

Predicting Stock Volatility Using After-Hours Information

Chun-Hung Chen¹

KPMG

Wei-Choun Yu²

Winona State University

Eric Zivot³

University of Washington

January 12, 2009

Abstract

We use realized volatilities based on after hours high frequency returns to predict next day volatility. We extend GARCH and long-memory 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) we find that the inclusion of the preopen variance can improve the out-of-sample forecastability of the next day conditional day volatility. Additionally, we find that the postclose variance and the overnight squared return do not provide any predictive power for the next day conditional volatility. Our findings support the results of prior studies that traders trade for non-information reasons in the postclose period and trade for information reasons in the preopen period.

¹Corresponding 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.

³Department of Economics, University of Washington, Box 353330, Condon 401, 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 models are the GARCH models developed by Engle (1982) and Bollerslev (1986). This family of models can explain well the stylized facts of financial return volatility: persistence, mean reversion, and the leverage effect. Moreover, as Andersen and Bollerslev (1998) and others have shown, the GARCH model can produce good forecasts of daily conditional volatility. Good modeling and forecasting of volatility, however, relies on a useful information set. Until recently, the most commonly used information set for modeling daily volatility is historical daily closing prices. However, recent research (e.g., Andersen, and Bollerslev, 1998, and Andersen, Bollerslev, and Lange, 1999) has shown that the use of intra-day high frequency data can substantially improve the measurement and forecastability of daily volatility. The majority of these studies used intra-day data observed during normal trading hours. We add to this literature by considering intra-day data observed during after-hours periods.

Although several studies have documented the importance of after-hours information (e.g. Oldfield and Rogalski, 1980; Greene and Watts, 1996; Cao et al., 2000; Masulis and Shivakumar, 2002; Taylor, 2007; and Tsiakas, 2008) for volatility modeling, only a few actually employed

high-frequency data in the analysis. One such paper is Taylor (2007), which used high-frequency overnight S&P 500 futures volatility information to predict S&P 500 stocks volatility. In this paper, we utilize the availability of after-hours trading opportunities to the public and the recording of high frequency after-hours transaction data of the NASDAQ stocks to examine how this extended information set could be effectively used to improve the modeling and the forecasting of next day conditional volatility.

We use two volatility forecasting models in our analysis: (1) generalized autoregressive conditional heteroskedasticity (GARCH) models for daily returns with after-hours realized variance as an exogenous variable included in the conditional variance equation; (2) flexible trend semiparametric fractionally integrated autoregressive (SEMIFAR) models for realized variance. The high frequency data used in the paper are the historical quotes and trade prices of selected stocks listed on the NASDAQ that have active overnight trading.

We use realized variance measures as proxies for unobserved volatility in our forecasting evaluation. Our results show that the inclusion of realized volatility for the whole night in the information set does not provide better forecasting of next day volatility from either the GARCH or the SEMIFAR models. When breaking up the whole after-hours period into three sub periods, we find that only the inclusion of the preopen period realized variance significantly improves the forecastability of future daily volatility.

Our study contributes to the existing literature in the following ways. First, we use high-frequency intranightly transaction data that has not yet been systematically exploited for the modeling and forecasting of daily volatility. 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, past research has often focused on in-sample forecasting evaluation while we also evaluate our models' predictive ability out-of-sample. Finally, we employ two different types of volatility models: a parametric model (GARCH) for squared returns and a

semiparametric long memory model (SEMIFAR) for realized variance. The employment of models under different forms and assumptions serves as a robustness check for our results.

The rest of the paper proceeds as follows. Section 2 reviews the volatility literature using after hours information. Section 3 explains our data and realized volatility construction. The results of modeling and forecasting conditional volatility based on the GARCH and the SEMIFAR models are provided in Sections 4 and 5. Section 6 contains our concluding remarks.

2. Literature Review

Conditional return volatility models such as GARCH demonstrate that past return shocks and volatilities contain information about the evolution of future volatilities, and therefore can be used for forecasting purposes. One explanation for this result is the Mixture of Distribution Hypothesis (MDH), suggested by Clark (1973), Tauchen and Pitts (1983), and Kalev, Liu, Pham, and Jarnecic (2004). They attributed dependence in volatilities to the serial correlation of the news arrival rate, and the highly autocorrelated nature of volatility comes from the persistence in information arrival rates.

Lamoureux and Lastrapes (1990) used the MDH proposed by Clark (1973) to explain the persistent nature of daily conditional volatility in the GARCH model⁴. They assumed that a stochastic model can be derived by considering the daily return in day t , ε_t , as a sum of *i.i.d.* $(0, \sigma^2)$ intraday price increments, δ_{it} ,

$$\varepsilon_t = \sum_{i=1}^{n_t} \delta_{it}, \quad (1)$$

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

⁴ For purpose of this study, we define day as the regular trading hours, i.e. from 9:30am to 4:00pm EST.

$$\varepsilon_t | n_t \sim N(0, \sigma^2 n_t). \quad (2)$$

When the arrival rate of information is serially correlated, n_t can be expressed as

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

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

$$\Omega_t = E[\varepsilon_t^2 | n_t] = \sigma^2 n_t = a\sigma^2 + b(L)\Omega_{t-1} + \sigma^2 u_t \quad (4)$$

Which demonstrates the persistence in the conditional variance captured in the GARCH model.

