An Empirical Study of

Intra-day Stock Return Volatility

BY

JIAN SU
M.S. Biostatistics, the University of Illinois at Chicago, 2006
M.S. Computer Science, the University of Illinois at Chicago, 2003
M.A. Economics, the University of Illinois at Chicago, 2001
B.S. Business Administration, Capital University of Economics and Business, 1995

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Business Administration
in the Graduate College of the
University of Illinois at Chicago, 2011

Chicago, Illinois

Defense Committee:
Lan Zhang, Finance, Chair
Gilbert W. Bassett, Finance
Dibyen Majumdar, Mathematics Statistics and Computer Science
Stanley Sclove, Information and Decision Sciences
Fangfang Wang, Information and Decision Sciences

To My Parents

ACKNOWLEDGMENTS

First and foremost I offer my sincerest gratitude to my advisor, Prof. Lan Zhang, who has supported me throughout my thesis with her patience and knowledge while allowing me the room to work in my own way. I attribute the level of my PhD degree to her encouragement and effort and without her this thesis, too, would not have been completed or written. I also learned a lot from Prof. Fangfang Wang as she can always provide me with valuable insight in statistical analysis problems. My appreciation and thanks also go to my committee members: Prof. Gib Bassett, Prof. Dibyen Majumdar, and Prof. Stan Sclove for providing support and encouragement for my dissertation.

I have been blessed with a friendly and cheerful group of fellow students. I would like especially to thank Inna Khagleeva for sharing her thought and literature resource with me. Finally, I am very grateful to my mother, my father and my sister, as without their love and care I would not have made it this far.

June 2011

Jian Su

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

TABLE **PAGE**

LIST OF ABBREVIATIONS

ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller
ARFIMA	Autoregressive Fractionally Integrated Moving Average
FARIMA	fractional ARIMA
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GPH	Geweke and Porter-Hudak
MLE	Maximum Likelihood Estimate
OLS	Ordinary Least Squares
PACF	Partial Autocorrelation Function
PP	Phillips and Perron
QQ	Quantile-Quantile
SEMIFAR	Semiparametric Fractional Autoregressive
SFARIMA	Seasonal Fractional ARIMA
TSRV	Two Scale Realized Volatility

SUMMARY

It is well known that microstructure noise could have substantial impact on volatility estimation of high frequency asset returns. The Two Scale Realized Volatility (TSRV) estimator makes use of all the available data and at the same time corrects the effect of market microstructure noise. In this study, 30-minute TSRV series is constructed from tick-by-tick Dow Jones 30 stock prices. Our results show that the 30-minute volatility estimate series has the stylized characteristics, including volatility clustering, long memory and displaying U-shape within the day. Also, the volatility for stocks during earning announcement period is significantly higher than that in non-announcement period. This phenomenon is particularly striking at the opening hour of the announcement day. Time series model is built on the periodic and long memory features with rolling window size of one month. We forecast the out-of-sample 30-minute volatility one day ahead based on Semiparametric Fractional Autoregressive model and modified HAR-RV linear regression model.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

In finance, volatility refers to the variations of the continuously compounded asset returns within a specific time horizon. Although volatility is inherently unobservable, understanding volatility and its characteristics lies at the center of asset pricing. As the primary measure of risk, volatility drives the construction of optimal portfolios, the hedging and pricing of options and other derivative securities based on Black-Scholes option pricing formula. It also plays a critical role in discovering trading and investment opportunities which provide an attractive risk-return trade-off.

Estimation of asset volatility attracts enormous interests in theoretical and empirical finance research. There are numerous widely-used volatility models, include parametric econometric models such as generalized autoregressive conditional heteroscedasticity (GARCH)-class models proposed by Engle (1982), Bollerslev (1986), Nelson (1991) and others, implied volatility from Black-Scholes model and Hull White model, stochastic volatility (SV) model as Heston model (1993), and the non-parametric volatility estimator taking the sum of squared ex post asset returns.

In recent years, as electronic trading becomes popular, security prices are quoted and traded at higher frequency, and the tick-by-tick data becomes available and contains a wealthy of pricing information of the market. Merton (1980) said that spot volatility may be inferred perfectly if the asset price follows a diffusion process and a continuous record of price is available.

However, there exist a great challenge to estimate the spot volatility and the integrated volatility in the high frequency setting.

Suppose S_t is asset transaction price at time t, and the logarithm of transaction price $X_t = logS_t$ follows an $It\hat{o}$ process,

$$dX_t = \mu_t dt + \sigma_t dB_t \tag{1.1}$$

where B_t is a standard Brownian motion, μ_t and σ_t are the drift and instantaneous volatility of the stochastic process X_t, respectively. (1.1) is typically referred to as the continuous-time semi-martingale for asset prices.

The parameter of our interest is the integrated volatility for a given time period [0,T]. It is also known as the continuous quadratic variation $\langle X, X \rangle$.

$$\langle X, X \rangle = \int_0^T \sigma_t^2 dt \tag{1.2}$$

The integrated volatility can simply be estimated by the sum of squared returns,

$$[X, X]_T = \sum_{i=1}^{n} (X_{t_{i+1}} - X_{t_i})^2 \tag{1.3}$$

where $[X, X]_T$ is the quadratic variation, n is the number of the observations from time 0 to time T, t_i's are the ith time point, X_{t_i}'s are all of the observations for the return process in [0,T]. The estimator $\sum_{i=1}^{n} (X_{t_{i+1}} - X_{t_i})^2$ is commonly used and usually called *realized volatility.*

Based on probability theory, large sample size will make this estimator approximate the true integrated volatility over a given time period. This is justified by the theoretical result from stochastic processes that if the sampling frequency increases, i.e., if for a given time period [0,T], we decrease the sampling intervals, we will have

$$plim \sum_{i=1}^{n} (X_{t_{i+1}} - X_{t_i})^2 \approx \int_0^T \sigma_t^2 dt \quad 0 < t_1 < \cdots < t_n < T \tag{1.4}$$

So the realized volatility estimator based on high frequency sample should be a precise estimator for the integrated volatility $\int_0^T \sigma_t^2 dt$.

However, in empirical finance research, the reality turns out to be in the opposite scenario. This realized volatility estimator diverges from the integrated volatility if we increase the sampling frequency. This is known as the effect of market microstructure noise in high frequency data analysis. The realized volatility signature plot popularized by Andersen et al. (2000) showed the dependence of volatility on sampling frequency by plotting the realized volatility as a function of the sampling frequency of the underlying intra-day returns.

When there exists market microstructure noise, an observed security price at time t_i, $logS_{t_i}$, is not the true latent asset price X_{t_i}, but contaminated with market microstructure noise. Suppose observed security price has the form

$$logS_{t_i} = X_{t_i} + \epsilon_{t_i} \tag{1.5}$$

where ϵ_{t_i} is the error caused by contamination of market microstructure noise. ϵ_{t_i} is assumed to be independent and identically distributed with $E\epsilon_{t_i} = 0$, $var(\epsilon_{t_i}) = E\epsilon^2$, and ϵ independent from X process. Market microstructure noise captures a variety of frictions inherent in the trading process: bid-ask bounces, discreteness of price changes, differences in trade sizes or informational content of price changes, gradual response of prices to a block trade, strategic component of the order flow, inventory control effects, etc. Although the market microstructure noise is of small magnitude, it progressively dominate the signal of true latent volatility as we increase the sampling frequency. It is therefore not surprising that volatility estimation and inference on high frequency data has received substantial attention in the financial econometric and statistical literature.

To mitigate the effect of microstructure noise, most of the empirical finance literature suggests not sampling too frequently. A general approach in finance literature is to construct a volatility estimator based on sparse sampling, which is arbitrary and suboptimal. The typically adopted sampling length in the literature usually ranges from 5 to 30 minutes. For instance, Ederington and Lee (1993) estimated standard deviation by sampling the option contract price every 5 minutes over the trading day; Zhou (1996)'s realized volatility estimator makes adjustment for the first-order autocorrelation in the high frequency returns, leading to unbiased but inconsistent estimator of the integrated volatility. Anderson, Bollerslev, Diebold, and Labys (2001) estimated daily realized exchange rate volatility based on 5-minute returns; Ait-Sahalia, Mykland and Zhang (2005) and Bandi and Russell(2006) used reduced form models to show that optimally-sampled realized volatility outperforms the realized volatility constructed

by ad-hoc sampling sparsely at 5- to 30-minutes intervals; Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) proposed the realized kernel estimators. Anderson, Bollerslev, Meddahi (2010) compared these various estimators for daily realized volatility based on sampling intervals 1-minute, 5-minute, 15-minute, 30-minute and daily returns, assuming the spot volatility process follows a GARCH Diffusion and a two-factor affine model in Anderson, Bollerslev, Meddahi(2004). They used the 1-minute and 5-minute sampling intervals to construct the Two Scaled Realized Volatility (TSRV), and found out that the forecasts generated by the average estimators including TSRV and kernel estimators are uniformly best. Patton and Sheppard (2009) evaluates and compares volatility forecast, and the volatility estimator is constructed based on 5-minute, 30-minute, and daily returns.

All of the above research constructed the volatility estimators based on a sparse subset of data. However, one of the basic guidelines in statistics is that we should not throw away the available data by sparse sampling. Zhang, Mykland and Ait-Sahalia (2005) proposed the Two Scale Realized Volatility (TSRV) estimator, which makes use of all available high frequency data and at the same time quantifies and corrects the microstructure noise. It is the first nonparametric unbiased and consistent volatility estimator. In their work, they incorporated the microstructure noise explicitly into the estimation procedure for the integrated volatility rather than skipping the microstructure noise problem.

Currently there are three main approaches to estimate volatility in the nonparametric case while using all the available data: Two Scaled Realized Volatility estimator and Multiple Scale Realized Volatility estimator, which are linear combination of realized volatilities obtained by

subsampling; realized kernel estimator, which is a linear combination of autocovariances; pre-averaging approach by Jacod, Li, Mykland, Podolskij and Vetter (2008).

Based on the time series of volatility estimation, we can make forecasting for intra-day volatilities. The list of recent literatures on volatility forecasting on high frequency data keeps on growing. Engle and Gallo (2006) proposed Multiplicative Error Model (MEM), their model specifies a GARCH structure for each of the realized measures, so that an additional latent volatility process is introduced for each realized measure in the model. MEM model has a total of three latent volatility processes. Shephard and Sheppard (2010) devised multivariate High-frEquency-bAsed VolatilitY (HEAVY) model, which is nested in the MEM framework. It includes at least two latent volatility processes. Unlike the traditional GARCH models, these models operate with multiple latent volatility processes. Chen, Ghysels, Wang (2009) proposed a class of High Frequency Data-based Projection-Driven GARCH, or HYBRID GARCH models that allows a mixture of frequencies in terms of prediction horizons and conditioning information. They distinguish three cases: 1)data-driven HYBRID processes, 2)structural HYBRID processes, and 3)HYBRID filtering processes.

Now we will get back to Zhang, Mykland and Ait-Sahalia (2005), they compared five realized volatility estimators summarized as follows:

1. Realized volatility using all of the data

$$[Y,Y]_T^{(all)} = 2nE\epsilon^2 + O_p(n^{1/2}) \tag{1.6}$$

where n is the sample size over time period [0,T]. The realized volatility estimates not the true integrated volatility but the variance of the contamination noise term. The magnitude of the realized volatility increases linearly with the sample size. Scaled by $(2n)^{-1}$, it estimates consistently the variance of microstructure noise, $E\epsilon^2$, rather than $\langle X, X\rangle$, as

$$\widehat{E\epsilon^2} = \frac{1}{2n}[Y,Y]_T^{all} \tag{1.7}$$

and

$$n^{1/2}(\widehat{E\epsilon^2} - E\epsilon^2) \to N(0, E\epsilon^4) \tag{1.8}$$

In this case, market microstructure noise totally swamps the variance of the price signal.

2. Realized volatility based on sparse sampling

$$[Y,Y]_T^{(sparse)} \approx < X,X >_T + 2n_{sparse}E\epsilon^2 + a^2Z \tag{1.9}$$

3. Realized volatility based on optimally determined sparse sampling

$$[Y,Y]_T^{(sparse,optimal)} \approx < X,X >_T + 2n_{optimal}E\epsilon^2 + b^2Z \tag{1.10}$$

4. Realized volatility based on subsampling and averaging

$$[Y,Y]_T^{(avg)} \approx < X,X >_T + 2\bar{n}E\epsilon^2 + c^2Z \tag{1.11}$$

5. Realized volatility based on subsampling and averaging on two scales

$$[Y,Y]_T^{(TSRV)} \approx < X,X >_T + d^2 Z \tag{1.12}$$

where n_{sparse}, $n_{optimal}$, $\bar{n}$ are sample sizes over time period [0,T]. a^2, b^2, c^2, d^2 is the total variance due to noise and discretization, and Z is a standard normal N(0,1) random variable. The symbol "≈" denotes that when multiplied by a factor, the convergence is in law. The realized volatility estimators in (1.9), (1.10) and (1.11) are biased. The bias caused by microstructure noise increases linearly with the sample size in time period [0,T]. The Two Scaled Realized Volatility estimator in (1.12) is centered at the $< X,X >_T$, i.e., unbiased, free from the effect of microstructure noise. It asymptotically follows a normal distribution, and the convergence is in law. In this study, we employ TSRV estimator on the trade security prices, document the dynamics of TSRV series, and predict the intra-day volatility.

The rest of the thesis proceeds as follows. The mathematical model for TSRV is described in Chapter 2; Chapter 3 describes the data and studies the stylized facts of the 30-minute volatility; Chapter 4 provides two alternative ways to forecast 30-minute volatility one day ahead; Chapter 5 investigates the impact of quarterly earning announcement to intra-day volatility.

CHAPTER 2

TWO SCALE REALIZED VOLATILITY METHODOLOGY

Zhang, Mykland and Ait-Sahalia (2005) proposed a model free approach, Two Scale Realized Volatility (TSRV) estimator, to estimate volatility at the highest frequencies, taking advantage of the rich tick-by-tick data and correcting the adverse effects of microstructure noise on volatility estimation. Instead of selecting a subsample arbitrarily or optimally, this approach is based on selecting a number of subgrids of the original grid of observation times, $\mathcal{G} = t_0, \cdots, t_n$, and then averaging the estimators derived from the subgrids. It makes use of all data by extending the estimator to each subgrid partition.

TSRV is built on aggregating the observations on two different sampling scales: a slow-scale (low frequency) estimator achieves the purpose of increasing signal-to-noise ratio, while the fast scale (high frequency) realized volatility is used as a bias correction device. The best results can be achieved by combining the estimators from two different scales. The final TSRV estimator is asymptotically normal, and converges at the rate of $n^{-1/6}$. It is the first consistent estimator for integrated volatility when microstructure exists in high frequency security price.

We follow the notation in Zhang, Mykland and Ait-Sahalia (2005). Suppose the full grid $\mathcal{G}$, $\mathcal{G} = \{t_0, \cdots, t_n\}$ is partitioned into K nonoverlapping subgrids $\mathcal{G}^{(k)}, k = 1, \cdots, K$;

$$\mathcal{G} = \bigcup_{k=1}^{K} \mathcal{G}^{(k)}, \quad where\ \mathcal{G}^{(k)} \bigcap \mathcal{G}^{(l)} = \emptyset\ when\ k \neq l. \tag{2.1}$$

A natural way to select the kth subgrid $\mathcal{G}^{(k)}$ is to start with t_{k-1} and pick every Kth observation until T. The subsample is

$$\mathcal{G}^{(k)} = \{t_{k-1}, t_{k-1+K}, t_{t-1+2K}, \ldots, t_{k-1+n_k K}\} \tag{2.2}$$

for $k = 1, \cdots, K$, where n_k is the integer that makes $t_{k-1+n_k K}$ the last element in $\mathcal{G}^{(k)}$. We let $n_k = |\mathcal{G}^{(k)}|$, where $|\mathcal{G}^{(k)}|$ = (# of points in grid $\mathcal{G}^{(k)}$)-1, that is, $|\mathcal{G}^{(k)}|$ is the number of time increments.

The realized volatility based on all observations points $\mathcal{G}$ is denoted as $[Y,Y]^{(all)}$. If we use only the subsampled observations $Y_t, t \in \mathcal{G}^{(k)}$, the realized volatility is denoted as $[Y,Y]^{(k)}$. It has the form

$$[Y,Y]_T^{(k)} = \sum_{t_j, t_{j,+} \in \mathcal{G}^{(k)}} (Y_{t_{j,+}} - Y_{t_j})^2 \tag{2.3}$$

where $t_j \in \mathcal{G}_{(k)}$ and $t_{j,+}$ is the next element after t_j in $\mathcal{G}_{(k)}$. We average this estimator to reduce the variation of realized volatility across subgrids, then we have

$$[Y,Y]_T^{(avg,K)} = \frac{1}{K} \sum_{k=1}^{K} [Y,Y]_T^{(k)} \tag{2.4}$$

This is a natural competitor to $[Y,Y]_T^{(all)}$, $[Y,Y]_T^{(sparse)}$ in (1.9), and $[Y,Y]_T^{(sparse,optimal)}$ in (1.10). The sample size $barn_k$ is averaged across k. We define that,

$$\bar{n}_K = \frac{1}{K} \sum_{k=1}^{K} n_k = \frac{n - K + 1}{K} \tag{2.5}$$

.

The TSRV estimator $\widehat{\langle X, X\rangle}_T$ is defined as

$$\widehat{\langle X, X\rangle}_T = [Y, Y]_T^{(avg,K)} - \frac{\bar{n}_k}{\bar{n}_j}[Y, Y]_T^{(avg,J)} \tag{2.6}$$

where $\bar{n}_K = \frac{n-K+1}{K}$ and $\bar{n}_J = \frac{n-J+1}{J}$, with $J \ll K$. The right hand side of (2.6) is a linear combination of realized volatilities obtained by subsampling. The first component is slow-scale estimator, and the second component is fast-scale estimator used to correct the bias. For small sample size, we adjust the TSRV estimator in the form of

$$\widehat{\langle X, X\rangle}_T^{adj} = (1 - \frac{\bar{n}_k}{\bar{n}_j})^{-1}\widehat{\langle X, X\rangle}_T \tag{2.7}$$

Empirically we set K=200 for liquid securities such as Microsoft and IBM stocks, and set lower value of K depending on the liquidity of stocks. For empirical reason, we set J=4 for the correction components for all securities. This estimator is asymptotically normal and is an unbiased and consistent estimator for the latent integrated volatility.

CHAPTER 3

FEATURES OF 30-MINUTE REALIZED VOLATILITY

3.1 Data

Our analysis is performed on the tick-by-tick stock trading prices of the 30 companies included in Dow Jones Industrial Average index, which are attractive candidates for examination because they have much attention from investors and very liquid in the equity market.

The tick-by-tick transaction data for Dow Jones 30 stocks was obtained from Wharton Research Data Services (WRDS: http://wrds.wharton.upenn.edu/) NYSE Trade and Quote (TAQ) database. The intra-day transaction data for the 30 stocks covers from the fourth quarter of 1999 through the third quarter of 2005. We also studied the stock quarterly earning announcement effect on the intra-day stock volatility. The actual dates of earning announcement on each of the 30 stocks are different. The actual earning announcement date information was obtained from Institutional Brokers Estimate System (I/B/E/S) database at WRDS.

In this study, 25 out of the 30 stocks cover the entire 24 quarters, the other five companies have earning information available over shorter period, as listed on I/B/E/S database, we accommodate this situation by using the shorter period for these five stocks. The tickers for the five companies are CVX, HPQ, JPM, KFT and VZ. The time line covered by the 30 stocks is listed in table I.

