Measuring and Forecasting the Daily Variance Based on High-Frequency Intraday and Electronic Data

Fatemeh Behzadnejad

Supervisor: Benoit Perron

Abstract

For the 24-hr foreign exchange market, Andersen and Bollerslev use intraday returns rather than daily returns to obtain a measure for the realized variance (RV). In equity markets, where trading is done during a part of the day, Hansen and Lunde suggest some estimators that use intraday returns during the active part of the day and close to open return for the inactive part of the day. In some markets such as futures market for S&P500, trading is done electronically when the real market is closed. Using this electronic data, we provide a new measure for the RV and then compare it with the variance estimators of Hansen and Lunde. If the measure that uses electronic data (RV_total) is considered as a reference, the optimal linear combination of open to close realized variance and squared close to open return, which is the third estimator of Hansen and Lunde, more corresponds to RV_total. Having access to such measure, forecasting the future variance values can be done exclusive of other variance estimators.

List of Tables and Figures

Table of Contents

Introduction

A precise and reliable measure of variance is useful for a range of applications. One of these applications is the evaluation of variance models. For example Hansen and Lunde (2005a) show that a noisy measure of variance can result an inconsistent ranking of variance models. Another study by Andersen and Bollerslev (1998) shows that in order to evaluate the performance of autoregressive conditionally heteroskedastic (ARCH)-type model, a precise variance estimator is necessary.

One of the variance measures is realized variance (RV) which is the sum of intraday squared returns. This measure can be used as a more precise proxy for theoretical quantities such as integrated variance (IV). Andersen and Bollerslev (1998) showed that daily squared return as a standard variance model is extremely noisy although it is an unbiased estimator. They argued that both theoretically and empirically, the sum of the intraday squared return is the best measure for realized variance.

Realized variance is constructed from high frequency intraday returns. High-frequency data are increasingly being used to address a wide range of problems in econometrics because of the information they contain about population parameters. But high-frequency data have been mainly used to estimate financial variance.

To estimate RV for a full day one needs high-frequency data for 24 hours of the day. Andersen and Bollerslev's results are for 24 hours foreign exchange market. The difference between exchange market and stock market is that most equities are traded for a part of a day such as six or seven hours per day. Since a part of daily variance may take place during inactive part of the day, using only the daily intermittent data does not reflect the variance for whole day.

A number of measures for stock market variance combine intraday returns (open to close) and overnight return (close to open). For example Hansen and Lunde (2005) propose three estimators for daily variance that are based on the realized variance for the active part of the day, $RV_{2,t}$, and the squared return of the inactive period , $r_{1,t}^2$. They characterize the assumptions that justify using each of these estimators. Their first estimator simply adds up the returns of the active part of the day and the squared overnight return while the second estimator is the scaled open to close return and so the overnight return is not considered. Finally, the third estimator is the optimal linear combination of $RV_{2,t}$ and $r_{1,t}^2$ which is obtained by mean squared error (MSE) method.

In some markets, although the real market is not active during close to open period, but electronic transactions are being done in this time interval. As an example S&P 500-index futures are being traded on the electronic overnight trading system (GLOBEX) since 1994. Making use of these electronic data, we suggest another estimator for the whole day variance. This new estimator is based on realized variance for the active part of the day, $RV_{2,t}$, and realized variance of electronic data for the period that the real market is closed.

Having access to 24 hours data and computing such measure of variance (named as RV_total here) will be helpful. It can be used as a reference for comparing other variance estimators in order to recognize which variance measure represents better RV_total to be used in the case that RV_total is not at hand. Using our data, the results of comparing the estimators of Hansen and Lunde with RV_total show that their optimal variance estimator more corresponds to RV_total. It will also be interesting to see if this variance measure is accessible, can it exclusively be used in the further operations on volatility data like forecasting, or on the contrary, the alternative measures may improve the performance of forecasting. To do that, we should introduce a forecasting model that corresponds better to our high frequency data. Consequently, by applying this model to different variance measures, we can obtain different series of predicted data that can be used to evaluate the forecasting performance of the variance measures. To forecast one day ahead volatility, we use AR model for each variance estimator separately. Applying AR on optimal estimator of Hansen and Lunde as well as RV_total and comparing the forecasted values show a better forecasting by RV_total.

This paper is organized as follows. Next section explains the concept of realized variance. Section 2 explains how to calculate whole day variance by using electronic data. Section 3 discusses the three estimators of Hansen and Lunde. In section 4, a model is presented to forecast one day ahead volatility in the market. The results of sections 2 to 4 are applied to two years of five minutes returns from S&P 500 index futures followed by discussion of their time series, statistical properties and forecasting process.

1. Definition of Realized Variance

We consider $\{p^*(t)\}_{t\in[0,\infty[}$ as logarithmic efficient price process which may differ from the observed price process p because of the market microstructure noise. So, we define $p = p^* + u$ where u is a noise process. If we use trading day as time unit and t as market closing time, close to close return will be $r_t = p(t) - p(t-1)$.

We shall assume the following continuous time model for the price process:

$$dp^*(t) = \mu_t dt + \sigma_t dw_t$$

where w_t is the standard Brownian motion and μ_t and σ_t denote to drift and volatility terms. We shall assume that $\mu_t = 0$ for all t to simplify the problem. In fact, the drift term is of order dt which is smaller than $(dt)^{1/2}$ of the volatility term and so is negligible at high-frequency data (see e.g. Andersen, Bollerslev and Diebold (2002) for more details). After simplification, the model will be:

$$dp^*(t) = \sigma_t dw_t$$

where σ_t is in general smooth time varying stochastic process that is independent of w_t . Close to close return can be considered as:

$$r_t = p(t) - p(t-1) = \int_{t-1}^{t} \sigma_u dw_u \quad \text{t=1,2,...}$$

Dividing each day to the intervals with lengths h, intraday return for any horizon h can be defined as follow:

$$r_i \equiv p(t_i) - p(t_{i-1}) = \int_{(i-1)h}^{ih} \sigma_u dw_u \quad \text{For i=1... 1/h (1/h integer)}$$

We eliminate the dependence of r_i on the horizon h. Intraday returns are $i.i.d.N(0,\sigma^2 h)$ if σ is constant. In other words we have:

$$r_i \equiv p(t_i) - p(t_{i-1}) = \int_{(i-1)h}^{ih} \sigma_u dw_u = \sigma(W_{ih} - W_{(i-1)h}) \equiv \sigma u_i \sim i.i.d.N(0,\sigma^2 h)$$

Where $u_i \sim i.i.d.N(0,h)$ for i=1...1/h.

