

Modeling and Forecasting TEDPIX using Intraday Data in the Tehran Securities Exchange

Alireza Heidarzadeh Hanzaee*, Leila Barati

Department of Financial Management, North Tehran Branch, Islamic Azad University, Tehran-Iran

*E-mail: a_heidarzadeh@iau-tnb.ac.ir

Abstract

The aim of the research is conditional and unconditional models performance comparison of volatility forecasting on Tehran dividend and price index (TEDPIX), using the intraday data, based on root mean square error in Tehran Securities Exchange (TSE). In research, it is tried to analyze the total price index behavior by conditional (Arch, Garch, Egarch, Gloston Garch and Rankle) and unconditional (Moving Average) and mixed models to determine the best forecasting model for price and dividend index of active companies in Tehran Securities Exchange. Indeed, the research results will be an analytical review on which kinds of variance dissimilarity models have the more accurate forecasting. The research population focuses on Iran capital market and includes dividend and price index (TEDPIX) data of Tehran Securities exchange. The sample contains 10624 observed days from 2009 until 2015 with 30 minutes sampling interval which have been analyzed.

The research results indicate that for the reason of smaller error on mean square error the mixed model, designed based on conditional model, is more accurate than other reviewed models. Also the return fluctuations are more influenced by closer data because in the mixed model the moving average, which uses data from the 60 and 120 past hours, has more accurate prediction on return fluctuation. Finally, Diebold and Mariano test statistics was used to determine predictive accuracy in both models that has the lowest root mean square error (RMSE), and as a result there was no significant difference between accuracy of these two models.

Keywords: Intraday Data, TEDPIX, Volatility, Root Mean Square Error, Forecasting, EGARCH

Introduction

Intra-day Volatility

Intraday Volatility is a factor that should not be overlooked. It is very difficult to trade simple unhedged positions on stocks that exhibit high levels of intra-day volatility. Complex positions are often designed to absorb price swings within a specified range, but simple positions are not. The intraday volatility question has many dimensions. Some analysts measure high-low as a percent of the closing price and chart this number to watch for periods of instability. The most common approach involves calculating the average for some period of time (such as 20 days) and using a sliding window to roll the calculation forward. More sophisticated calculations use weighted average to emphasize the most recent changes. However, for our purpose the method falls short, because our principal concern is that intraday volatility could be substantially higher than the classical volatility that underlies stock prices. More specifically, we are concerned that high levels of intraday volatility could result in large price swings that are not comprehended in the prices of the stocks we are trading. We need a method for accurately comparing intra-day and closing price volatilities.

Literature Review

The concept of volatility estimators is widely researched in financial literature. Scientists try to find the best estimator of true volatility, which is not observed/rather latent process, through numerous researches on daily, weekly or high frequency financial data.

Contemporarily, the most frequently used estimator is still classical volatility estimator (the sum of squared differences of its return and the mean return over the analyzed period of time) which is the part of many kinds of financial models and which is frequently treated as sufficient estimator of true volatility process. Although this estimation is to a large extent successful, we are aware of the fact that it is possible to find better, more efficient, still unbiased and consistent estimators. The most important disadvantage of Standard deviation is that it is calculated on the daily basis, not revealing intraday fluctuations and that it is supposed to have low efficiency in comparison with other volatility estimators.

Since the concept of volatility has grown in importance through the last forty years, many new concepts of volatility estimators focused on gaining on efficiency and being robust to all existing microstructure biases (bid-ask spread, the opening jump effect, non-trading bounce, etc.) have been invented. Therefore we have thoroughly and chronologically studied the most influential works concerning the issue of volatility estimators and their properties, in order to place our research as the natural consequence of the contemporary state-of-the-art and focus on the most important details which were not sufficiently explained in the previous works.

Merton (1980), who was the first to propose realized volatility concept (the sum of squared returns over the analyzed period of time measured in equidistant periods) as the unbiased and consistent estimator of daily variance on condition that the returns have a zero mean and are uncorrelated. He agreed that RV is the true volatility estimator when returns are sampled as often as possible. This concept was later heavily researched by Taylor and Xu (1997) and Andersen et al.³ (1998, 2000, 2001a and 2001b) as well as others, who additionally paid close attention to microstructure bias which unfortunately grows in importance as the sampling frequency increases.

