Predicting Stock Volatility Using After-Hours Information

Chun-Hung Chen[1]

KPMG

Wei-Choun Yu[2]

Winona State University

Eric Zivot[3]

University of Washington

November 2, 2008

Abstract

This study uses realized volatilities based on the high frequency information during the after-hours period to predict daily volatility. We extend the GARCH and the SEMIFAR forecasting models to include additional information: the whole night, the preopen, the postclose realized variance, and the overnight squared return for four NASDAQ stocks, MSFT, AMGN, CSCO, and YHOO from 2001 through 2004. We find that the inclusion of the preopen variance can improve the out-of-sample forecastability of the conditional day volatility. The postclose variance and the overnight squared return, on the other hand, do not exhibit any predictive power on the future conditional volatility. The evidence supports the results of prior studies that traders trade for non-information reasons in the postclose period, while they do for information reasons in the preopen period.

[1]Corresponding Author: Economic and Valuation Service, KPMG LLP, 500 East Middlefield Rd., Mountain View, CA 94043.Email: chunhungchen@kpmg.com. Tel: +1-650-404-4704. Fax: +1-650-523-4340. We thank Mike Wenz for helpful comments.

[2] Economics and Finance Department, Winona State University, Somsen 319E, Winona, MN 55987, USA. Email: wyu@winona.edu. Tel: +1-507-457-2982. Fax: +1-507-457-5697.

[3]Department of Economics, University of Washington, Box 353330, Savery 302, Seattle, WA 98195. Email: ezivot@u.washington.edu. Tel: +1-206-543-6715. Fax: +1-206-685-7477.

Predicting Stock Volatility Using After-Hours Information

1. Introduction

Volatility modeling has received much attention over the past two decades in the finance literature not only because it relates directly to the profits of traders, but its importance to the valuation of derivative instrument. The goals for the modeling and forecasting of volatility are to have better risk management, more accurate derivative prices, and more efficient portfolio allocations. A good financial decision-making relies on an accurate prediction of the second moment of the underlying financial instrument.

Among various volatility-modeling techniques, the most popular volatility models are the GARCH models developed by Engle (1982) and Bollerslev (1986). This family of models can explain well the common features of volatility processes, i.e. persistence, mean reversion, and the leverage effect. Moreover, it is well known that the GARCH model can produce good forecastability of the conditional volatility. In addition to modeling techniques, a good modeling and forecasting of volatility relies on a useful information set. Until recently, the most commonly used information set for modeling daily volatility is historical daily closing prices. However, employing high-frequency intradaily data has gradually become appealing in the recent literature because it can significantly improve measurement and forecastability of the models (Andersen, and Bollerslev, 1998, and Andersen, Bollerslev, and Lange, 1999).

Although the previous literature has documented the importance of after-hours information (for example, Oldfield and Rogalski, 1980; Greene and Watts, 1996; Cao et al., 2000; Masulis and Shivakumar, 2002; Taylor, 2007; and Tsiakas, 2008), few of them analyze the high-frequency data. Taylor (2007) did use high-frequency information but he uses overnight S&P 500 futures volatility to predict S&P 500 stocks volatility. In contrast, we use after-hours stock volatility to predict the following day volatility. The availability of after-hours trading opportunities to the public and the recording of after-hours transaction data of the NASDAQ stocks allow us to examine how this extended information set could affect the modeling and the forecasting of the conditional daily volatility by utilizing high frequency intranightly data.

The paper uses two volatility forecasting models: (1) GARCH based on daily returns with after-hours realized variance as an exogenous variable in conditional variance equation; (2) SEMIFAR based on historical realized variance. The high frequency data used in the paper are stock prices of Microsoft, Amgen, Cisco, and Yahoo listed on the NASDAQ. We use realized variance measures as proxies for unobserved volatility in our forecasting evaluation. The results show that the inclusion of the volatility for the whole night in the information set does not provide a better forecast performance for either the GARCH or the semiparametric fractional autoregressive (SEMIFAR) volatility model employed in this study. When breaking up the whole after-hours period into three sub periods, we find that the inclusion of preopen period variance significantly improves the forecastability both in terms of the variability and accuracy for1-step and 5-step ahead conditional daily volatilities, while the other two sub periods do not.

Our study contributes to the literature in four aspects. First, we use high-frequency intranightly transaction data that is newly available and have not yet been explored in the literature. Second, we completely use the after-hours information by segmenting the whole after-hours period into sub periods based on their different information densities. Third, most past research focuses on in-sample forecasting evaluation while we evaluate our model's predictive ability out-of-sample, which is more practical and useful. Finally, we employ two different volatility models: a parametric model (GARCH) and a semiparametric model (SEMIFAR). The latter model is generally free from functional form assumption and therefore is more flexible and consistent (Andersen, Bollerslev, and Diebold, 2002). This is especially important when we have less knowledge on after-hours volatility. Furthermore, the employment of models under different forms and assumptions can serve as a robustness check for our hypothesis.

The rest of the paper proceeds as follows. In Section 2, we review the after-hours literature. Section 3 explains data and realized volatility measurement. The methods and results of modeling the conditional volatility are based on the GARCH and the SEMIFAR techniques and their forecast evaluations are provided in Section 4 and 5. Section 6 concludes.

2. Literature Review

Conditional volatility models, such as GARCH and long memory models, demonstrate that past shocks and volatilities contain the information about future volatilities in stock prices, and therefore can be used for forecasting purposes. One explanation for this is the Mixture of Distribution Hypothesis (MDH), suggested by Clark (1973), Tauchen and Pitts (1983), and Kalev, Liu, Pham, and Jarnecic (2004). They attribute this long-run dependence of volatilities to the serial correlation of the news arrival rate. The MDH postulates that the volatility can be attributed to varying information arrival rate, and thereby the highly autocorrelated nature of volatility comes from the persistence in information arrival rates contained in the past daily shock or volatility process.

Lamoureux and Lastrapes (1990) use the MDH proposed by Clark (1973) to explain the persistent nature of the conditional day volatility in the GARCH model. They assume that a stochastic model can be derived by considering the daily return in day t, ε_t, as a sum of *i.i.d.* intraday price increments, δ_{it},

$$\varepsilon_t = \sum_{i=1}^{n_t} \delta_{it} \tag{1}$$

where i denotes the ith intraday price movement, and the random variable n_t is a mixed variable that denotes the arrival rate of information in day t. Clark (1973) assumes that ε_t is drawn from a mixture of distributions, of which the variances depend on n_t,

$$\varepsilon_t \mid n_t \sim N(0, \sigma^2 n_t) \tag{2}$$

if δ_i is *i.i.d.* with mean zero and variance σ^2, and n_t is sufficiently large. When the arrival rate of information is serially correlated, we can write n_t as

$$n_t = a + b(L)n_{t-1} + u_t \tag{3}$$

where a is a constant and u_t is a white noise. The conditional variance is

$$\Omega_t = E(\varepsilon_t^2 \mid n_t) = \sigma^2 n_t \tag{4}$$

substitute the above equation into the AR representation of n_t, we have

$$\Omega_t = \sigma^2 a + b(L)\Omega_{t-1} + \sigma^2 u_t \tag{5}$$

and this demonstrates the persistence in the conditional variance in the GARCH model.

To examine this hypothesis, Lamoureux and Lastrapes (1990), Sharma, Mougoue, and Kamath (1996), and Brooks (1998) use the trading volume as a proxy for information arrival rate and include it as an exogenous variable in the GARCH (1,1) specification for daily volatility. They show that the inclusion of volume greatly reduces the persistence parameter in the GARCH regression. Moreover, Brooks (1998) shows that it does not improve the volatility forecasting because no new information is provided to the forecasting of daily volatility by the inclusion of volume variable, which is simply a perfect substitute for the past conditional volatility process.

After hours is the time period from the previous market closing time through the next market opening time, which is also called the close-to-open or whole night period. If the after-hours volatility could provide additional information rather than a substitute for the persistence parameter such as volume mentioned above, it would improve the forecasting ability. It is well known in the microstructure literature that information and announcements could occur during the night, regardless of the existence of trading opportunity during that time. This occurrence and accumulation of information during the close-to-open period should contribute to the upcoming day (open-to-close) volatility. When after-hours trading is not available, the information will be realized at the opening hours. The occurrence of larger-than-normal after-hours news implies higher-than-usual volatility during the following regular trading hours.

Even when trading is available for all or part of the night, we can still expect information to have an impact in the following regular trading hours for two reasons. The first reason is the spillover effect. If the market is not fully efficient, it would take some time for the information to be incorporated into prices. This could be due to the highly illiquid nature of after-hours trading environment. Since it takes trades to facilitate price discovery (Barclay and Hendershott (2003)), the information might not be fully incorporated into the price until the regular trading hour, when the trading volume is much higher. The second reason is the informed nature of trades in after hours. Barclay and Hendershott (2004) show that the traders in after hours are mainly professional and institutional. Many of them trade for short-lived private information. It is likely that they trade for private or scheduled news that has yet to be announced. Therefore, it is rational to expect a highly volatile night trading would lead to a highly volatile day trading in the next day.