The parameter of interest is integrated variance over a day that we assume it to be finite:

$$IV = \int_0^1 \sigma_u^2 du$$

An empirical estimator of integrated variance is realized variance which is sum of the squared intraday returns:

$$RV = \sum_{i=1}^{1/h} r_i^2$$

When $h \to 0$ or the number of intraday observations tends to infinity, RV is a consistent estimator of integrated variance under certain assumptions including absence of microstructure noise.

We define realized variance for an interval [a,b] as follow:

$$RV_{[a,b]}^{\Xi} \equiv \sum_{i=1}^{m} \{p(t_i) - p(t_{i-1})\}^2 \qquad (1)$$

Note that Ξ is a partition of [a,b] and $RV_{[a,b]}^{\Xi}$ is the realized variance of this partition.

2. Calculating Realized Variance by Whole Day Data

In order to calculate realized variance for a full day, we need the data of entire 24 hours of a day. For example Andersen and Bollerslev (1998) calculated realized variance upon 24-hours foreign exchange rate market data. The problem is that future markets do not trade on a 24-hours basis.

While it seems that the markets are normally open for a fraction of a day such as 7 or 8 hours a day, most equities are electronically traded during the hours the real market is closed. In this situation that we have access to high frequency data during day and electronic data over night, realized variance for the whole day can be estimated by

$$RV_t = RV_{[a,b]}^{\Xi} \equiv \sum_{i=1}^{m} \{p(t_i) - p(t_{i-1})\}^2$$

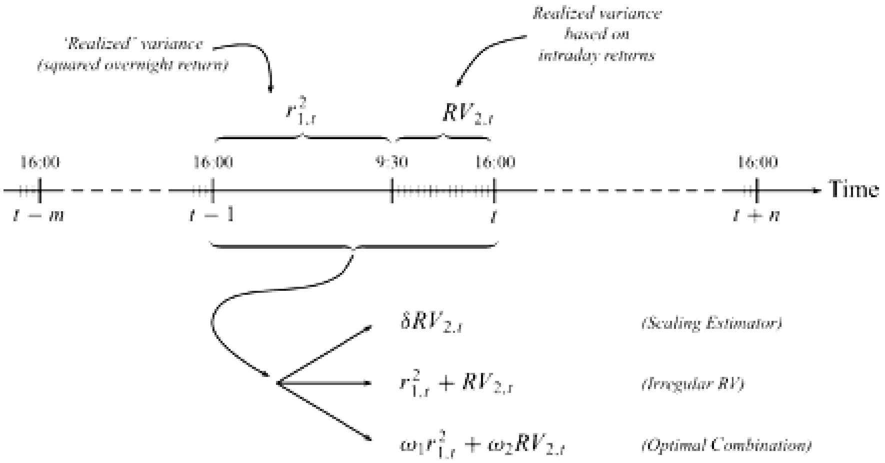
Where [a,b] is the interval of one complete day. $p(t_i)$ represents either the asset price during the active part of the day or the electronic asset price over night when the real market is closed.

3. Estimating RV by Open Market Hours Data

Hansen and Lunde (2005), presented three estimators to calculate whole day variance if high frequency data are available only for the active period of a day. Note that here active period represents the hours of opening the market and so does not include electronic trades while the market is closed.

Defining Δ_0 as the interval of time in which the market is active, $IV_{1,t} \equiv IV_{[t-1,t-\Delta_0]}$ and $IV_{2,t} \equiv IV_{[t-\Delta_0,t]}$ represent integrated variance of inactive and active part of the day, respectively. We also write $r_t = r_{1,t} + r_{2,t}$ that $r_{1,t} \equiv p(t-\Delta_0) - p(t-1)$ is close to open return and $r_{2,t} = p(t) - p(t-\Delta_0)$ is the open to close return. Open to close period is exactly the time when the high frequency data are available. We let $RV_{2,t} = RV_{[t-\Delta_0,t]}$
the RV measure of this active part of day.

Three estimators of Hansen and Lunde (2005), are based on realized variance of active part of the day, $RV_{2,t}$ and the square of close to open return, $r_{1,t}^2$ .Their first estimator is a scaled value of $RV_{2,t}$ while the second one uses the value of $r_{1,t}^2$ and is the sum of $RV_{2,t}$ and $r_{1,t}^2$. Finally, they define the third estimator as $\omega_1 r_{1,t}^2 + \omega_2 RV_{2,t}$ in which ω_1, ω_2 are the weights that minimize the mean-squared error (MSE).



3.1 Scaling estimator of IV

The first estimator of IV presented by Hansen and Lunde (2005) is constructed scales the $RV_{2,t}$ by constant value of $\hat{\delta}$, so they consider $RV_t^{scale} = \hat{\delta}.RV_{2,t}$. If $\bar{r} = n^{-1}\sum_{t=1}^{n} r_t$, then $\hat{\delta} = \sum_{t=1}^{n}(r_t - \bar{r})^2 \Big/ \sum_{t=1}^{n} RV_{2,t}$ is a consistent estimator of $\delta \equiv E[IV_t]/E[RV_{2,t}]$.

The condition needed to justify this simple scaling are completely characterized through theorem 1 of Hansen and Lunde (2005).