To examine this hypothesis, Lamoureux and Lastrapes (1990), Sharma, Mougoue, and Kamath (1996), and Brooks (1998) used trading volume as a proxy for the information arrival rate and included it as an exogenous variable in the GARCH (1,1) specification for daily volatility. They showed that the inclusion of volume greatly reduced the persistence parameter of the estimated GARCH model. Moreover, Brooks (1998) showed that including trading volume in a GARCH model does not improve volatility forecasts because no new information is provided which is not already captured by past conditional volatility.

The after hours time period is from the previous market closing time through the next market opening time⁵. If the after-hours volatility provides additional information rather than substitutes for information already incorporated in past conditional volatility or volume, it could be used to improve forecasts of next day volatility. It is well known in the microstructure literature that information and announcements frequently occur during after normal trading hours, regardless of the existence of trading opportunities 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.

⁵ We use after hours, close-to-open, and whole night period interchangeably in this study.

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 the 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) indicated 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 that a highly volatile night trading would lead to a highly volatile day trading in the next day.

Gallo and Pacini (1998) studied 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 using a GARCH (1,1) model with the close-to-open returns as an exogenous variable. Martens (2002) studied 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) found that the inclusion of close-to-open returns improves forecastability of conditional volatility for some stock indices, while Martens (2002) found that the inclusion of the close-to-open squared returns cannot improve forecastability. This mixed evidence could come from the poor exploitation of after-hours information. We utilize more information in the after hours period than was used in previous studies. In particular, we segment the after-hours period based on information density. This segmentation of information was proposed by Barclay and Hendershott (2003, 2004) and we are the first to apply to forecasting volatility.

The precise measurement of volatility is important for evaluating volatility forecasts. Past studies, such as Cumby, Figlewski and Hasbrouck (1993), Figlewski (1997), and Jorion (1995), have shown that standard volatility models such as GARCH perform poorly in terms of out-of-sample forecasting when squared returns are used to proxy volatility. Andersen and Bollerslev (1998) pointed out that while the squared return is an unbiased estimate for unobserved volatility, it is a very noisy estimate and this can explain why volatility models appear to produce poor forecasts. They showed that realized volatility, which is defined as the sum of squared returns sampled at high intradaily frequency, provides a much more reliable ex post volatility measure than squared returns and that GARCH forecasts evaluated against realized volatility can be very accurate. Furthermore, Andersen, Bollerslev, and Lange (1999) showed that the forecasting performance of standard volatility models can be greatly improved by utilizing high frequency data. Indeed, Andersen, Bollerslev, Diebold, and Laybs (2003) used daily realized volatility as data and constructed an autoregressive fractional integrated moving average (ARFIMA) model to forecast future daily realized volatility. They found this model to be superior to many other volatility models in terms of out-of-sample forecasting performance. In our analysis, we expand on the previous literature by employing realized volatility measures that utilize intradaily and intranightly data.

3. Data and Volatility Measurement

In this section we describe our data set, how we segment information in the after hours period, and how we construct our volatility measures. Unlike Taylor (2007) and Tsiaks (2008), we study individual stocks instead of market indices. As mentioned in Campbell et al. (2001), there are several motives for studying 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.

3.1. Data

Our high frequency data is taken from the Trade and Quote (TAQ) database, which provides data on tick-level transaction prices and quotes from 8 am until 6:30 pm EST, when the NASDAQ Trade Dissemination Service (NTDS) is on. Since trading volume is relatively low for stocks in after hours, we have chosen stocks that show the highest liquidity during the after-hours period. The stocks we consider are Microsoft (MSFT), Amgen (AMGN), Cisco (CSCO), and Yahoo (YHOO) listed on the NASDAQ. We use MSFT as our benchmark stock and focus most of our analysis and discussion on results pertaining to MSFT. We mainly use the other three stocks for a robustness check. The sampling period is from January 2001 to December 2004, during which time 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.

The TAQ data typically contain a lot of recording errors. We remove any recorded trades that have a change of positive or negative 25% from their immediately prior trades in a day⁶. We also remove dates in which either the preopen, postclose, or day transaction data is missing as well as the occurrence of stock splits.

3.1. After-Hours Subperiods

Barclay and Hendershott (2003, 2004) broke the entire 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 investigated the information structure of the postclose and preopen and found that the probability of an informed trade is much higher in the latter period than the former period. They found 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.⁷ This implies

⁶ The daily volume weighted price for MSFT is \$39.44 for the sample period, and 25% of which would be about \$10.

⁷ Note, however, that this trading at close activity only represents 15 percent of trades in postclose.

that traders tend to trade for liquidity demands right after the regular trading hour is closed. Furthermore, they used 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 NTDS. Barclay and Hendershott (2003) used their proprietary dataset and found that only 1% of total after-hours trades occur during that period.

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.

3.3. Volatility Measurement

Realized variance is a more accurate measure of conditional variance than the squared return. We use it to measure trading day variance and variance during the after-hours periods. We use it to evaluate our volatility predictions and we use it as a historical volatility series from which we build a direct volatility forecasting model. Following Bollerslev and Wright (2001), Andersen *et al.* (2001), and Andersen *et al.* (2003) we construct realized variance by summing up intra-period high frequency squared returns:

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

where p denotes the logarithmic stock price; i denotes either the regular hour, the preopen, or the postclose period; r is the intraday return; $1/\Delta$ is the number of observations for each of the periods (Δ is 5 minutes in regular hours and is 15 minutes in after-hours); and $\sigma_{i,t}^2$ is the estimated realized variance for period i in day t . Realized volatility is computed as the square root of realized variance. 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:

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

Andersen *et al.* (2001) showed that as sampling frequency increases, realized variance accurately measures integrated variance, which is the actual realized return variation over a given horizon for a continuous time diffusion process and is an unbiased estimate of conditional variance.