We study the tick level trade data, which is the real transaction stock price from NYSE TAQ database. The data was cleaned according to the following rules:

1. Transactions earlier than 9:30:00 and later than 16:00:00 were discarded, in another words, we analyze the intra-day transactions from 9:30:00 to 16:00:00, 6.5 trading hours per day.

2. We remove all trading days that are shorter than normal trading hours covering from 9:30:00 through 16:00:00, including the inactive trading days right before and right after national holidays such as New Year, July Fourth, Thanksgiving, Christmas, and other inactive trading days such as September 11, 2002, and June 8, 2001.

3. Outliers significantly deviated from the time series plot of price trend are treated as data error, and therefore are removed. The trend is by time series plot on daily basis.

This data cleaning procedure is not optimal. Since TSRV estimator is robust and the magnitude of the high frequency data size is sufficiently large, the data cleaning procedure will not affect our findings and conclusion. Other automatic cleaning procedure such as Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) could also be adopted in high frequency data.

3.2 Stylized Facts of TSRV Time Series

In practice, researchers have uncovered many so-called "stylized facts" about the volatility of financial time series; Bollerslev, Engle and Nelsen (1994) give a complete list of these facts. The TSRV estimator is capable of modelling volatility for high frequency data and capturing some important stylized facts of the volatility dynamics based on observations in high frequency financial time series.

Trading time is from 9:30 to 16:00, total 6.5 trading hours per day. We built 30-minute realized volatility time series for each of the Dow Jones 30 stocks, that is, a TSRV volatility is estimated on transaction prices within each 30-minute interval. There are 13 TSRV estimates per day. Tick-by-tick data is voluminous, with as much as roughly 1000 to more than 100,000 observations in one trading day.

We transformed the stock tick price to the logarithm form to construct the 30-minute TSRV time series. The continuously compounded returns are simply the difference of the logarithm price between the post and ex time points. Applying (2.6), we use various grid sizes K for different stocks, empirically based on the trading frequency of stocks. The values of K for the 30 stocks in Dow Jones Industrial average are listed in Table II.

The intra-day time series plots of TSRV series for all of the Done Jones 30 stocks are in similar pattern, suggesting that they share the same characteristics. Since the five stocks with tickers CVX, HPQ, JPM, KFT and VZ have shorter periods, we align and chop the TSRV estimates of 30 stocks to the same period of time, i.e., from August 13, 2002 through November 18, 2005. The resulting TSRV time series has 10634 observations in 818 days. We will investigate the characteristics and stylized facts for TSRV time series of each of the 30 stocks in this period of time.

Figure 1 shows the time series plot of 30-minute volatility for each of 30 stocks, the horizontal axis is the time line, and the vertical axis is the 30-minute volatility. It shows that the volatility is higher for many stocks before the third quarter of 2003, when the stock market is still in the mire of dot com bubble recession, and becomes stable after that, indicating the stock

market improved afterwards. The time series plot also shows pronounced persistence. The large values of volatility tend to be followed by large values, and small values of volatility tend to be followed by small values. The visual impression of strong persistence in the volatility measure is confirmed by the highly significant Ljung-Box test, and the test statistics is 27350.21. It is a manifestation of the well-documented return volatility clustering of financial time series.

Figure 2 is the zoom-in version of the time series plot for an arbitrary selected 21 days. The horizontal axis is the time line for 21 days and the vertical axis is the 30-minute volatility. The 30-minute volatility per day is denoted by different colors. We can see clearly the intra-day pattern of the 30-minute volatility generally demonstrates U-shape, or mirror image of J-shape. The volatility for the opening market and closing market are higher than the volatility in the mid-day. The open market volatility looks much higher than the closing market volatility.

In table III, We performed two sample t-test and paired t-test on the opening 30-minute volatility and the closing 30-minute volatility, the t statistics are 32.6338 and 74.2584 and p-values are close to zero. Therefore the opening 30-minute volatility is significantly higher than the closing 30-minute volatility.

Table IV shows the descriptive statistics of the 30-minute volatility for 818 days. the mean is 0.0028, the median is 0.0024, the standard deviation is 0.0012, the skewness is 1.7366, and the kurtosis is 4.0684. The skewness is greater than 0 and kurtosis is greater than 3, suggesting that the distribution of 30-minute volatility skews to the right, and it is a fat right-tailed, or leptokurtic distribution. As shown in figure 3, The probability density function of 30-minute volatility usually skewed to the right and have fatter right tail, compared with standard normal

distribution. Figure 4 shows the probability density function of logarithm of 30-minute volatility against standard normal distribution. The density functions of logarithm volatility are closer to normal distribution. The descriptive statistics in Table IV shows that the logarithm of 30-minute volatility has smaller skewness than original volatility, and it is now close to a normal distribution.

Figure 5 shows the autocorrelation functions (ACF) of 30-minute volatility with lag as 200. There are layers for different trading time in ACF plot. The highest autocorrelation spikes start from more than 0.8, and decay very slowly to roughly 0.6 at a displacement of about 15 trading days. The high bars are the autocorrelations of 30-minute volatility for the same time period at different days. The bottom panel of Figure 6 is the zoom-in view of ACF of 30-minute volatility, with lag as 91, which is 7 trading days. The autocorrelations forms a neat U-shape pattern, and repeats this pattern for every 13 lags on daily basis. The slow decay in ACF seems to indicate that the 30-minute volatility has the property of long memory process. This is consistent with many empirical studies. See Lobato and Savin (1998), Ray and Tsay (2000), Andersen, Bollerslve, Diebold and Labys (1999) and others.

Based on the aforementioned features of 30-minute volatility, we found that the volatility stylized facts demonstrated by TSRV time series are as follows:

1. volatility clustering
2. Skew to the right
3. leptocurtic, or fat right tails

4. U-shape or inverse J-shape pattern in the daily time frame, volatility in the opening 30 minutes is significantly higher than that in closing 30 minute

5. periodic long memory process

CHAPTER 4

FORECASTING

Volatility is one of the key elements in pricing and hedging financial derivatives such as options. It also plays an active role in both traditional portfolio management and modern arbitrage based trading strategies. Improved volatility estimate and forecast should help better decision making in pricing, hedging and trading. Therefore, high frequency trading players develop trading strategies heavily based on intra-day volatility estimate and forecast. Investigating intra-day volatility forecast becomes essential in equity and options trading environment. After we studied the features of 30-minute TSRV volatility time series, we can build statistical models on these estimates in order to make forecasting on the future intra-day stock volatility.

The findings of slow autocorrelation decay may seem to indicate the presence of a unit root, as in the integrated GARCH model of Engle and Bollerslev (1986). In the bottom panel of figure 4, the intra-day ACF pattern repeats for every 13 observations, as there are 13 30-minute volatility estimates within daily trading hours. Two alternative linear regressions can be employed to capture the daily periodic pattern. First method, a linear regression can be built on the 12 indicator variables corresponding to the first 12 30-minute volatility, having the closing 30-minute volatility as a comparison benchmark. Second method, a linear regression can be built on polynomial time variables. Unit root tests are then performed on the de-seasonal 30-minute volatility time series across 30 stocks to determine whether it is a random walk, I(1) or stationary, I(0) process.

Since the distribution of average 30-minute volatility is skewed to the right, we use the logarithm of volatility as dependent variable in order to correct some skewness. The regression models to capture seasonality are as follows.

Suppose $Y_t = log(\widehat{\langle X, X \rangle}_T)$, we have

Model I:

$$Y_t = \alpha_0 + \sum_{i=1}^{12} \beta_i d_{it} + \varepsilon_t \quad d_{it} = \begin{cases} 1 & t\%13 = i \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \cdots, 12 \tag{4.1}$$

d_{it}s are the 12 indicator variables.

Model II:

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \varepsilon_t \tag{4.2}$$

where α_0, α_1, α_2, α_3 are the intercept, linear, quadratic, and cubic time coefficient respectively.

To find out if the 30-minute volatility time series is an I(1) or I(0) process, unit root tests are then performed on the residuals extracted from the above regressions. We have used Augmented Dickey-Fuller (ADF) unit root test. (Said and Dickey 1984). The ADF tests the null hypothesis that a time series is I(1) against the alternative that it is I(0), assuming that the dynamics in the data have an ARMA structure. The test regression in the ADF test is formulated as

$$Y_t = \beta' D_t + \phi Y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta Y_{t-j} + \varepsilon_t \tag{4.3}$$

or

$$\Delta Y_t = \beta' D_t + \pi Y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta Y_{t-j} + \varepsilon_t \tag{4.4}$$

where D_t is a vector of deterministic terms (constant, trend etc.) The p lagged difference terms, ΔY_{t-j}, are used to approximate the ARMA structure of the errors, and the value of p is set so that the error ε_t is serially uncorrelated. The error term is assumed to be homoskedastic. Under the null hypothesis, Y_t is I(1) which implies that $\phi = 1$ or $\pi = 0$. The hypothesis are

$H_o : \phi = 1 \Leftrightarrow \pi = 0 \Leftrightarrow Y_t \sim I(1)$

$H_a : \phi < 1 \Leftrightarrow \pi < 0 \Leftrightarrow Y_t \sim I(0)$.

The ADF t-statistic is -12.88, and the corresponding p-value is close to 0, thus we reject the null hypothesis of unit root at the significant level of 0.05. The very slow autocorrelation decay combined with the negative signs and slow decay of the estimated augmentation lag coefficients, which are listed in table VI, suggest that long memory may be present.

In addition, Phillips and Perron (1988) (PP) developed another unit root tests that are different from the ADF test in how they deal with serial correlation and heteroskedasticity in the errors. The ADF test use a parametric autoregression to approximate the ARMA structure of the errors in the test regression, the PP test ignore any serial correlation in the test regression. The test regression for the PP test is

$$\Delta Y_t = \beta' D_t + \pi Y_{t-1} + u_t \tag{4.5}$$

where u_t is I(0) and may be heteroskedastic. The PP tests make a nonparametric correction for any serial correlation and heteroskedasticity in the errors u_t of the test regression by directly modifying the t-test statistics. See Phillips, P.C.B and P. Perron (1988). The test is robust with respect to unspecified autocorrelation and heteroscedasticity in the disturbance process of the test equation. The PP test t-statistic is close to -55.27 and p-value is close to 0. The result is consistent with the ADF test, thus we conclude that the de-seasoned 30-minute volatility time series is not an unit root.

Based on the above parametric model, we conclude that the TSRV time series in the period between August 13, 2002 and November 18, 2005 has periodic intra-day pattern and is a long memory process. We model these properties and forecast out-of-sample 30 minute volatility using two different approaches for each of the 30 Dow Jones composite stocks. The first approach is a rolling analysis on Semiparametric Fractional Autoregressive (SEMIFAR) model. The second approach is a rolling analysis on HAR-RV model, which is simply a linear regression on the 30-minute volatility estimates at previous 30-minute, day, week and month. As the 30-minute TSRV time series has similar pattern for all 30 stocks, in the next section we only demonstrate how to build forecasting model on the stock Alcoa Aluminum (ticker: AA).

Rolling analysis of a time series model is often used to assess the model's stability over time. As the volatility are changing, it is not reasonable to assume that the model's parameters are constant. A common technique to assess the constancy of a model's parameters is to compute parameter estimates over a rolling window of a fixed size through the sample. We use rolling analysis to backtest our forecasting model on historical volatility to evaluate stability and

predictive accuracy. Backtesting generally works in the following way. We split our a section of volatility series into the estimation sample and a prediction sample. The model is then fit using the estimation sample and h-step ahead predictions are made for the prediction sample. We can also call it out-of-sample forecasting. Since the volatility for which the predictions are made are estimated by TSRV, h-step ahead prediction errors can be found. The estimation sample is then rolled ahead one day, and the estimation and prediction exercise is repeated until it is not possible to make any more h-step predictions. The statistical properties of the collection of h-step ahead prediction errors are then summarized using mean squared error to evaluate the adequacy of our statistical model.

Our motivation is to forecast intra-day 30-minute volatility. In this study, we arbitrarily pick April 2005 as our forecasting month. Only one to three months volatility time series is needed to build our estimation models on which the intra-day volatility forecasts are based. Rolling analysis on SEMIFAR model and HAR-RV model is specified in the following subsections.

4.1 SEMIFAR model

We use the 30-minute volatility starting from March 2005 to predict the 30-minute volatility in April 2005. The rolling window starts from 3/1/2005 to 3/31/2005 with window size of 22 days containing 286 observations. We estimate this training data with Semiparametric Fractional Autoregressive (SEMIFAR) model, forecast the 13 out-of-sample 30-minute volatility one day ahead, and roll window with increment of one day with 13 observations, and repeat the process of estimation and forecasting. Here we only illustrate the stylized facts and estimation model for the 30-minute TSRV series in the first rolling window.

Figure 7 shows the autocorrelation function (ACF) of 30-minute volatility with lag 1300 and 91 for stock AA. Figure 8 is the time series plot of our in-sample 30-minute volatility in March 2005. Figure 9 shows the density distribution of the 30-minute volatility and that of its logarithm form, overlaying against the normal distribution. The distribution is skewed to the right, with thinner left tail and fatter right tail. It is consistent with the stylized facts specifying that the volatility is skewed to the right. We transform the volatility estimates into logarithm form in later presentation and analysis.

Figure 10 shows the ACF and PACF plot on logarithm 30-minute volatility with lag 200. The autocorrelation of the 30-minute volatility decay very slowly and are highly persistent indicating the long memory feature exists in the 30-minute volatility series. As expected, the ADF test and PP test applied on the residuals from model (4.1) and (4.2) both reject the random walk hypothesis, indicating that the long memory property does exist. In residuals from model (4.1), the t-statistics on the ADF test is -3.811 with p-value 0.003169, and the t-statistics on PP test is -11.73 with p-value close to 0. In residuals from model (4.2), the t-statistics on the ADF test is -3.833 with p-value 0.002941, and the t-statistics on PP test is -11.78 with p-value close to 0.

The logarithm volatility time series has the characteristics of periodic pattern and long memory. We model these properties separately in two steps. First step, we use a linear regression on 12 indicator variables or polynomial time variables to remove periodic pattern; Second step, we employ SEMIFAR to model the long memory property of the 30-minute volatility based on the residuals extracted from the first step.

In the first rolling window, model (4.1) is built on volatility estimates in the period 3/1/2005 - 3/31/2005. The estimates of the parameters are given in table VII. We set the the last 30-minute volatility as bench mark. The values of coefficients for the 12 indicator variables is highest at the first 30-minute trading, and then the values go down until 11:30am, after which the volatility in the rest of the day do not significantly differ from that in the closing 30 minutes, the bench mark value. The intercept α_0 is significant with p-values close to zero. The indicator variables β_1-β_3 are statistically significant at level 0.05, indicating that the 30-minute volatility at 9:30-11:30 are statistically different from the volatility at the closing 30 minutes. The set of coefficient estimates successfully model the inversed J-shape pattern of intra-day volatility. The adjusted R^2 is 0.4298, indicating that about 43% of the variation in TSRV time series can be explained by intra-day periodicity.

Figure 11 is the residual plots after fitting the model (4.1). The autocorrelation of the residuals decay very slowly and are highly persistent as the long memory feature exists in the 30-minute volatility series. We then proceed to the next step of long memory investigation.

In the past twenty years, more applications have evolved using long memory processes, which lie halfway between stationary I(0) processes and the non-stationary I(1) processes. There is substantial evidence that long memory processes can provide a good description of many highly persistent time series.

Granger and Joyeux (1980) and Hosking (1981) showed that a long memory process can be modeled parametrically by extending an integrated process to a fractionally integrated process. A time series based on fractional integration is modeled as follows:

$$(1 - B)^d(y_t - \mu) = u_t \tag{4.6}$$

where y_t is a long memory process, B denotes the lag operator, d is the fractional integration or fractional differenced parameter, μ is the mean of y_t, and u_t is a stationary short memory disturbance with zero mean.

In practice, when a time series is highly persistent or appears to be non-stationary, we difference the time series once to achieve stationarity, in this case we let d=1. Otherwise We set d to be fractional to allow for long memory process. It is known that when $|d| > 1/2, y_t$ is non-stationary; when $0 < d < 1/2$, y_t is stationary and has long memory; when $-1/2 < d < 0$, y_t is stationary and has short memory.

Obtaining the estimate of the long memory parameter d is of our interest. There are many methods to test for long memory and estimate d, such as R/S statistic by Hurst (1951), GPH test by Geweke and Porter-Hudak (1983), Periodogram Method, and Whittle's Method etc. More flexible models include FARIMA model and SEMIFAR model etc. FARIMA model is capable of modeling both the long memory and the short memory dynamics in a stationary time series.

When we roll our training window, the volatility series in the window could contain a trend and be non-stationary. FARIMA model is not capable of modeling non-stationary time series, its function is only limited to stationary long memory time series. A more flexible model, SEMIFAR method, is able to estimate the long memory parameter d, and the number of integer difference m to make 13-step-ahead (1-day-ahead) forecast. SEMIFAR model allow for a possible deterministic trend in a time series, in addition to a stochastic trend, long memory and short memory components. Beran, Feng and Ocker (1998), Beran and Ocker(1999), and Beran and Ocker(2001) propose the Semiparametric Fractional Autoregressive (SEMIFAR) model. The SEMIFAR model is based on the following extension to FARIMA(p,d,0) model, and it is written as:

$$\phi(B)(1-B)^{\delta}[(1-B)^{m}y_t - g(i_t)] = \varepsilon_t \tag{4.7}$$

where $g(i_t)$ is a smooth trend function on [0,1], with $i_t = t/T$. SEMIFAR model is estimated based on a nonparametric kernel estimate of $g(i_t)$. We allow $0 \leq p \leq 2$, and select the model with p=0 based on the minimum BIC. Refer to Beran, Feng, and Ocker(1998) for details of the algorithm.

Here we only show the result of the SEMIFAR model fit for the first rolling window. The result of SEMIFAR model is shown in Figure 12. The top left panel is the residual series from linear model (4.1); The top right panel is the nonparametric smoothed trend; The bottom left panel is the fitted values after fitting SEMIFAR model, and it has similar pattern as the plot in top left panel; The bottom right panel is the residuals extracted from SEMIFAR model.

The descriptive statistics of the residuals is reported in Table VIII. The mean of the residuals are close to zero; The variance of the residuals is 0.31; the distribution is slightly skewed to the left with skewness -0.33; the kurtosis is 0.59. In Figure 13, the top left panel is the residual plot; The top right panel is the ACF of the residual; The bottom left panel is the PACF of the residual. As very few ACF and PACF bars stick out of the 95 % confidence interval boundary, we believe SEMIFAR model is adequate; The bottom right panel is the QQ plot of the residual, which shows the residuals close to normal. Ljung-Box test for the residuals shows that the test statistics is 157.9891 with p-value 0.0479, with lag 130, equivalent to 10 days. This is shows that there is marginally no autocorrelation in the residual time series.

In order to forecast the volatility in the period 4/1/2005 - 4/29/2005 for 21 days, we roll the window 20 times with increment of 13 volatility estimates for one day. The estimated $\hat{m}$ and $\hat{d}$ are shown in table IX. When $\hat{m}$=0, the volatility time series in that specific window is stationary, and the $\hat{d}$ is in the range of 0.24-0.41. When $\hat{m}$=1, the volatility series in a specific window is not stationary, which does not occurred in the 20 rolling windows in this case.