3.2 Incorporating the over night return

While scaling of $RV_{2,t}$ seems interesting to obtain whole day variance, an alternative is to make use of over night return, $r_{1,t}$ as well. Hansen and Lunde (2005) presented two ways to combine $RV_{2,t}$ and $r_{1,t}$ in order to calculate daily variance. In the first approach, $r_{1,t}$ is simply added to the high frequency intraday return and thus the estimator is given by

$$RV_t^{+on} = r_{1,t}^2 + RV_{2,t}$$

Note that the two terms of $RV_{2,t}$ and $r_{1,t}^2$ can be considered as estimators of IV during the active period (open to close) and inactive period (close to open), respectively.

The second approach is to consider general linear combination of $r_{1,t}^2$ and $RV_{2,t}$, eventually

$$RV_t(\omega) = \omega_1 r_{1,t}^2 + \omega_2 RV_{2,t} \qquad (2)$$

where $\omega = (\omega_1, \omega_2)'$. Thus the two estimators of RV_t^{scale} and RV_t^{on} are the special case of $RV_t(\omega)$ using the weights of $(0, \overset{\wedge}{\delta})$ and (1, 1), respectively.

In equation (2) the optimal value of $\omega = (\omega_1, \omega_2)$ is the solution of the following optimization problem:

$$\min_{\omega_1, \omega_2} \operatorname{var}(\omega_1 r_{1,t}^2 + \omega_2 RV_{2,t}), \text{ s.t. } \omega_1\mu_1 + \omega_2\mu_2 = \mu_0 \qquad (3)$$

where μ_0, μ_1, μ_2 are defined as $\mu_0 \equiv E(IV_t)$, $\mu_1 \equiv E({r_{1,t}}^2)$ and $\mu_2 \equiv E(RV_{2,t})$. Under some assumptions, $E_\sigma[RV_t(w)] = IV_t$ for all ω that satisfy $\omega_1\mu_1 + \omega_2\mu_2 = \mu_0$ (see Hansen and Lunde (2005) for further details).

''Let $\eta_1^2 \equiv \operatorname{var}(r_{1,t}^2), \eta_2^2 \equiv \operatorname{var} RV_{2,t}$ and $\eta_{12} \equiv \operatorname{cov}(r_{1,t}^2, RV_{2,t})$. The solution to equation (3) is given by

$$\omega_1^* \equiv (1-\varphi)\frac{\mu_0}{\mu_1} \text{ and } \omega_2^* \equiv \varphi\frac{\mu_0}{\mu_2} \qquad (4)$$

Where φ is a relative importance factor, defined by

$$\varphi \equiv \frac{\mu_2^2\eta_1^2 - \mu_1\mu_2\eta_{12}}{\mu_2^2\eta_1^2 + \mu_1^2\eta_2^2 - 2\mu_1\mu_2\eta_{12}} \qquad (5)$$

This solution is intuitive and is particularly simple to interpret if $\eta_{12} = 0$. In this special case we have $\omega_1^* / \omega_2^* \equiv \frac{\mu_2/\mu_1}{\eta_2^2/\eta_1^2}$ which shows that an increase in volatility during the active period (relative to the inactive period) has a positive impact on the relative weight $\partial(\omega_1^*/\omega_2^*)/\partial(\mu_2/\mu_1) \geq 0$, whereas the opposite is the case for an increase in the relative noise, $\partial(\omega_1^*/\omega_2^*)/\partial(\eta_2^2/\eta_1^2) \leq 0$ ''

The result of this theorem can be easily used in practice by replacing the quantities of $\mu_1, \mu_2, \eta_1, \eta_2$ and $\eta_{1,2}$ by their sample average.

4. Forecasting Realized Variance

Financial market volatility is a key factor in risk management theory and asset pricing. As an example, investor's assessment of the stock variance over the life of the option is a crucial parameter in most pricing models. Thus accurate volatility forecast is necessary to successfully determine the price of derivative securities. Many statistical methods have been suggested to describe volatility dynamic in the financial markets, including ARMA, different versions of GARCH models and many other models that are based on the daily return.

The preceding models are mostly based on the daily return volatility. Andersen, Bollerslev, Diebold and Labys (ABDL) (2003) have proposed a framework for forecasting the realized volatility where high- frequency intra day returns are available. This model is motivated from following regularities which are the results of experimental analysis by ABDL. First, the distribution of logarithms of realized volatility is approximately Gaussian, although the distribution of realized volatility is right skewed. Second, a fractionally-integrated long run process can provide a good estimation for the long run memory of the logarithms of realized volatility.

Regarding the mentioned distributional features, they consider this simple vector auto regressive model for the logarithm of realized variance or VAR-RV.

$$\phi(L)(1-L)^d(y_t-\mu)=\varepsilon_t \qquad (6)$$

Where y_t is the logarithm of realized volatility and ε_t is a white noise process. After determining the degree of fractional differencing operator or d, the model can be easily estimated by applying OLS.

This method efficiently makes use of the information in the intraday returns without having to present a model for this intraday data. On the other hand, comparing with the other currently popular models that are relied on the daily returns, VAR-RV makes a significant improvement in forecasting performance.

ABDL (2003) have compared their results of forecasting the exchange rate market volatility with a wide variety of models. For example, they have compared VAR-RV with the long memory filtered daily logarithmic absolute return. This model is identical to (6) except for the volatility proxy which is the daily absolute return instead of the realized volatility.

VAR-RV forecasts are also compared with the GARCH model of Engle (1982) and Bollerslev (1986) which is the most popular procedure in academic applications. They have also considered FIEGARCH that is a variant of the GARCH model that incorporate long memory. Another model considered by ABDL is RiskMetrics by J.P.Morgan's (1997) which is the most widespread model used by practitioners.

The results of forecasting variance by all the above models and also VAR- RV have been striking. The regressing for forecast evaluation has the following form

$$\left(\{v_{t+1}\}_i\right)^{1/2}=b_0+b.\left(\{v_{t+1|t,VAR-RV}\}_i\right)^{1/2}+b_2.\left(\{v_{t+1|t,Model}\}_i\right)^{1/2}+u_{t+1,i}$$

If this regression includes only one variance measure, R^2 is always the highest for VAR-

RV in the data used by ABDL. On the other hand, for almost none of the VAR-RV forecast, can they reject the hypothesis that $b_0 = 0$ and $b_1 = 1$ in the corresponding t-test. They reject the hypothesis that $b_0 = 0$ and/or $b2 = 1$ for most of the models. Furthermore, if the regression includes both VAR-RV and another alternate variance forecast, for most of the cases, the estimation of b_1 and b_2 is close to 1 and 0, respectively.

