred to usedbackpropagation algorithm. Backpropagation algorithm is a supervised learning algorithm in which the parametersof the network is updated iteratively to fit a set of

given outputs via a setof input vectors (training data). Backpropagation algorithm requires a costfunction to minimize; root mean squared error (RMSE) is a common andmost of the time appropriate choice.
As an example the output of a network which has one hidden layer is;

$$\boldsymbol{h} = \eta\left(\omega^{\boldsymbol{h}}\boldsymbol{x}\right) + \boldsymbol{b}^{\boldsymbol{h}} \tag{19}$$

$$\boldsymbol{y} = \theta(\omega^{\boldsymbol{o}}\boldsymbol{x}) + \boldsymbol{b}^{\boldsymbol{o}} \tag{20}$$

where η and θ are activation functions of hidden and output layers respectively, **x** is input vector, **h** is output of hidden layer, ω is weight matrices, and **b** bias vectors.

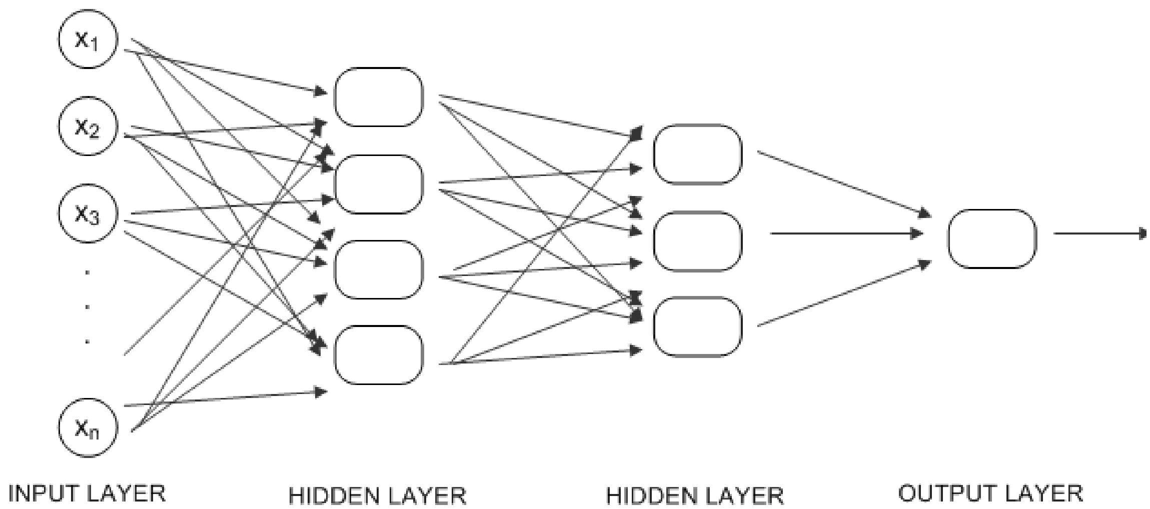

Figure 2 : A simple neural network

RMSE,then, can be written as;

$$RMSE = \sqrt{(\boldsymbol{y} - \widetilde{\boldsymbol{y}})^T(\boldsymbol{y} - \widetilde{\boldsymbol{y}})} \tag{21}$$

where$\widetilde{\boldsymbol{y}}$ is the vector of desired output. Using numerical optimization techniques, the parameter matrices of synapses can be found by minimizing RMSE.
Another important issue is network design. Unfortunately, there is no common rule for designing a neural network. However, earlier literature provides some suggestions. Kruschke (1989) and Looney(1996) suggested that succeeding layers should be smaller than the preceding ones. In most of the studies a two-layer network is used, however, the determination of the appropriate layer size requires many trial and error methods. Another alternative is to select network size by using a constructive approach in which the network starts with a minimal size and grows gradually with respect to RMSE performance.
In this study, I used a 4-layer network with 2 hidden layers. The input layer is consisted of 11 inputs. All synapses are full synapse in which each neuron is connected with all neurons of neighboring layers. The first hidden layer has 5 neurons and the second has 2 neurons. Finally, the output layer has only one neuron whose output shows the probability of having a positive return. In all layers logistic sigmoidal function is used as activation function.

4. Empirical Analysis

4.1 Data

Distribution connection method uses only exchange rate data itself, however ANN method uses several explanatory inputs. Daily values for the period 2003:01 to 2007:06 of USD/TRL exchange rate is collected from Electronical Data Distribution System (EDDS) of Central Bank of Turkey. The data is based on interbank exchange rate o ers. The original data is intradaily and consists of 6 one-hour data for each trading date starting from 10:30 AM to 3:30 PM. The realized volatility calculations based on this data.