In summary, the time series model for 30-minute TSRV series is additive model of periodic and long memory pattern as follows.

$$Y_t = \alpha_0 + \sum_{i=1}^{12} \beta_i d_{it} + \varepsilon_t, \quad d_{it} = \begin{cases} 1 & t\%13 = i \\ 0 & \text{others} \end{cases} \tag{4.8}$$

$$\phi(B)(1-B)^{\delta}[(1-B)^{m}\varepsilon_t - g(i_t)] = a_t \quad a_t \sim i.i.d.\ \ N(0, \sigma_a) \tag{4.9}$$

We make a 13-step-ahead forecasting for 30-minute volatility in each rolling window. The 13-step-ahead volatility is the sum of fitted value of (4.8) and 13-step-ahead prediction from (4.9). Figure 14 shows the rolling forecasting plot for 30-minute volatility based on the SEMIFAR model compared with the TSRV estimates in the period 4/1/2005 - 4/29/2005. The forecasting mean square error is 0.0009.

As the 30-minute volatility displays U-shape or inverse J-shape pattern, an alternative way to model the periodic pattern is regression on polynomial time variables, as in (4.2)

In table X, the regression (4.2) from the first rolling window shows that the intercept, linear, quadratic, cubic time variables are all significant with p-values close to zero. The adjusted R^2 is 0.4202. It indicates that about 42% of the variation in TSRV time series can be explained by daily periodicity. Figure 15 is the residual plots after fitting the model (4.10). The ACF plot on the top right panel shows that the autocorrelation of the residuals decay very slowly and are highly persistent. We then again use SEMIFAR to model long memory property.

The result of the SEMIFAR model fitting for the first window is shown in Figure 16. The top left panel is the time series plot of the residual from (4.2); The top right panel is the nonparametric smoothed trend of 30-minute volatility; The bottom left panel is the fitted values after fitting SEMIFAR model, and it has similar pattern as the original time series plot; The bottom right panel is the residual extracted from SEMIFAR model fit.

The descriptive statistics of the residuals a_t is reported in table XI. The mean of the residuals is 0.00048; The standard deviation of the residuals is 0.318; the distribution is slightly skewed to the left with skewness -0.3; the kurtosis is 0.582. In Figure 17, the top left panel is the residual

plot; The top right panel is the ACF of the residual; The bottom left panel is the PACF of the residual. As very few ACF and PACF bars stick out of the 95% confidence interval boundary, we believe the long memory feature has been captured by SEMIFAR model; The bottom right panel is the QQ plot of the residual. Ljung-Box test with lag 130 for the residuals shows that the test statistics is 147.8296 with p-value 0.1357, indicating that there is no autocorrelation in the residual time series.

we roll the window 20 times with increment of 13 estimates to make one-day-ahead 30-minute volatility forecast in the period 4/1/2005 - 4/29/2005 for 21 days, The estimated $\hat{m}$ and $\hat{d}$ are shown in table XII. All $\hat{m}$=0, the volatility time series in all rolling window are stationary, and the $\hat{d}$ is in the range of 0.23-0.39.

The alternative time series model for 30-minute TSRV series captures periodical long memory pattern, is formulated as

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \varepsilon_t \tag{4.10}$$

$$\phi(B)(1-B)^{\delta}[(1-B)^m \varepsilon_t - g(i_t)] = a_t \quad a_t \sim i.i.d.\ \ N(0, \sigma_a) \tag{4.11}$$

The forecasting root mean square error is 0.000877. Figure 18 shows the rolling forecasting plot for 30-minute volatility based on the polynomial time regression and SEMIFAR model compared against the TSRV estimates in the period 4/1/2005 - 4/29/2005.

4.2 HAR-RV model

A simpler model than the fractionally integrated long-memory form is Heterogeneous Autoregressive model of the Realized Volatility (HAR-RV) proposed by Corsi (2003). The HAR-RV model is an AR-type model in the realized volatility considering volatilities realized over different time horizons. In spite of the simplicity of its structure, simulation results seem to confirm that the HAR-RV model successfully achieves the purpose of reproducing the main empirical features of volatility (long memory, fat tail, self-similarity) in a very simple and parsimonious way. Although the HAR structure does not formally include long memory, the mixing of relatively few volatility components reproduces a remarkably slow volatility autocorrelation decay.

We assume that the 1-step-ahead forecast of 30-minute volatility is dependent on the current volatility, the same-time volatility one day ago, the same-time volatility one week (5 days) ago, and the same-time volatility one month (22 days) ago. To capture the periodic daily pattern, we include the dummy variables or polynomial time components to the HAR-RV model. The HAR-RV model is applied on 30-minute volatility with the scale of logarithm standard deviation. Again we use rolling analysis on HAR-RV model. The 30-minute volatility estimates in the period of 01/02/2005 - 3/31/2005 is our first estimation window on OLS regressions, which are performed according to the following two models.

$$log(RV_{t+1}) = \alpha_0 + \sum_{i=1}^{12} \alpha_i d_{it} + \beta_1 log(RV_t) + \beta_2 log(RV_{t-1D}) + \beta_3 log(RV_{t-5D}) + \beta_4 log(RV_{t-22D}) + \varepsilon_{t+1} \tag{4.12}$$

$$log(RV_{t+1}) = \alpha_0+\alpha_1 t+\alpha_2 t^2+\alpha_3 t^3+\beta_1 log(RV_t)+\beta_2 log(RV_{t-1D})+\beta_3 log(RV_{t-5D})+\beta_4 log(RV_{t-22D})+\varepsilon_{t+1} \tag{4.13}$$

where $d_{it} = \begin{cases} 1 & t\%13 = i \\ 0 & \text{others} \end{cases}$; t, t^2, t^3 are the linear, quadratic, and cubic time covariates respectively. t = T-22*13, ..., T.

The OLS regression performed on dummy variables are shown in table XIII. The 30-minute volatility one day ago, one week ago and one month ago do not have significant impact on volatility for tomorrow. OLS regression on polynomial time covariates are show in table XIV. Similarly the 30-minute volatility one day ago, one month ago and one week ago do not have impact on tomorrow's volatility at the 0.05 significant level. The residual time series, ACF, and PACF plot from (4.12) and (4.13) are shown in Figure 19 and Figure 20. We can see that residuals from OLS regression on dummy variables and polynomial time components are both autocorrelated in the ACF plot of Figure 19 and Figure 20. This is verified by Ljung-Box test on residuals. Table XIV shows that autocorrelation exist in residual from HAR-RV regression with dummy variables and polynomial time variables both have p-value close to 0.

Although HAR models (4.12) and (4.13) does not completely remove autocorrelation, we still keep these two models to forecast the out-of-sample 30-minute volatility in 4/1/2005 - 4/29/2005. The overlaying 30-minute volatility forecasting plots against the TSRV estimates in 4/1/2005 - 4/29/2005. are shown in Figure 21. The forecasting root mean square error is 0.0009 for model (4.12) and 0.000988 for model (4.13).

Up to now we have used rolling analysis on SEMIFAR model with periodic components based on dummy variables and polynomial time variables, and rolling analysis on HAR-RV

model on logarithm volatility with dummy variables and polynomial time variables. It seems that all four models are good at forecasting 30-minute volatility.

Figure 22 shows the overlaying 30-minute volatility forecasting by all four models in April, 2005. Both SEMIFAR model and HAR-RV model seem to be very successful at modeling and forecasting the 30-minute volatility.

4.3 Forecasting Comparison

We employ the SEMIFAR model with dummy and polynomial time variables, and HAR-RV model with dummy and polynomial time variables to each of the Dow Jones 30 stock. The 30-minute volatility forecasting from the four models for 4/1/2005 - 4/29/2005 are documented in Figure 23 to Figure 27. Table XVI shows the mean square forecasting errors in the period of 4/1/2005 - 4/29/2005. We can see that the performance of the four models are similar. We can conclude that all four models can be used to predict the 30-minute volatility time series having periodic long memory pattern.

CHAPTER 5

IMPACT OF QUARTERLY EARNING ANNOUNCEMENT TO INTRA-DAY VOLATILITY

This chapter examines the effect of the scheduled quarterly earning announcements on the intra-day volatility of stock prices. Many market participants believe that the announcement of a significant economic event, such as earning announcements, stock split, dividend announcement, employment report, consumer price index (CPI), or producer price index (PPI), will have a great impact on the abnormal rate of return and volatility of stock prices. We found out that the 30-minute volatility during the period of earning announcement is higher than that in the non-announcement period, and the volatility is significantly higher during the opening 30 minutes in the earning announcement day.

There are a handful of literatures about the association of volatility with earning announcements. Hillmer and Yu (1979) used a cumulative sum technique to measure the speed of adjustment of any "market behavior variable", and found out that there is a significant increase in stock return variance which lasted from three to seventeen hours with the effects sometimes beginning prior to the disclosure. They measured the adjustment interval for each of the five specific events, including two earning reports and three defense contract announcements. Their test statistic is based on parametric assumption that is appropriate for daily or longer sampling intervals but is restrictive in applying intraday data. They assume that the consecutive price changes are independently and identically distributed. Patell and Wolfson (1984), Jennings

and Starks (1985) studied the effect of earnings and dividend announcements on the intra-day behavior of stock prices, using non-parametric procedures on data from Chicago Board Options Exchange / Berkeley Options transactions Database. They also found out that the earnings announcements are associated with large increases in the variance of intraday returns, which persist for up to four hours after the disclosure, and the significant variance increase extend into the following day. Acker (2002) examined post-earnings announcement volatility using implied standard deviations (ISDs) derived from option prices to construct daily volatility within the announcement period. The author found that ISDs tend to rise before the earning announcement date and fall after it.

Unlike their studies of earning and dividend announcement on the stock return volatility based on low frequency data, we estimate the intra-day volatility based on high frequency data. We applied TSRV to estimate 30-minute volatility on tick-by-tick data, and compared the volatility in earning announcement period with that in non-announcement period.

Our empirical analysis is based on the tick-by-tick stock price of the 30 companies included in Dow Jones Industrial Average index. In this study, We defined the earning announcement period as the 10 trading days before the actual announcement date, the announcement day, through the 10 trading days after. All the rest of the trading days will fall into the non-announcement period. For each of the 30 stock, we have 13 estimates for 30-minute volatility each day. As there are 21 days in the quarterly earning announcement period, therefore we have 173 volatility estimates in total in one quarter. Stocks normally have different actual quarterly earning announcement dates, thus different earning announcement periods both in

stocks and in year dimension. We align the earning announcement periods based on time line relative to the actual quarterly earning announcement date. So that we can take the average and the median of the 173 volatility estimates across all available, from 19 to 20, quarters for one stock. Similarly we take the average and the median of the 13 estimates across all the non-announcement days, around 800 days for 5 years depending on a given stock, we can get the 13 average or median hourly volatility for one stock during non-announcement period. The volatility patterns for all 30 stocks look similar, so we integrate the volatility time series in the stock dimension. We take the average of the average and median 30-minute volatility across the 30 Dow Jones stocks for both earning announcement and non-announcement period. We have 173 mean or median volatility estimates for earning announcement period and 13 mean or median volatility estimates for non-announcement period. The 13 mean or median volatility estimates for non-announcement period are then replicated to 21 days in order to be compared with the 30-minute volatility for announcement period.

Figure 28 and Figure 30 are the mean and median 30-minute volatility in quarterly earning announcement period vs. in non-announcement period, respectively. The horizontal axis is the number of days relative to the actual quarterly earning announcement day, which is label as 0; the days before earning announcement date is negative, and the days after that is positive. The vertical axis is the mean and median 30-minute volatility respectively. The mean and median 30-minute volatility is higher in earning announcement day and the following day, the effect is especially pronounced in the beginning trading hours of the day. The mean and median 30-minute volatility in non-announcement days of the earning announcement period are similar

to that in non-announcement period. Figure 29 and 31 are the the zoomed-in view of the mean and median 30-minute volatility in quarterly earning announcement period versus non-announcement period. In table XVII, to compare the mean 30-minute volatility in the earning announcement period versus non-announcement period, the two sample t-test indicates that the t-statistic is 3.6633, and the p-value is 0.0003; The paired t-test has the t-statistics 19.7495 with p-value close to 0. To compare the median 30-minute volatility in the earning announcement period versus non-announcement period, the two sample t-test indicates that the t-statistic is 3.4825, and the p-value is 0.0005; The paired t-test has the t-statistics 20.3820 with p-value close to 0. Therefore the mean and median 30-minute volatility is significantly higher in quarterly earning announcement period than in non-announcement period.

The former literatures on the impact of earning announcements on stock volatility is consistent with our findings from estimating 30-minute volatility by TSRV from tick-by-tick trading data.

CHAPTER 6

CONCLUSION

Theoretically, the non-parametric TSRV estimator is an unbiased and consistent estimator for the integrated volatility. What differentiates TSRV estimator from other volatility estimators lies in that fact that it makes use of all of the observations from high frequency data and at the same time corrects the market mircrostructure noise. We empirically investigate the dynamics of the 30-minute TSRV estimator on each of the 30 stocks in Dow Jones Industrial Average Index. We found out that the distribution of the logarithm 30-minute TSRV time series demonstrate some stylized facts that usually observed in volatility series. The sylized facts include volatility clustering, skewed to the right, leptocurtic; It has U-shape or inverse J-shape pattern in the daily time frame, volatility in the opening 30 minutes is significantly higher than closing 30 minute. The ACF of logarithm TSRV indicates periodic intra-day patterns. TSRV series does not have unit root, thus it is considered as a stationary long memory process. Rolling analysis on SEMIFAR model and HAR-RV model are both successful in predicting 30-minute volatility. We have also shown that the event of the quarterly earning announcement has a significant impact on the stock intraday volatility, especially at the beginning of the earning announcement day and the following day.

Future work will further study the performance of volatility estimators on pricing options and options trading strategies, and empirically compare our forecasting model with. Also we can investigate how TSRV performs when the volatility signature plot has a positive slope.

APPENDICES

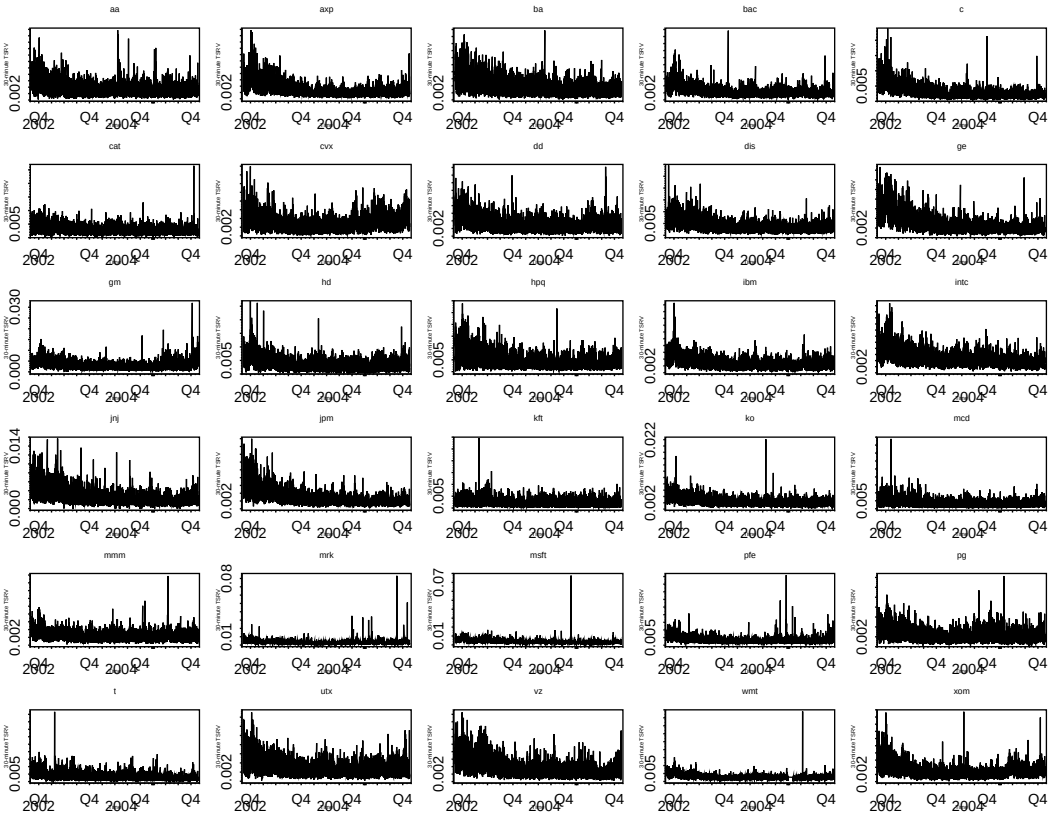

Figure 1. The time series plot of 30-minute TSRV from August 13, 2002 through November 18, 2005

Figure 2. Zoom in: Intraday 30-minute TSRV for 20 days

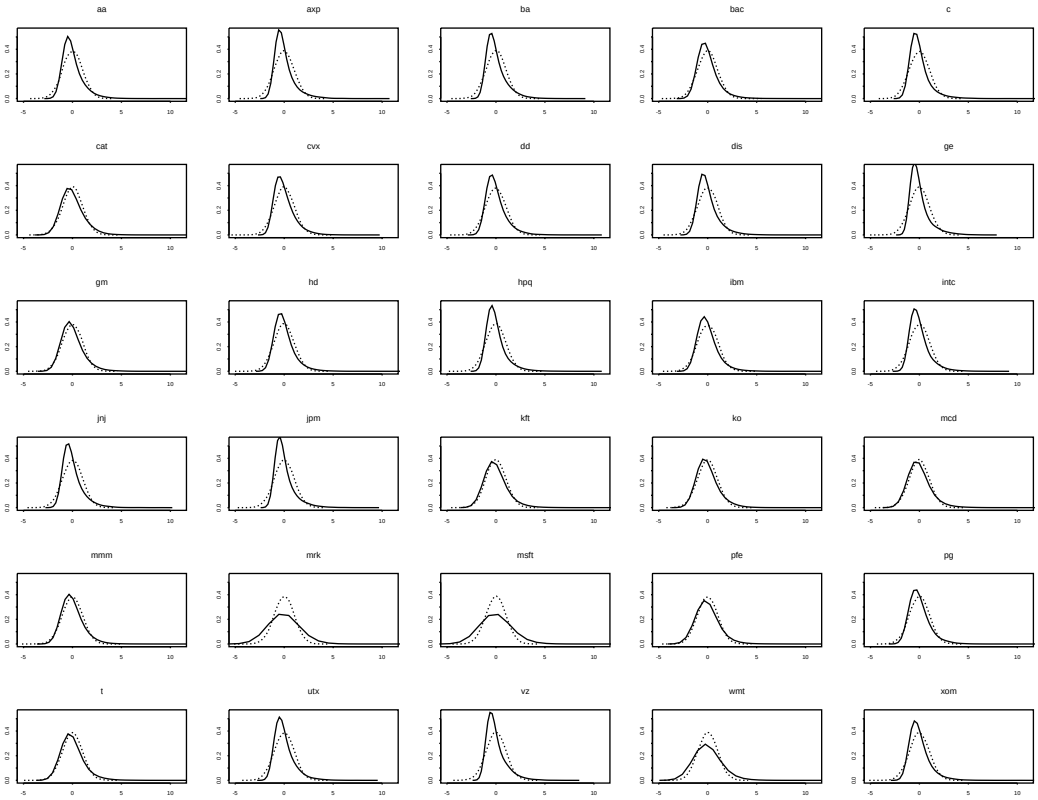

Figure 3. The distribution of 30-minute volatility from August 13, 2002 through November 18, 2005