As it was mentioned, the long run dependence in financial market volatility can be modeled by fractionally integrated processes such as the VAR-RV model explained previously or integrated ARCH; see, e.g., Baillie, Bollerslev and Mikkelsen (1996). In order to obtain parameter d in fractionally integrated processes, the implied hyperbolic decay rate k^{-2d-1} can be used. Using the Geweke and Porter-Hudak (1983) log-priodogram regression, called GPH technique, the value of d can be estimated. In essence, if we estimate the logarithm of correlation by the lag logarithms, the estimated linear estimation of logarithm has the slope of $-2d-1$.ABDL (2003) applied multivariate extension of GPH estimator to the sample of autocorrelation of the realized logarithmic volatility in exchange rate market out to lag of 70 days which resulted on an estimation of d equal to 0.401, which is a common value in such markets.

Implementing the degree of fractional differencing equal to 0.401 to filter our data does not result in a good forecast. Since the size of available data is not big enough, estimating d by our sample does not lead to the best value for d. However, as it can be seen in figures 1 and 2 it suggests that the values of d which are closer to zero can provide a better prediction for y_t. Choosing d equal to zero transforms the model (6) to an AR process for which is the model finally applied to our data. Here, $y_t = (1/2)\log(V_t)$ or the logarithm of volatility, where V_t is the variance of returns at day t and. It is assumed that the order of lag polynomial is one day. Therefore the AR process that is considered is as follow

$$\phi(L)(y_t - \mu) = \varepsilon_t \tag{7}$$

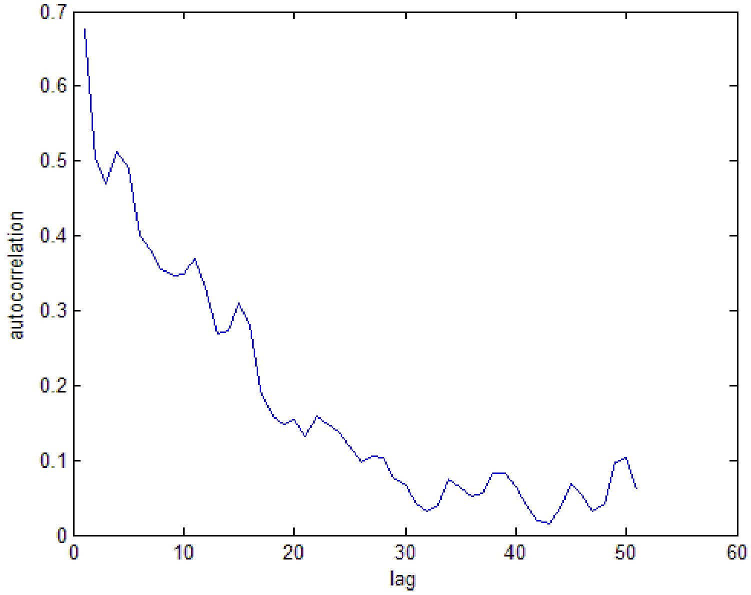
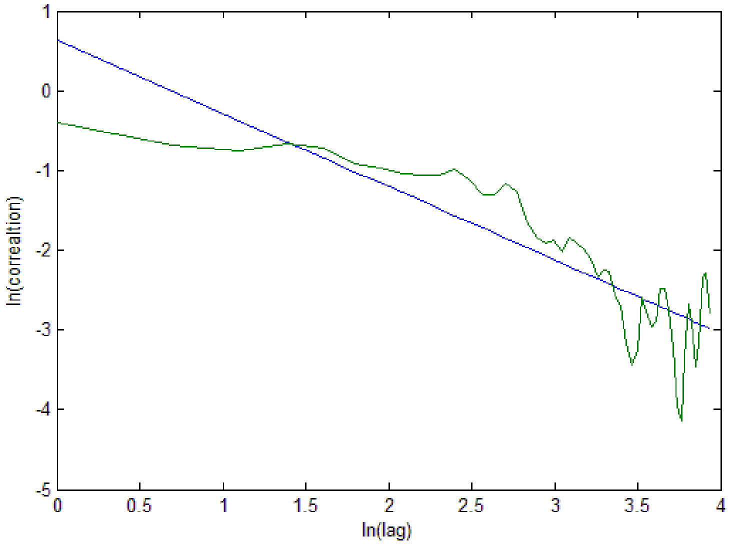


Figure 1 sample autocorrelation



Logarithmic correlation ——

Figure2 sample logarithmic autocorrelation

Fitted value ——

This simple autoregressive model can be applied to RV_total and also to different variance measures of Hansen and Lunde to forecast one day ahead volatility. Consequently, the forecasting done by each of the variance estimators can be evaluated by comparing different forecasts. In fact, there is no generally accepted method to evaluate the performance of competing forecasts and many statistical procedures have been used to do that (see Andersen, Bollerslev, and Lange (1999). Here the alternative forecasts are evaluated by projecting the volatility logarithm on a constant and the different model forecasts.

$$y_{t+1} = \alpha_0 + \alpha_1 (y_{t+1|t}) + \alpha_2 (z_{t+1|t}) + u_{t+1} \tag{8}$$

The relative weight of coefficients and the statistics of the regression can be used to evaluate the different forecasts.

5. Empirical application to S&P 500 futures

In this section, the results of the previous parts are applied to the data of S&P 500 index-futures transaction prices. The estimated RV's are calculated for S&P 500 index futures prices. S&P 500 index futures have traded electronically on GLOBEX during night when the stock market is inactive since 1994. The chosen period for this study is from January 2006 to December 2007. The sample period contains n=514 trading days.