Inputs of ANN model for the same time period are logarithmic return of USD/TRL, logarithmic return of 30-day, 90-day, 180-day, and 360-day treasury rates, logarithmic return of Istanbul Stock Exchange's (ISE) ISE-100 index, realized volatility of USD/TRL returns, Riskmetrics volatility forecast of USD/TRL returns and percentage change in term spread between 360-day and 30-day, 360-day and 90-day, and 360-day and 180-day interest rates. Yield curves for the treasury rates are constructed by daily price data of treasury bills and bonds. The required data for yield curves is collected from the ISE's daily reports. ISE-100 data is again obtained EDDS of Central Bank of Turkey.

All data series are consists of 1086 daily observations, first 724 observations are used for model estimation and ANN training, the last 362 observations are used for out-of-sample forecasting.

For parametric distribution connection models the volatility estimation is central. One-step head volatility forecasts of four volatility models are presented in Figure 3.

4.2 Model Evaluation

I will compare the results of the models by using three different approaches. First of all, I will apply a likelihood ratio test to examine whether the results are statistically likely to occur under the null hypothesis. The null hypothesis for our case is that the probability of having a positive return is 50%. If a model fails to reject the null, then the results are not better than coin tossing. In other words, a failure in the rejection means that tossing a coin may possible generate similar performance. The likelihood test is constructed as follows. Let the resulting sequence be defined;

$$\epsilon_t = \begin{cases} 1, & if\,(\aleph_t = f_t) \\ 0, & otherwise \end{cases} \tag{22}$$

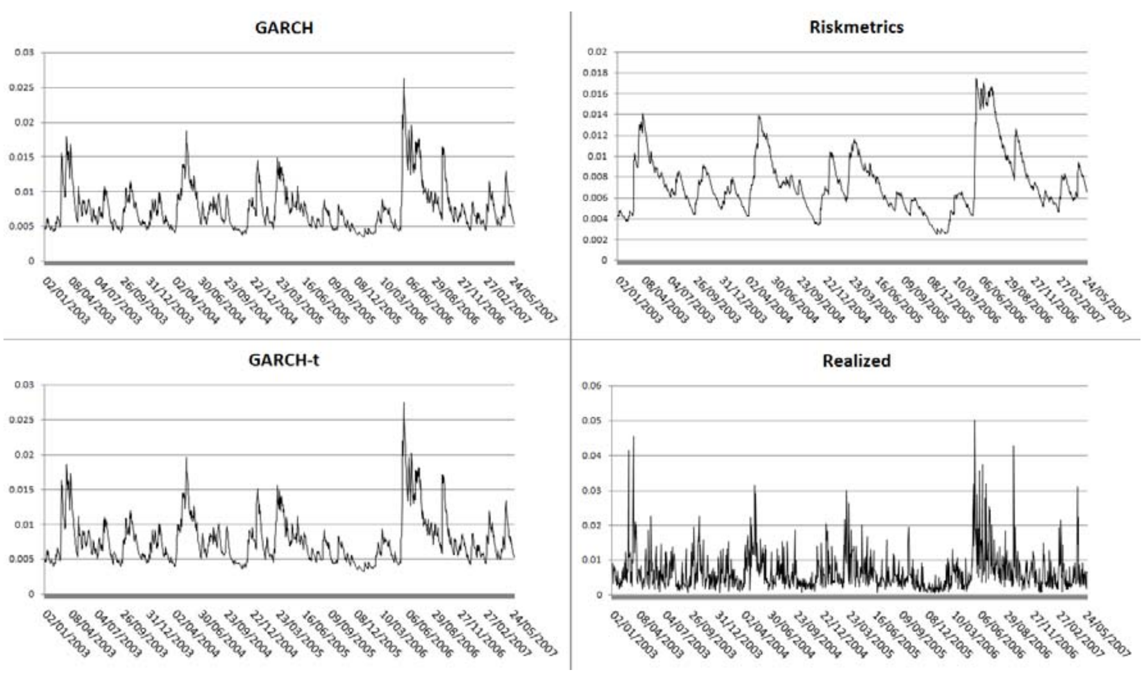

Figure 3 : Volatility graphs

where $\aleph_t$ is an indicator which is equal to 1 if the excess return at time t is bigger than zero, otherwise it is zero. f_t is another indicator that represents the forecast of the model in the same manner. Then, *ith* element of ϵ_t sequence will be 1, if *ith* element of $\aleph_t$ and f_t are equal, in other words, when the forecast is correct, and it will be zero otherwise.

The sequence, ϵ_t, can be modeled as independent draws from a Bernoulli distribution with 50% probability of having a positive return. Christoffersen(1998) suggest a likelihood ratio test for a similar problem. Under the null hypothesis that;

$$H_0: \hat{\alpha} = \alpha \tag{23}$$

where$\hat{\alpha}$ ML estimate of α. Likelihood of an i.i.d. Bernoulli distributedsequence can be written as;

$$L(\alpha) = \prod_{t=1}^{T} (1-\alpha)^{1-\epsilon_t} \alpha^{\epsilon_t} = (1-\alpha)^{T_0} \alpha^{T_1} \tag{24}$$

whereT_0 is the number of incorrect forecasts and T_1 is the number of correct forecasts. ML estimate of is;

$$\hat{\alpha} = \frac{T_1}{(T_0 + T_1)} \tag{25}$$

Now, we can easily find the likelihood of the sample by plugging the MLestimate into equation 23;

$$L(\hat{\alpha}) = \left(1 - \frac{T_1}{T}\right)^{T_0} \left(\frac{T_1}{T}\right)^{T_1} \tag{26}$$

Then, likelihood ratio test is;

$$LR = 2\big(l(\hat{\alpha}) - l(\alpha)\big) \backsim \chi_1^2$$

where$l(.)$is the log-likelihood function which defined asln (L).