Gallo and Pacini (1998) study the impact of close-to-open returns, which are measured as the difference of the previous daily closing price and current daily opening price, on the following day (open-to-close) volatility for the six major market indices by using a GARCH (1,1) model with the close-to-open returns as an exogenous variable. Martens (2002) studies whether GARCH (1,1) models that include different functional forms of the after-hours volatility can improve the forecasts of the following day volatility for the S&P 500 index futures transactions. Gallo and Pacini (1998) find that the inclusion of close-to-open returns improves forecastability of conditional volatility for some stock indices, while Martens (2002) finds that the inclusion of the close-to-open squared returns cannot improve the forecastability. This mixed evidence could come from the poor exploitation of after-hours information. Our paper bridges the gap.

Because of the newly available intranightly transaction data, we are able to segment the after-hours period based on its different information density. The idea has been proposed by Barclay and Hendershott (2003, 2004) but yet to be applied on forecasting evaluation. By and large, the measurement of daily volatility can be improved by utilizing high-frequency intranightly data. Past studies, such as Cumby, Figlewski and Hasbrouck (1993), Figlewski

(1997), and Jorion (1995), have shown that the standard volatility models perform poorly in the out-of-sample forecast using the squared returns as a proxy for volatility. Andersen and Bollerslev (1998) point out that the problem comes from the improper use of the daily squared return as the ex post volatility measure. Even though the squared return is unbiased for unobserved volatility, it is very noisy due to the large idiosyncratic error term. They show that the realized volatility, which is defined as the sum of squared returns sampled at high intradaily frequency, provides a more reliable ex post volatility measure. Andersen, Bollerslev, and Lange (1999) also show that the forecasting performance of standard volatility models can be greatly improved by utilizing high frequency data. Andersen, Bollerslev, Diebold, and Laybs (2003) treat each estimated daily realized volatility as an observation, and construct an autoregressive fractional integrated moving average (ARFIMA) model for daily volatility. They find that this model is superior to many other volatility models in terms of the forecasting performance. Therefore, we will employ realized volatility measures that utilize intradaily and intranightly data to provide more reliable results.

The previous papers have mainly used the GARCH model for volatility modeling and forecasting analysis. Here we propose a way to include after-hours volatility in the semiparametric fractional autoregressive (SEMIFAR) volatility model, a long memory model proposed by Beran and Ocker (1999, 2001). We find that both GARCH and SEMIFAR models give exactly the same qualitative result, and therefore provide robustness for our hypothesis.

3. Data and Volatility Measurement

Unlike Taylor (2007) and Tsiaks (2008), we investigate individual stocks instead of market indices. As mentioned in Campbell et al. (2001), there are several reasons to be interested in the volatilities of individual stocks. For instance, many investors have large holdings of individual stocks, which have not been diversified and therefore are subject to idiosyncratic volatility. We are able to analyze the volatility induced from the firm-level event news in additional to macroeconomic news. Since trading volume is relatively low for stocks in after hours, we have chosen stocks that are liquid during the after-hours periods. The high-frequency data used in the paper are stock prices of Microsoft (MSFT), Amgen (AMGN), Cisco (CSCO), and Yahoo (YHOO) listed on the NASDAQ. We use MSFT as our benchmark stock and other three stocks for a robustness check. The sampling period is from January 2001 to December 2004, during which the after-hours trading information is available to the public and recorded. We choose the first three and a half years as the in-sample period for modeling volatility, and the later half a year as the out-of-sample period to evaluate forecasting performance.

3.1. After-Hours Subperiods

Barclay and Hendershott (2003, 2004) break the whole after-hours period into three subperiods: the postclose period (4:00 to 6:00 pm EST), the overnight period (6:00 pm to 8:00 am EST), and the preopen period (8:00 to 9:30 am EST). They investigate the information structure of the postclose and preopen and find that the probability of an informed trade is much higher in the latter than the former period. They find that about 80 percent of all trading volume in postclose occurs at the closing price or within the closing quotes at 4:00 pm EST.[4] This implies that traders tend to trade for liquidity demands right after the regular hour is closed. On the other hand, they use the probability of informed trade measure developed by Easley, Kiefer, and O'Hara (1997) to show that trading is highly informed during the preopen, which implies that traders are more likely to trade for information reasons in this period. Even though traders can still trade through an electronic communication network (ECN) or a market maker during the overnight period, there is no formal analysis on the information structure for this period. The overnight data is usually not available from the reporting service provided by the NASDAQ Trade Dissemination Service (NTDS). Barclay and Hendershott (2003) use their proprietary dataset and find that only 1% of total after-hours trades occur during that period.

[4] Note, however, that this trading at close activity only represents 15 percent of trades in postclose.

The uneven information in each after-hours subperiod leads us to hypothesize that the volatility in each subperiod should have different effects on the following day volatility. We expect that the postclose volatility contains little to no information, while the volatility in the preopen contains new and additional information about the following day volatility. This means that the inclusion of the preopen volatility in the information set may improve the forecastability of a volatility model. The impact of volatility in the overnight period on conditional day volatility, however, is less obvious. If the preopen trades have realized most or all of the information that occurred in the overnight period, or if the overnight squared return measure is very noisy, we would expect little or no effect on the day volatility.

3.2. Data

The main data source used is the Trade and Quote (TAQ), which provides data on tick-level transaction prices from 8 am until 6:30 pm EST, when the NTDS is on. NTDS is responsible for the reporting of trades and quotes (for details, see Barclay and Hendershott, 2004). Since the TAQ data usually contains a lot of recording errors, we have performed the following data-filtering algorithm. Since the unconditional daily volume weighted average duration between any two consecutive trades is about 0.35 seconds during the regular hours, 11.5 seconds during the postclose period, and 14.3 seconds during the preopen period, we do not expect a big price movement between any two consecutive trades within a day. Therefore, we remove any recorded trades that have a change of positive or negative 25%[5] from their immediate prior trades in a day. We also remove dates in which either the preopen, postclose, or day transaction data is missing as well as the occurrence of stock splits.

3.3. Volatility Measure

[5] The daily volume weighted price for MSFT is $39.44 for the sample period, and 25% of which would be about $10.

The realized volatility is a more accurate measure of conditional volatility than the squared return because it contains the information of population parameters. We can use it to measure day volatility and volatility during after-hours periods. It can be used to evaluate our volatility predictions and we can use it as a historical volatility series from which we can build direct volatility forecasting models. We follow the method developed by Bollerslev and Wright (2001), Andersen, Bollserslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2003) to construct the realized volatility, which utilizes high-frequency data:

$$\begin{aligned} r_{i,t+n\Delta} &= p_{i,t+n\Delta} - p_{i,t+(n-1)\Delta} \\ r_{i,t} &= \sum_{n=1}^{1/\Delta} r_{i,t+n\Delta} \\ \sigma_{i,t}^2 &= \sum_{n=1}^{1/\Delta} r_{i,t+n\Delta}^2 \end{aligned} \tag{6}$$

where *p* denotes the logarithmic stock price, *i* is denoted as either the regular hour, the preopen, or the postclose period; *r* is the intraday return; $1/\Delta$ is the number of sampling observations for each of the periods (Δ is 15 minutes in regular hours while is 15 minutes in after-hours period (explained in later text); and $\sigma_{i,t}^2$ is the estimated realized volatility period *i* in day *t*. Since there is no data for trades in the overnight period, we measure the variance based on the first trade of preopen and the last trade of previous day’s postclose:

$$\sigma_{Overnight,t}^2 = (p_{First\ Trade\ of\ \mathrm{Pr}eopen,t} - p_{Last\ Trade\ of\ Postclose,t-1})^2 \tag{7}$$

Andersen, Bollerslev, Diebold, and Labys (2001) show that as sampling frequency increases to a very large number, the realized variance accurately measures the implied integrated variance, which is the actual realized return variation over a given horizon for a continuous time diffusion process. In summary, realized volatility is a model-free approach to consistently estimate the squared return variation under the assumption of arbitrage-free condition. Note that the measurement of realized volatility serves two purposes in our paper. First, it will be used directly in the volatility process when we construct the long memory model (discussed in Section 5).

Second, it will be used as the ex post volatility measure to be the benchmark for forecasting evaluation.

Even though the theory demonstrates that the measurement error associated with the estimation of the realized volatility becomes very small as the sampling frequency increases to a large number, the microstructure frictions, including bid-ask bounce, price discreteness, and infrequent trading, become a non-negligible issue as the sampling interval becomes very small. To avoid this problem, Andersen et al. (2001) propose to sample the intradaily observations at 5 minutes. Since the trading environment after hours is known to have much larger microstructure friction than during regular hours, we follow Andersen et al. (2001) to sample observations at 5-minute frequency for regular-hours intradaily data, and at 15-minute frequency for after-hours intranightly data. Therefore the number of samples for the regular hours, the preopen, and the postclose is 78, 6, and 8, respectively. Figure 1 shows the daily, the preopen, and the postclose realized volatility time series, and the overnight square return time series for MSFT. Table 1 lists the descriptive statistics of realized volatility measures of MSFT. These measures represent the total amount of volatility per day in each period. Similar to the distribution of returns, the distributions of volatilities are all skewed to the right and have fat tails.

Barclay and Hendershott (2003) find that the price changes are larger in the preopen than the postclose. This indicates that there is more private information and less noise in the preopen period. Table 2 provides volatilities per hour and per trade for the preopen and the postclose periods. The average volatilities per hour for the preopen and postclose are 0.39% and 0.29%, and the average volatilities per trade are 0.0102% and 0.00592%. The numbers show that the volatility in the preopen is higher than the one in the postclose, which is consistent with the result of Barclay and Hendershott (2003). Both the median volatilities per hour and per trade provide the same qualitative results. The autocorrelation plots in the four periods are shown in Figure 2. The daily, preopen and postclose realized volatility series all exhibit the commonly known characteristic of long memory or persistence. On the other hand, we do not observe this feature in the overnight series.