At Chicago Mercantile Exchange (CME), futures floor trading is open from 8:30 a.m. to 3:15 p.m., Chicago time, for day time trading. GLOBEX overnight trading begins from 3:30 p.m. and lasts until 8:15 a.m. of the next day. 5-minutes intervals are selected to avoid market micro structure problem such as bid- ask bounce. In each 5-min interval, the chosen price is the last price of the interval or closing price.

In order to estimate daily variance accurately, a high number of intraday returns should be available. On the other hand if the chosen time intervals are too small or the returns are too frequent, then microstructure effects such as bid-ask spread may cause some biases. Finally, 5- minutes intervals are used to eliminate such biases.

Daily realized variance is calculated based on four estimators. The first estimator, RV_t, which was presented in section 2 uses both day time trading and electronic night time trading. Variance of each day is calculated by replacing $p(t_i)$ by logarithm of 5-minute interval prices. $p(t_i)$ could be either the real market price (in day time) or the electronic price (in night time). The realized variance calculated by this method is referred as RV_total.

The three remaining estimators were explained in section 3. All of these three estimators use the realized variance of active part of the day, $RV_{2,t}$ and the squared close to open return $r_{1,t}^2$. Therefore the night time electronic data are not considered in these estimators. As it was explained, to obtain the scaled estimator the parameter of $\overset{\wedge}{\delta}$ is needed which is calculated in section 3.1. We call this estimator as es_var1. The second estimator of daily variance is achieved simply by adding $RV_{2,t}$ and $r_{1,t}^2$. This estimator is referred as es_var2.

The third estimator is the optimum linear combination of $RV_{2,t}$ and $r_{1,t}^2$ suggested by Hansen and Lunde and was explained in section 3.2. The estimates of $\omega_1, \omega_2, \varphi$ are defined from (4) and (5). The following equations could be used to estimate the parameters needed in (4) and (5).

$$\hat{\mu}_0 \equiv \frac{1}{n}\sum_{t=1}^{n}(r_{1,t}^2 + RV_{2,t}), \; \hat{\mu}_1 \equiv \frac{1}{n}\sum_{t=1}^{n} r_{1,t}^2 , \hat{\mu}_2 \equiv \frac{1}{n}\sum_{t=1}^{n} RV_{2,t} , \; \hat{\eta}_1^2 \equiv \frac{1}{n}\sum_{t=1}^{n}(r_{1,t}^2 - \hat{\mu}_1)^2$$

$$\hat{\eta}_2^2 \equiv \frac{1}{n}\sum_{t=1}^{n}(RV_{2,t} - \hat{\mu}_2)^2, \; \hat{\eta}_{12} \equiv \frac{1}{n}\sum_{t=1}^{n} RV_{2,t}(r_{1,t}^2 - \hat{\mu}_1)$$

Table 1 summarizes the statistical properties of the four estimated variances. RV_total is the daily variance using electronic data and es_var1, es_var2 and es_var3 are the three estimators that use $RV_{2,t}$ and $r_{1,t}^2$. If RV_total is considered as a reference, it is possible to compare the three estimators of Hansen and Lunde together. Considering the average and variance of different variance measures in the table, the average of es_var3 is much closer to RV_total. Besides that this measure is more stable than es_var2 and es_var3.

Figure 3 shows the time series plot of four estimated daily variances. As it can be seen in the figure, if RV_total is the reference, es_var1 underestimates RV_total and as a result the position of es_var1 time series is lower than RV_total and the other estimators.

Table1 statistical properties of different variance measures

	Mean $\times 10^{-5}$	St.Dev. $\times 10^{-5}$	Skewness	kurtosis
RV_total	0.6921	0.8305	3.716	19.853
es_var1	0.6872	0.8160	3.577	18.532
es_var2	0.6567	0.8436	3.939	21.365
es_var3	0.6566	0.7783	3.500	17.383

correlation	RV_total	es_var1	es_var2	es_var3
RV_total	1			
es_var1	0.9372	1		
es_var2	0.9344	0.8913	1	
es_var3	0.9509	0.9972	09225	1

This table contains the statistical properties of RV_total, es_var1, es_var2 and es_var3 as different measures of variance. If RV_total is considered as the reference, es_var3 which is the optimal linear combination of $RV_{2,t}$ and $r_{1,t}^2$, is closer to RV_total.