Andersen and Bollerslev started to popularize the notion of realized volatility and correlation in 90s having written the numerous research papers (Andersen and Bollerslev, 1998, 1999a, 1999b) devoted to the techniques focusing on many possible aspects and dimensions of that issue, especially the properties of such estimator calculated on the high frequency data. They noticed that the realized volatility is a more efficient and unbiased estimator of volatility than the popular daily classical volatility estimator.

Moreover, it converges to the true underlying integrated variance when the length of the intraday interval goes to zero (Andersen et al. 2001a, 2001b). They found that the efficiency of the daily high-low range is between that of the realized variance computed using 3- and 6- hour returns. Estimating realized volatility of stock returns they noticed that the sampling frequency of 5- and 30-minute intervals strike a balance between the increasing accuracy of higher frequencies and the adverse effects of market microstructure frictions (Andersen et al., 2001a, 2003)⁴.

Zhang et al. (2005) went one step further and developed the estimator which combined realized variance estimator obtained from returns sampled at two different frequencies. The realized variance estimator obtained using a certain (low) frequency was corrected for bias due to microstructure noise using the realized variance obtained with the highest available sampling frequency.

When testing the relative performance of various historical volatility estimators that incorporate daily trading range Shu and Zhang⁵ (2006) found that the range estimators perform very well when asset price follows a continuous geometric Brownian motion. However, significant differences among various range estimators are detected if the asset return distribution involves an

opening jump or large drift. Nonetheless, the empirical result is supportive of the use of range estimators in estimating historical volatility.

Martens and Dijk (2007) tried to develop the concept of realized range by introducing scaled realized range which was additionally robust to microstructure noise. They noticed that realized range with their bias-adjustment procedure was more efficient than realized variance when using the same sampling frequency.

Data and Methodology

The study population consisted of index and prices in the Tehran Stock Exchange has during the years 2009 till 2015 studied based on six observation data intraday intervals of 30 minutes and the hours of 9:00 to 12:00 by the end of 2013 and seven View from 9:00 to 12:30 by the end of 2015 hours is used.

The calculation of price return will help us to make their used data homogeneous if it is heterogenic and make the statistical calculations and their possibilities simple and the prediction model for the research in the first step are ones which use from the interval historical information. In the research among the conditional models, four models of ARCH, GARCH, and GARCH AND GLOSTEN model and RANKLE model and one leveling model in the model explanation for evaluation of inter-day data were selected.

Result and Discussion

Unit Root Test

In order to explore the time series stagnation in the research, Augmented Diki fuller test were used. In the test, hypothesis is defined as following

The results from the test in the above table are observable.

Table 1: extended Dickey Fuller test

Significance level	t statistic	Augmented Diki Fuller Test	
0	0	Critical value	
-3,431,094			
-2,861,753			
		% 1	
		% 5	
		% 10	

Since the t statistic Absolute is larger than critical value Absolute and the significance level is equaled with zero the zero hypothesis in the 99 percent confidence level will be refused and hypothesis based on not unit root will be approved.

Investigation of Heteroskedasticity

In traditional econometric models, not varying of the variance of residuals always is considered as one of the major and classic assumptions of econometrics. In this part we used from the ARCH test for exploring the issue.

The findings from the test is observable in the following table.

Table 2: Variance Heterogeneity Exploration

0	Fisher significance level	-3.431094	Fisher statistic
0	Chi 2 significance level	-2.861753	Observations multiplied in determination coefficient

Since the significance level is less than 5 percent the zero hypothesis will be refused. Therefore the heterogeneity of the variance will be confirmed and the ARCH family models should be used in order to estimate and predict.