4. GARCH Modeling and Forecasting

The generalized autoregressive conditional heteroskedasticity (GARCH) model is arguably the most common approach to modeling and forecasting volatility. It is capable of capturing many of the stylized facts of the volatility behavior usually observed in empirical finance. The GARCH(p,q) is specified as

$$\begin{aligned} r_t &= \mu + \varepsilon_t \\ \varepsilon_t &= z_t h_t \\ h_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}^2 \end{aligned} \tag{8}$$

where μ and ω are constants in the conditional mean equation and the conditional variance equation, ε_t is the serially uncorrelated residual term, or news shock, with mean zero, z_t is assumed to be *i.i.d.* standard normal, and h_t^2 is the conditional variance at time t. The conditional variance term, h_t^2, is dependent on the past p periods of the squared residuals, ε_t^2, and the past q periods of the conditional variance term. By and large, a GARCH (1,1) model is usually sufficient for most financial time series applications (Andersen and Bollerslev (1998)). Table 3 shows the Akaike information criterion, Bayesian information criterion, and log-likelihood for GARCH (p,q) models with the lags of p and q set to be less than or equal to 2 for the daily MSFT return series in the in-sample period, which is from January 2001 to June 2004. The GARCH (1,1) with Student's t error distribution appears to be the appropriate model.

The first column of Table 4 (A) reports the coefficients for the daily model in regular hours. We see that the sum of α and β is 0.997, which shows that the conditional volatility is quite persistent. This result is very similar to 0.9986 reported by Martens (2002) for S&P 500 futures. The ARCH and Ljung-Box tests on the squared residuals are employed to check for the adequacy of the fitted model. We find that the GARCH (1,1) specification fits the in-sample return series of the MSFT well. GARCH(1,1) estimations of AMGN, CSCO, and YHOO are reported in Table 4 (B), (C), and (D), respectively.

4.1. GARCH Model for Day Returns with Night Variance

The GARCH model offers flexibility where additional exogenous variables that are thought to have impacts on conditional volatility can be included in the conditional variance equation. The modified GARCH (1,1) model can be specified as:

$$\begin{aligned} r_t &= \mu + \varepsilon_t \\ \varepsilon_t &= z_t h_t \\ h_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}^2 + \rho x_{t-1} \end{aligned} \tag{9}$$

where x_t is the additional exogenous variable. Both Gallo and Pacini (1998) and Martens (2002) use this approach by including the close-to-open squared return as the additional exogenous variable in the conditional variance equation. Martens (2002) finds that the coefficient for the additional variable is statistically insignificant. On the other hand, Gallo and Pacini (1998) find the coefficient to be statistically significant for most of the major market indices while the signs of the coefficient are positive for some, and negative for others. In our hypothesis, we expect the sign of the coefficient for the after-hours volatility to be significant and positive. If the impact of after-hours information on the volatility of regular hours is caused by the possibility of the informed traders trading private information before the news is publicly announced during the regular hours, a large after-hours volatility should lead to a large day volatility.

We propose the following four exogenous variables to be added to the conditional variance equation of the GARCH (1,1) models:

(1) All three subperiods together: $$x_t = \frac{1.5}{17.5}\sum_{n_{PO}=1}^{N_{PO}} r_{t,n_{PO}}^2 + \frac{2}{17.5}\sum_{n_{PC}=1}^{N_{PC}} r_{t,n_{PC}}^2 + \frac{14}{17.5} r_{t,n_{ON}}^2 \tag{10}$$

(2) The preopen period only: $x_t = \sum_{n_{PO}=1}^{N_{PO}} r_{t,n_{PO}}^2$ (11)

(3) The postclose period only: $x_t = \sum_{n_{PC}=1}^{N_{PC}} r_{t,n_{PC}}^2$ (12)

(4) The overnight period only: $x_t = r_{t,n_{ON}}^2$ (13)

where PO, PC, and ON each denotes the preopen, the postclose, and the overnight period.

The x_t defined in model (1) is a time-weighted average realized variance of the close-to-open (whole night) period. The second, third, fourth, and fifth column of Table 4 (A) show the coefficients and test statistics for the ARCH effect and the autocorrelation of residuals of each model for MSFT. The test statistics show that the models are well-fit. The coefficient for the whole night period realized variance, ρ, in model (1) is 0.065, which is positive but statistically insignificant. This result favors Martens (2002) in that the close-to-open variance does not significantly explain the conditional variance.

The coefficient for the postclose variance is negative and statistically insignificant as well. The result agrees with the hypothesis that the traders trade in the postclose for non-information reason, and therefore there is no information to be carried over into the next day. Another possibility is that if there was any information in the postclose period, it might have been spilled over to and removed during the following overnight period and the preopen period. The only explanatory variable that is statistically significant is the preopen realized variance, which has an estimated coefficient of 0.221. This means that a 1% increase in the preopen realized variance would lead to a 0.221% increase in the following regular hour conditional variance. This is consistent with our hypothesis in that the coefficient should be positive and significant.

Table 4 (B), (C), and (D) list the estimations for three other stocks. The results support significant and positive coefficients of the preopen variance as the additional variable in the day GARCH (1,1) model for all three stocks. The coefficients for the postclose variance, on the other hand, are all insignificant. The overnight squared return provides some explanatory power to day conditional volatility for CSCO, but not the other two stocks.

The estimated persistence parameters in our study are 0.997 and 0.971 for the ordinary GARCH (1,1) and the GARCH (1,1) with the preopen variance, respectively. The result shows only a slightly decrease in the persistence parameter, and therefore the explanatory variable appears to provide independent information from that contained in the past day returns. Along with the coefficient being statistically significant, this enhances our hypothesis that the addition of the preopen variance into the model would increase the forecasting performance of the conditional day volatility.

Like MSFT, we also do not observe much reduction in the persistence parameter for the day GARCH (1,1) model after the preopen variance is included as the additional explanatory variable for each of the other three stocks shown in Table 4 (B), (C), and (D). It is worth noting that we do believe that including leverage effects into the model would improve the volatility forecasting (e.g. EGARCH, GJR GARCH, or TGARCH). However, our main focus is on the after-hours information. Therefore for the purpose of the model simplicity, we do not consider this asymmetric effect.

4.2. Forecast Evaluation

Figure 3 shows the MSFT out-of-sample ex post volatility series and the forecasted conditional volatility series of the GARCH (1,1) with the preopen realized variance and with the whole night realized variance as exogenous variables. We evaluate forecasting performance of the out-of-sample using the parameters estimated from the in-sample. There are many forecast evaluation methods available for volatility models. We use the following six methods.

(i) Mincer-Zarnowitz Regression: $\left(\sigma_{t+k}^{2}\right)^{1/2} = a_0 + a_1\left(h_{t+k|t}^{2}\right)^{1/2} + u_{t+k}$ (14)

(ii) RMSE: $\left[\frac{1}{T}\sum_{t=1}^{T}\left(\sigma_{t+k}^{2}-h_{t+k|t}^{2}\right)^{2}\right]^{1/2}$ (15)

(iii) MAE: $\frac{1}{T}\sum_{t=1}^{T}\left|\sigma_{t+k}^{2}-h_{t+k|t}^{2}\right|$ (16)

(iv) HRMSE: $\left[\frac{1}{T}\sum_{t=1}^{T}\left(1-\sigma_{t+k}^{2}/h_{t+k|t}^{2}\right)^{2}\right]^{1/2}$ (17)

(v) HMAE: $\frac{1}{T}\sum_{t=1}^{T}\left|1-\sigma_{t+k}^{2}/h_{t+k|t}^{2}\right|$ (18)

(vi) LL: $\frac{1}{T}\sum_{t=1}^{T}\left[\log\left(\sigma_{t+k}^{2}/h_{t+k|t}^{2}\right)\right]$ (19)

where k denotes the k-step ahead prediction.

Mincer-Zarnowitz regression proposed by Mincer and Zarnowitz (1969) is considered one of the most popular metrics for evaluating volatility forecastability. If the conditional volatility model is correctly specified, we should have a_0 and a_1 equal to zero and one, respectively, with high statistical significances. However, Andersen and Bollerslev (1998) suggest that the coefficients suffer from a standard errors-in-variables problem, i.e. the estimates of the conditional volatility are subject to estimation error. The R^2 of the regression, nevertheless, can be used to evaluate the variability of the ex post volatility that is explained by the forecasted conditional volatility.

We employ several other forecasting criteria to assess the predictability of the volatility process based directly on the difference between the forecasted conditional variance and the ex post realized volatility measure. The root mean square error (RMSE) and the mean absolute error (MAE) are the two commonly seen criteria. To check whether the results are reliable or not in the highly nonlinear and heteroskedastic environment, we follow Andersen, Bollerslev, and Lange (1999) to use the heteroskedasticity adjusted RMSE (HRMSE), MAE (HMAE), and the logarithmic loss function (LL).