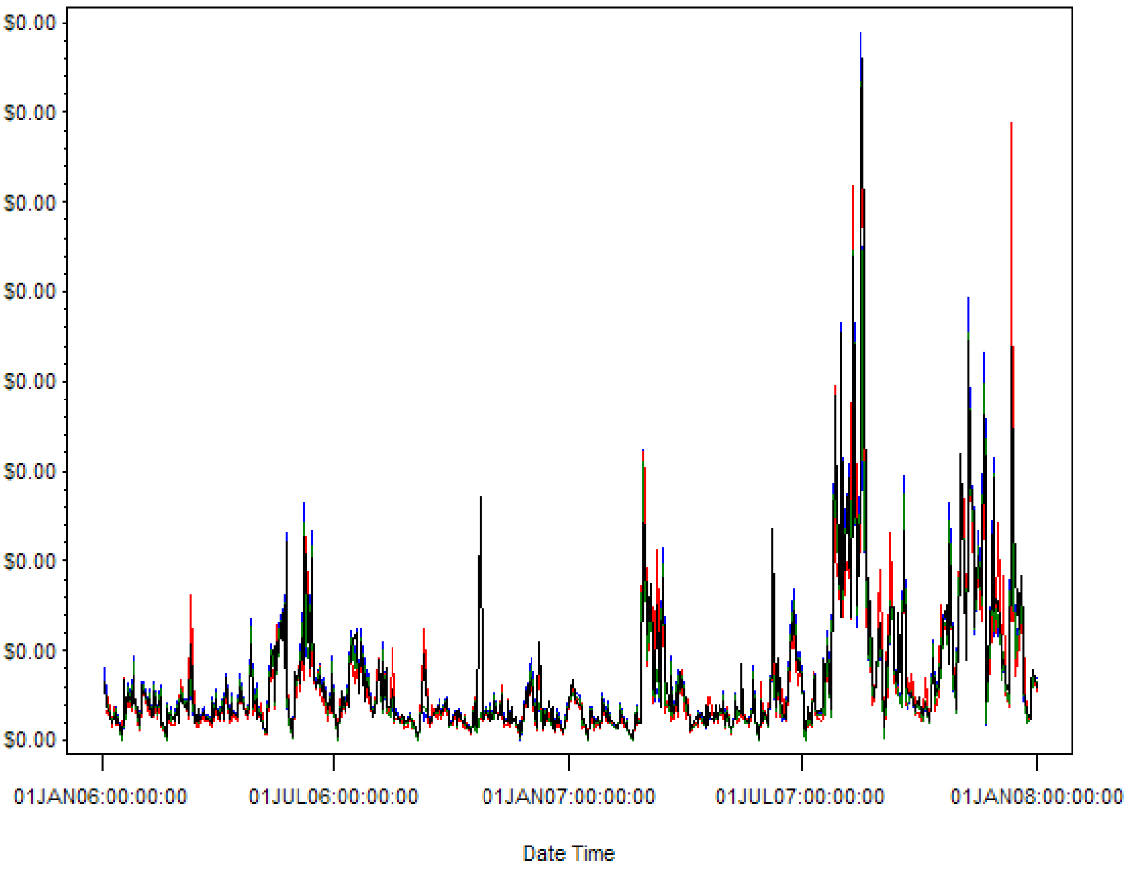


Figure3 Time series plot for RV_total, es_var1, es_var2 and es_var3

es_var1 ——

es_var1 ··········

es_var2 - - - - - -

es_var3 — — — —

There is a bias problem in the realized variance measure of equation (1). This bias is due to the autocorrelation in the intraday returns which is caused by market microstructure effects such as bid ask bounces, nonsynchronous trading, and rounding errors. [see, e.g., Andreou and Ghysels (2002)]

The intraday returns autocorrelation becomes more problematic when sample frequency increases. Bandi and Russell (2004) and Zhang, Mykland and Ait Sahalia(2005) found that under independent market microstructure noise the optimal sampling frequency is often between one and five minutes. In practice the frequency which corresponds to five minutes intraday returns is chosen. Figure 5, 6 and 7 shows the autocorrelation plot of different variance measures.

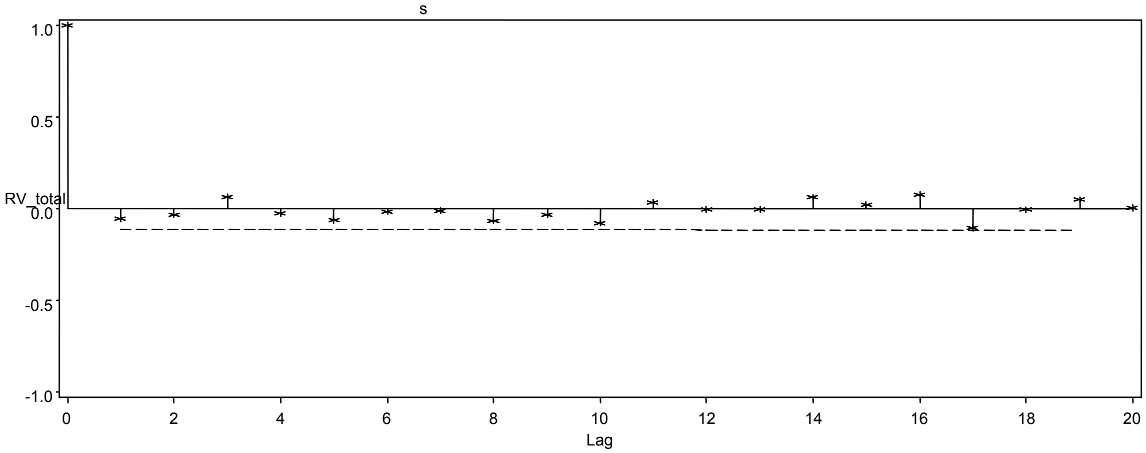


Figure4 Autocorrelations with 95% Confidence Limit (RV_total)

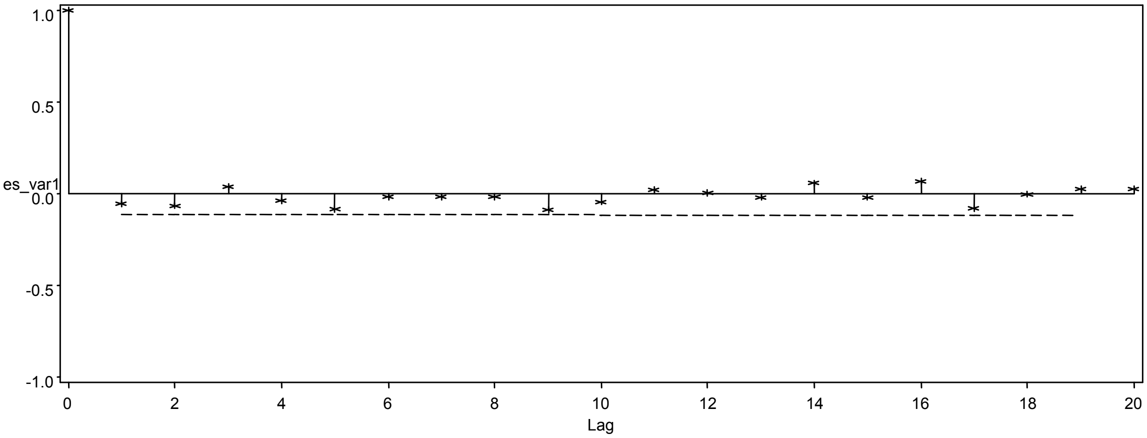


Figure5 Autocorrelations with 95% Confidence Limit (es_var1)