Although theory demonstrates that the measurement error associated with the estimation of the realized variance becomes very small as the sampling frequency increases, market microstructure frictions (e.g. bid-ask bounce, price discreteness, and infrequent trading) create large biases. To avoid this problem, Andersen *et al.* (2001) proposed sampling the intradaily observations at 5 minute intervals⁸. Since the trading environment after hours is known to have much larger microstructure frictions than during regular hours, we sample observations at a 5-minute frequency for regular-hours and at a 15-minute frequency for after-hours. Therefore the number of inter-period observation for the regular hours, the preopen, and the postclose is 78, 6, and 8, respectively.

⁸ This rule of thumb is often close to the optimal sampling frequency advocated by Bandi and Russell (2008).

Figure 1 shows the regular hours, the preopen, and the postclose realized volatility, and the overnight absolute return for MSFT from January 2001 to December 2004. Table 1 lists some descriptive statistics of realized volatility measures for 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. 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. In contrast, we do not observe this feature in the overnight absolute return.

Barclay and Hendershott (2003) found that 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%, respectively, and the average volatilities per trade are 0.0102% and 0.00592%, respectively. The numbers show that volatility in the preopen is higher than 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.

4. GARCH Modeling and Forecasting

The GARCH framework is the most common approach to modeling and forecasting volatility.

We use the GARCH(p, q) model⁹

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

⁹ Including leverage effects (e.g. EGARCH, GJR GARCH, or TGARCH) might improve the volatility forecasting. 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.

where μ and ω are constants in the conditional mean equation and the conditional variance equation, respectively; ε_t is a serially uncorrelated residual term (news shock) with mean zero; z_t is an *i.i.d.* random variable with mean zero and unit variance ; and h_t^2 is the conditional variance at time t . While there are many variations of the GARCH(p,q) model, a GARCH (1,1) model is usually sufficient for most financial time series applications (Andersen and Bollerslev, 1998; Hansen and Lunde, 2004).

Table 3 shows the Akaike information criterion, Bayesian information criterion, and log-likelihood for all GARCH (p,q) models with $p \leq 2$ and $q \leq 2$ 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 for z_t appears to be the most appropriate model. The first column of Table 4 (A) reports the coefficients for the daily GARCH(1,1) model in regular hours. The sum of the estimates 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. The results of the GARCH(1,1) model fit to the daily returns of AMGN, CSCO, and YHOO are reported in Table 4 (B), (C), and (D), respectively. These results are similar to those for MSFT.

4.1. GARCH Model for Day Returns with Night Variance

The GARCH model offers flexibility in that 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) is:

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

where x_t represents an additional exogenous variable in the conditional variance equation. Both Gallo and Pacini (1998) and Martens (2002) used this approach by including the close-to-open squared return as the additional exogenous variable in the conditional variance equation. Martens (2002) found the coefficient on the additional variable to be statistically insignificant. In contrast, Gallo and Pacini (1998) found the coefficient to be statistically significant for most of the major market indices but the sign of the coefficient was positive for some indices and negative for others. 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 higher after-hours volatility should lead to a higher following day volatility.

To investigate the impact of after hours information in the GARCH(1,1) model (9), we use the following four exogenous variables:

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

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

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

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

where PO, PC, and ON denote the preopen, the postclose, and the overnight period, respectively. The variable x_t defined in (10) is a time-weighted average realized variance of the close-to-open (whole night) period.

The second through fifth columns of Table 4 (A) show the estimation results of the modified GARCH(1,1) model (9) using the exogenous variables defined in (10) – (13) for MSFT. First, we find that the estimated coefficient (standard error) for the whole night period realized variance (10)

is 0.065 (0.064), which is positive but statistically insignificant. This result agrees with Martens (2002) in that the close-to-open variance does provide explanatory power for the next day conditional variance. Second, we find that the coefficient for the postclose variance (12) is negative and statistically insignificant as well. This result is consistent with the hypothesis that traders primarily trade in the postclose for non-information reasons, and therefore there is no information to be carried over into the next day volatility. Another possibility is that if there was any information in the postclose period, it could be spilled over to and removed during the following overnight period and the preopen period. Third, the only explanatory variable that we find to be statistically significant is the preopen realized variance (11), which has an estimated coefficient (standard error) of 0.221 (0.090). Hence, a 1% increase in the preopen realized variance would lead to a 0.221% increase in the following regular hour conditional variance. This result is consistent with our hypothesis that the coefficient should be positive and significant. Finally, we find that the estimated persistence parameters are 0.997 and 0.971 for the GARCH (1,1) model (8) and the modified GARCH (1,1) model (9) with the preopen variance, respectively. This slight decrease in the persistence parameter shows that the preopen variance appears to provide independent information from that contained in the past day returns. Along with the coefficient being statistically significant, this result enhances our hypothesis that the addition of the preopen variance into the model would improve forecasts of the next day conditional day volatility.

Tables 4 (B), (C), and (D) give the estimation results of the modified GARCH(1,1) model (9) for three other stocks. The results show significant and positive coefficients on the preopen variance (11) 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. 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.

4.2. Forecast Evaluation

Figure 3 shows the out-of-sample ex post realized volatility series and the forecasted 1-step ahead conditional volatility series of various fitted GARCH (1,1) models for MSFT. The forecasts with the pre-open variance as an exogenous variable in the conditional variance equation appear to track the realized volatility series the best.