It is important to choose the right ex post volatility measure that serves as the benchmark for the evaluation, since volatility is not directly observed. Several past studies, such as Figlewski (1997) and Jorion (1995, 1996), have used daily squared returns as the proxy for the ex post volatility measure, and conclude that standard volatility models explain little of the variability in the ex post volatility. Andersen and Bollerslev (1998) find that even though the daily squared return provides an unbiased estimate of the latent volatility, it is quite noisy because of the idiosyncratic error term. They propose to use the realized volatility measure, which employs high-frequency data, and demonstrate that it provides a more reliable and accurate measure of the true volatility. Therefore, we use the estimated realized volatility in the out-of-sample as the benchmark measure.

We first estimate the parameters of MSFT from the in-sample, and perform 1-step-ahead prediction on conditional volatility using a rolling window. Table 5(A) lists the 1-step ahead forecast evaluation results for the different GARCH models. When using the Mincer-Zarnowitz regression, we see that the GARCH (1,1) with the preopen variance provides the best forecasting performance both in terms of accuracy and explaining the variability in the ex post measures. The coefficient of the conditional volatility, 0.976, is significant in a GARCH (1,1) model with preopen, compared to 0.899 for that of the day GARCH (1,1). The GARCH (1,1) with the preopen variance also yields a R^2 of 0.157, which is much higher than 0.091 for the day GARCH (1,1). The coefficients for day GARCH (1,1) and day GARCH (1,1) with preopen variance are not significantly different from one, but they suffer from the errors-in-variables issue discussed earlier. For other evaluation statistics, the forecasts provided by the GARCH (1,1) with the preopen variance are always superior to day GARCH (1,1). We also see that the forecasts from the GARCH (1,1) with the close-to-open period always perform relatively poor compared to the day GARCH (1,1). Since the coefficient for the realized volatility measure during the close-to-open period is not statistically significant in the GARCH(1,1) model, it is not surprising to see the forecast provides relatively poor performance and no improvement in terms of R^2. This implies that the after-hours volatility does not have any predictive power or the ability to improve the forecastability of conditional day volatility if one only compares day GARCH (1,1) to GARCH (1,1) with the close-to-open realized variance as the additional explanatory variable. Since the coefficients of the postclose and the overnight variances in the GARCH (1,1) are not statistically significant, we do not perform the forecasting evaluation for the two periods. Table 5 (B) provides evaluation results of 5-step-ahead prediction, and the outcomes are qualitatively identical.

Table 6 shows the forecast evaluation results of the other three stocks. It is interesting to see that the R^2 of YHOO is much higher than those of the other two stocks. However, the R^2 of the day GARCH(1,1) with the preopen variance is always higher than the usual day GARCH (1,1) model for each of the three stocks. The results of the five forecast measures, for the most part, also support our hypothesis that the information contained in the preopen variance can improve the forecastability of the conditional day volatility.

We follow Andersen, Bollerslev, Diebold, and Labys (2003) to perform an encompassing regression to further understand the effect of including the additional explanatory variable on the ex post volatility process. The encompassing regression takes the form,

$$\left(\sigma^2_{t+h,d}\right)^{1/2} = a_0 + a_1\left(h^2_{t+h|t,d}\right)^{1/2} + a_2\left(h^2_{t+h|t,d+n_i}\right)^{1/2} + u_{t+h} \qquad (20)$$

where the first explanatory variable on the right-hand-side is the *k*-step ahead forecast based on the day GARCH (1,1), and the second explanatory variable is the *k*-step ahead forecast based on day GARCH (1,1) with the additional explanatory variable, while the dependent variable is the *k*-step ahead realized volatility. We should observe statistically insignificant coefficient and significant coefficient for the first and second explanatory variables, respectively, if the latter can explain better the variation in dependent variable. In that case, the R^2 should also be higher than that of the regression with only the first explanatory variable.

Table 7 shows the result of the regressions for both conditional forecasts as the second explanatory variable. It is clear to see that when including the forecast of day GARCH (1,1) with the preopen variance, the R^2 increases from 0.098 to 0.171; while the R^2 for the addition of forecast based on day GARCH (1,1) with the overnight variance does not change. This is additional evidence that indicates the inclusion of preopen variance in the GARCH model would improve the forecastability of the day volatility process in terms of explaining the variability, while the inclusion of postclose variance does not. The lack of predictive power of the overnight squared return can be explained by Martens (2002), who concludes that the close-to-open squared return, like the day squared return, is a very noisy measure.

5. SEMIFAR Modeling and Forecasting

The semi- or nonparametric volatility measurements are data-driven without any specific functional form assumption. In this section, we use an alternative long memory model called the semiparametric fractional autoregressive (SEMIFAR) model proposed by Beran and Oker (1999). The model allows parametric estimation of a local spurious trend generated by stationary process with long-range dependence, a stochastic trend, and a nonparametric estimation of a deterministic trend. Parameters of the fractional differencing process can be estimated by maximum likelihood, while the deterministic trends can be estimated by kernel smoothing (Beran and Feng, 2002). As discussed in Section 3.3, we followed Andersen, Bollerslev, Diebold, and Labys (2001) and created the realized volatility time series. Based on the time series, which are shown in Figure 1, the SEMIFAR model appears to be a good candidate for modeling the volatility, since the time series exhibits both features of long memory and slightly downward historical trend.

5.1. The SEMIFAR Modeling

From the daily MSFT volatility time series in Figure 1, it can be seen that the volatility has become smaller in the later period. The SEMIFAR model is suitable for our study because it allows an estimation of a possible deterministic trend in the time series.

The specification of a SEMIFAR(p,d) is:

$$\phi(L)(1-L)^d[\sigma_t^2 - g(i_t)] = \varepsilon_t \tag{21}$$

for $t = 1,\ldots,T$, where $\phi(L)$ denotes the autoregressive component of order p, σ_t is the realized volatility, $g(i_t)$ is the deterministic trend function, i_t is t/T, and ε_t is the *i.i.d.* normal error term. The fractional difference filter, $(1-L)^d$ was introduced by Granger and Joyeux (1980) and Hosking (1981).

Hansen and Lunde (2005) suggest that the optimal linear combination of active realized volatility and inactive squared overnight returns is empirically solvable. Extending their work, we linearly combine realized volatilities for both regular and after-hours periods. To construct the time series, we treat each daily realized variance estimate as an observation. The procedure to construct the time series for modeling is described as follows. We first construct individual realized variance time series for the day, the preopen, the postclose, and the night periods. The night period is defined the same way as in the previous section. Since the magnitudes of realized variance during the day and the night are due to differing sampling frequency and fundamentally their differing environment, we need to normalize each time series by going through the transformation:

$$\varphi_{i,t}^2 = \frac{\upsilon_{i,t}^2 - \overline{\upsilon_{i,t}^2}}{SE\left(\upsilon_{i,t}^2\right)} \sim N(0,1) \tag{22}$$

where $\nu_{i,t}^2$ is the logarithm of realized variance time series and $\varphi_{i,t}^2$ is the normalized log realized variance time series for period i and time t. The objective is to construct the following four time series:

(i) A time series consisting of the day realized variance observations.

(ii) A time series consisting of alternating the day and the night realized variance observations.

(iii) A time series consisting of alternating the day and the preopen realized variance observations.

(iv) A time series consisting of alternating the day and the postclose realized variance observations.

Since the modeling of SEMIFAR assumes a normal distribution of the series and Andersen, Bollerslev, Diebold, and Ebens (2001) find that the distribution for the log realized variance series is approximately Gaussian, we take the logarithm of all the adjusted time series to obtain the log adjusted realized variance time series. We do not include a time series of the day realized variance with the overnight squared return for two reasons. The first is that since some values of overnight squared return are zero, taking the logarithm transformation of the series is impossible. The second is that the night time series that combine all three time series of the subperiods closely match that of the overnight squared return, since the weight during that period is much higher.

Table 8 shows the estimated parameters of *d* and the AR components. We see that the values of *d* for the models are all between 0 and 0.5, which indicate that there is long-range dependence in each of the series. The estimated *d* values based on the GPH test all tend to be higher, while those based on the Whittle test all tend to be lower, than the estimates from the SEMIFAR. The Ljung-Box test statistics all show that the residuals are all serially uncorrelated.

5.2. Forecast Evaluation

Table 9 (A) and (B) show the 1-step and 5-step-ahead predictions of various SEMIFAR models. The R^2 of the model with day and night variance is almost the same as that of the model with only day variance, which indicates that the information contained in the overnight squared return does not help to improve the forecastability of the day volatility process. The SEMIFAR with the day and the preopen provides the highest R^2 and the lowest statistics of the other forecasting criteria. This is consistent with the result from the GARCH models. Figure 4 also plots both the out-of-sample ex post realized volatility and 1-step-ahead forecast volatility.

Table 10 lists the results for encompassing regressions. The R^2 increases from 0.146 to 0.192 when the preopen variance is included in the SEMIFAR. The coefficient of the SEMIFAR with the day only is insignificant and that of the SEMIFAR with both the day and the preopen is significant, which indicates that the latter contains more information that the former. We also see that the SEMIFAR with the postclose does not contribute to any improvement in forecasting of the conditional day series.

6. Conclusion

Most of the volatility forecast literature has focused on comparing the forecast performance of different volatility models. In this study, we concentrate on whether an expanded information set can increase the forecastability of a day conditional volatility model. The usual information set other papers have used is the daily return and/or variance measures while the additional information we include in the information set is the after-hours variance.