Secondly, we will assess the performance by using Brier scores of forecasts. The Brier score can be defined as;

$$Brier = \frac{1}{T} \sum_{t=0}^{T} 2\big(\hat{P}(r_{t+1} > 0) - \aleph_t\big)^2 \tag{27}$$

where$\hat{P}(r_{t+1} > 0)$ is the forecasted probability of having positive return and $\aleph_t$ is again the same indicator that is defined as $\aleph_t = 1(r_{t+1} > 0)$. The best Brier score is zero that occurs when the forecasts consist of 0 or 1 and each time correct. The worst Brier score is 2 and it occurs only if the forecasts consist of 0 or 1 again, but this time each of them is incorrect. If we use the conventional cut-off value of 0.5, the correct forecasts will have a Brier score between 0 and 0.5, whereas incorrect forecasts have scores between 0.5 and 2. Thus, the Brier score of a successful model should be somewhere between 0 and 0.5, while the score of a bad model will be between 0.5 and 2. When comparing different models, the model with lowest Brier score is the best among them.

Finally I will compare out-of-sample trading performances of the models. The basic idea is the following; as long as we know the sign of tomorrow's exchange rate return, we may buy or sell the foreign exchange and gain profit. Then, a good model should generate higher profit than the buy and hold strategy that is suggested by the market efficiency theory. The best model among different models is the one that generate the highest profit. Assuming short selling is possible and avoiding the transaction costs, the profit generated by the model is;

$$\Pi = \sum_{t=0}^{T} \epsilon_t |r_t| \tag{28}$$

whereϵ_t is 1 when the forecast is correct, and otherwise it is zero.

4.3 Results

Figures 4 to 10 shows the probability forecasts of the models. For all models except ANN, the sign forecast changes rarely. Although the volatility choice affect the level of probability, as we passed to sign forecast, the volatility choice becomes unimportant. Therefore all of the volatility connection models -including not only different volatility models but also different distribution assumptions- created the same sign forecast sequence.

Table 1 shows the results of likelihood ratio test. As it is obvious in the table, ANN model clearly outperforms the others. Moreover, all the models except ANN are failed to reject the null hypothesis. The proportion of correct forecast to all forecasts is less than 0.5 for volatility connection models. Nonparametric distribution connection models performed a bit better but their correct forecast ratio is not statistically different from 0.5 either. ANN, on the other hand has forecasted correctly 66.30% of all out-of-sample forecasting period. Although I did not report the in-sample results, ANN has forecasted correctly 77.62% of the data, while correct forecast ratio of other models quite similar to their out-of-sample performances.

Brier scores of the out-of-sample forecasts are presented in table 2. Robustness of ANN model over other techniques is again obvious. All methods except ANN had Brier scores around 0.5, which is nearly equal to our cut-off value and that means all methods but ANN had poor performances. Finally, let us take look at the cumulative profit that models generate. Table 3 shows the cumulative profits of each model. Unsurprisingly ANN model is the best. All other model causes losses those are even worse than the result of buy and hold strategy. The loss of buy and hold strategy is 1.41% for this period. Interestingly nonparametric models caused higher losses than parametric models, although they -even the difference insignificant- outperformed the others in statistical tests.

All the results show that ANN model is a better choice for sign forecasting. The main disadvantage of the models that uses volatility or distribution dependence is their sluggish changes in the direction. This problem may be overwhelmed by applying the model to less frequent data, like monthly or quarterly data. ANN on the other hand, is a powerful tool for the sign forecasting because of its nonlinear nature. Especially, the activation functions like we used provides the required nonlinearity in sign forecasts.

5. Conclusion

In this study, I examined the forecastability of sign-direction. I consider mainly two approaches; first is based on the volatility connection of Christoffersen& Diebold (2006) and second is ANN.

Although volatility connection and the models using the same idea based on a strong theoretical explanation, in practice the forecasts of these models are not able to mimic frequent changes in the direction. They follow very stable direction patterns that are unlikely in daily return series.

ANN models, on the other hand, best fits the problem, they are flexibleand nonlinear models. Thus their forecasts can change direction very frequently comparing to volatility connection based models. Moreover, ANNmodels require neither modeling nor assumptions for the data generating process. The similar result can be obtained with many other network designs.