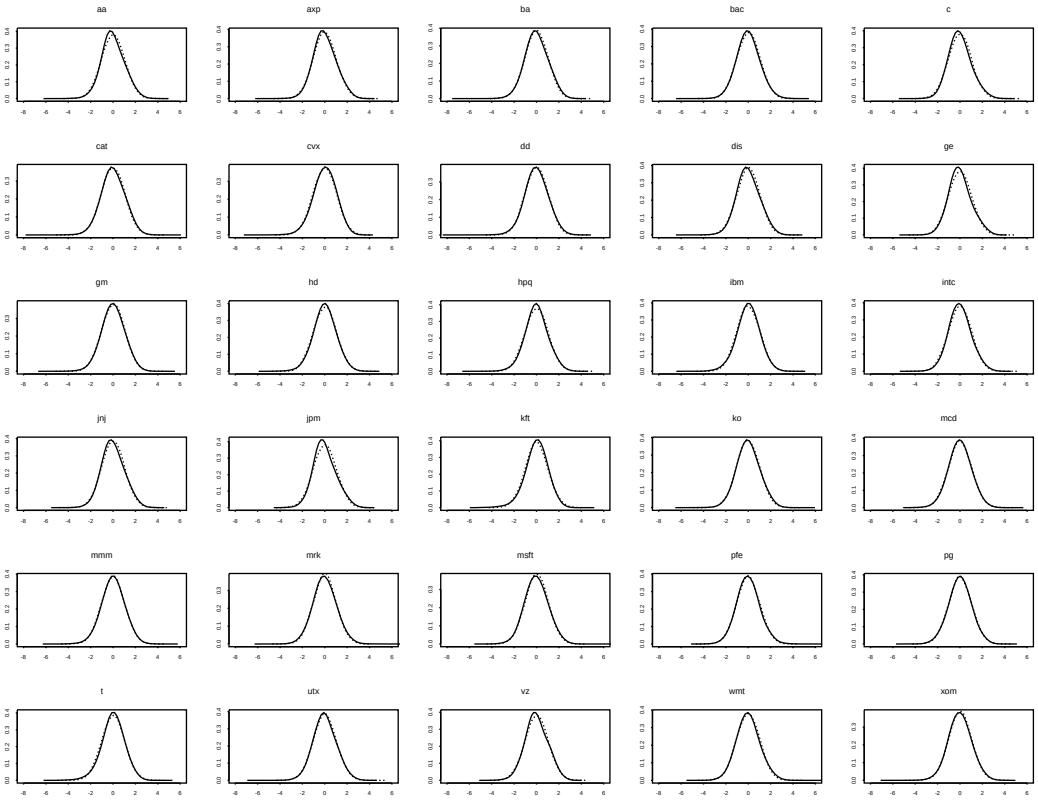

Figure 4. The distribution of logarithm of 30-minute volatility from August 13, 2002 through November 18, 2005

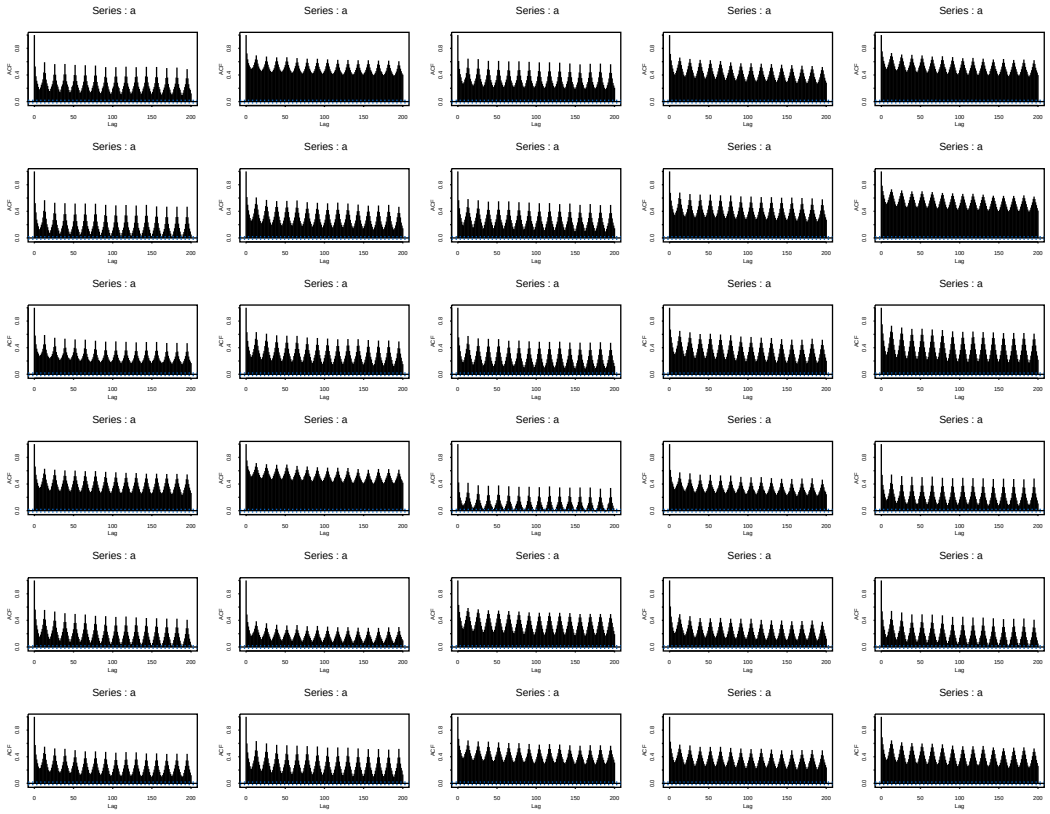

Figure 5. The autocorrelation function (ACF) of 30-minute volatility with lag 200

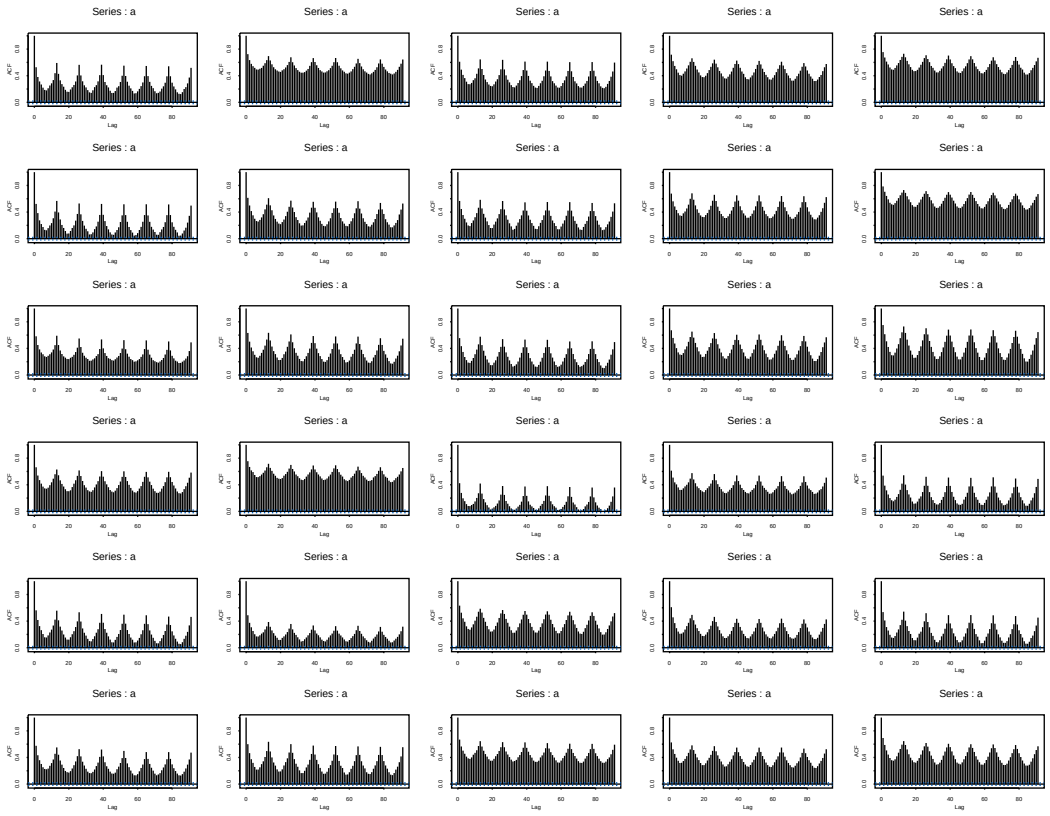

Figure 6. The autocorrelation function (ACF) of 30-minute volatility with lag 91, which is equivalent to 7 days.

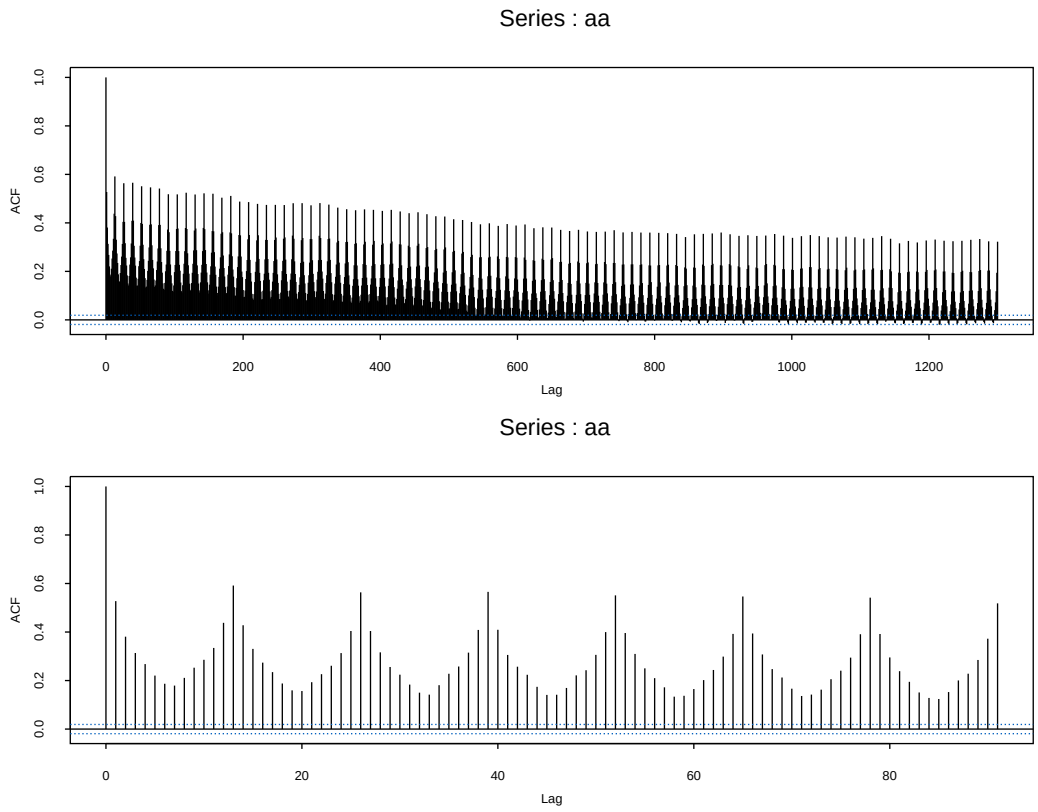

Figure 7. The autocorrelation function (ACF) of 30-minute (Ticker:AA)

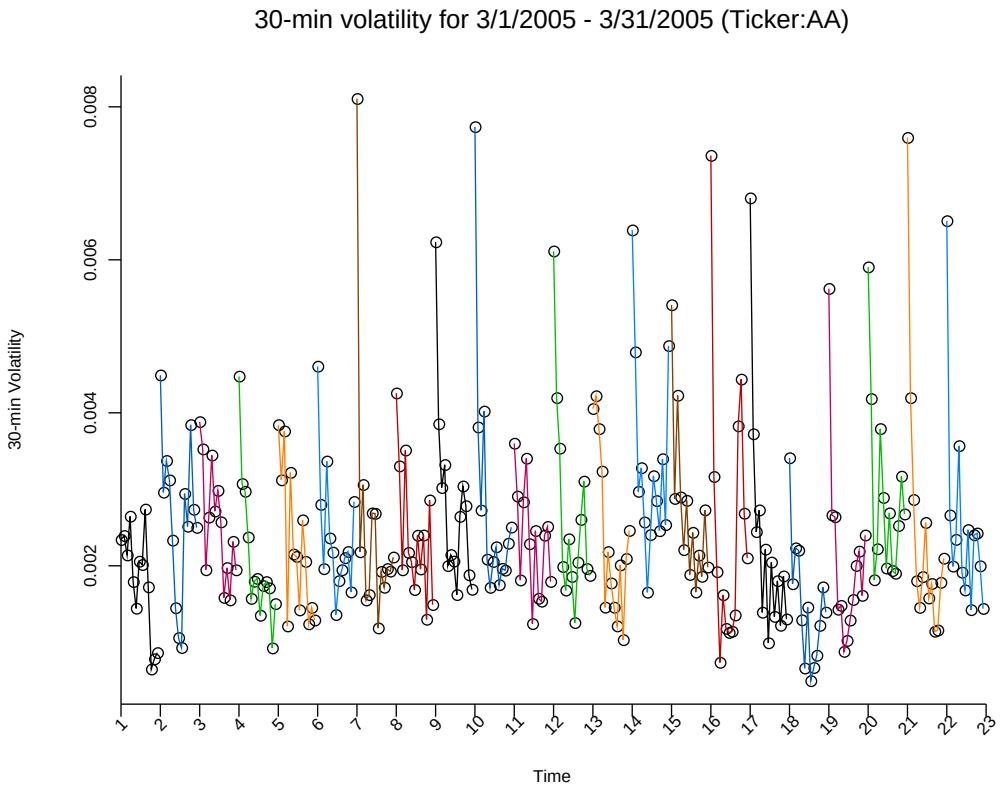
30-min volatility for 3/1/2005 - 3/31/2005 (Ticker:AA)
30-min Volatility
0.002
0.004
0.006
0.008
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Time

Figure 8.

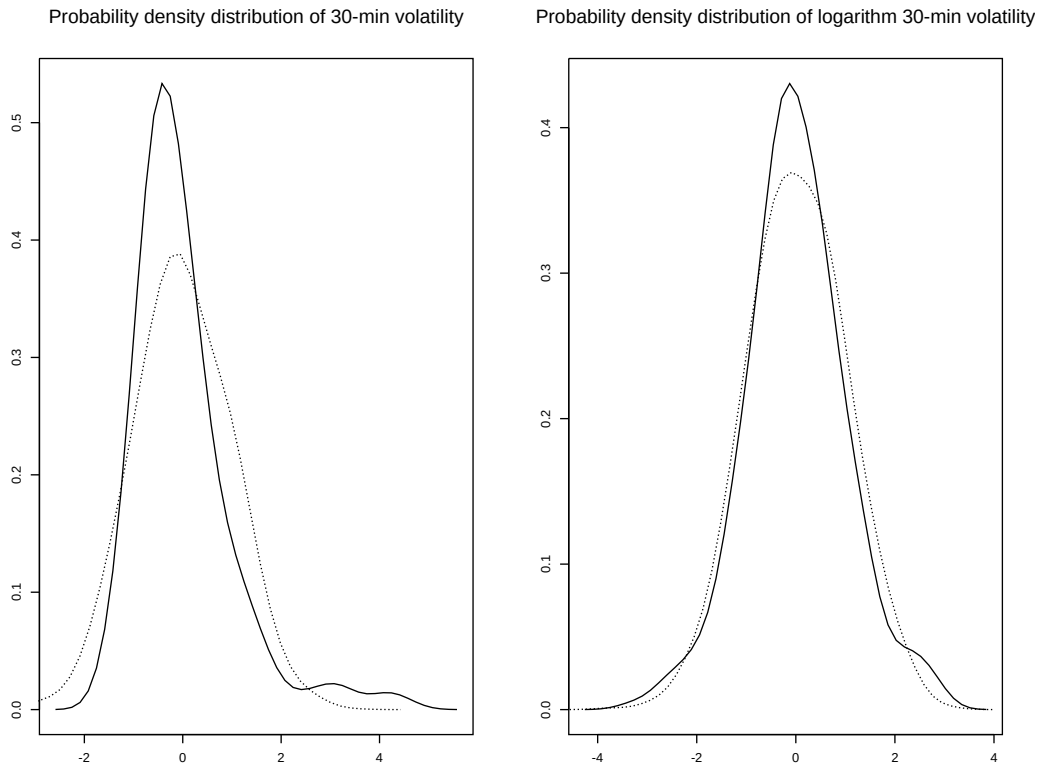

Figure 9. Density of log 30-minute volatility 3/1/2005-3/31/2005 (Ticker:AA)

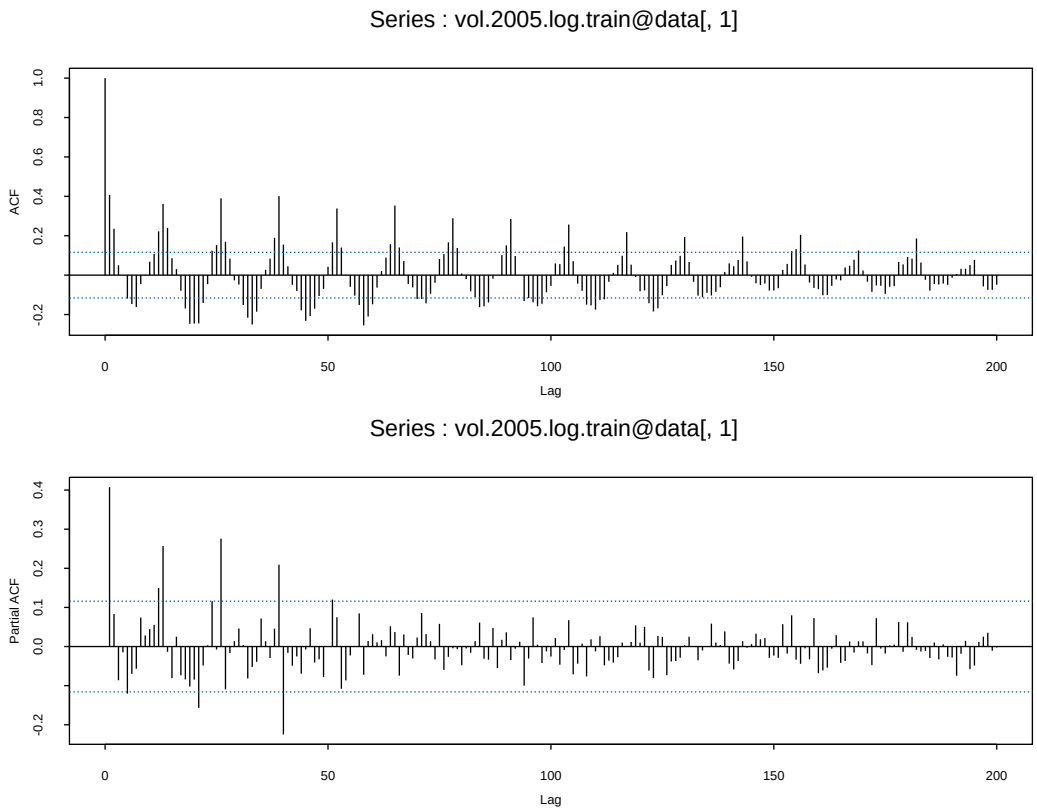

Figure 10. Top panel: ACF plot of mean 30-minute volatility.
Bottom panel: PACF plot of mean 30-minute volatility.