We expand the daily GARCH and SEMIFAR models by including the additional information: the combined whole night, the preopen, the postclose variance, and the overnight squared return. By examining four NASDAQ stocks, MSFT, AMGN, CSCO, and YHOO, we find that the inclusion of the preopen variance can improve the out-of-sample forecastability of the conditional day volatility. The postclose variance and the overnight squared return, on the other hand, do not exhibit any predictive power on the future conditional volatility. The evidence supports the results of prior studies that traders trade for non-information reasons in the postclose period, while they trade for information reasons in the preopen period.

Moreover, we propose two reasons for why the preopen variance can be used to improve the predictability of the model. The first is the spillover effect, and the second is the possibility of the informed traders trading private information that is yet to be released during the following regular hours. One possible extension is to examine how the preopen variance affects the volatilities in different intraday periods. If the predictive power of the preopen variance comes from the spillover information from the peropen period to the regular hours, we can expect the highest impact to occur in the opening hours. If the time of day affected appears to be random, it is more likely due to the second conjecture.

The SEMIFAR model with the inclusion of night variance employed in this study offers another possibility of future extension. Mutual and hedge funds traders, international investors, and after-hours traders might be interested in how and whether the volatility outside of regular trading hours can be predicted. The SEMIFAR and other long memory models with the night variance constructed by using the proposed procedure in this paper can be used for such an analysis.

REFERENCES

Andersen, Torben G. and Tim Bollerslev, "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review* 39 (1998), 885-905.

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold, "Parametric and Nonparametric Volatility Measurement," *Working Paper*, NBER (2002).

Andersen, Torben G, Tim Bollerslev, Francis X. Diebold, and Heiko Ebens, "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics* 61 (2001), 43-76.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association* 96 (2001), 42-55.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, "Modeling and Forecasting Realized Volatility," *Econometrica* 71 (2003), 579-625.

Andersen, Torben G., Tim Bollerslev, and Steve Lange, "Forecasting Financial Market Volatility: Sample Frquency vis-à-vis Forecast Horizon," *Journal of Empirical Finance* 6 (1999), 457-477.

Barclay, Michael J. and Terrence Hendershott, "Price Discovery and Trading After Hours," *Review of Financial Studies* 16 (2003), 1041-73.

Barclay, Michael J. and Terrence Hendershott, "Liqudity Externalities and Adverse Selection: Evidence from Trading After Hours," *Journal of Finance* 59 (2004), 681-710.

Beran, Jan and Yuanhua Feng, "SEMIFAR Models: A Semiparametric Framework for Modelling Trends, Long-Range Dependence and Nonstationarity," *Computational Statistics and Data Analysis* 40 (2002), 393-419.

Beran, Jan and Dirk Ocker, "SEMIFAR Forecasts, with Applications to Foreign Exchange Rates," *Journal of Statistical Planning and Inference* 80 (1999), 137-153.

Beran, Jan and Dirk Ocker, "Volatility of Stock-Market Indexes – An Analysis Based on SEMIFAR Models," *Journal of Business & Economic Statistics* 19 (2001), 103-116.

Bollerslev, Tim, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* (1986), 307-327.

Bollerslev, Tim and Jonathan H. Wright, "High-Frequency Data, Frequency Domain Inference, and Volatility Forecasting," *The Review of Economics and Statistics* 83 (2001), 596-602.

Brooks, Chris, “Predicting Stock Index Volatility: Can Market Volume Help?” *Journal of Forecasting* 17 (1998), 59-80.

Cao, Charles, Eric Ghysels, and Frank Hatheway “Price Discovery without Trading: Evidence from the Nasdaq Preopening” *Journal of Finance* 55 (2000), 1339-1365.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu, “Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Rick,” *Journal of Finance* 56:1 (2001), 1-43.

Clark, Peter K., “A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices,” *Econometrica* 41 (1973), 135-155.

Cumby, Robert E., Stephen Figlewski, and Joel Hasbrouck, “Forecasting Volatility and Correlations with EGARCH Models,” *Journal of Derivatives* Winter (1993), 51-63.

Easley, David, Nicholas M. Kiefer, and Maureen O’Hara, “The Information Content of the Trading Process,” *Journal of Empirical Finance* 4 (1997), 159-186.

Engle, Robert F., “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation,” *Econometrica* 50 (1982), 987-1008.

Figlewski, Stephen, “Forecasting Volatility,” *Financial Markets, Institutions and Instruments* 6 (1997), 1-88.

Gallo, Giampiero M. and Barbara Pacini, “Early News is Good News: The Effects of Market Opening on Market Volatility,” *Studies in Nonlinear Dynamics and Econometrics* 2 (1998), 115-131.

Granger, Clive W.J. and Roselyne Joyeux, “An Introduction to Long-Memory Time Series Models and Fractional Differencing,” *Journal of Time Series Analysis* 1 (1980), 15-29.

Greene, Jason T. and Susan G. Watts, “Price Discovery on the NYSE and the Nasdaq: The Case of Overnight and Daytime News Releases,” *Financial Management* 25 (1996), 19-42.

Hansen, Peter R. and Asger Lunde, “A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data,” *Journal of Financial Econometrics* 3:4 (2005), 525-554.

Hosking, Johnathan RM, “Fractional Differencing,” *Biometricka* 68 (1981), 165-176.

Jorion, Philippe, “Predicting Volatility in the Foreign Exchange Market,” *The Journal of Finance* 50 (1995), 507-528.

Kalev, Petko S., Wai-man Liu, Peter K. Pham, and Elvis Jarnecic, “Public Information Arrival and Volatility of Intraday Stock Returns,” *Journal of Banking and Finance* 28 (2004), 1441-1467.

Lamoureux, Christopher G. and William D. Lastrapes, “Heterskedasticity in Stock Return Data: Volume versus GARCH Effects,” *The Journal of Finance* 45 (1990), 221-229.

Martens, Martin, “Measuring and Forecasting S&P 500 Index-Futures Volatility Using High-Frequency Data,” *The Journal of Futures Markets* 22 (2002), 497-518.

Masulis, Ronald W. and Lakshmanan Shivakumar, “Does Market Structure Affect the Immediacy of Stock Price Responses to News?” *Journal of Financial and Quantitative Analysis* 37 (2002), 617-648.

Mincer, Jacob, and Victor Zarnowitz, “The Evaluation of Economic Forecasts,” *Economic Forecasts and Expectations*, New York: NBER (1969).

Oldfield, George S. and Richard J. Rogalski, “A Theory of Common Stock returns Over Trading and Non-trading Periods,” *Journal of Finance* 35 (1983), 729-751.

Sharma, Jandhyala L., Mbodja Mougoue, and Ravindra Kamath, “Heteroscedasticity in Stock Market Indicator Return Data: Volume versus GARCH Effects,” *Applied Financial Economics* 6 (1996), 337-342.

Tauchen, George E. and Mark Pitts, “The Price Variability-Volume Relationship on Speculative Markets,” *Econometrica* 51 (1983), 485-505.

Taylor, Nicholas, “A Note on the Importance of Overnight Information in Risk Management Models,” *Journal of Banking and Finance* 31 (2007), 161-180.

Table 1. Summary Statistics of Daily Realized Return and Volatility for MSFT

	Min.	Mean	Median	Max.	St. dev.	Skew.	Kurt.
Return	-0.0771	0.0001	-0.0007	0.1058	0.0190	0.4731	5.3961
Volatility							
Reg. Hour	0.4183	1.7717	1.5852	6.4343	0.8706	1.4916	6.3789
Preopen	0.0367	0.5818	0.4700	5.7256	0.4269	3.2429	26.8472
Postclose	0.0330	0.5888	0.3473	12.5415	0.8427	5.9118	58.0698
Overnight	0.0000	0.0053	0.0008	0.2190	0.0155	7.4075	77.7413

The realized volatilities are all in percentage terms.

Table 2. Average Volume and Volatility for MSFT

	Volume (daily)	Volume (hourly)	Volatility (hourly)	Volatility (per trade)
Reg. Hour	978959	150609	0.273	0.000181
Preopen	5701	3801	0.388	0.0102
Postclose	9925	4963	0.294	0.00592

All reported volumes are in terms of trades, and all reported volatilities are in percentage term.

Table 3. GARCH Model Selection

	Normal Error Distribution			
	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	-4459	-4457	-4458	-4453
BIC	-4440	-4433	-4434	-4424
Likelihood	2233	2234	2234	2232
	Student's t Error Distribution			
	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	-4463	-4461	-4461	-4442
BIC	-4439	-4433	-4433	-4409
Likelihood	2237	2237	2237	2228

The GARCH selection is based on the Akaike information criterion (AIC), Bayesian information criterion (BIC), and log-likelihood of the model.