We formally evaluate the forecasting performance of the different models using the following metrics:

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

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

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

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

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

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

where σ_{t+k}^2 denotes realized variance in day $t+k$, and $h_{t+k|t}^2$ denotes the conditional variance forecast for day $t+k$ based on information available in day t . In the Mincer-Zarnowitz regression, proposed by Mincer and Zarnowitz (1969), 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 significance. However, Andersen and Bollerslev (1998) pointed out that the coefficients suffer from a standard errors-in-variables problem which makes interpretation difficult. Nonetheless, they argue that the R^2 of the regression can be used to evaluate the variability of the ex post

volatility that is explained by the forecasted conditional volatility. The root mean square error (RMSE) and the mean absolute error (MAE) are two commonly used criteria. To check whether the results are reliable in the nonlinear and heteroskedastic environment, we follow Andersen, Bollerslev, and Lange (1999) and use the heteroskedasticity adjusted RMSE (HRMSE), MAE (HMAE), and the logarithmic loss function (LL).

As stressed by Andersen and Bollerslev (1998) and Hansen and Lunde (2006), it is crucial to choose the right ex post volatility measure to serve as the benchmark for the forecast 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 concluded that standard volatility models explain little of the variability in the ex post volatility. Andersen and Bollerslev (1998) and Hansen and Lunde (2006) used realized volatility and demonstrated that it provides a more reliable and accurate measure of the true volatility and its use in forecast evaluation statistics leads to more accurate inferences regarding forecasting accuracy.

To perform the forecast evaluations, we first estimate the parameters of the GARCH(1,1) models (9) and (10) from the in-sample data, and then compute 1-step-ahead predictions for conditional volatility over 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 (14), we see that the GARCH (1,1) with the preopen realized variance provides the best forecasting performance both in terms of accuracy and explaining the variability in the ex post measures. The estimated coefficient on a_1 is 0.976 when forecasts are computed from a GARCH (1,1) model with preopen realized variance, compared to 0.899 for forecasts computed from the day GARCH (1,1) model. The GARCH (1,1) with the preopen realized variance has a substantially higher $R^2 = 0.157$ than the $R^2 = 0.091$ from the day GARCH (1,1). Based on the forecast evaluation statistics (15)-(18), the forecasts provided by the GARCH (1,1) with the preopen realized variance are always superior to those from the day GARCH (1,1). We also see that the forecasts from the

GARCH (1,1) with the close-to-open period realized variance always perform relatively poorly compared to those from the day GARCH (1,1). Because the coefficient for the realized variance during the close-to-open period is not statistically significant in the GARCH(1,1) model, it is not surprising to see that the forecast shows relatively poor performance and no improvement in terms of R^2 . We conclude that the after-hours volatility does not have any 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. Because the coefficients of the postclose and the overnight realized variances in the GARCH (1,1) are not statistically significant, we do not perform the forecasting evaluation for these two models. Table 5 (B) provides forecast evaluation results of 5-step-ahead predictions, and the outcomes are qualitatively identical.

Table 6 shows the forecast evaluation results for 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. The R^2 values of the day GARCH(1,1) with the preopen realized 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 also support our hypothesis that the information contained in the preopen realized variance improves the forecastability of the conditional day volatility.

In addition to the Mincer-Zarnowitz regression, we follow Andersen et al. (2003) and perform an encompassing regression to evaluate the importance of including the after hours realized variance measures as explanatory variables in the GARCH(1,1) model. The encompassing regression takes the form

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

where the first explanatory variable on the right-hand-side of (20) is the k -step ahead forecast based on the day GARCH (1,1) model, and the second explanatory variable is the k -step ahead forecast based on the day GARCH (1,1) model with an after hours realized variance measure as

an explanatory variable. We should observe a statistically insignificant coefficient and a significant coefficient on the first and second explanatory variables in (20), respectively, if the latter can explain better the variation in dependent variable. In that case, the adjusted R^2 from the regression should also be higher than that of the regression with only the first explanatory variable. Table 7 shows the estimation results of (20) with conditional forecasts using (11) and (13) as the second explanatory variable, respectively. Including the forecast of day GARCH (1,1) with the preopen variance increases the R^2 from 0.098 to 0.171, whereas including the forecast based on day GARCH (1,1) with the overnight variance does not change the R^2 . The lack of predictive power of the overnight squared return, as explained by Martens (2002), can be attributed to the fact that the close-to-open squared return, like the day squared return, is a very noisy measure of overnight conditional variance.

5. SEMIFAR Modeling and Forecasting

In this section, we evaluate the use of after hours information for predicting next day volatility by modeling and forecasting the observed series of realized volatility directly. Motivated by the results in Andersen et al. (2001), we model log realized volatility as a long memory process. In particular, we use the SEMIFAR model proposed by Beran and Oker (1999, 2001) and Beran and Feng (2002). The SEMIFAR model allows for a smooth deterministic trend, a stochastic trend, long memory and short memory components. From Figures 1 and 2, the SEMIFAR model appears to be a good candidate for modeling realized volatility as the time series exhibits features of long memory as well as a displaying a slightly downward historical trend.

5.1. The SEMIFAR Modeling

The SEMIFAR(p, d) model for a time series y_t is given by:

$$\phi(L)(1-L)^d [(1-L)^m y_t - g(i_t)] = \varepsilon_t, t = 1, \dots, T \quad (21)$$

where $-\frac{1}{2} < d < \frac{1}{2}$ is the fractional difference parameter, m is an integer giving the number of times that y_t must be differenced to achieve stationarity, $\phi(L)$ is a stationary autoregressive polynomial of order p , $g(i_t)$ is a smooth deterministic trend function on $[0,1]$ with $i_t = t/T$, and ε_t is an *i.i.d.* normal error term. In (21), the trend function $g(i_t)$ is estimated using a nonparametric kernel estimate.