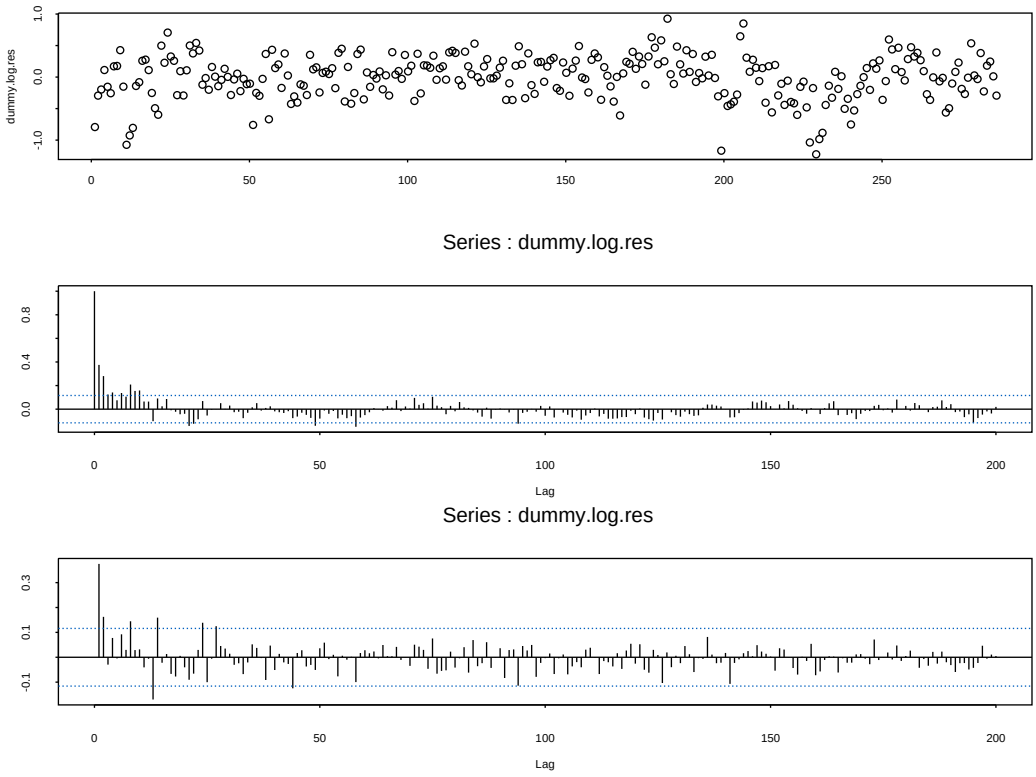

Figure 11. Residual plot after regression on dummy variables (Ticker:AA)
Top panel: time series plot of residuals
Middle panel: ACF plot of residuals
Bottom panel: PACF plot of residuals

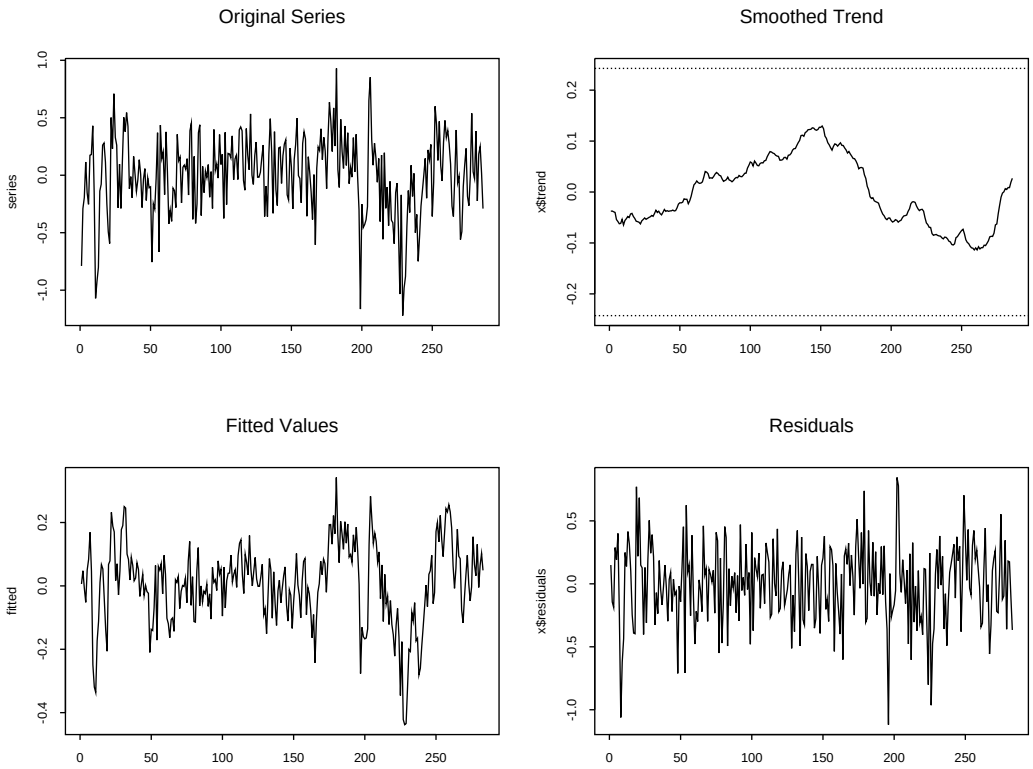

Figure 12. Dummy-SEMIFAR model plot (Ticker:AA)

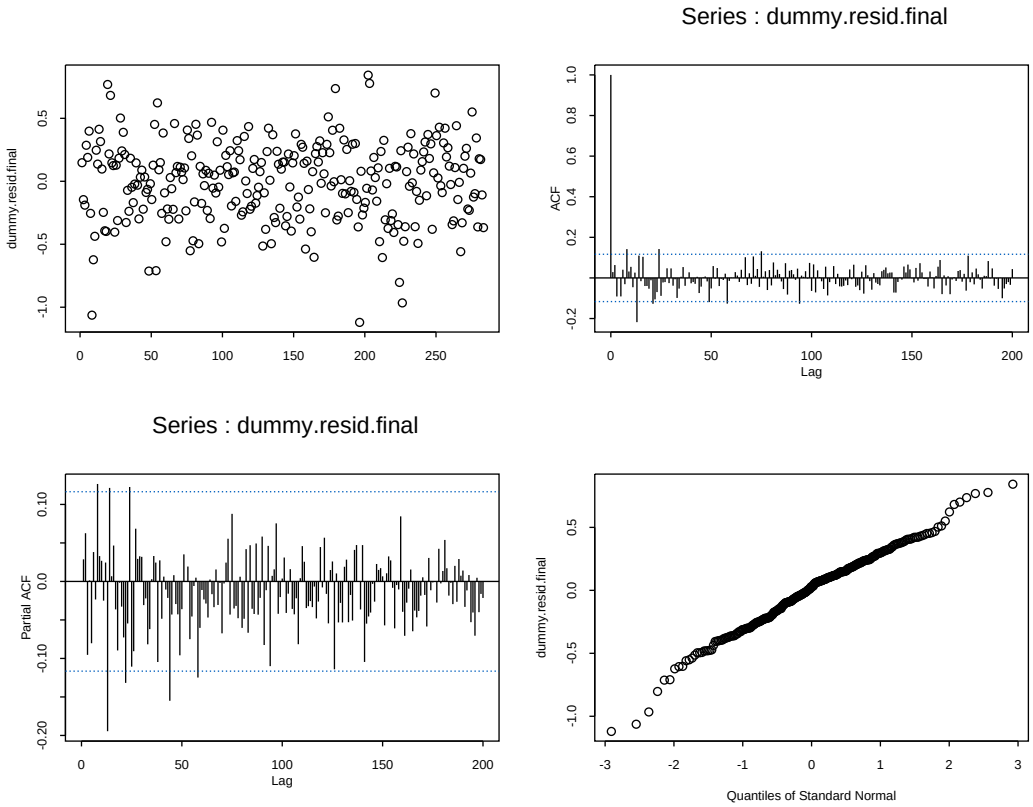

Figure 13. Dummy-SEMIFAR Residual plot (Ticker:AA)
Top left panel: Scatter plot of residuals
Top right panel: ACF plot of residuals
Bottom left panel: PACF plot of residuals
Bottom right panel: QQ plot of residuals

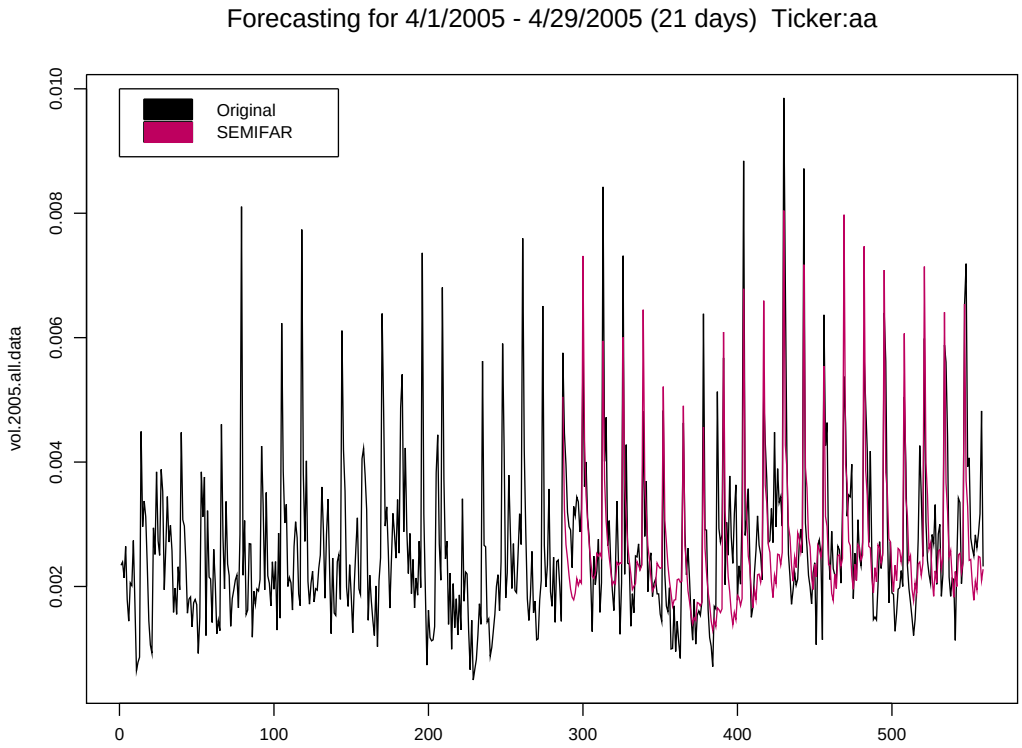

Figure 14. Forecasting based on Dummy-SEMIFAR model (Ticker:AA)

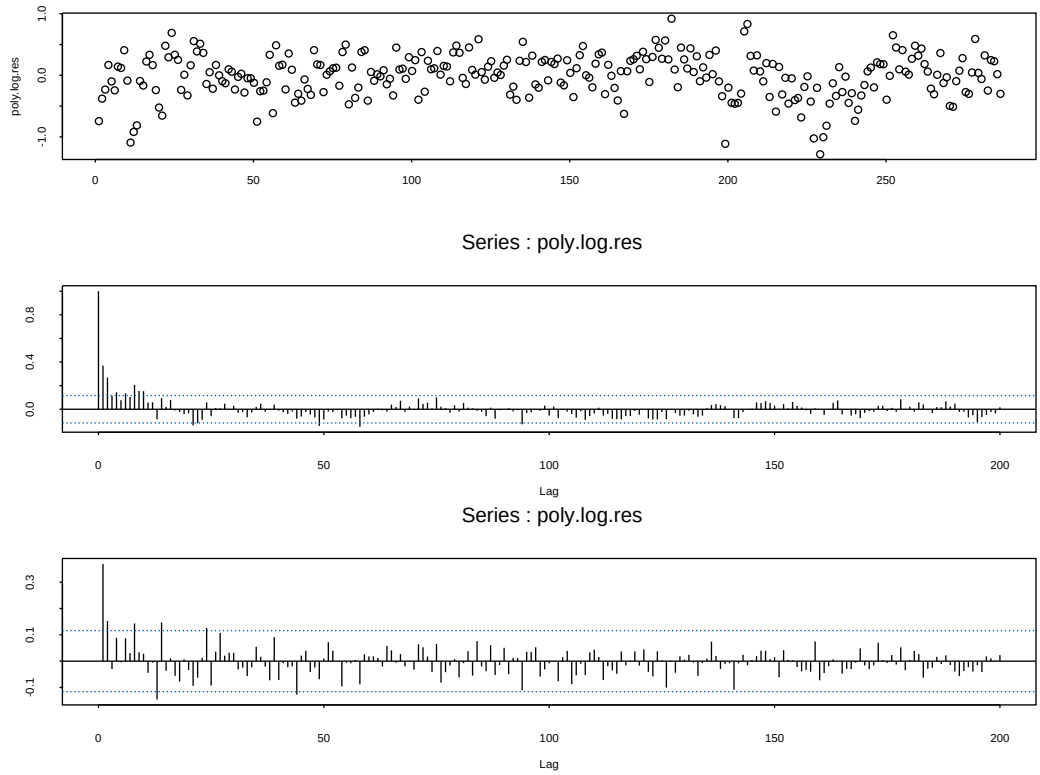

Figure 15. Residual plot from regression on polynomial time variables (Ticker:AA)
Top panel: Scatter plot of residuals
Middle panel: ACF plot of residuals
Bottom panel: PACF plot of residuals

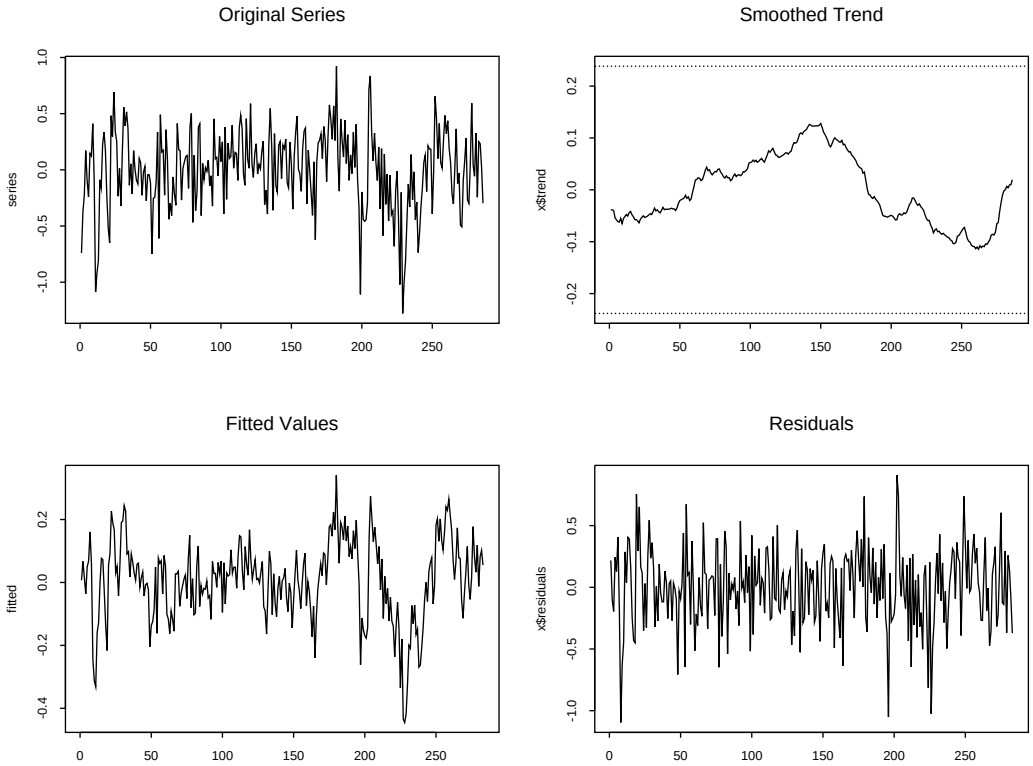

Figure 16. Polynomial-SEMIFAR model plot

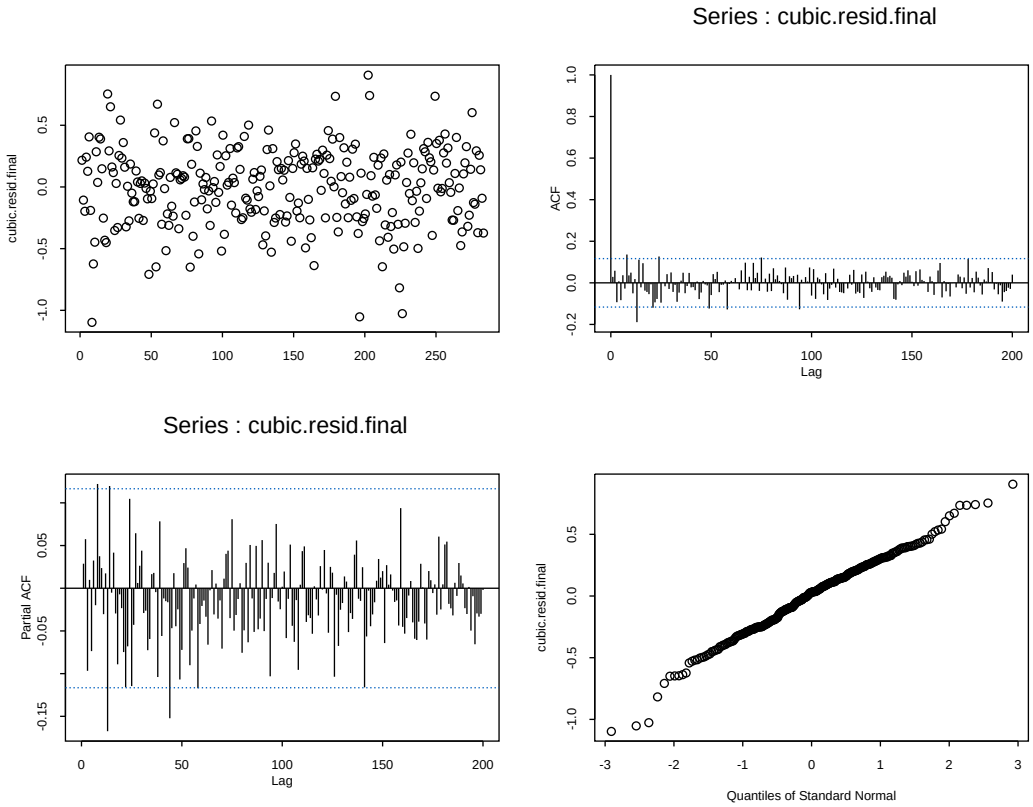

Figure 17. Polynomial-SEMIFAR Residual plot
Top left panel: Scatter plot of residuals
Top right panel: ACF plot of residuals
Bottom left panel: PACF plot of residuals
Bottom right panel: QQ plot of residuals