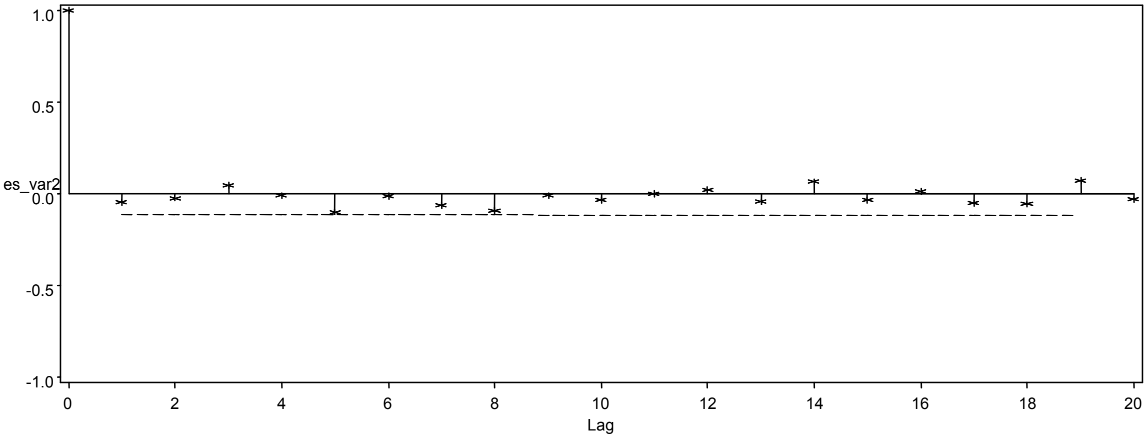


Figure6 Autocorrelations with 95% Confidence Limits (es_var2)

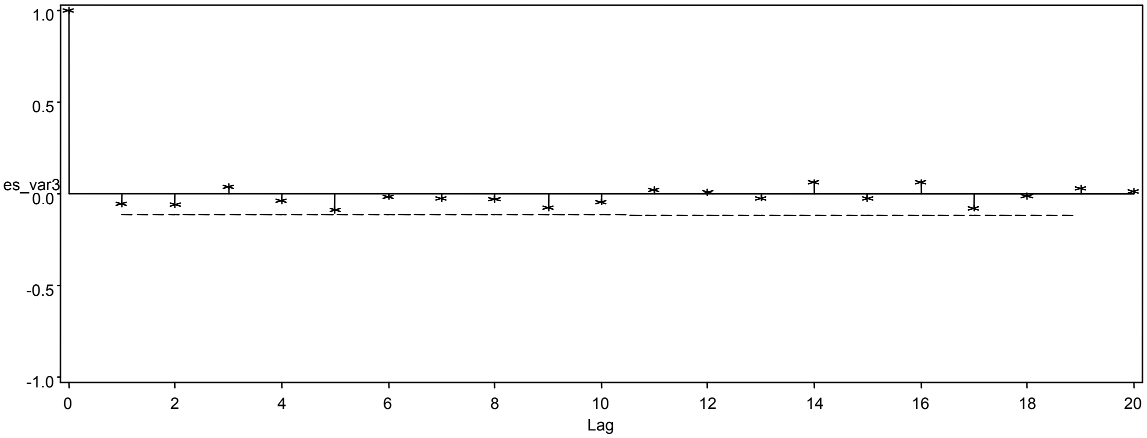


Figure7 Autocorrelations with 95% Confidence Limits (es_var3)

In the following part we implement the results of part 4 to forecast one day ahead logarithm of volatility or y_t for in-sample data. As it was mentioned in part 4, the model which is considered to is AR for y_t. The degree of lag polynomial, selected by Bayesian information criterion (BIC), is equal to 2. Therefore we have

$$\phi(L)(y_t - \mu) = \varepsilon_t$$

or $$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$$

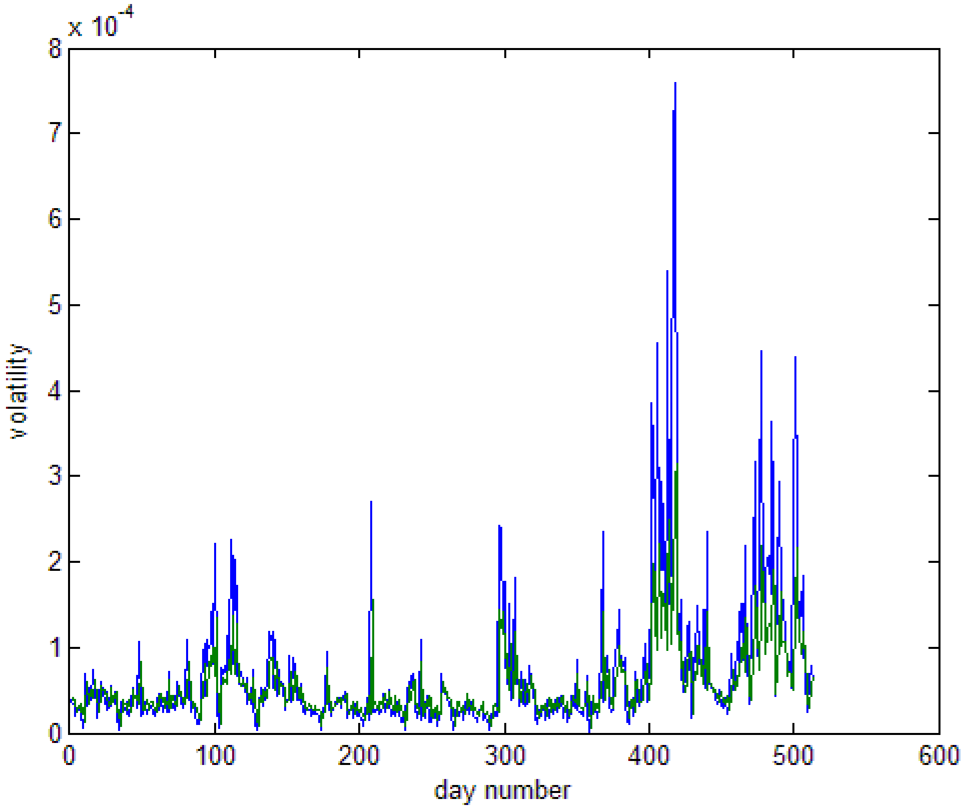
This AR model can be applied to any measure of variance. We apply this model to RV_total as the measure of variance that uses electronic data during nights and also to the es_var3 which is the optimum variance measure of Hansen and Lunde. Figure 8 shows the time series of in sample forecast for RV_total and es_var3. Table 2 shows the statistical parameters of the AR(2) model.