To construct the realized volatility time series y_t to be fit by the SEMIFAR model, we treat each daily realized variance estimate as an observation. We first construct individual realized variance time series for the day, the preopen, the postclose, and the night periods, respectively, as described in sub-section 4.1. Because the magnitudes of realized variance during the day and the night are different due to differing sampling frequency and fundamentally their differing environment, we normalize each time series using the transformation:

$$\varphi_{i,t}^2 = \frac{v_{i,t}^2 - \overline{v_{i,t}^2}}{SE(v_{i,t}^2)} \sim N(0,1) \quad (22)$$

where $v_{i,t}^2$ is the logarithm of realized variance during period i at time t , giving the following four time series¹⁰:

- (i) A time series consisting of the day realized variance observations.
- (ii) A time series consisting of alternating the day and the night (close-to-open) 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.

We then follow Hansen and Lunde (2005) and linearly combine the normalized realized variance series for both the regular and after-hours periods to create a full day realized variance measure.

Because the logarithm of realized variance is approximately normal (Andersen et al., 2001), the ratio in (22) should be approximately standard normal which makes it appropriate for use in the SEMIFAR model. The moment statistics for the unconditional distributions of the time series

¹⁰ We did 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 weighted average 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.

constructed using (22), shown in Table 8, indicate that the transformed realized variance series are all very close to the standard normal distribution.

Table 9 summarizes the estimation results for the SEMIFAR model for MSFT. 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 estimator all tend to be higher, while those based on the local Whittle estimator all tend to be lower, than the estimates from the SEMIFAR model. For all series, there are also significant short-memory AR components. The Ljung-Box test statistics all show that the residuals are all serially uncorrelated.

5.2. Forecast Evaluation

Tables 10 (A) and (B) show the 1-step and 5-step-ahead predictions from the 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 model with the day and the preopen variance provides the highest R^2 and the lowest statistics of the forecasting evaluation 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 11 lists the results for the encompassing regressions. The R^2 increases from 0.146 to 0.192 when the preopen variance is included in the SEMIFAR model. The coefficient of the SEMIFAR forecast with the day variance is insignificant and that of the SEMIFAR forecast with both the day and the preopen variance is significant, which indicates that the latter contains more information than the former. We also see that the SEMIFAR forecast with the postclose variance 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 is the daily return and/or variance measures while the additional information we include is a measure of after-hours variance. We augment GARCH and SEMIFAR models for daily volatility by including various measures of after-hours 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 substantially 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 for 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.

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 extension of our analysis 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 preopen 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.

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 Frequency vis-à-vis Forecast Horizon," *Journal of Empirical Finance* 6 (1999), 457-477.
- Bandi, Federico and Jeffery Russell. "Microstructure Noise, Realized Variance, and Optimal Sampling." *Review of Financial Studies*, 79 (2008), 339-369.
- 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, "Liquidity 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 Risk," *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 Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1) Model?," *Journal of Applied Econometrics*, 20 (2004), 873-889.
- 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.

- Hansen, P., and Lunde, A. "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131 (2006), 97-121.
- 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.
- Tsiakas, Ilias, "Overnight Information and Stochastic Volatility: A Study of European and US Stock Exchanges," *Journal of Banking and Finance* 32 (2008), 251-268.

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) with Close-to-Open	GARCH (1,1) with Preopen	GARCH (1,1) with Postclose	GARCH (1,1) with 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 estimation of a Student's t 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) with Close-to-Open	GARCH (1,1) with Preopen	GARCH (1,1) with Postclose	GARCH (1,1) with 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) with Close-to-Open	GARCH (1,1) with Preopen	GARCH (1,1) with Postclose	GARCH (1,1) with 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) with Close-to-Open	GARCH (1,1) with Preopen	GARCH (1,1) with Postclose	GARCH (1,1) with 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) with Close-to-Open	Day GARCH (1,1) with 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) with Close-to-Open	Day GARCH (1,1) with 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) with Preopen	Day GARCH (1,1)	Day GARCH (1,1) with Preopen	Day GARCH (1,1)	Day GARCH (1,1) with 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 Close-to-Open	Day Garch(1,1) with Preopen
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 preopen 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. The Moment Statistics for the Unconditional Log Adjusted Realized Variance, v_{it}^2 , for MSFT

	Min.	Mean	Median	Max.	St. dev.	Skew.	Kurt.
Log Variance (daily)							
Day only	-3.1030	-0.4308	-0.4385	2.3633	0.9217	0.1074	2.7532
Day with Close-to- Open	-6.0777	-0.8632	-0.7941	3.5127	1.3651	-0.2513	3.2557
Day with Preopen	-5.9591	-0.6337	-0.6238	4.1427	1.1273	-0.1430	3.6035
Day with Postclose	-6.8792	-1.2015	-1.0862	5.0035	1.5591	-0.1953	3.1762

Table 9. SEMIFAR Parameter Estimates for MSFT In-Sample Period

	SEMIFAR with Day only	SEMIFAR with Day and Close-to- Open	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 estimator	0.678	0.684	0.688	0.563
Whittle estimator	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 10A. Forecasting Evaluation Methods for 5-Step Ahead Prediction of SEMIFAR Models for MSFT

	SEMIFAR with Day only	SEMIFAR with Day and Day and Close-to-Open	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 10B. Forecasting Evaluation Methods for 5-Step Ahead Prediction of SEMIFAR Models for MSFT

	SEMIFAR with Day only	SEMIFAR with Day and Close-to-Open	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 11. Encompassing Regression for Different SEMIFAR Models

	Day only	Day with Close-to-Open	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

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

where d denotes day period, n_i denotes for the whole night or the preopen 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.