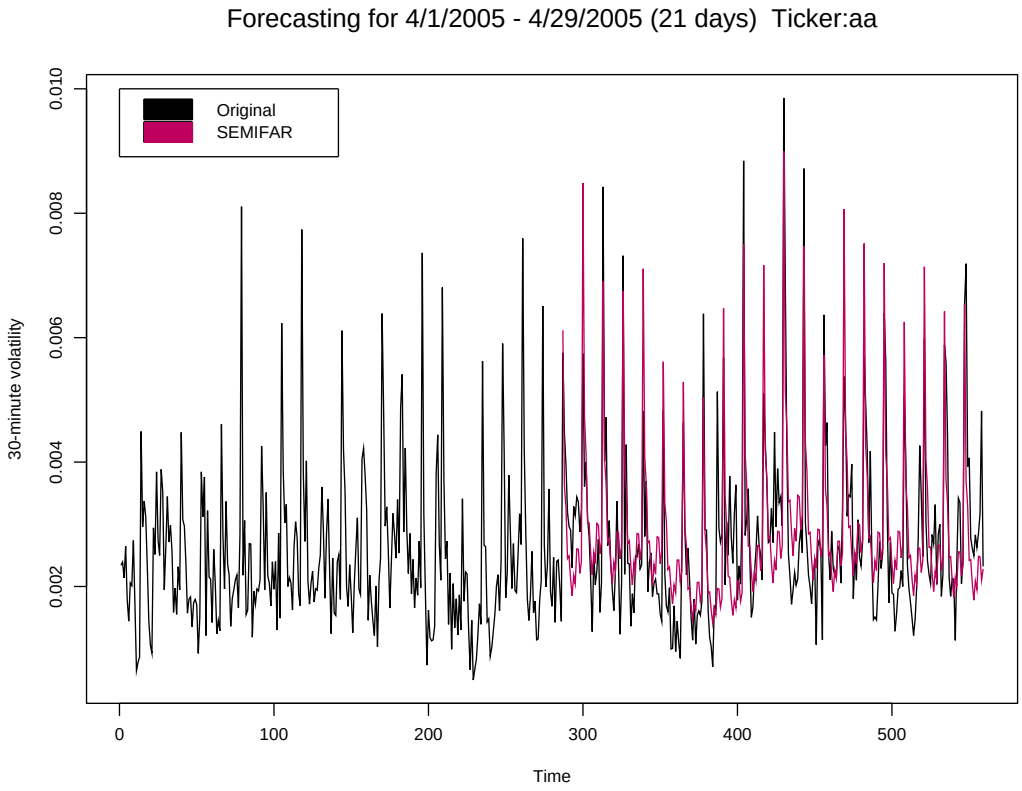

Figure 18. Forecasting based on Polynomial-SEMIFAR model (Ticker:AA)

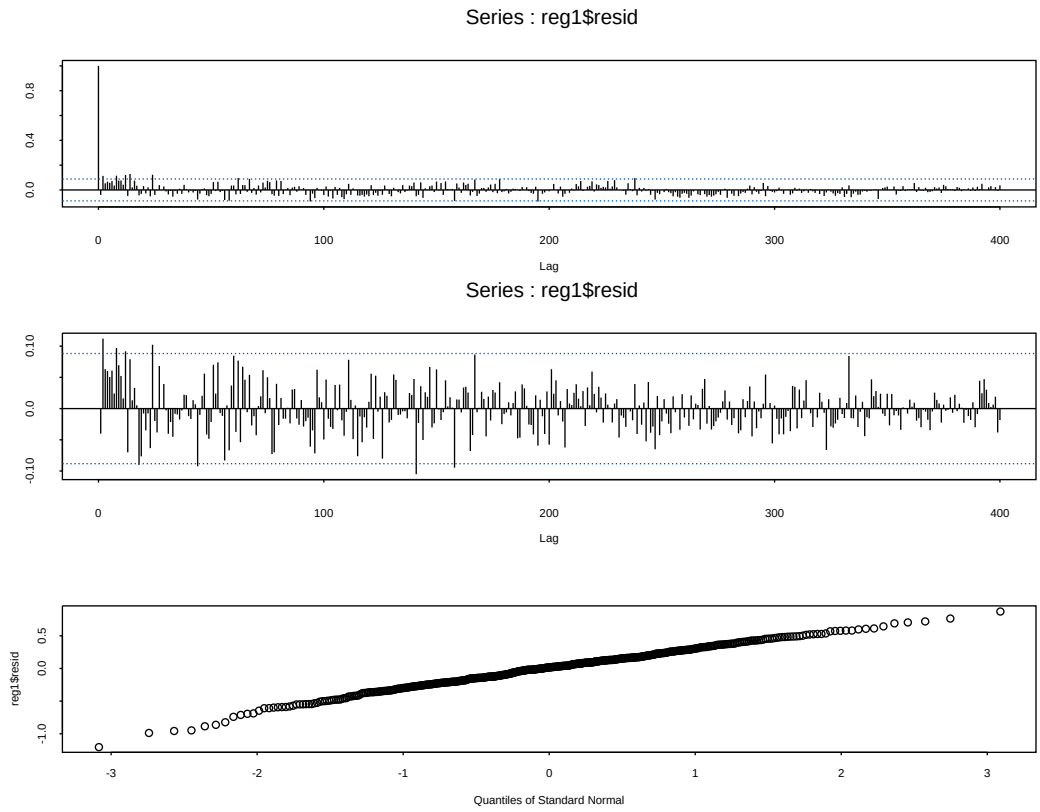

Figure 19. Residual plot from dummy-HAR-RV model Ticker:AA)

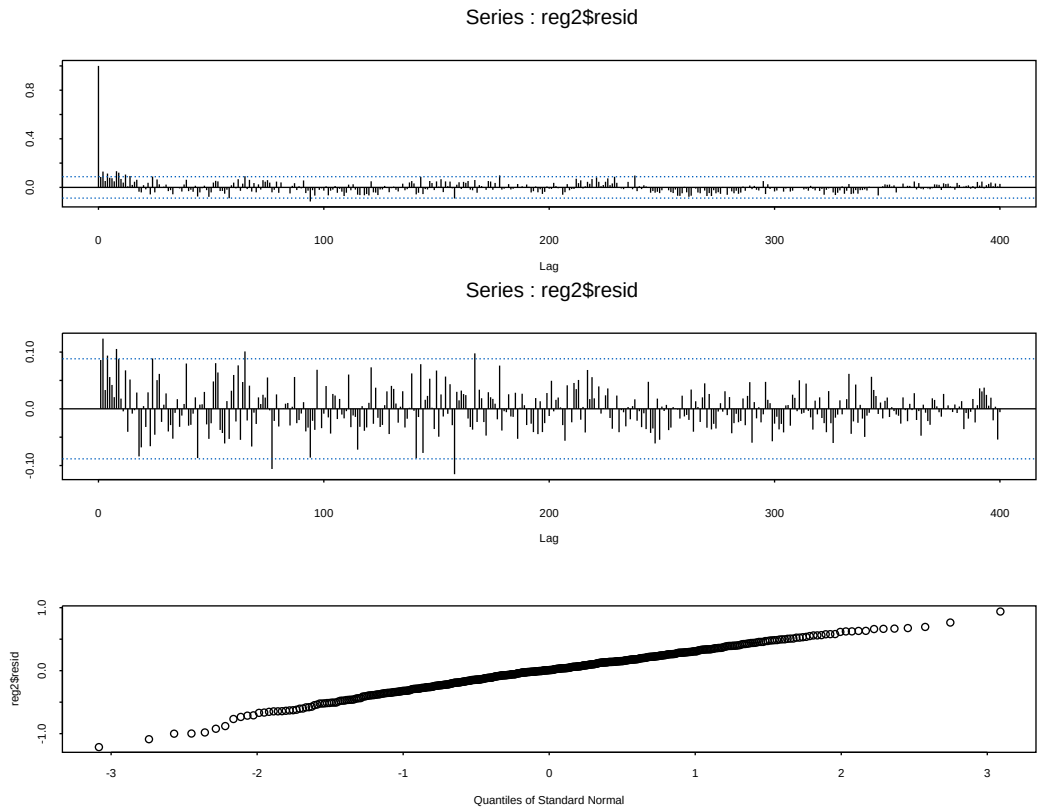

Figure 20. Residual plot from polynomial-HAR-RV model (Ticker:AA)

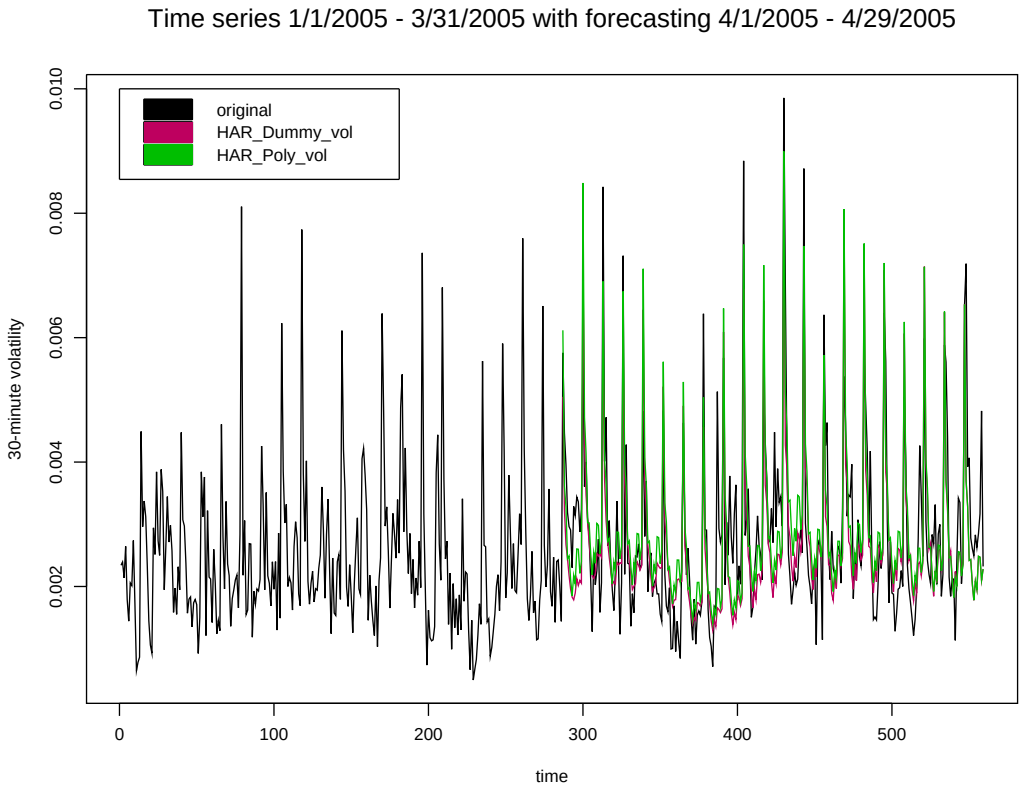

Figure 21. Forecasting plot based on HAR-RV models (Ticker:AA)