ARCH Model with Residuals Normal Distributions

Findings from the model is observable in the following table

Table 3: ARCH Model with Residuals Normal Distributions

Variable	Coefficient	Standard Deviation of Error	Z Statistic	Significance Level
Constant	4.00E-06	7.83E-07	5.104807	0.0000
Before period Fluctuations	0.058413	0.040931	1.427119	0.1535
Variance Equation				
Constant	1.29E-09	1.84E-12	700.6491	0.0000
RESID(-1)^2	0.171429	0.015397	11.13371	0.0000
Determination Coefficient	0.003412	Akaik Criterion		-17.17004
Adjusted determination Coefficient	0.003315	Shwartz Criterion		-17.16724
Durbin Watson	1.999810			

Since the significance level is higher than 5 percent the model always does not have a serial correlation problem of residuals.

Table 4: variance heterogeneity in the ARCH model with residuals normal distributions

9903*0	Fisher significance level	000149*0	Fisher statistic
9903*0	Chi 2 significance level	000149*0	Observation multiplied in determination coefficient

Based on the above table, the significance level is higher than 5 percent and it means that hypothesis based on the heterogeneity of variance cannot be refused. Due to above table the model is efficient.

ARCH model with residuals t student distributions.

Findings from the model are observable in the below table:

Table 5: ARCH Model With Residuals T Student Distribution

variable	coefficient	Standard deviation of error	Z statistic	Significance level
Constant	7.52E-07	3.93E-08	19.12570	0.0000
Before Period Fluctuations	0.058413	0.003618	16.14604	0.0000
Variance Equation				
Constant	4.42E-12	5.01E-14	88.18838	0.0000
Resid(-1)^2	0.171429	0.006054	28.31434	0.0000
Determination	-0.001927	Akaik Criteria		-21.74501

Coefficient			
Adjusted Determination Coefficient	-0.002024	Showartz Criteria	-21.74221
Durbin Watson	1.989154		

Since the significance level is higher than 5 percent the model don't has a serial correlation problem of residuals.

Table 6: variance inconsistency in the ARCH model with residuals t students' distribution

9880.0	Fisher significance model	000225.0	Fisher statistic
9880.0	Chi 2 significance level	000225.0	Determination multiple

Since the significance level is higher than 5 percent and it means that the zero hypotheses based on the heterogeneity of variance cannot be refused and based on the above results, the model is efficient.

ARCH model with residuals general distribution.

The results are observable in the following table:

Table 7: ARCH model with residuals general distribution

Variable	Coefficient	Standard deviation of error	Z statistic	Significance level
Constant	1.08E-06	4.48E-07	2.418823	0.0156
Before period fluctuation	0.058413	0.026055	2.241909	0.0250
Variance Equation				
Constant	4.30E-10	4.63E-13	929.1896	0.0000
RESID(-1)^2	0.171429	0.014614	11.73048	0.0000
Determination Coefficient	-0.000891	Akaike Criteria	-19.17039	
Adjusted Determination Coefficient	-0.000988	Shwartz Criteria	-19.16758	
Durbin Watson	1.991213			

Since the significance level is higher than 5 percent the model has no serial correlation problem of residuals

Table 8: Heterogeneity of variance in ARCH model with residuals general distribution

9887.0	Fisher significance level	000201.0	Fisher statistic
9887.0	Chi 2 significance level	000201.0	Determination coefficient multiple in observations

Since the significance level is higher than 5 percent and it means that zero hypothesis based on heterogeneity of variance cannot be refused the model is efficient

Due to above findings, all three models are efficient but due to the fact that akaike and shwartz models are less for ARCH model with t student distribution of residuals, ARCH model with t student distribution of residuals is more suitable and the error square average root values from the method is equaled with 0.00003862.

GARCH Model

In this section we try to predict the return Fluctuations in Tehran exchange using GARCH model. In order to explore the issue, three GARCH models with residual normal distribution, GARCH model with t student distribution of residuals and GARCH model with general distribution of residuals are used and due to akaik and shwartz criteria, the suitable model will be selected. It means each model that has lesser akaik and shwartz will be more suitable.

GARCH Model with Residuals Normal Distribution

The results can be observable in the following table.