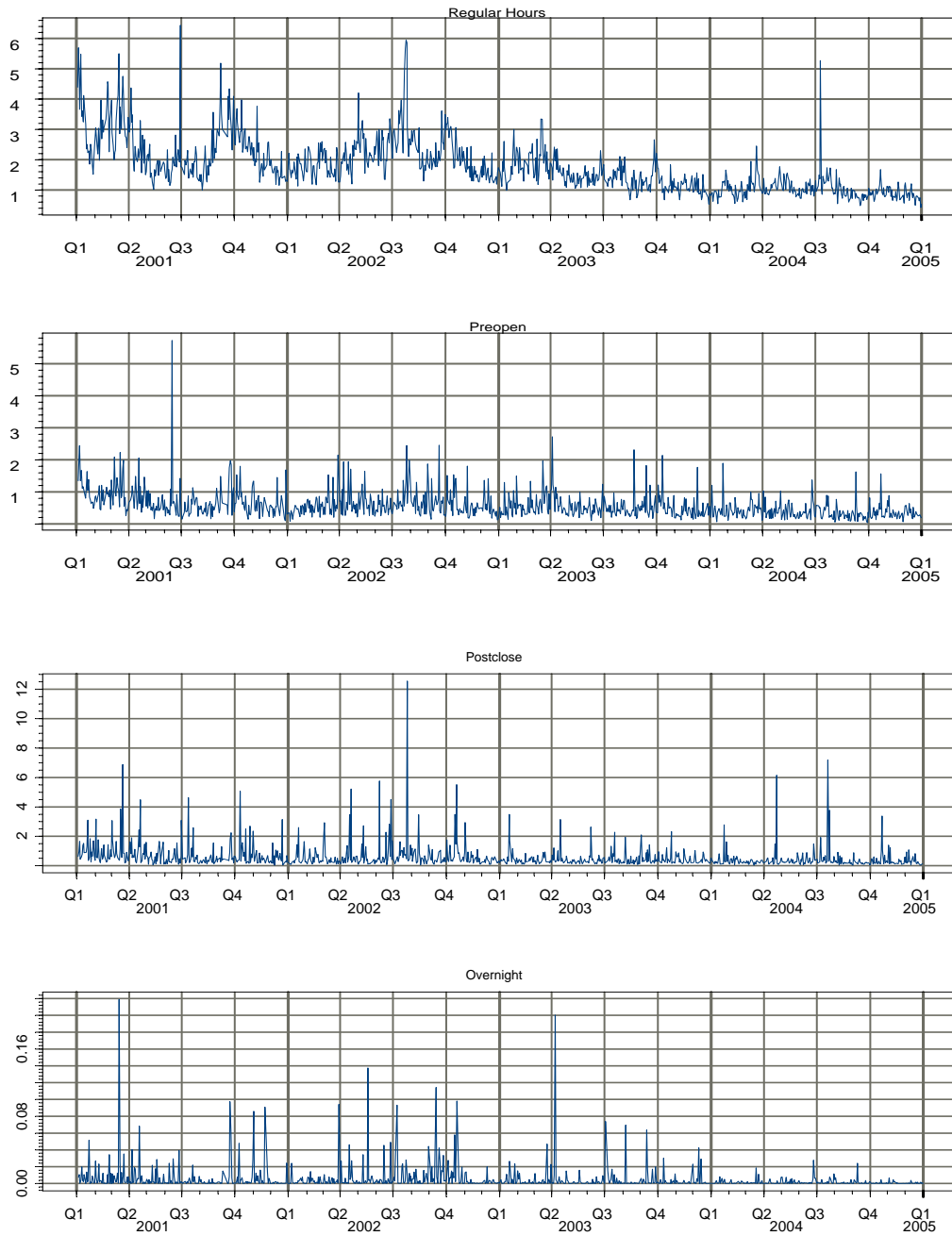


Figure 1. Time Series of the MSFT Volatilities for Different Time Periods

These are realized volatilities for regular Hours, preopen, and postclose, and square root of overnight returns. The volatilities are in percentages.

