Realized Volatility or Price Range: Evidence from a discrete simulation of the continuous time diffusion process

Degiannakis, Stavros and Livada, Alexandra

Department of Statistics, Athens University of Economics and Business, Postgraduate Department of Business Administration, Hellenic Open University

2013

Realized Volatility or Price Range: Evidence from a discrete simulation of the continuous time diffusion process

Stavros Degiannakis*, Alexandra Livada

Department of Statistics, Athens University of Economics and Business, 76 Patission str., Athens, 10 434, Greece

Postgraduate Department of Business Administration, Hellenic Open University, Aristotelous 18, Patras, 26 335, Greece

Abstract

The study provides evidence in favour of the price range as a proxy estimator of volatility in financial time series, in the cases that either intra-day datasets are unavailable or they are available at a low sampling frequency.

A stochastic differential equation with time varying volatility of the instantaneous log-returns process is simulated, in order to mimic the continuous time diffusion analogue of the discrete time volatility process. The simulations provide evidence that the price range measures are superior to the realized volatility constructed at low sampling frequency. The high-low price range volatility estimator is more accurate than the realized volatility estimator based on five, or less, equidistance points in time. The open-high-low-close price range is more accurate than the realized volatility estimator based on eight, or less, intra-period log-returns.

JEL classification: C15, C53, G17.

Keywords: Integrated Volatility, Intra-day Volatility, Price range, Realized volatility, Stochastic Differential Equation.

*Corresponding author e-mail: Stavros.Degiannakis@gmail.com, tel: 0030 210 8203521.

1. *Introduction*

Realized volatility, introduced by Andersen and Bollerslev (1998), is an alternative measure of daily volatility in financial markets. The modeling of realized volatility is based on the idea of using the sum of squared intraday returns to generate more accurate daily volatility measures. Merton (1980) was the first who noted the idea of using high frequency data to compute measures of volatility at a higher frequency, whereas French et al. (1987), Schwert (1989, 1990) and Schwert and Seguin (1990) computed the monthly variance by summing the variance of the daily log-returns. Nowadays there is a growing literature in constructing daily realized volatility from ultra-high frequency log-returns, i.e. intraday asset prices per minute. Andersen and Benzoni (2009), Andersen et al. (2001a, 2001b, 2003, 2010), McAleer and Medeiros (2008), among others, have provided comprehensive reviews for the estimation and the distributional properties of the realized volatility.

The realized volatility is a less noisy and more accurate estimate of volatility in financial time series than the squared daily log-returns[1]. However, the estimation of the realized volatility requires the availability of intra-day datasets. On the other hand, the price range, i.e. the difference between the highest and the lowest log-prices, can be constructed even when detailed intra-day datasets are not available, as the daily high and low prices are recorded in business newspapers and Japanese candlestick charting techniques[2].

The purpose of the present study is to provide evidence in favour of the use of the price range as a proxy estimator of volatility in financial time series, in the cases that either intra-day time series datasets are unavailable or they require a high cost of data collection and processing.

The price range can be constructed based on either two-data-points or four-data-points. The two-data-points price range estimator is based the highest and the lowest prices of the asset over a specific time interval, whereas the four-data-points price range requires, additionally, the first and the last prices of the asset. The simulations provide evidence that the price range measures are superior/inferior to the realized volatility constructed at low/high sampling frequency. Specifically, the two-data-points price range estimator provides more accurate volatility estimates that the realized volatility constructed with 8 equidistance points

[1] According to Oomen (2001), the average daily return variance is estimated more accurately by summing up squared intra-daily returns rather than calculating the squared daily return.

[2] A candlestick chart is a bar-chart that displays the open, close, high and low prices of the trading day (Nison, 2001).

in time. In example[3], for a daily trading period of 16 hours and 40 minutes, the price range can provide more accurate risk estimate than the sum of squared intraday returns at a sampling frequency of 125 minutes. Additionally, the four-data-points price range provides more accurate volatility estimates that the realized volatility constructed with 10 equidistance points in time; i.e. the price range is a more accurate volatility estimator than the realized volatility at a sampling frequency of 100 minutes (for a daily trading period of 16 hours and 40 minutes).

This paper is organized as follows. Section 2 illustrates the notion of *integrated volatility* as well as its relation to the *realized volatility*. Section 3 provides a brief description of the *price range* estimators, whereas section 4 provides the framework of the relative simulation. The last section concludes the paper.