As for future work, the empirical analysis can be replicated for other markets and longer time-periods, to assure that the results are the same across different markets, different assets and different time-periods. Another direction of further study is to compare sign direction forecast of ANN models against several trading strategies by considering the transaction costs. ANN models may be better alternatives for these kinds of strategies, since they provide more complex structures.

6. References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307-327.

Breen, W., Glosten, L., & Jagannathan, N. (1989). Economic Significance of Predictable Variations in Stock Index Returns. *Journal of Finance, 44*(5), 1177-1189.

C.M., K., & T., L. (1995). Forecasting Exchange Rates Using Feed-forward and Recurrent Neural Networks. *Journal of Applied Econometrics, 10*, 347-364.

Christoffersen, P., & Diebold, F. (2006). Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics. *Management Science, 52*(8), 1273-1287.

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance, 25*(2), 383-417.

Fama, E., & French, K. (1988). Permanent and Temporary Components of Stock Prices. *Journal of Financial Economics, 22*(1), 3-25.

Gencay, R. (1998). Optimization of technical trading strategies and the profitability in security markets. *Economics Letters, 59*(2), 249-254.

Gencay, R. (1998). The predictability of security returns with simple technical trading rules. *Journal of Empirical Finance, 5*(4), 347-359.

Gencay, R. (1999). Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics, 47*(1), 91-107.

Jegadeesh, N. (1990). Evidence of Predictable Behavior of Security Returns. *Journal of Finance, 45*(3), 881-898.

Mark, N. C. (1995). Exchange Rates and Fundamentals: Evidence on. *American Economic Review, 85*(1), 201-218.

Pesaran, M., & Timmerman, A. (1995). Predictability of Stock Returns: Robustness and Economic Significance. *Journal of Finance, 50*(4), 1201-1228.

Zhang, G. P., & Y., H. M. (1998). Neural network forecasting of the British Pound/US Dollar exchange rate. *Omega, 26*, 495-506.

7. Appendix Graphs & Tables

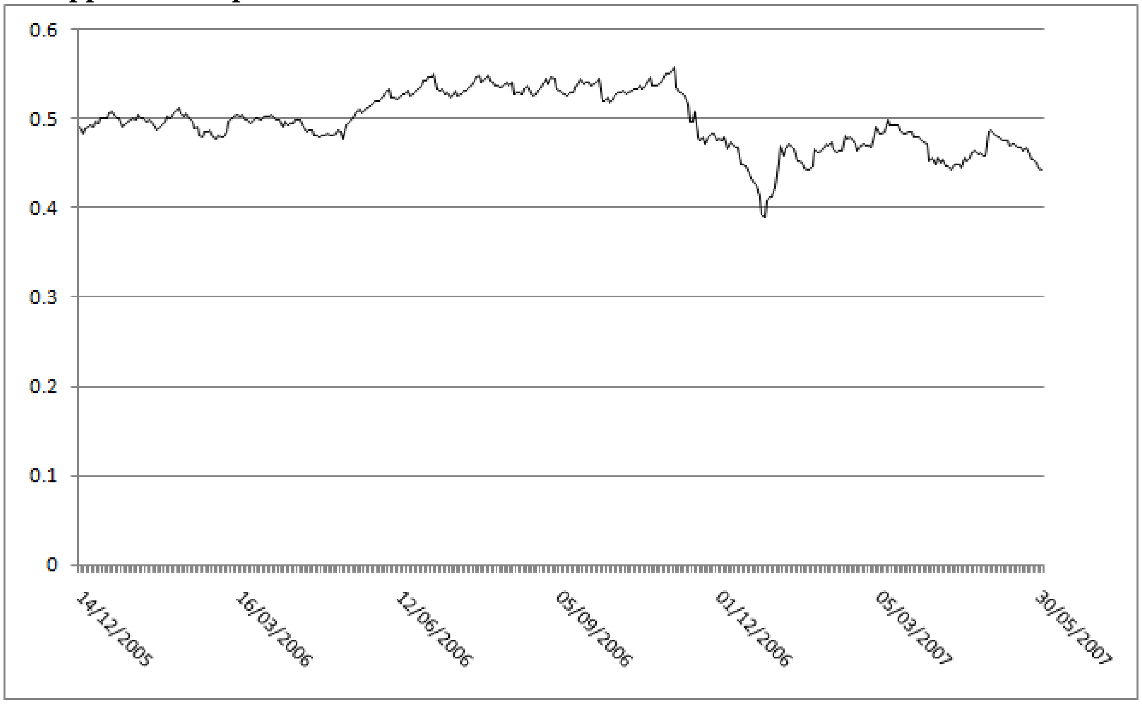

Figure 4 : Probability forecast under volatility connection with GARCH(1,1)and Normal distribution