To evaluate the forecasting performance of each model, we use the regression (8) or

$$y_{t+1} = \alpha_0 + \alpha_1 (y_{t+1|t}) + \alpha_2 (z_{t+1|t}) + u_{t+1}$$

Where y_t is the logarithm of volatility obtained from electronic data or $(1/2)\log(RV\_total_t)$ and z_t is the logarithm of the optimal volatility estimated by Hansen and Lunde or $z_t = (1/2)\log(es\_\text{var}3_t)$. Table 3 shows the statistics of this regression.

The results show the coefficients of y_t and z_t are close to 1 and 0 respectively. From this result and also negative sign of α_2, it appears that including the optimal variance of Hansen and Lunde or es_var3 doesn't add any new information in the forecasting process and RV_total measure of variance can be effectively used for forecasting, if it is available.



In sample volatility ——

1-day prediction ——

Figure8 in sample volatility and 1 day prediction

Table2 coefficient estimation for AR(2)

	β_1	β_2	R^2
RV_total	0.529	0.232	0.500
es_var3	0.461	0.202	0.361

Table3 coefficient estimation for forecasting evaluation regression

α_1	α_2	$p-value(\alpha_1)$	$p-value(\alpha_2)$	R^2
1.215	-0.236	0.000	0.194	0.501

6. Conclusion

In this study the idea of using intraday returns to measure the daily variance which was presented by Andersen and Bollerslev (1998) is applied for measuring the stock market variance. The stock markets are usually active for a fraction of the day, but in some markets electronic trading is active when the real market is closed overnight. In this situation, we can use the real market data for the active part of the day and electronic data for the remaining part of the day in order to estimate the realized variance.

Another alternative which was suggested by Hansen and Lunde (2005) does not consider the electronic data. In their method the whole day variance is declared by three estimators that make use of the intraday returns for the active part of the day and the squared close to open return for the inactive part of the day.

To compare the different variance measures, the results of the study are applied to two years S&P500 index futures data. In the empirical analysis it can be seen that the estimator that has the form of scaled active part variance underestimates the daily variance if the estimator which uses the electronic data is considered as reference. In addition, we can see that the optimal linear combination of intraday returns and squared overnight return declares a better estimation for the whole day variance comparing to the first and second estimator of Hansen and Lunde.

An AR(2) model could be effectively used to forecast one day ahead logarithmic volatility using all different measures of variance. Forecasting evaluation of RV_total that uses 24 hours data (electronic during nights) and the optimal variance measure of Hansen and Lunde, shows that RV_total has a better forecasting performance.

References

Andersen, T.G. and Tim Bollerslev, “Answering the skeptics: Yes, Standard Volatility Model to provide Accurate forecasts”, International Economic Review, vol.39, 1998, 885-905.

Andersen, T.G., Tim Bollerslev and F. X. Diebold, “Parametric and Nonparametric Measurements of Volatility, in Ait Sahalia and L.P. Hansen (Eds.), Handbook of financial Econometrics, forthcoming.

Andersen, T.G., Tim Bollerslev, F. X. Diebold, and Heiko Ebens, “The Distribution of Stock Return Volatility”, Journal of Financial Economics, vol. 61, 2001, 43-76.

Andersen, T.G., Tim Bollerslev, F. X. Diebold and Paul Labys, “Modeling and Forecasting Realized Volatility”, Econometrica, vol. 71, no. 2, 2003, 579-625.

Andersen, T.G., Tim Bollerslev, and S. Lange, “Forecasting Financial Market Volatility: Sample Frequency vis-a-vis Forecast Horizon”, Journal of Empirical Finance, vol. 6, 1999, 457-477.

Andreou, E., and E. Ghysels, “Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation, and Empirical Results” Journal of Business and Economic Statistics, vol. 20, 2002, 363-376.

Areal, Nelson M.P.C and Stephen Taylor, “The Realized Volatility of FTSE-100 Futures Prices”, Journal of Futures Markets, vol. 22, no. 7, 2002, 627-648.

Baillie, R.T., Tim Bollerslev, and H. O. Mikkelsen, “Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity”, Journal of Econometrics, vol. 74, 1996, 3-30.

Bandi, F. M., and J. R. Russell, “Microstructure Noise, Realized Volatility, and Optimal Sampling”, Review of Economics studies, vol.75, 2008, 339-369.

Bollerslev, Tim, “Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics”, vol. 31, 1986, 307-327.

Engle, R. F., “ Autoregression Conditional Heteroskedasticity with Estimates of Variance of U.K. inflation”, Econometrica, vol. 50, 1982, 987-1008.

Geweke, J., and S. Porter-Hudak, “The Estimation and Application of Long Memory Time Series Models”, Journal of Time Series Analysis, vol. 4, 1983, 221-238.

Hansen, R.H., and Asger Lunde, “A Realized Variance for the Whole day Based on Intermittent High Frequency data”, Journal of financial Econometrics, vol. 3, 2005, 525-554.

Hull, John, Options, Futures and other derivatives, Prentice Hall, 2000.

Martens, Martin, “Measuring and Forecasting S&P500 index Futures Volatility using High-Frequency Data “, vol. 22, no. 6, 2002, 497-518.

Zhang, L., P. A. Mykland, and Y. Ait-Sahalia, “A Tale of Two Time Scales: Determining Integrated Volatility with noisy High Frequency Data”, Journal of the American Statistical Association, vol. 100, 2005, 1394-1411.