Table 4A. Day GARCH (1, 1) Parameter Estimates of MSFT

	GARCH (1,1)	GARCH (1,1) Close-to-Open	GARCH (1,1) Preopen	GARCH (1,1) Postclose	GARCH (1,1) Overnight
μ	-6.37e-4 (5.16e-4)	-5.85e-4 (5.18e-4)	-4.52e-4 (5.29e-4)	-6.24e-4 (5.21e-4)	-6.02e-4 (5.15e-4)
ω	1.32e-6 (1.25e-6)	8.85e-7 (1.56e-6)	-1.06e-6 (2.02e-6)	1.47e-6 (1.32e-6)	1.05e-6 (1.48e-6)
α	0.065*** (0.017)	0.066*** (0.018)	0.068*** (0.019)	0.066*** (0.017)	0.066*** (1.75e-2)
β	0.932*** (0.017)	0.922*** (0.020)	0.903*** (0.022)	0.932*** (0.017)	0.923*** (0.020)
ρ		0.065 (0.064)	0.221*** (0.090)	-0.007 (0.015)	0.057 (0.057)
Degree of Freedom	19.23	28.85	28.84	21.28	19.50
Likelihood	2237	2245	2240	2239	2238
ARCH test (P-value)	0.121	0.228	0.154	0.143	0.222
Ljung-Box Test (P-value)	0.638	0.660	0.535	0.645	0.657

The reported coefficients are based on quasi-maximum likelihood GARCH(1,1) model estimated from in-sample period:

$$r_t = \mu + \varepsilon_t$$

$$\varepsilon_t = z_t h_t$$

$$h_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}^2 + \rho x_{t-1}$$

The standard error is in the parenthesis. ARCH test and Ljung-Box Test are performed to check for the ARCH effect and autocorrelation of the residuals. *** denotes significance at 1% level, ** denotes significance at 5% level, and * denotes significance at 10% level.

Table 4B. Day GARCH (1, 1) Parameter Estimates of AMGN

	GARCH (1,1)	GARCH (1,1) Close-to-Open	GARCH (1,1) Preopen	GARCH (1,1) Postclose	GARCH (1,1) Overnight
μ	-7.82e-4 (6.13e-4)	-7.07e-4 (6.186e-4)	-7.00e-5 (6.51e-4)	-6.71e-4 (6.13e-4)	-6.80e-4 (6.18e-4)
ω	3.46e-6 (2.46e-6)	3.83e-6 (2.72e-6)	4.28e-6 (3.12e-6)	3.91e-6 (2.52e-6)	4.04e-6 (2.75e-6)
α	0.062*** (0.016)	0.062*** (0.017)	0.061*** (0.019)	0.065*** (0.016)	0.062*** (0.017)
β	0.929*** (0.017)	0.921*** (0.020)	0.907*** (0.023)	0.929*** (0.017)	0.920*** (0.019)
ρ		0.048 (0.046)	0.14* (0.081)	-0.014 (0.015)	0.050 (0.042)

Table 4C. Day GARCH (1, 1) Parameter Estimates of CSCO

	GARCH (1,1)	GARCH (1,1) Close-to-Open	GARCH (1,1) Preopen	GARCH (1,1) Postclose	GARCH (1,1) Overnight
μ	-8.64e-4 (7.87e-4)	-9.51e-4 (7.95e-4)	-8.45e-4 (7.98e-4)	-8.70e-4 (8.95e-4)	-9.08e-4 (7.92e-4)
ω	2.89e-6 (2.53e-6)	5.39e-6 (3.64e-6)	3.23e-6 (3.92e-6)	2.21e-6 (3.89e-6)	5.94e-6 (3.64e-6)
α	0.038*** (0.012)	0.024 (0.014)	0.041*** (0.015)	0.037** (0.015)	0.030 (0.014)
β	0.958*** (0.013)	0.934*** (0.020)	0.929*** (0.021)	0.949*** (0.016)	0.937*** (0.020)
ρ		0.202** (0.079)	0.155** (0.073)	0.063 (0.046)	0.144** (0.060)

Table 4D. Day GARCH (1, 1) Parameter Estimates of YHOO

	GARCH (1,1)	GARCH (1,1) Close-to-Open	GARCH (1,1) Preopen	GARCH (1,1) Postclose	GARCH (1,1) Overnight
μ	1.53e-3 (9.94e-4)	1.62e-3 (9.87e-4)	1.47e-3 (9.94e-4)	1.64e-3 (9.92e-4)	1.62e-3 (9.87e-4)
ω	5.028e-6 (4.20e-6)	5.57e-6 (4.44e-6)	7.26e-6 (5.98e-6)	1.75e-6 (3.75e-6)	5.57e-6 (4.43e-6)
α	0.048*** (0.013)	0.049*** (0.013)	0.046*** (0.016)	0.036*** (0.011)	0.049*** (0.013)
β	0.948*** (0.013)	0.946*** (0.014)	0.920*** (0.022)	0.957*** (0.011)	0.946*** (0.014)
ρ		-4.31e-4 (7.79e-4)	0.163** (0.083)	0.037 (0.026)	-3.49e-4 (6.21e-4)

Table 5A. Forecasting Evaluation Methods for 1-Step Ahead Prediction by GARCH Models for MSFT

	Day GARCH (1,1)	Day GARCH (1,1) Close-to-Open	Day GARCH (1,1) Preopen
MZ Regression			
a_0	0.097 (0.122)	0.166 (0.124)	0.033 (0.141)
a_1	0.899 (0.130)	0.759** (0.134)	0.976 (0.175)
Adj. R^2	0.091	0.091	0.157
RMSE	0.440	0.452	0.424
MAE	0.206	0.244	0.202
HRMSE	0.442	0.399	0.391
HRMAE	0.208	0.223	0.204
LL	-0.049	-0.133	-0.029

Table 5B. Forecasting Evaluation Methods for 5-Step Ahead Prediction by GARCH Models for MSFT

	Day GARCH (1,1)	Day GARCH (1,1) Close-to-Open	Day GARCH (1,1) Preopen
MZ Regression			
a_0	0.210 (0.201)	0.293 (0.183)	0.088 (0.048)
a_1	0.703 (0.180)	0.586*** (0.166)	0.850*** (0.056)
Adj. R^2	0.031	0.029	0.073
RMSE	0.476	0.502	0.456
MAE	0.275	0.320	0.250
HRMSE	0.437	0.420	0.401
HRMAE	0.251	0.271	0.234
LL	-0.164	-0.230	-0.118

Values for RMSE, MAE, HRMSE, and HRMAE are in percentage term. *** denotes significance at 1% level and ** denotes significance at 5% level. H_0: $a_0 = 0$, $a_1 = 1$.
Reported in parenthesis are the White's heteroskedasticity-consistent standard deviation.

Table 6. Forecast Evaluation Statistics for 1-Step Ahead Prediction By GARCH (1,1) for Amgen, Cisco, and Yahoo

	AMGN		CSCO		YHOO	
	Day GARCH (1,1)	Day GARCH (1,1) Preopen	Day GARCH (1,1)	Day GARCH (1,1) Preopen	Day GARCH (1,1)	Day GARCH (1,1) Preopen
MZ Regression						
a0	0.676*** (0.076)	0.343 (0.335)	-0.403 (0.796)	-0.499 (0.335)	0.148 (0.107)	-0.219 (0.069)
a1	0.512*** (0.049)	0.741 (0.245)	1.149 (0.461)	1.217 (0.245)	0.782*** (0.020)	0.940*** (0.013)
Adj. R2	0.017	0.061	0.039	0.054	0.209	0.373
RMSE	0.456	0.442	0.673	0.664	0.597	0.551
MAE	0.303	0.288	0.376	0.409	0.487	0.450
HRMSE	0.336	0.318	0.391	0.338	0.251	0.234
HRMAE	0.229	0.212	0.219	0.231	0.207	0.193
LL	-0.015	-0.043	-0.137	-0.126	-0.196	-0.193

*** denotes significance at 1% level, while * denotes significance at 5% level. H_0: $a_0 = 0$, $a_1 = 1$.

Table 7. Encompassing Regression of GARCH(1,1) Models for MSFT

	Day GARCH(1,1)	Day Garch(1,1) with (1)	Day Garch(1,1) with (2)
a_0	0.097 (0.122)	0.122 (0.138)	0.170 (0.121)
a_1	0.899*** (0.130)	0.411 (1.070)	-0.511 (0.617)
a_2		0.424 (0.945)	1.353*** (0.599)
R^2	0.098	0.099	0.171
Adj. R^2	0.091	0.085	0.158

The regression is of the form

$$\left(\sigma_{t+k,d}^{2}\right)^{1/2} = a_0 + a_1\left(h_{t+k|t,d}^{2}\right)^{1/2} + a_2\left(h_{t+k|t,d+n_i}^{2}\right)^{1/2} + u_{t+k}$$

where d denotes day period, n_i denotes for the whole night or the preoen period. All values are in percentage term. *** denotes significance at 1% level and ** denotes significance at 5% level.
In parenthesis are the White's heteroskedasticity-consistent standard deviation.