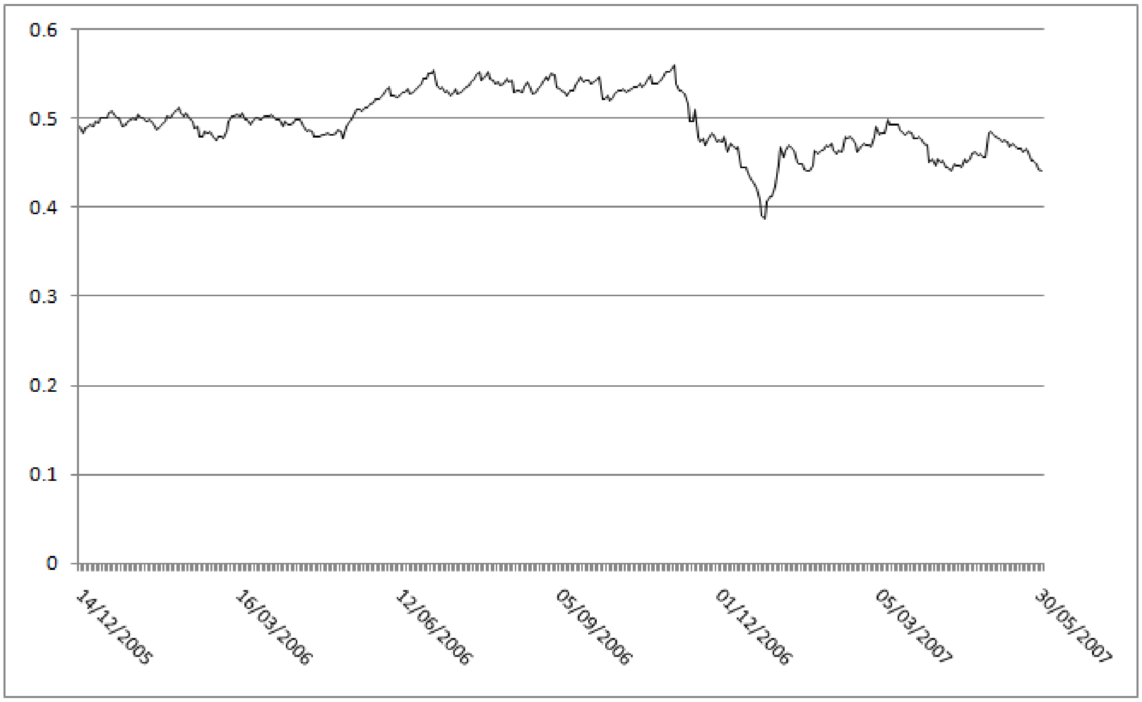

Figure 5 : Probability forecast under volatility connection with GARCH(1,1)and t distribution

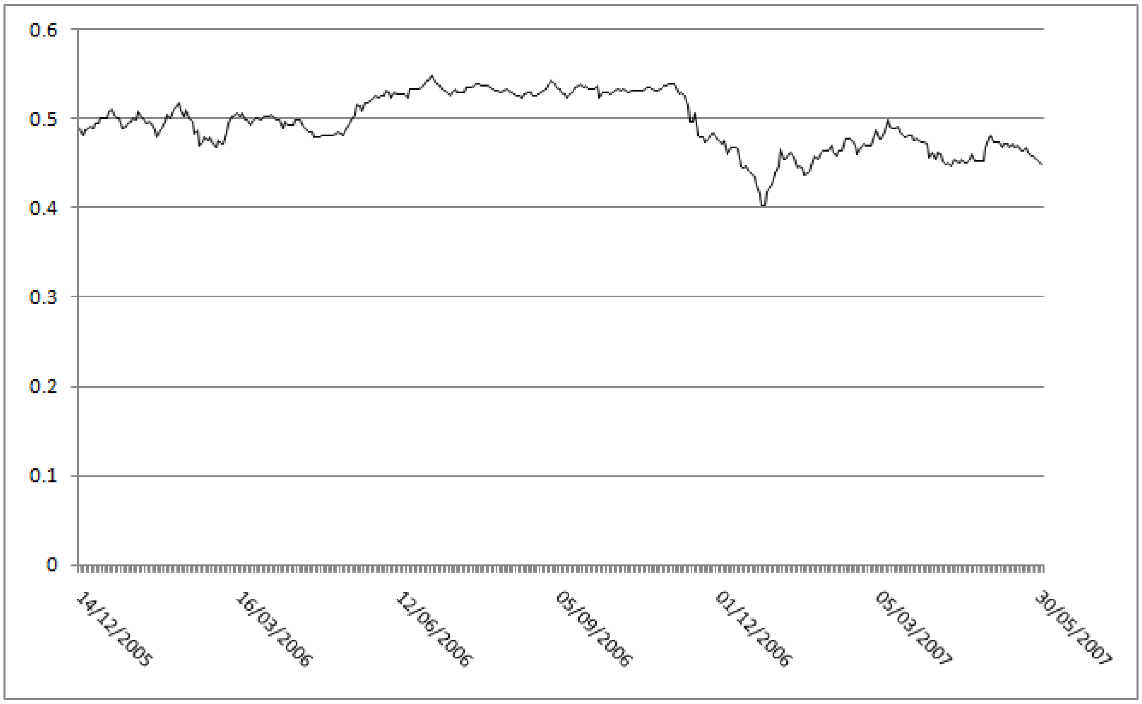

Figure 6 : Probability forecast under volatility connection with Riskmetricsvolatility and Normal distribution