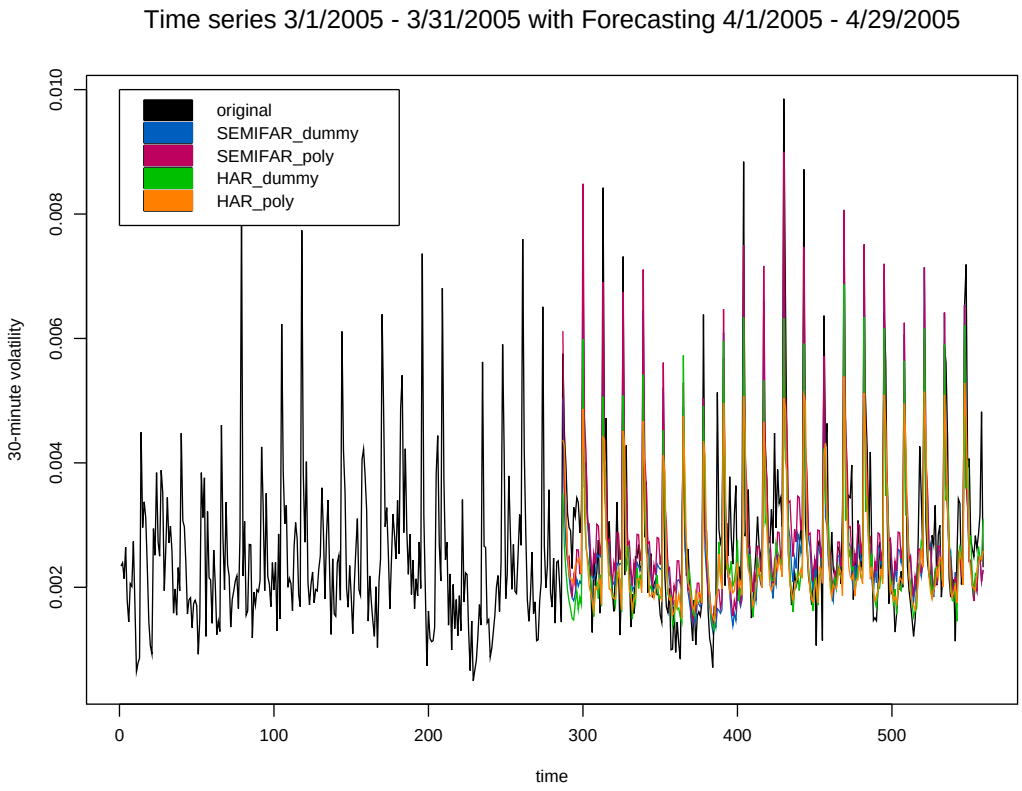

Figure 22. Forecasting plot

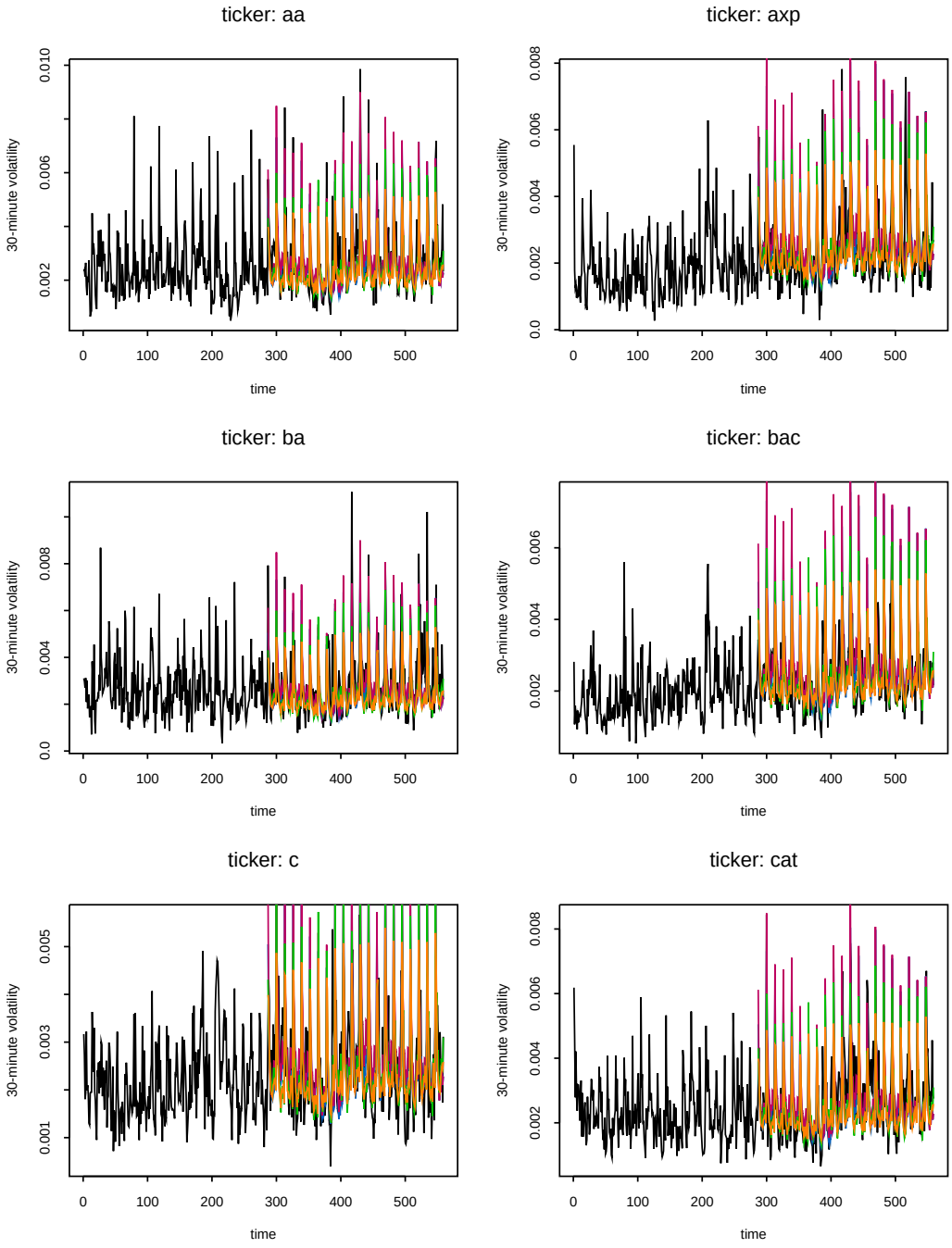

Figure 23. Forecasting plot

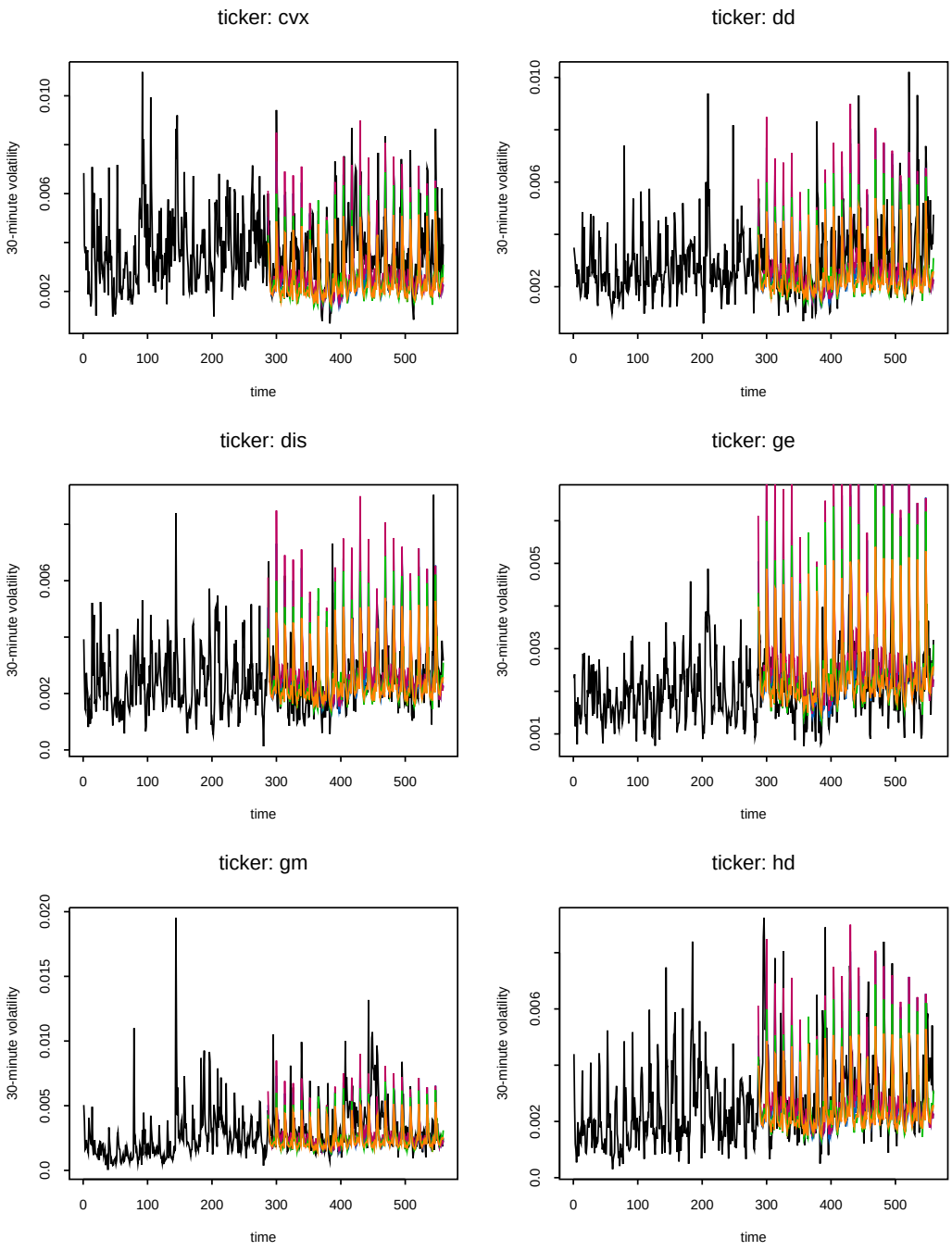

Figure 24. Forecasting plot

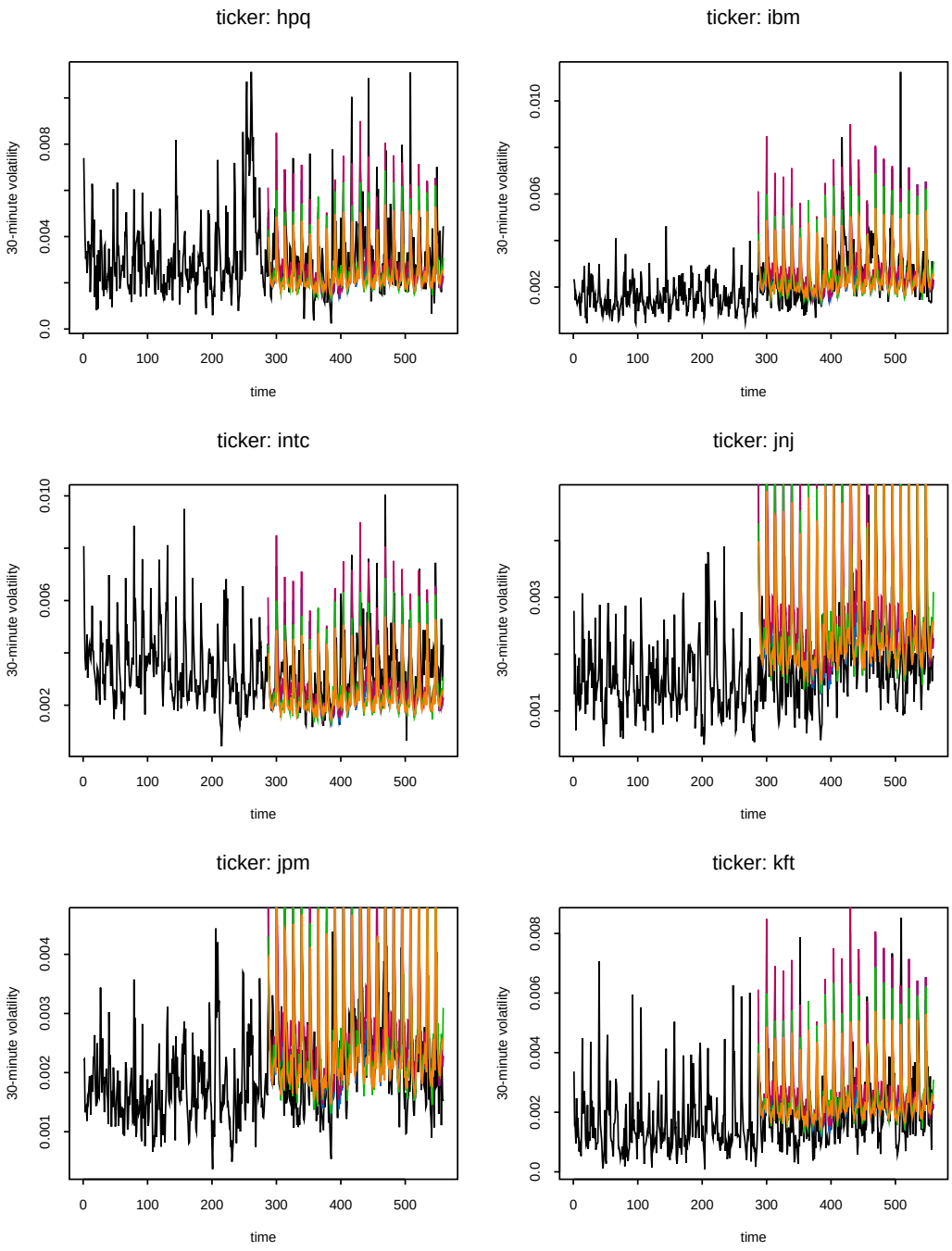

Figure 25. Forecasting plot

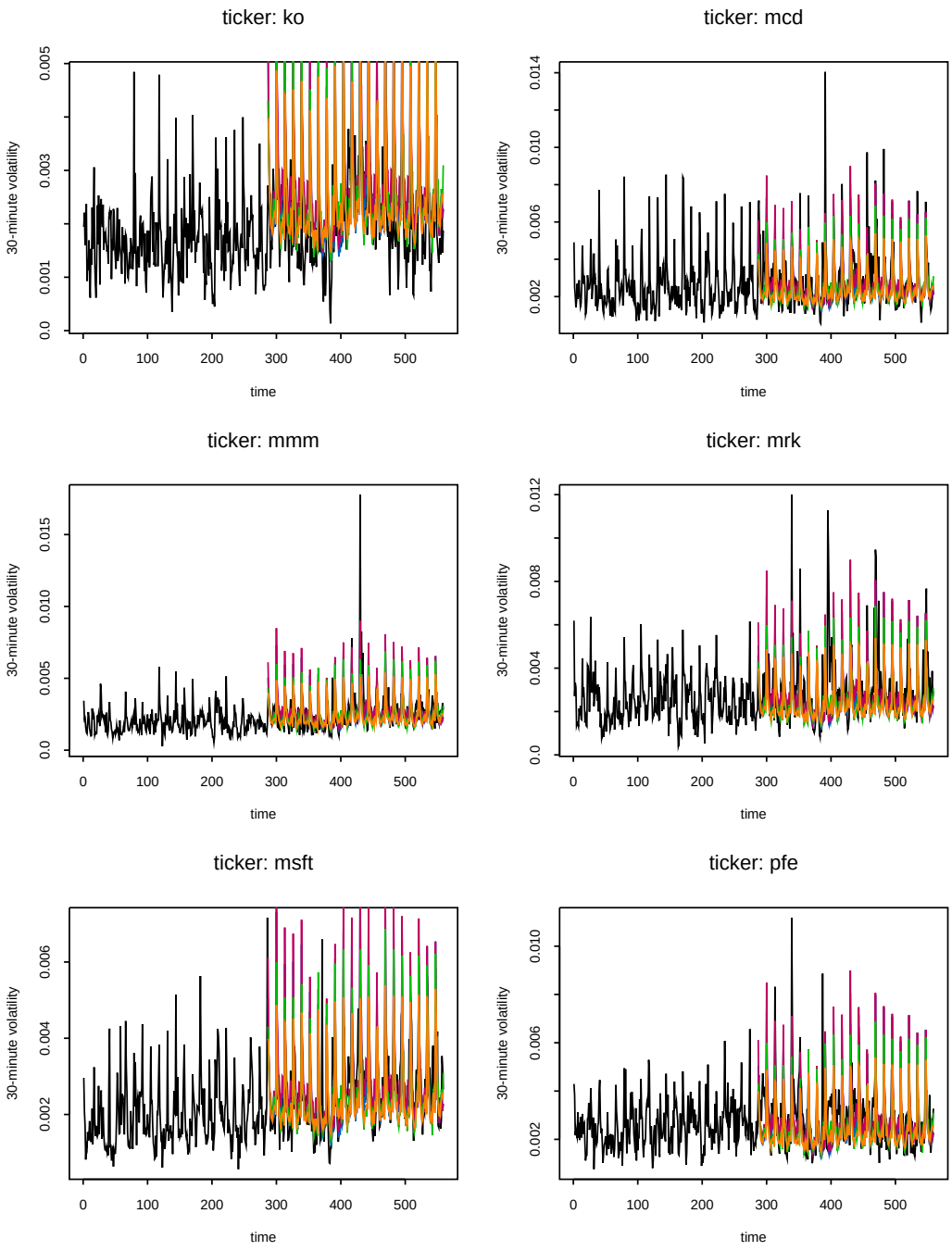

Figure 26. Forecasting plot

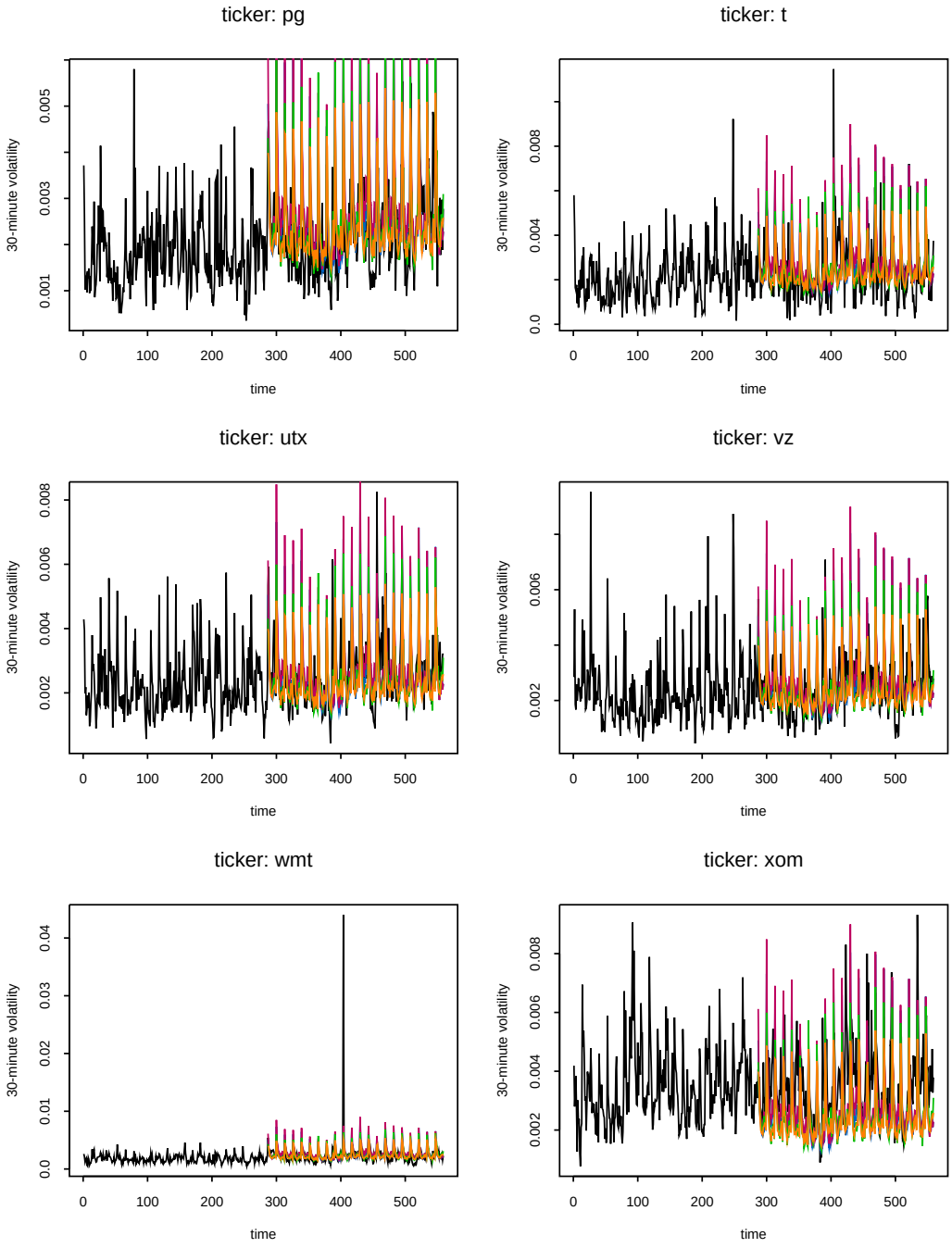

Figure 27. Forecasting plot

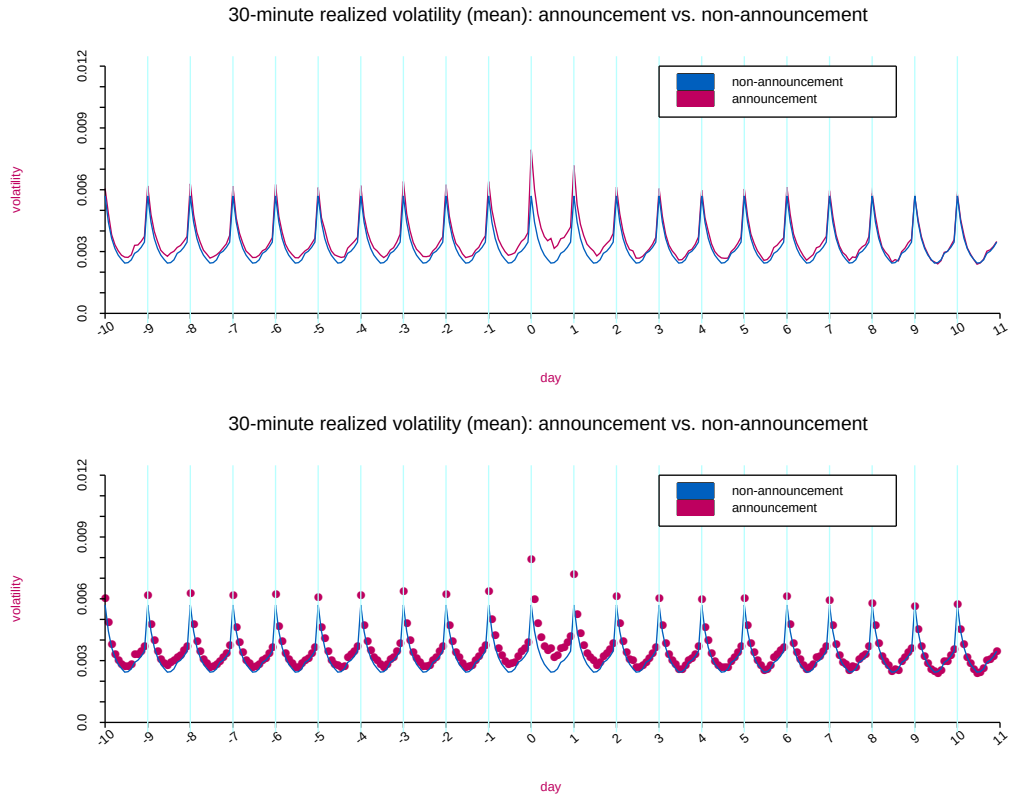

Figure 28. Mean 30-minute volatility in quarterly earning announcement period vs. in non-announcement period

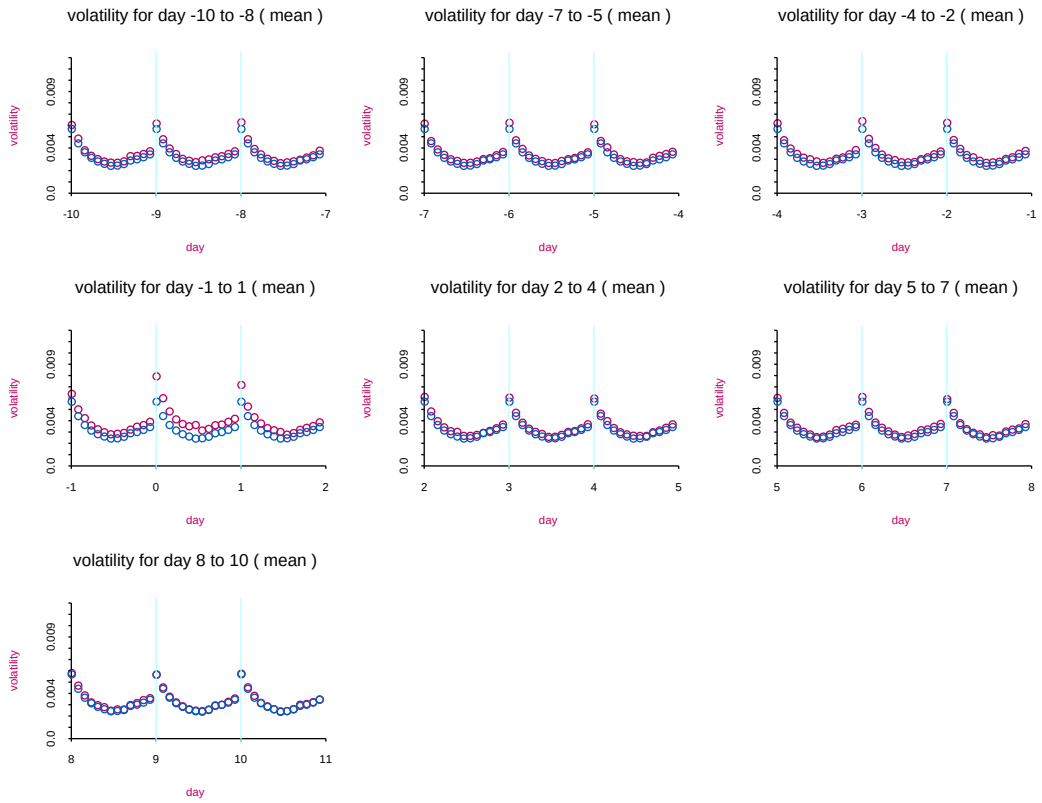

Figure 29. Zoom-in mean 30-minute volatility in quarterly earning announcement period vs. in non-announcement period

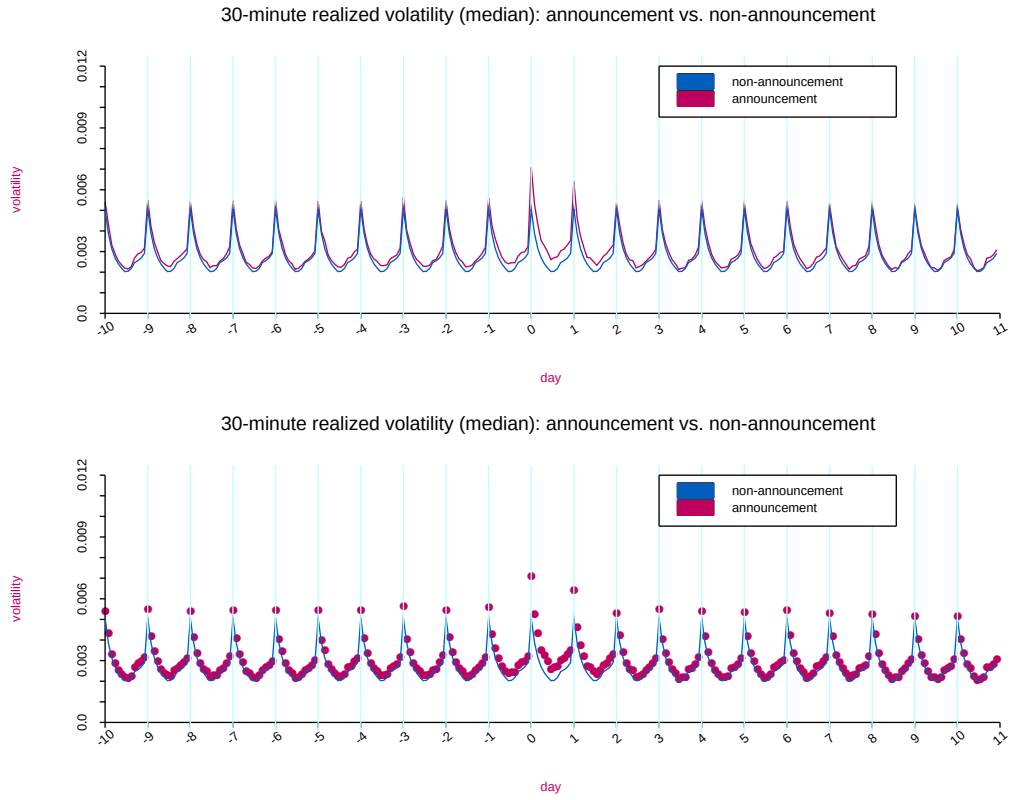

Figure 30. Median 30-minute volatility in quarterly earning announcement period vs. in non-announcement period