2. *Integrated and Realized Volatility*

The instantaneous prices $p(t)$ represent the continuous time prices of the asset generated by the true data generated mechanism. Financial literature assumes that the instantaneous logarithmic price, $\log p(t)$, of a financial asset follows a simple diffusion process[4]:

$$d\log p(t) = \sigma(t)dW(t). \tag{1}$$

The $\sigma(t)$ is the volatility of the instantaneous log-returns process and the $W(t)$ is the standard Wiener process. Over the time interval $[a,b]$ the aggregated volatility, $\sigma^{2(IV)}_{[a,b]}$, is:

$$\sigma^{2(IV)}_{[a,b]} = \int_a^b \sigma^2(t)dt. \tag{2}$$

The *integrated variance*, $\sigma^{2(IV)}_{[a,b]}$, is the actual, but unobservable, variance we would like to estimate.

As the actual volatility is not observed, we require a proxy measure for the $\sigma^{2(IV)}_{[a,b]}$. Although the integrated volatility is a latent variable, according to the theory of quadratic variation of semi-martingales (Barndorff-Nielsen and Shephard, 2001, 2002, 2005), it can be consistently estimated by the realized volatility. The time interval is partitioned in τ

[3] A trading day of 16 hours and 40 minutes, i.e. the market is open from 07:00 to 23:40, is divided in $\tau = 1.001$ one-minute points in time.

[4] Although a jump-diffusion model (Barndorff-Nielsen and Shephard, 2004) is highly important for financial economics (asset allocation, derivatives pricing, risk management), we focus in simulating a continuous time diffusion without jumps, in order to concentrate on the comparison between price range and realized volatility measures.

equidistance points in time $t_1, t_2, ..., t_\tau$. At each point in time t_j, the integrated variance is decomposed to:

$$\sigma^{2(IV)}_{[a,b]} = \int_{t_1}^{t_2} \sigma^2(t)dt + \int_{t_2}^{t_3} \sigma^2(t)dt + ... + \int_{t_{\tau-1}}^{t_\tau} \sigma^2(t)dt, \tag{3}$$

For the length of each sub-interval tending to zero, $dt \approx t_j - t_{j-1}$, and the number of equidistance points in time tending to infinity, $\tau \rightarrow \infty$, the *realized volatility* is a consistent estimator for $\sigma^{2(IV)}_{[a,b]}$:

$$RV_t^{(\tau)} = \sum_{j=1}^{\tau} \left(\log P_{t_j} - \log P_{t_{j-1}}\right)^2, \tag{4}$$

The realized volatility converges in probability to the integrated volatility, as $\tau \rightarrow \infty$,

$$p \lim_{\tau \rightarrow \infty} \left(RV_t^{(\tau)}\right) = \sigma^{2(IV)}_{[a,b]}, \tag{5}$$

and is asymptotically normally distributed:

$$\frac{\sqrt{\tau}\left(RV_t^{(\tau)} - \int_a^b \sigma^2(t)dt\right)}{\sqrt{\int_a^b 2\sigma^4(t)dt}} \xrightarrow{d} N(0,1). \tag{6}$$

The asymptotic volatility of volatility, $\sigma^{2(IQ)}_{[a,b]}$, is termed *integrated quarticity*:

$$\sigma^{2(IQ)}_{[a,b]} = \int_a^b 2\sigma^4(t)dt. \tag{7}$$

The $RV_t^{(\tau)}$ would be an ideal estimate of volatility[5], over any time interval $[a,b]$, under the assumptions that i) the logarithmic prices follow the diffusion process and ii) there are no microstructure frictions[6].

[5] Consider the realized volatility for n days defined as the sum of squared returns observed over one-minute time intervals. Each trading day, the asset is pricing in the time interval $[a,b] = [09:00, 15:00]$, or, in other words, the market is open from 09:00 to 15:00. The five-days realized volatility defined as the sum of squared log-returns observed over one-minute time intervals is denoted as: $RV_t^{(n\cdot\tau)} = \sum_{t=1}^{n}\sum_{j=1}^{\tau}\left(\log P_{t_j} - \log P_{t_{j-1}}\right)^2$, where P_{t_j} are the financial asset prices for the trading day t, which is divided in $\tau = 360$ equidistance intra-day log-returns. The $RV_t^{(5*360)}$ denotes the five-days realized volatility from the trading day t up to the trading day $t+5$, based on $\tau = 360$ log-returns for each trading day.