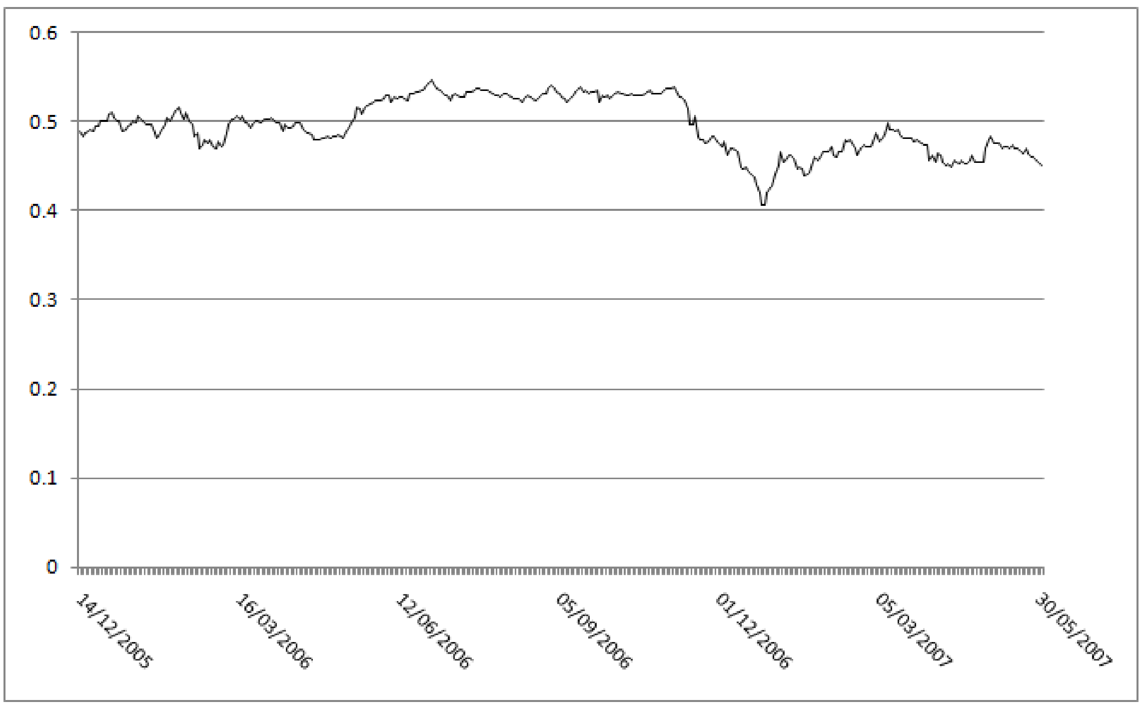

Figure 7 : Probability forecast under volatility connection with Riskmetricsvolatility and t distribution