Table 8. SEMIFAR Parameter Estimates for MSFT In-Sample Period

	SEMIFAR with Day only	SEMIFAR with Day and Night	SEMIFAR with Day and Preopen	SEMIFAR with Day and Postclose
d	0.468*** (0.040)	0.349*** (0.044)	0.412*** (0.044)	0.290*** (0.053)
AR(1)	-0.107** (0.050)	-0.340*** (0.048)	-0.335*** (0.049)	-0.320*** (0.057)
AR(2)		-0.031 (0.044)	-0.098** (0.044)	0.018 (0.049)
AR(3)		-0.152*** (0.032)	-0.172*** (0.034)	-0.144*** (0.034)
AR(4)		0.029 (0.033)	0.008 (0.034)	0.109*** (0.036)
AR(5)		-0.139*** (0.025)	-0.111*** (0.026)	-0.102*** (0.027)
AR(6)				0.022 (0.030)
AR(7)				-0.114*** (0.025)
BIC	1118	5290	4415	5604
GPH Test	0.678	0.684	0.688	0.563
Whittle Test	0.419	0.184	0.249	0.109
ADF Test (P-value)	0.010	6.36e-16	3.93e-12	1.11e-18
LB Test (P-value)	0.722	0.137	0.681	0.176

*** denotes significance at 1% level and ** denotes significance at 5% level.

Table 9A. Forecasting Evaluation Methods for 5-Step Ahead Prediction of SEMIFAR Models for MSFT

	SEMIFAR with Day only	SEMIFAR with Day and Day and Night	SEMIFAR with Day and Preopen	SEMIFAR with Day and Postclose
a_0	-0.067 (0.227)	0.036 (0.210)	-0.006 (0.239)	2.086*** (1.017)
a_1	1.082 (0.263)	0.902 (0.199)	1.054 (0.287)	-1.356** (1.207)
Adj. R^2	0.092	0.035	0.095	0.018
RMSE	0.439	0.450	0.422	0.499
MAE	0.211	0.219	0.207	0.246
HRMSE	0.296	0.292	0.260	0.283
HMAE	0.214	0.215	0.199	0.217
LL	-0.038	-0.007	-0.004	0.114

Values for RMSE, MAE, HRMSE, and HRMAE are in percentage term. *** denotes significance at 1% level and ** denotes significance at 5% level. H_0: $a_0 = 0$, $a_1 = 1$.
Reported in parenthesis are the White's heteroskedasticity-consistent standard deviation.

Table 9B. Forecasting Evaluation Methods for 5-Step Ahead Prediction of SEMIFAR Models for MSFT

	SEMIFAR with Day only	SEMIFAR with Day and Night	SEMIFAR with Day and Preopen	SEMIFAR with Day and Postclose
a_0	-0.067 (0.227)	0.036 (0.210)	-0.006 (0.239)	2.086*** (1.017)
a_1	1.082 (0.263)	0.902 (0.199)	1.054 (0.287)	-1.356** (1.207)
Adj. R^2	0.092	0.035	0.095	0.018
RMSE	0.439	0.450	0.422	0.499
MAE	0.211	0.219	0.207	0.246
HRMSE	0.296	0.292	0.260	0.283
HMAE	0.214	0.215	0.199	0.217
LL	-0.038	-0.007	-0.004	0.114

Table 10. Encompassing Regression for Different SEMIFAR Models

	Day only	Day with Night	Day with Preopen	Day with Postclose
a_0	0.098 (0.282)	-0.060 (0.229)	-0.259 (0.231)	0.124 (0.244)
a_1	0.919** (0.324)	0.554 (0.348)	-0.061 (0.415)	0.954** (0.284)
a_2		0.541 (0.424)	1.321** (0.496)	-0.062 (0.364)
R^2	0.146	0.157	0.192	0.147
Adj. R^2	0.139	0.144	0.179	0.133

The regression is of the form

$$\left(\sigma^2_{t+k,d}\right)^{1/2} = a_0 + a_1\left(h^2_{t+k|t,d}\right)^{1/2} + a_2\left(h^2_{t+k|t,d+n_i}\right)^{1/2} + u_{t+k}$$

where d denotes day period, n_i denotes for the whole night or the preoen period. All values are in percentage term. ** denotes significance at 1% level, while * denotes significance at 5% level.
In parenthesis are the White's heteroskedasticity-consistent standard deviation.

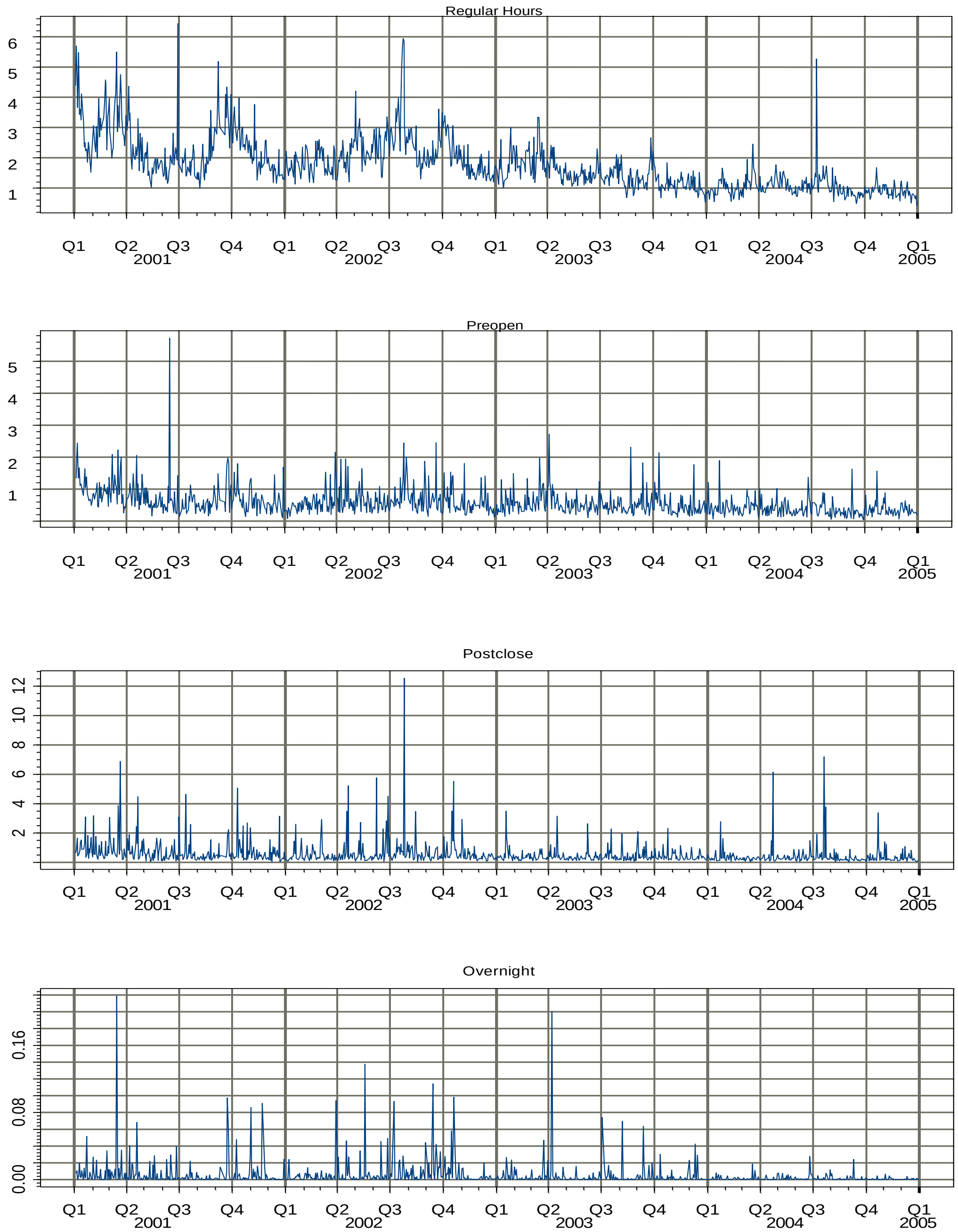


Figure 1. Time Series of the MSFT Volatilities for Different Time Periods
These are realized volatilities for regular Hours, preopen, and postclose, and square root of overnight returns. The volatilities are in percentages.