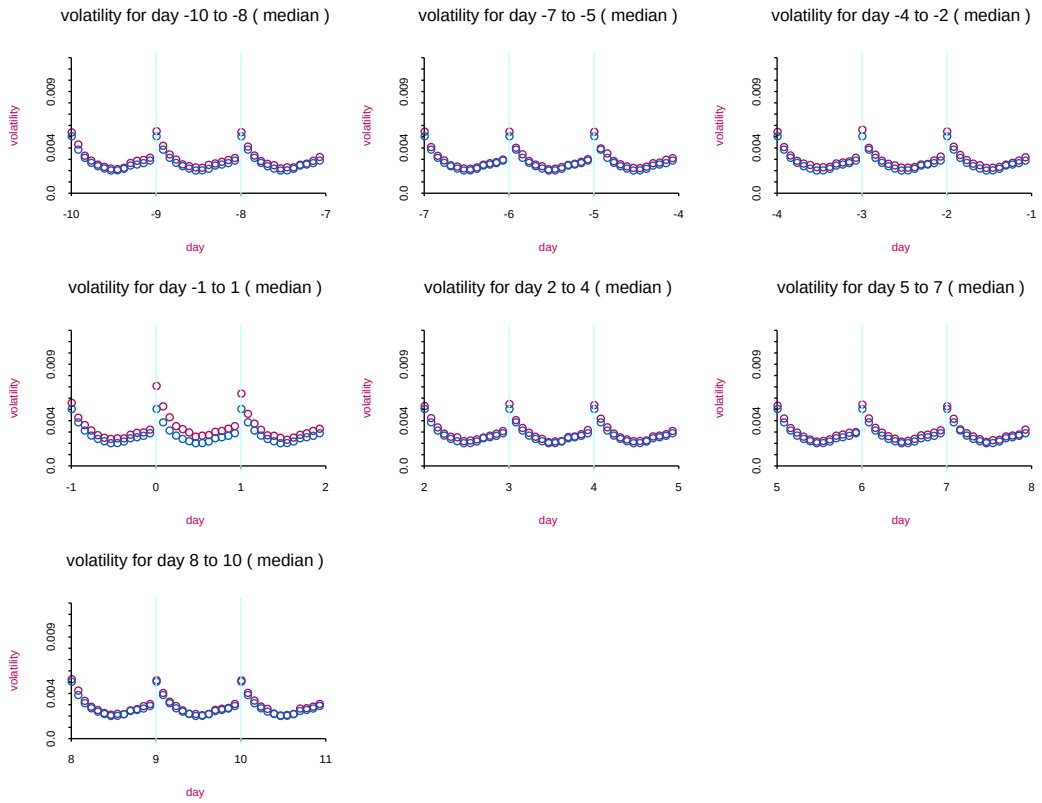

Figure 31. Zoom-in median 30-minute volatility in quarterly earning announcement period vs. in non-announcement period

TABLE I

TIME LINE OF THE 30 STOCKS IN DOW JONES INDUSTRIAL AVERAGE INDEX

Ticker	Time	Ticker	Time
AA	12/23/1999 - 12/30/2005	JNJ	01/10/2000 - 12/30/2005
AXP	01/07/2000 - 12/30/2005	JPM	01/03/2000 - 12/30/2005
BA	01/04/2000 - 12/30/2005	KFT	10/02/2001 - 12/30/2005
BAC	01/03/2000 - 12/30/2005	KO	01/11/2000 - 12/30/2005
C	01/03/2000 - 12/30/2005	MCD	01/11/2000 - 12/30/2005
CAT	01/06/2000 - 12/30/2005	MMM	01/11/2000 - 12/30-2005
CVX	01/14/2002 - 12/30/2005	MRK	01/11/2000 - 12/30/2005
DD	01/11/2000 - 12/30/2005	MSFT	01/08/2000 - 12/30/2005
DIS	01/07/2000 - 12/30/2005	PFE	01/03/2000 - 12/30/2005
GE	01/05/2000 - 12/30/2005	PG	01/10/2000 - 12/30/2005
GM	01/05/2000 - 12/30/2005	T	01/10/2000 - 12/30/2005
HD	02/07/2000 - 12/30/2005	UTX	01/04/2000 - 12/30/2005
HPQ	08/13/2002 - 12/30/2005	VZ	10/16/2000 - 12/30/2005
IBM	01/14/2000 - 12/30/2005	WMT	01/02/2000 - 12/30/2005
INTC	12/29/1999 - 12/30/2005	XOM	01/10/2000 - 12/30/2005

TABLE II

VALUE OF GRID SIZE K IN THE 30 STOCKS IN DOW JONES INDUSTRIAL AVERAGE 1999-2005

ticker	K	ticker	K
AA	12	JNJ	25
AXP	25	JPM	12
BA	25	KFT	12
BAC	25	KO	12
C	50	MCD	25
CAT	12	MMM	12
CVX	50	MRK	25
DD	25	MSFT	200
DIS	25	PFE	50
GE	50	PG	25
GM	12	T	25
HD	50	UTX	12
HPQ	50	VZ	25
IBM	50	WMT	25
INTC	200	XOM	50

TABLE III

VOLATILITY COMPARISON FOR OPENING 30-MINUTE VS. CLOSING 30-MINUTE

	t	P-value
Two Sample T Test	32.6338	0.0000
Paired T Test	74.2584	0.0000

TABLE IV

DESCRIPTIVE STATISTICS OF MEAN 30-MINUTE TSRV (TICKER:AA)

	vol	log(vol)
Mean	0.0030	-5.9352
Median	0.0026	-5.9709
Standard Deviation	0.0018	0.528
Skewness	2.0654	0.0714
Kurtosis	7.8038	0.4012

TABLE V

UNIT ROOT TEST AFTER REMOVING PERIOD PATTERN

	t	P-value
Augmented DF Test	-12.89	0.0000
Phillip-Perron Test	-55.27	0.0000

TABLE VI

AUGMENTED DICKEY-FULLER TEST LAG COEFFICIENTS

lag1	lag2	lag3	lag4	lag5	lag6
-0.1205	-0.4605	-0.3567	-0.2928	-0.2482	-0.2171
lag7	lag8	lag9	lag10	lag11	lag12
-0.1699	-0.1559	-0.1201	-0.1063	-0.0730	-0.0497

TABLE VII

REGRESSION TO REMOVE PERIODIC PATTERN ($R^2 = 0.4298$)

	estimate	t	P-value
α_0	-6.2528	-83.0632	0.0000
β_1	0.9849	9.2518	0.0000
β_2	0.5083	4.7751	0.0000
β_3	0.2995	2.8130	0.0053
β_4	0.2029	1.9057	0.0577
β_5	0.0781	-0.7338	0.4637
β_6	-0.0352	-0.3306	0.7412
β_7	-0.1070	-1.0049	0.3159
β_8	-0.1347	-1.2650	0.2070
β_9	-0.0775	-0.7276	0.4675
β_{10}	0.0371	0.3486	0.7277
β_{11}	-0.0178	-0.1675	0.8671
β_{12}	0.0218	0.2051	0.8376

TABLE VIII

DESCRIPTIVE STATISTICS OF RESIDUALS IN THE FIRST ROLLING WINDOW

	SEMIFAR
Mean	0.0004731635
Median	0.03165559
Standard Deviation	0.3142394
Skewness	-0.3292028
Kurtosis	0.5918353

TABLE IX

SEMIFAR MODEL ESTIMATES AFTER REGRESSION ON DUMMY VARIABLES

Window number	m	d
1	0	0.27940
2	0	0.26007
3	0	0.29888
4	0	0.27890
5	0	0.24833
6	0	0.24688
7	0	0.26286
8	0	0.26917
9	0	0.31061
10	0	0.33562
11	0	0.32043
12	0	0.39819
13	0	0.41432
14	0	0.34090
15	0	0.40253
16	0	0.34679
17	0	0.32286
18	0	0.34258
19	0	0.37916
20	0	0.38075
21	0	0.34522

TABLE X

LINEAR REGRESSION ON POLYNOMIAL TIME ($R^2 = 0.4202$)

	estimate	t	P-value
α_0	-4.8790	-42.8091	0.0000
α_1	-0.4890	-7.2029	0.0000
α_2	0.0527	4.7662	0.0000
α_3	-0.0018	-3.4247	0.0007

TABLE XI

DESCRIPTIVE STATISTICS OF RESIDUALS IN THE FIRST ROLLING WINDOW

	FARIMA
Mean	0.0004836245
Median	0.03347719
Standard Deviation	0.3182325
Skewness	-0.3060191
Kurtosis	0.5823561

TABLE XII

SEMIFAR MODEL ESTIMATES AFTER REGRESSION ON POLYNOMIAL TIME VARIABLES

Window number	m	d
1	0	0.27490
2	0	0.25446
3	0	0.29984
4	0	0.28009
5	0	0.23992
6	0	0.23269
7	0	0.25088
8	0	0.25255
9	0	0.28704
10	0	0.30109
11	0	0.28591
12	0	0.30641
13	0	0.38553
14	0	0.29463
15	0	0.34182
16	0	0.31078
17	0	0.27882
18	0	0.29884
19	0	0.33946
20	0	0.34333
21	0	0.29677

TABLE XIII

HAR-RV + DUMMY VARIABLE REGRESSION

	$log(RV_{t+1})$	
	estimate	P-value
α_0	-4.7358	0.0000
α_1	1.0516	0.0000
α_2	0.2213	0.0147
α_3	0.1403	0.0733
α_4	0.0449	0.5445
α_5	0.0529	0.4699
α_6	-0.0342	0.6395
α_7	-0.1342	0.0698
α_8	-0.0433	0.5563
α_9	-0.0354	0.6295
α_{10}	0.1163	0.1118
α_{11}	-0.0764	0.2961
α_{12}	0.0318	0.6626
tm30m	0.3523	0.0000
t	-0.0665	0.1191
tm5d	-0.0144	0.7470
tm22d	-0.0251	0.5714
	R^2= 0.4961	

TABLE XIV

HAR-RV + POLYNOMIAL TIME REGRESSION

	$log(RV_{t+1})$	
	estimate	P-value
α_0	-3.6332	0.0000
t	-0.5338	0.0000
t_2	0.0655	0.0000
t_3	-0.0025	-0.0000
tm30m	0.2227	0.0000
t	-0.0254	0.5590
tm5d	0.0192	0.6726
tm22d	-0.0021	0.9636
	$R^2 = 0.4507$	

TABLE XV

LJUNG-BOX TEST ON RESIDUALS

Model	p-value
HAR-RV + dummy	0.0217
HAR-RV + Polynomial time	0.0138

TABLE XVI

MEAN SQUARE ERROR OF FORECASTING IN 4/1/2005 - 4/29/2005 FOR DOW JONES 30 STOCKS

Ticker	SEMiFAR-dummy	SEMIFAR-polynomial	HAR-RV-dummy	HAR-RV-polynomial
AA	0.000903	0.000899	0.000905	0.000988
AXP	0.000994	0.000991	0.000985	0.001008
BA	0.001131	0.001132	0.001127	0.001185
BAC	0.000802	0.000788	0.000802	0.000818
C	0.000863	0.000860	0.000763	0.000764
CAT	0.000963	0.000966	0.000904	0.000931
CVX	0.001301	0.001308	0.001135	0.001120
DD	0.001188	0.001190	0.001208	0.001220
DIS	0.001057	0.001052	0.001024	0.001023
GE	0.000668	0.000680	0.000650	0.000655
GM	0.001786	0.001772	0.001587	0.001651
HD	0.001433	0.001438	0.001312	0.001368
HPQ	0.001331	0.001319	0.001244	0.001299
IBM	0.001007	0.001003	0.001099	0.001130
INTC	0.001003	0.001023	0.000904	0.000914
JNJ	0.000656	0.000655	0.000634	0.000644
JPM	0.000683	0.000686	0.000646	0.000652
KFT	0.001044	0.001044	0.001054	0.001108
KO	0.000647	0.000660	0.000607	0.000633
MCD	0.001183	0.001171	0.001153	0.001230
MMM	0.001197	0.001197	0.001224	0.001263
MRK	0.001479	0.001474	0.001338	0.001365
MSFT	0.000752	0.000747	0.000781	0.000794
PFE	0.001150	0.001142	0.001069	0.001080
PG	0.000832	0.000831	0.000773	0.000791
T	0.001260	0.001291	0.001222	0.001279
UTX	0.000944	0.000939	0.000850	0.000873
VZ	0.000981	0.000973	0.000958	0.000963
WMT	0.002531	0.002527	0.002550	0.002575
XOM	0.001170	0.001165	0.001031	0.001016

TABLE XVII

ANNOUNCEMENT PERIOD VS. NON-ANNOUNCEMENT PERIOD

		t	P-value
Mean	Two Sample T Test	3.6633	0.0003
	Paired T Test	19.7495	0.0000
Median	Two Sample T Test	3.4825	0.0005
	Paired T Test	20.3820	0.0000

CITED LITERATURE

1. Acker, D.: Implied standard deviations and post-earnings announcement volatility. Journal of Business and Finance & Accounting, 29:3–4, May 2002.

2. Andersen, T., Bollerslev, T., Diebold, F., and Labys, P.: Great realizations. Risk, 13:105 – 108, 2000.

3. Andersen, T., Bollersleve, T., and Meddahi, N.: Realized volatility forecasting and market microstructure noise. Journal of Econometrics, 2010.

4. Andersen, T. G., Bollerslev, T., Diebold, F., and Labys, P.: The distribution of realized exchange rate volatility. Journal of American Statistical Association, 96:42–55, 2001.

5. Andersen, T.G., B. T. D. F. and Labys, P.: (understanding, optimizing, using and forecasting) realized volatility and correlation,. Manuscript, Northwestern University, Duke University and University of Pennsylvania., 1999.

6. Bandi, F. and Russell, J.: Separating microstructure noise from volatility. Journal of Finance Economics, 79:655–692., 2006.

7. Barndorff-Nielsen, O., Hansen, P., Lunde, A., and Shaphard, N.: Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica, 76:1481–1536, 2008.

8. Beran, J. and Ocker, D.: Volatility of stock indices - an analysis based on semifar models. Journal of Business and Economic Statistics, 19(1):103–116, 2001.

9. Beran, J. and Ocker, D.: Semifar forecasts, with applications to foreign exchange rates. Journal of Statistical Planning and Inference, 80:137–153, 1999.

10. Beran, J., Feng, Y., and Ocker, D.: Semifar models. Technical report, SFB 475 University of Dortmund., 1998.

11. Bollersleve, T.: Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31:307–327, 1986.

12. Bollersleve, T., Engle, R., and Nelson, D.: Arch models. Handbook of Econometrics, 4, 1994.

13. Chen, X., Ghysels, E., and Wang, F.: Hybrid-garch: A generic class of models for volatility predictions using mixed frequency data. 2011. Working Paper.

14. Ederington, L. and Lee, J.: How markets process information: News releases and volatility. Journal of Finance, 48:1161–1191, 1993.

15. Engle, R.: Autoregressive conditional heteroskedasticity with estimates of variance of the united kingdom inflation. Econometrica, 50:987–1008, 1982.

16. Engle, R. and Gallo, G.: A multiple indicators model for volatility using intra-daily data. Journal of Econometrics, 131:3–27, 2006.

17. Geweke, J. and Porter-Hudak, S.: The estimation and application of long memory time series models. Journal of Time Series Analysis, 4:221–237, 1983.

18. Granger, C. and Joyeux, R.: An introduction to long memory time series models and factional differencing. Journal of Time Series Analysis, 1:15–29, 1980.

19. Hillmer, S. and Yu, P.: The market speed of adjustment to new information. Journal of Financial Economics, 7:312–345, 1979.

20. Hosking, J.: Fractional differencing. Biometrika, 68:165–176, 1981.

21. Jacod, J. and Shiryaev, A.: Limit Theorems for Stochastic Processes. Springer-Verlag, 2003.

22. Jacod, J., Li, Y., Mykland, P., Podolskij, M., and Vetter, M.: Microstructure noise in the continuous case: The pre-averaging approach. Stochastic Processes and Their Applications, 119:2249–2276, 2009.

23. Jacquier, E., Polson, N., and Rossi, P.: Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics, 12:371–389, 1994.

24. Lobato, I. and Savin, N.: Real and spurious long-memory properties of stock-market data. Journal of Businesss and Economic Statistics, 16:261–268, 1998.

25. Merton, R.: On estimating the expected return on the market: An exploratory investigation. Journal of Financial Economics, 8:323–361, 1980.

26. Nelson, D.: Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59:347–370, 1991.

27. Patell, J. and Wolfson, M.: The intraday speed of adjustment of stock prices to earnings and dividend announcements. Journal of Financial Economics, 13:223–252, 1984.

28. Patton, A. and Sheppard, K.: Good volatility, bad volatility: Signed jumps and the persistence of volatility. 2009.

29. Phillips, P. and Perron, P.: Testing for unit roots in time series regression. Biometrika, 75:335–346, 1988.

30. Jennings, R. and Starks, L.: Information content and the speed of stock price adjustment. Journal of Accounting Research, 23, 1985.

31. Ray, B. and Tsay, R.: Long range dependence in daily stock volatilities. Journal of Business and Economic Statistics, 18:254–262, 2000.

32. Said, S. and Dickey, D.: Testing for unit roots in autoregressive moving-average models with unknown orde. Biometrika, 71:599–607, 1984.

33. Shephard, N. and Shephard, K.: Realising the future: Forecasting with high frequency based volatility (heavy) models. Journal of Applied Econometrics, 2010.

34. Zhang, L., Mykland, P., and Ait-Sahalia, Y.: A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of American Statistical Association, 100:1394–1411, 2005.

35. Zhou, B.: High-frequency data and volatility in foreign-exchange rates. Journal of Business and Economics Statistics, 14:45–52, 1996.

VITA

JIAN SU

EDUCATION

Ph.D in Business Statistics, December 2011
University of Illinois at Chicago, Chicago, Illinois
M.S. in Biostatistics, May 2006
University of Illinois at Chicago, Chicago, Illinois
M.S. in Computer Sciences, July 2003
University of Illinois at Chicago, Chicago, Illinois
M.A. in Economics, July 2001
University of Illinois at Chicago, Chicago, Illinois
B.S. in Business Administration, July 1995
Capital University of Economics and Business, Beijing, China

INDUSTRY EXPERIENCE

Financial Engineer Intern, Spot Trading, Chicago, May 2011 - October 2011
Application Developer Intern, MorningStar Inc., Chicago, January 2004 - July 2004

RESEARCH EXPERIENCE

Research Assistant, Methodology Research Core, Institute of Health Research and Policy, UIC, 2006-2007
Research Assistant, College of Nursing, UIC, 2005
Research Assistant, Department of Neurology and Rehabilitation, UIC, 2001-2003

TEACHING EXPERIENCE

Lecturer, Business Statistics (undergraduate level)
Department of Information Decision Sciences, UIC, Fall 2007, Spring/Fall 2008, Spring 2009

HONORS AND PROFESSIONAL ACTIVITIES

- UIC Graduate Student Travel Award 2010
- UIC Graduate Student Council Travel Award 2010
- Visiting Student, Oxford-Man Institute of Quantitative Finance, University of Oxford, United Kingdom. October, December 2009

PUBLICATIONS

"Porting a Handheld Cognitive Assessment Form to a Mental Expert System: Using PDA to Assist Screening Patient's Performance in Clinical Examination", Neuroinformatics, Summer 2003, Volumn 1, and Issue 2, pp. 203-206. Jao CS, Dollear W, Su J.

COMPUTER SKILLS

- Language: C/C++/C#, Java, Visual Basic
- Quantitative: Matlab, R, S-Plus, SAS, SPSS
- Database: Oracle, SQL Server, MS Access, XML