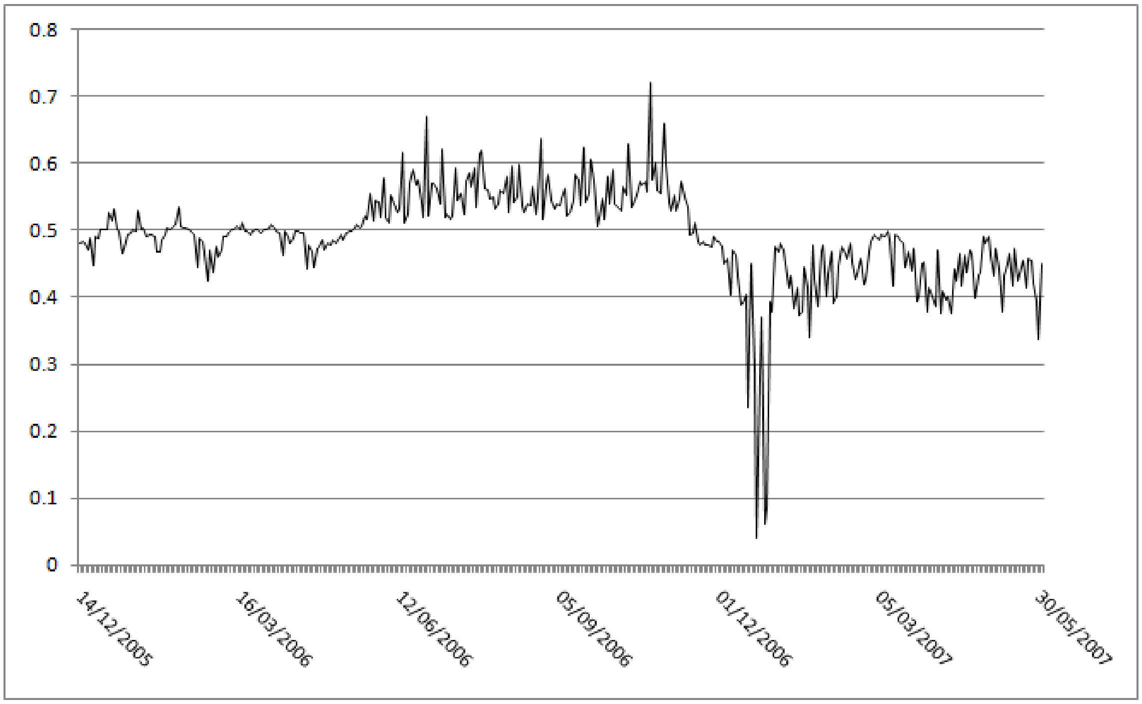

Figure 8Probability forecast under volatility connection with realizedvolatility and Normal distribution

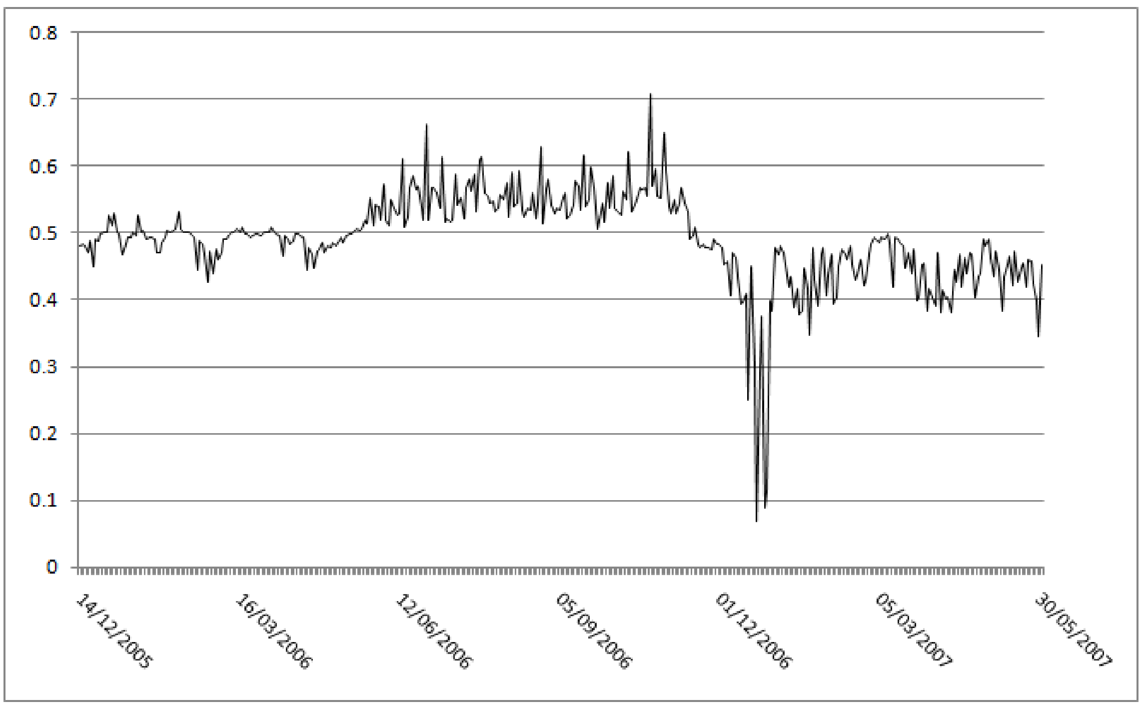

Figure 9 : Probability forecast under volatility connection with realizedvolatility and t distribution