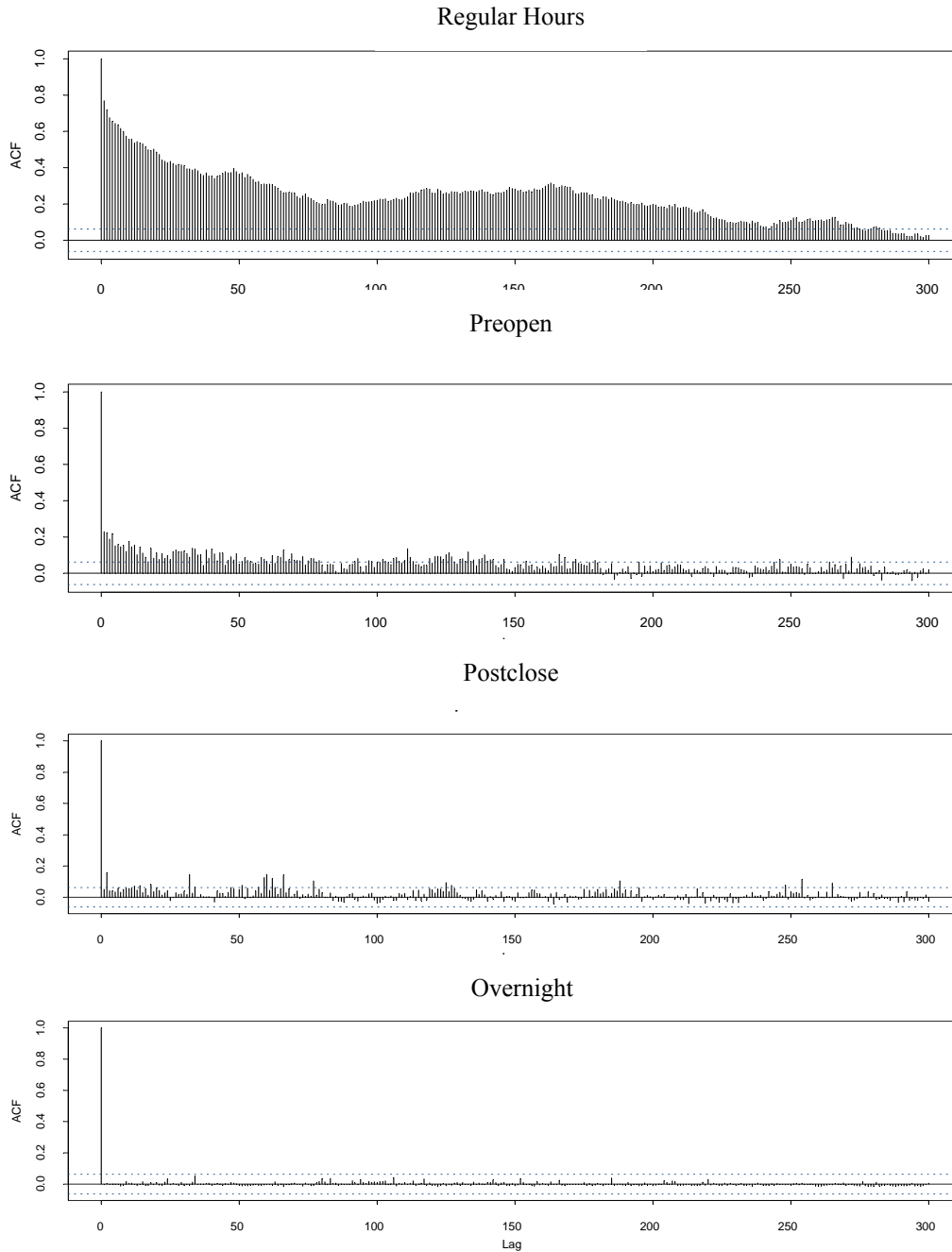


Figure 2. Autocorrelations of the MSFT Volatilities for Different Time Periods

These are the autocorrelations of realized volatilities for regular Hours, preopen, and postclose, and square root of overnight returns.

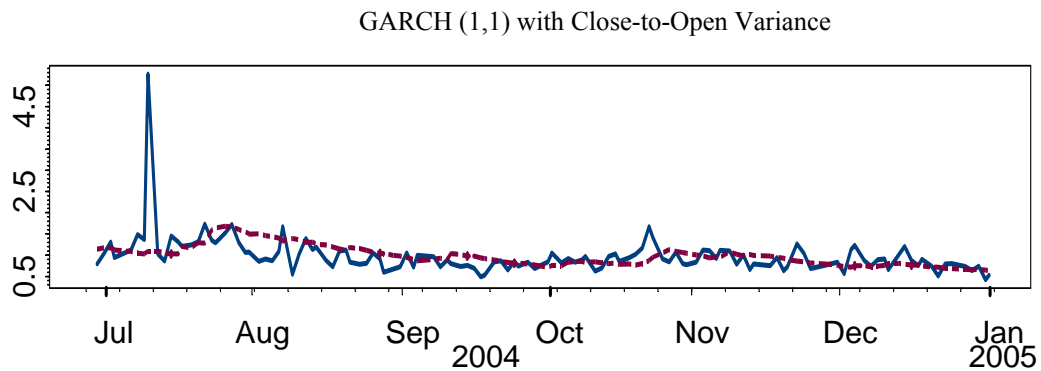
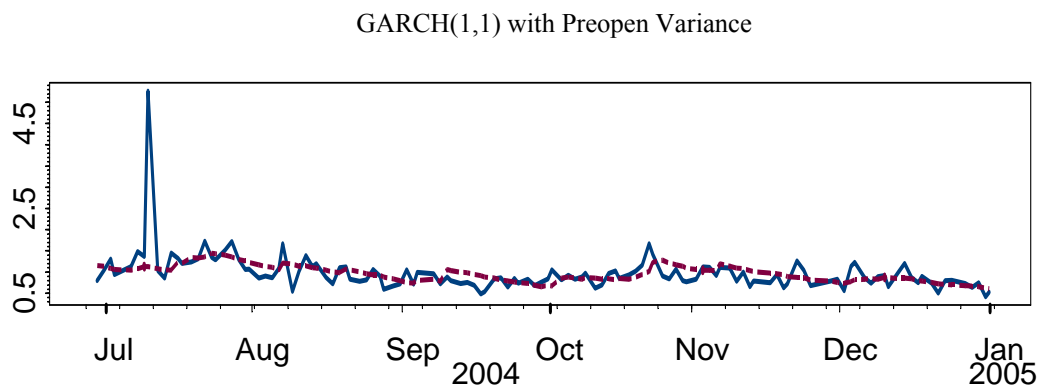
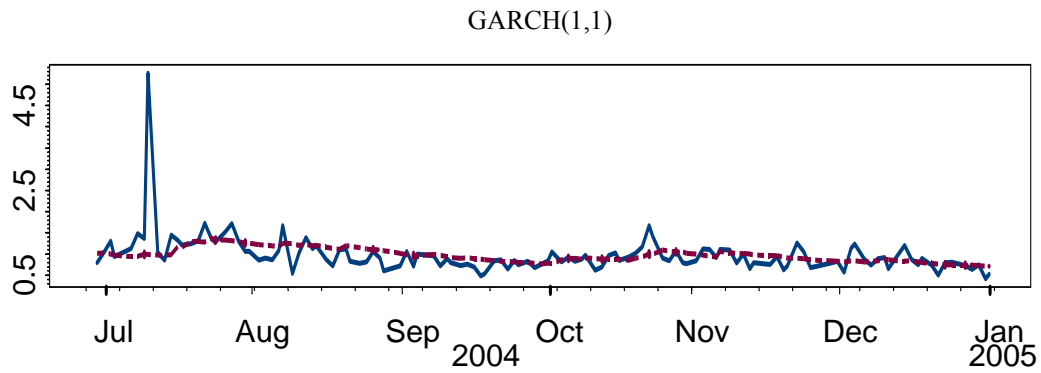