Table 9 GARCH model with residual normal distribution

variable	coefficient	Standard deviation of error	Z statistic	Significance level
Constant	9.88E-07	1.78E-06	0.554390	0.5793
Last period Fluctuations	0.058412	0.105860	0.551789	0.5811
Variance Equation				
Constant	8.46E-10	4.91E-11	17.23488	0.0000
RESID(-1)^2	0.150004	0.021633	6.934017	0.0000
GARCH(-1)	0.600005	0.023350	25.69562	0.0000
Determination coefficient	-0.001177	akaike	-17.33449	
Adjusted determination coefficient	-0.001274	shwartz	-17.33099	
Watson durbin	1.990643			

Since the significance level is higher than 5percent the model has not the serial correlation problem with residuals

Table 10: variance heterogeneity in GARCH model with residual normal distribution

0.9929	Fisher significance level	0.000078	Fisher statistic
0.9929	Chi 2significance level	7.82E-05	Determination coefficient multiplied in observations

Since the significance level in above table is higher than 5 percent and it means that zero hypothesis based on variance heterogeneity cannot be refused the model is efficient.

GARCH Model with Residual T Student Distribution

Results from the model are observable in following table:

Table 11 GARCH Model With T Student Distribution of Residuals

Variable	coefficient	Standard deviation of error	Z statistic	Significance level
Constant	4.63E-07	2.54E-08	18.24609	0.0000
Before Period Fluctuations	0.058413	0.013901	4.202084	0.0000
Variance Equation				
Constant	4.27E-13	8.50E-15	50.28354	0.0000
Resid(-1)^2	0.150000	0.003603	41.63286	0.0000
Garch(-1)	0.600000	0.002856	210.0869	0.0000
Determination Coefficient	-0.002923	Akaike criteria	-21.84636	
Adjusted Determination Coefficient	-0.003020	Shwartz criteria	-21.84285	
Durbin Watson	1.987178			

since the significance level is higher than 5 percent the model has no serial correlation problem with residuals

Table 12 variance heterogeneity in GARCH model with residuals t student distribution

9563*0	Fisher significance level	003000*0	Fisher statistic
9563*0	Chi 2 significance level	003001*0	Determination coefficient multiplied in observation

Since the significance level is higher than 5 percent and it means the zero hypothesis based on variance heterogeneity cannot be refused the model is efficient

GARCH Model With Residual General Distribution

Findings are observable in following table

Table 13: GARCH Model with Residual General Distribution

Variable	coefficient	Standard deviation of error	Z static	Significance level
Constant	7.93E-07	1.63E-07	4.877173	0.0000
Before Period Fluctuations	0.058645	0.017818	3.291375	0.0010
Variance Equation				
Constant	3.48E-11	4.95E-13	70.31565	0.0000
Resid(-1)^2	0.151756	0.006360	23.86106	0.0000
Garch(-1)	0.604536	0.005674	106.5451	0.0000
Determination Coefficient	-0.001789	Akaike criterion	-19.82416	
Adjusted Determination Coefficient	-0.001886	Shwartz criteria	-19.82065	
Watson Durbin	1.989891			

Since the significance level in above table is higher than 5 percent the model has no serial correlation problem with residuals

Table 14 variance heterogeneity in GARCH model with residual general distribution

9727.0	Fisher significance level	001170.0	Fisher statistic
9727.0	Chi 2 significance level	001170.0	Determination coefficient multiplied in observations

Due to above table the significance level is higher than 5 percent and it means that zero hypothesis based on the variance heterogeneity cannot be refused. Due to above result the model is efficient. Likewise due to results three models are efficient. But due to the fact that akaike and shwartz criterion are less for GARCH model with t student distribution of residuals the GARCH model with t student distribution of residuals are more suitable.