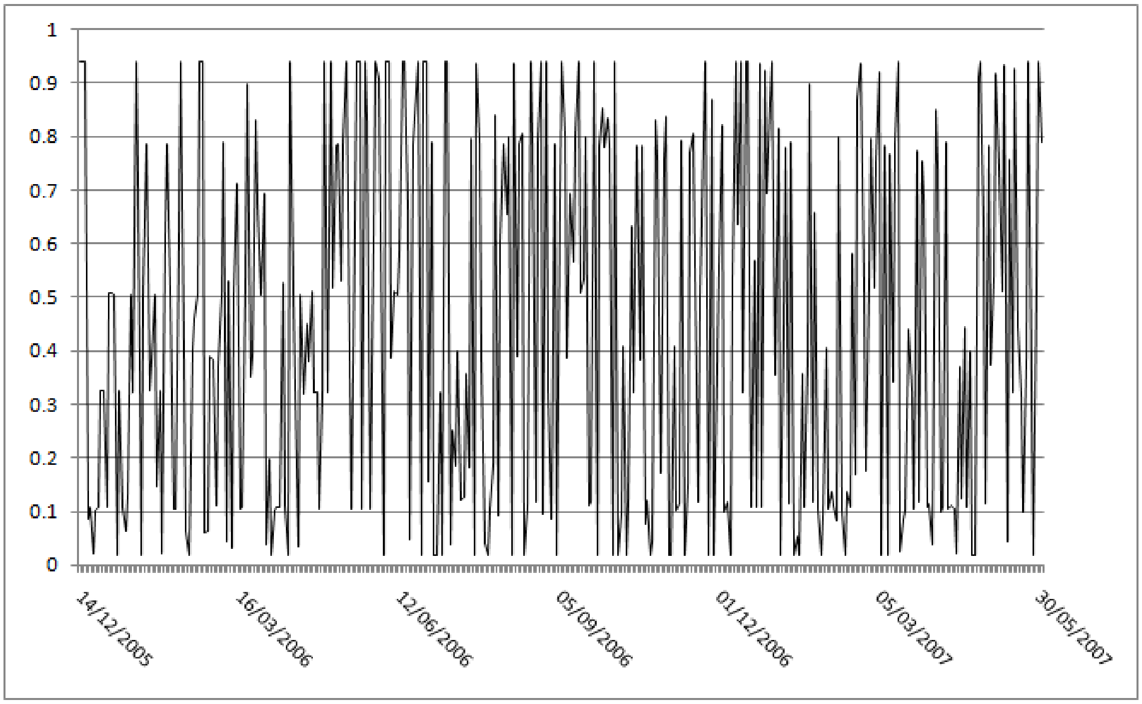

Figure 10 : Probability forecast of ANN

Model	$\widehat{p}$	# of Corrects	LR	p − value
V-Conn-1	48.90%	177	0.1768	67.41%
V-Conn-2	48.90%	177	0.1768	67.41%
V-Conn-3	48.90%	177	0.1768	67.41%
V-Conn-4	48.90%	177	0.1768	67.41%
V-Conn-5	48.90%	177	0.1768	67.41%
V-Conn-6	48.90%	177	0.1768	67.41%
ANN	66.30%	240	39.1759	0.00%
Kernel	53.59%	194	1.8690	17.16%
TW-Nonparam	52.76%	191	1.1055	29.31%

Table 1 : Table 1: Table shows the results of Likelihood ratio test applied of theout-of-sample forecasts. V-Conn-1 is the volatility connection model withGARCH(1,1) and normality assumption, V-Conn-2 is the volatility connection model with GARCH(1,1) and t-distribution assumption, V-Conn-3 isthe volatility connection model with Riskmetrics volatility and normalityassumption, V-Conn-4 is the volatility connection model with Riskmetricsvolatility and t distribution assumption, V-Conn-5 is the volatility connection model with realized volatility and normality assumption, V-Conn-6is the volatility connection model with realized volatility, and t distribution assumption. ANN is the neural network model, Kernel is the Gaussian kernel based non-parametric distribution connection model and finallyTW-Nonparam is non-parametric distribution connection model with time dependent weighting scheme.

Model	Brier Score
V-Conn-1	0.4989
V-Conn-2	0.4990
V-Conn-3	0.4980
V-Conn-4	0.4980
V-Conn-5	0.5047
V-Conn-6	0.5038
ANN	0.1232

Table 2 : Table shows the results of Brier score of the out-of-sample forecasts. V-Conn-1 is the volatility connection model with GARCH(1,1) andnormality assumption, V-Conn-2 is the volatility connection model withGARCH(1,1) and t-distribution assumption, V-Conn-3 is the volatility connection model with Riskmetrics volatility and normality assumption, V-Conn-4 is the volatility connection model with Riskmetrics volatility andt distribution assumption, V-Conn-5 is the volatility connection model withrealized volatility and normality assumption, V-Conn-6 is the volatility connection model with realized volatility, and t distribution assumption. ANNis the neural network model.

Model	Cumulative Profit
V-Conn-1	-3.79%
V-Conn-2	-3.79%
V-Conn-3	-3.79%
V-Conn-4	-3.79%
V-Conn-5	-3.79%
V-Conn-6	-3.79%
ANN	68.12%
Kernel	-8.04%
TW-Nonparam	-7.96%

Table 3 : Table shows the profit provided by the out-of-sample forecasts. V-Conn-1 is the volatility connection model with GARCH(1,1) and normalityassumption, V-Conn-2 is the volatility connection model with GARCH(1,1)and t-distribution assumption, V-Conn-3 is the volatility connection modelwith Riskmetrics volatility and normality assumption, V-Conn-4 is thevolatility connection model with Riskmetrics volatility and t distribution assumption, V-Conn-5 is the volatility connection model with realized volatilityand normality assumption, V-Conn-6 is the volatility connection model withrealized volatility, and t distribution assumption. ANN is the neural network model, Kernel is the Gaussian kernel based non-parametric distributionconnection model and finally TW-Nonparam is nonparametric distributionconnection model with time-dependent weighting scheme.