Figure 3. MSFT One-Step Ahead Volatility Forecasting for Day GARCH(1,1) by Different Models

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

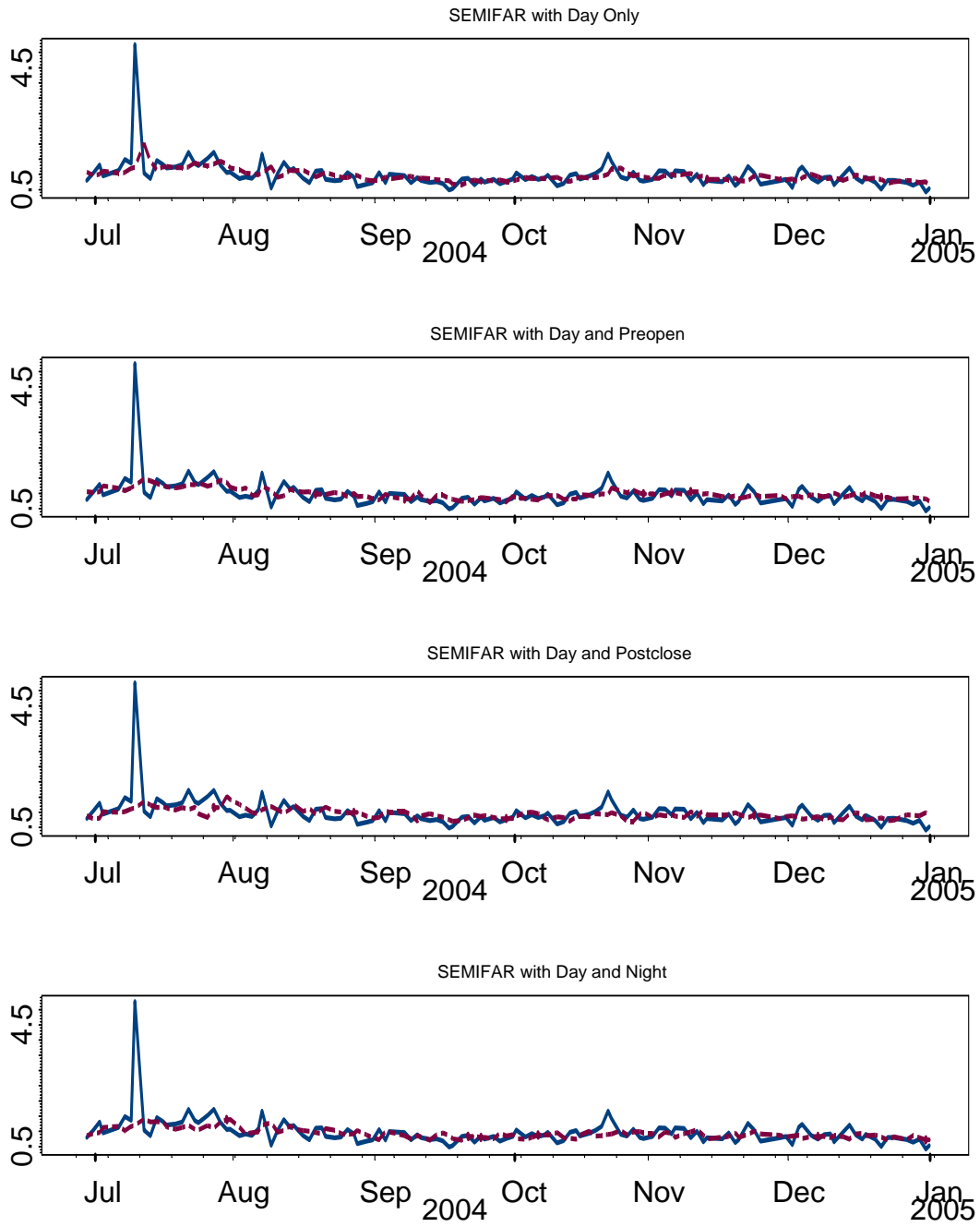


Figure 4. MSFT One-Step Ahead Forecasting for Different SEMIFAR Models

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