[6] Microstructure frictions include discreteness of the pricing data, trading liquidity, transaction and regulatory costs, taxes, properties of the trading mechanism and protocols, the bid-ask spreads, etc. For a comprehensive explanation you are referred to the excellent reviews of Alexander (2008) and Madhavan (2000).

Barndorff-Nielsen and Shephard (2005), based on the realised power variation theory, examined the finite sample performance of the asymptotic approximation to the distribution of the realised variance. The realized power variation of order 2q is defined as:

$$RV_t^{(\tau)[2q]} = \sum_{j=2}^{\tau} \left(\log P_{t_j} - \log P_{t_{j-1}} \right)^{2q} . \tag{8}$$

They studied the finite sample behaviour of the realized variance

$$\frac{RV_t^{(\tau)} - \sigma_{[a,b]}^{2(IV)}}{\sqrt{\frac{2}{3} RV_t^{(\tau)[4]}}} \xrightarrow{d} N(0,1), \tag{9}$$

and the logarithmic realized variance

$$\frac{\log\left(RV_t^{(\tau)}\right) - \log\left(\sigma_{[a,b]}^{2(IV)}\right)}{\sqrt{\frac{2}{3} RV_t^{(\tau)[4]} RV_t^{-2}}} \xrightarrow{d} N(0,1), \tag{10}$$

as well. The asymptotic normality holds for $\log\left(RV_t^{(\tau)}\right)$ even for moderately small values of τ, whereas for the case of $RV_t^{(\tau)}$ a much higher value of τ is required. Barndorff-Nielsen and Shephard (2005) provided simulated evidence where the quantity:

$$\frac{\log\left(RV_t^{(\tau)}\right) - \log\left(\sigma_{[a,b]}^{2(IV)}\right) + 0.5 \max\left(\frac{2}{3} RV_t^{(\tau)[4]} RV_t^{-2}, \frac{2}{\tau} \right)}{\sqrt{\max\left(\frac{2}{3} RV_t^{(\tau)[4]} RV_t^{-2}, \frac{2}{\tau} \right)}} \xrightarrow{d} N(0,1) \tag{11}$$

improves the finite sample behaviour.

3. *Price Range Estimators of Volatility*

The two-data-points price range, introduced by Parkinson (1980), for the time interval $[a,b]$, is the difference between the highest and the lowest log-prices:

$$Range_{[2],[a,b]} = \frac{1}{4\log(2)} \left(\log\left(\max\left(P_{t_j}\right)\right) - \log\left(\min\left(P_{t_j}\right)\right) \right)^2 . \tag{12}$$

The advantage of the price range proxy is its construction due to the availability of the high and low prices. Even when detailed intra-day datasets are not available, intra-day high and low prices are recorded in business newspapers and Japanese candlestick charting techniques.

Under the assumption that the instantaneous logarithmic price, $\log p(t)$, of a financial asset follows the diffusion process in equation (1), Parkinson (1980) showed that

$$E\left(\log\left(\frac{\max(p(t))}{\min(p(t))}\right)\right) = \sqrt{8/\pi}\sigma(t). \tag{13}$$

and

$$E\left(\left(\frac{\max(p(t))}{\min(p(t))}\right)^2\right) = 4\log(2)\sigma(t)^2. \tag{14}$$

The computation of the price range is based on two data points; the highest and the lowest prices over the time interval. Garman and Klass (1980) proposed an extension of the price range, incorporating information for the opening and the closing prices, as well. The four-data-points price range estimator, or $Range_{[4],[a,b]}$, is computed as:

$$Range_{[4],[a,b]} = \frac{1}{2}\left(\log\left(\frac{\max\left(P_{t_j}\right)}{\min\left(P_{t_j}\right)}\right)\right)^2 - (2\log(2)-1)\left(\log\left(\frac{P_{t_\tau}}{P_{t_1}}\right)\right)^2, \tag{15}$$

where P_{t_1} and P_{t_τ} are the open and close prices for the time interval $[a,b]$, which is partitioned in τ equidistance points, respectively.

4. Simulations

We simulate a stochastic differential equation, by relaxing the assumption of constant volatility of the instantaneous log-returns process in equation (1). We undertake a time varying volatility of the instantaneous log-returns process in order to mimic the continuous time diffusion analogue of the GARCH(1,1) process[7].