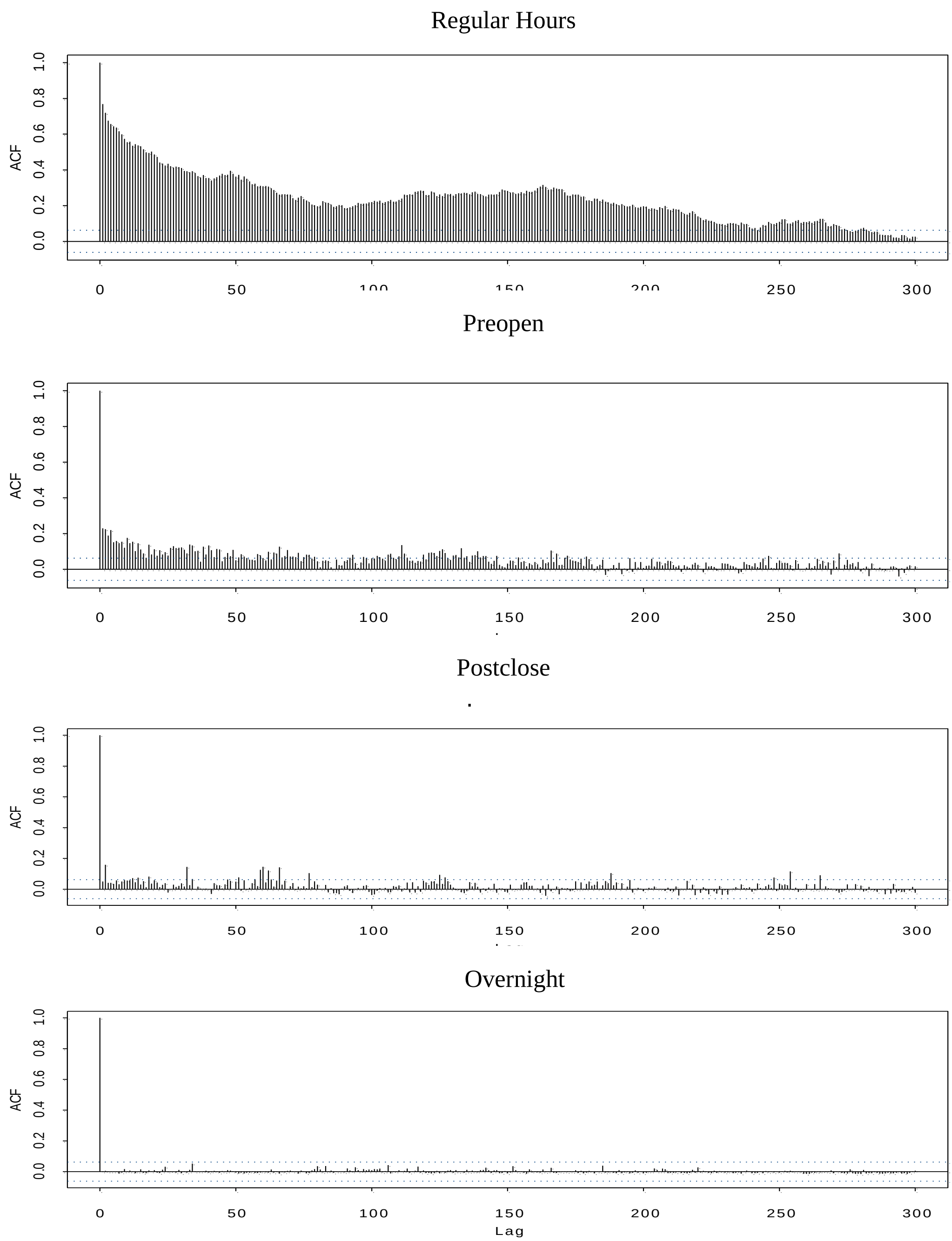
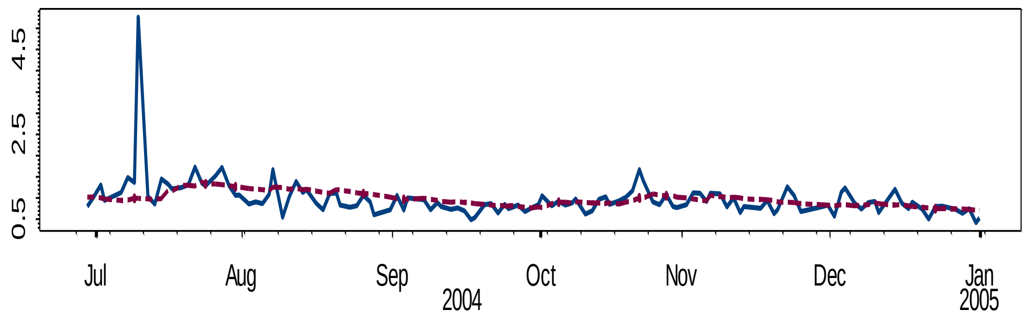
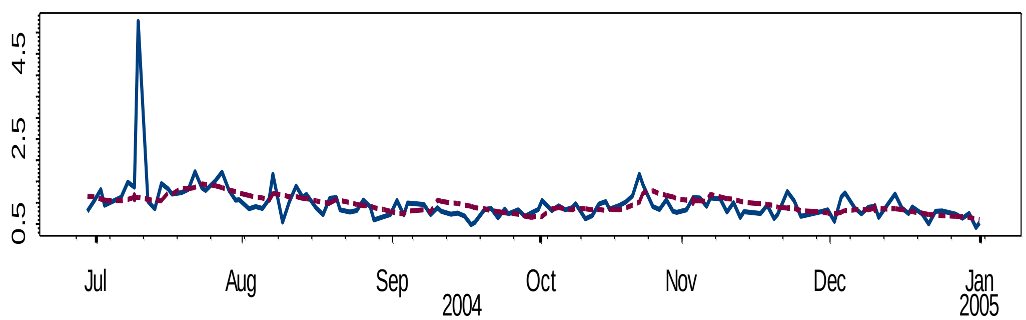


Figure 2. Autocorrelations of the MSFT Volatilities for Different Time Periods
These are the autocorrelations of realized volatilities for regular Hours, preopen, and postclose, and square root of overnight returns.

GARCH(1,1)



GARCH(1,1) with Preopen Variance



GARCH (1,1) with Close-to-Open Variance

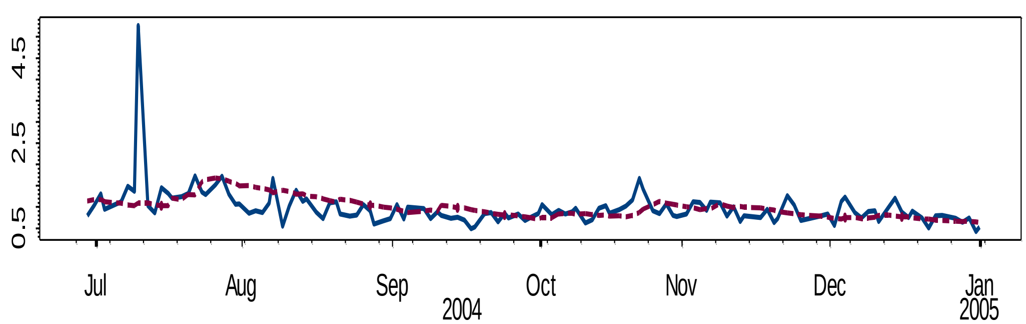


Figure 3. MSFT One-Step Ahead Volatility Forecasting for Day GARCH(1,1) by Different Models

The solid line represents the ex post realized volatility series, and the break line represents the forecast conditional volatility.

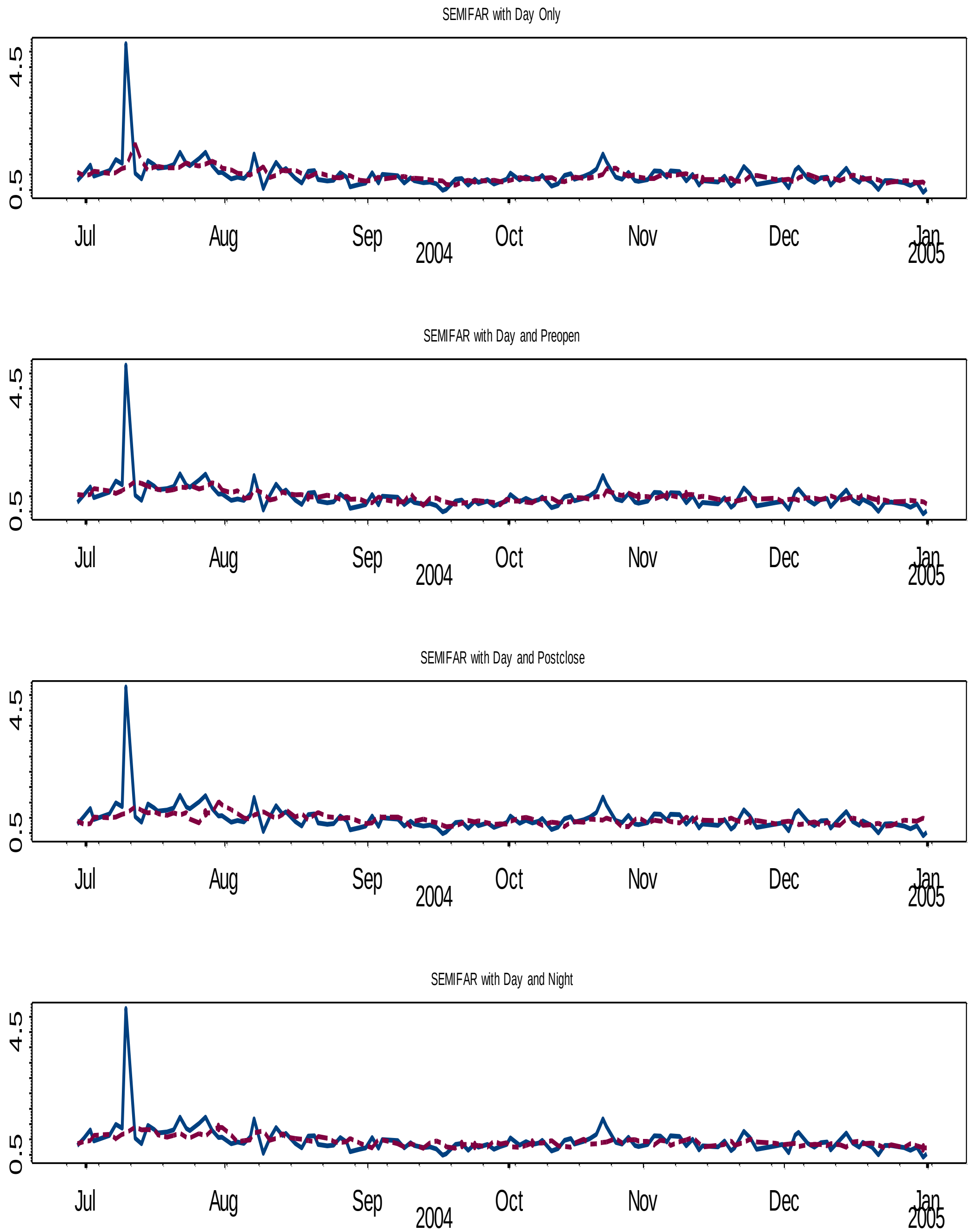


Figure 4. MSFT One-Step Ahead Forecasting for Different SEMIFAR Models

The solid line represents the ex post realized volatility series, and the break line represents the forecast conditional volatility.