Square average root value of the error in the method is equaled with 0.000038686

Results of Mariano and Daibold Test Statistic

As we said in the previous chapter, in order to compare the best models, the Mariano and Diabold test will be used. Since we can be certain which conditional model (120 hours moving average with error square average root value 0.00003790) and conditional combinational model (60hourse moving average and 120 hours moving average with error square average root value 0.000037895) performance that have the lowest error square average root value among the models are better, the statistic are used for comparison that hypothesizes are as followed

H0: equation of model prediction strength

H1: not equation of models prediction strength

Above hypothesis is done because there is high emphasis on examined models difference in fluctuations prediction. Since the software output for statistic of test is 0.408 and critical value in confidence level of 95 percent is 1.67 and the significance level is 0.68 that is higher than 5 percent. The zero hypothesis cannot be refused therefore the prediction strength of both models are same.

Conclusion and Suggestions

In the end after exploration we answer to 3 research questions as following

- 1) Among the studies conditional models , Exponential GARCH model of the for modeling and prediction of return fluctuations is more suitable
- 2) Among the non-conditional models of the research the moving average model 120 is more suitable for prediction of return fluctuation prediction
- 3) Among the conditional and unconditional models and their combination the unconditional combinational model is more suitable for modeling and prediction of return fluctuation prediction

As we said in pervious section for determination of prediction strength, 2 premier models (120 hours moving average conditional model with error square moving root value 0.0003790 and 60 hours and 120 hours moving average conditional model with error square average root value that has lowest error square average root level among the explored model in the research, the Daibold and Mariano test has been used that is observed with value of 0.408 finally. There is no significance difference between prediction strength of both models. Because in facet difference of the error square average root value of both models are not too high

Table 15: Results Of Research Model Error Average Square

Selected Model	Standard Deviation Of Error	Research Model	
Exponential GARCH	0.00003862	ARCH	conditional
	0.000038686	GARCH	
	0.000038687	GARCH and GLOSTEN and rankle	
	0.000038008	Exponential GARCH	
120 hours moving average	0.000038796	20 hours moving average	unconditional
	0.000038008	60 hours moving average	
	0.00003790	120 hours moving average	
	0.000038262	250 hours moving average	
	0.000049832	Exponential leveling	
60 and 20 hours moving average	0.000037895	20and 60 hours moving average	
	0.000038083	Exponential GARCH and ARCH	
0.408		Daibold And Mariano Model	

Research suggestions

There are some suggestions due to results of the research that are categorized in 2 category and they include suggestions based on the research results and suggestions for future suggestions.

Suggestions Based on the Research Results

Tehran exchange should create a safe bed for the investors that can calculate the values exposed fluctuations daily and for the various industries. Due the findings of the research and more studies about this field we can design suitable models for prediction of return fluctuations and make them available for users in order to make correct decision in investment domains planning and policy making. It lead to higher efficiency of the capital market

Suggestions for the future

- ✓ Our method for modeling was single variable but it can be extended to one environment with multi variables
- ✓ Using from the current research method for fluctuations prediction in the various indicators of Tehran exchange
- ✓ Using from the other conditional models that are not studied in the research in order to predict indicator of Tehran exchange
- ✓ Using from the conditional and unconditional model and both of them in order to predict the price for shares and derivatives
- ✓ Using from the wavelets and superficial nervous network in order to predict the return fluctuations and its comparisons with current research results
- ✓ In the study 4 family model of ARCH and 5 moving average models are used that there are so many models in the field that can be applied and different results to be gained

References

Anderson, T.G., Bollerslev, T., Diebold, F.X., Ebens H., (2001a). The Distribution of Realized Stock Return Volatility, Journal of Financial Economics, 43-76

- Anderson, T.G., Bollerslev, T., Diebold, P., Labys, (2003). Modeling and Forecasting Realized Volatility, *Econometrica* a. 579-625.
- Shu, J., Zhang, J., (2006). Testing Range Estimators of Historical Volatility of S&P 500 Index. *Journal of Empirical Finance*. 297-313.
- Slepaczuk, R., Zakrzewski, J. (2010). High Frequency and Model-Free Volatility Estimators, Conference Paper - 17th International Conference: FORECASTING FINANCIAL MARKETS. Hannover, Germany.