The GARCH(1,1) process is defined as:

$$\begin{aligned} y_t &= \sigma_t z_t, \\ \sigma_t^2 &= a_0 + a_1\varepsilon_{t-1}^2 + b_1\sigma_{t-1}^2, \\ z_t &\overset{i.i.d.}{\sim} N(0,1), \end{aligned} \tag{16}$$

According to Andersen and Bollerslev (1998) and Drost and Werker (1996), the discrete time GARCH(1,1) process with parameters a_0, a_1 and b_1 is related to the continuous time GARCH(1,1) diffusion:

[7] Literature has provided an extensive number of ARCH type processes that model the properties of financial assets. In example, the FIGARCH model captures the long memory property of volatility (Baillie et al., 1996), the regime switching ARCH model allows the modelling of regimes in markets (Hamilton and Susmel, 1994), etc. However, the GARCH(1,1) is the most widely applied discrete time volatility process which captures the property of volatility clustering in asset returns; see also Hansen and Lunde (2005).

$$
\begin{aligned}
d\log(p(t)) &= \sigma(t)dW_1(t) \\
d\sigma^2(t) &= a_0^*\left(a_1^* - \sigma^2(t)\right)dt + \sqrt{2a_0^* b_1^* \sigma(t)}\,dW_2(t),
\end{aligned} \tag{17}
$$

with $W_1(t)$ and $W_2(t)$ denoting independent standard Wiener processes and with parameters a_0^*, a_1^*, b_1^* relating to those of the discrete time model as

$$
a_0^* = -\log(a_1 + b_1),
$$

$$
a_1^* = {a_0}\Big/{(1 - a_1 - b_1)},
$$

$$
b_1^* = \frac{2(\log(a_1 + b_1))^2}{\dfrac{\left(1-(a_1+b_1)^2\right)(1-b_1)^2}{a_1\left(1-b_1(a_1+b_1)\right)} + \left(6\log(a_1+b_1) + 2(\log(a_1+b_1))^2 + 4(1-a_1-b_1)\right)}.
$$

In general, as the length of the discrete time intervals goes to zero, the stochastic difference ARCH process convergences to a stochastic differential equation. For technical details see Nelson (1990).

We assume a generated data process of 1.000 trading days for each of which there will be 1.000 intraday log-returns[8]. Therefore, the simulated process P_{t_j}, where $j = 0,1,...,\tau$ and $t = 1,...,T$, for $\tau = 1.000$ equidistance points in time and $T = 1.000$ days is observed at sampling frequency $m = \dfrac{b-a}{\tau} = \dfrac{1-0}{1001}$, or $dt = t_j - t_{j-1} = 1/1.000$. Therefore, there are $\tau = 1.000$ intra-day *log-returns* over the daily intervals, $[a,b] = [0,1]$.

Hence, we generate 1.000.000 observations from the continuous time GARCH(1,1) diffusion in framework (17). The discrete presentation for $a_0 = 0{,}001$, $a_1 = 0{,}12$ and $b_1 = 0{,}80$ in equation (16) is[9]:

$$
\begin{aligned}
\log(p(t+dt)) &= \log(p(t)) + \sigma(t)\sqrt{dt}\,W_1(t), \\
\sigma^2(t+dt) &= 0{,}00108dt + \sigma^2(t)\left(1 - 0{,}083dt + \sqrt{0{,}084dt}\,W_2(t)\right),
\end{aligned} \tag{18}
$$

where $W_1(t)$ and $W_2(t)$ denote independent standard normal variables. Then, we simulate the $T = 1.000$ daily log-returns, y_t, as $y_t = \left(\log p_{t_\tau} - \log p_{t-1_\tau}\right)$. Note that under the ideal situations of the simulated framework; $\log p_{t_\tau} \equiv \log(p(t\tau))$. Our purpose is to estimate the discrete time GARCH(1,1) model for the 1.000 simulated daily log-returns as:

---

[8] In the simulated framework there are no market frictions. Thus we do need to take into consideration any frictions, such as the bid-ask spread, the time interval that the market is closed, etc.

[9] The values of the parameters reflect the representative estimates of the parameters of a GARCH(1,1) process for stock indices.

$$
\begin{aligned}
y_t &= \sigma_t z_t \ , \\
\sigma_t^2 &= a_0 + a_1 y_{t-1}^2 + b_1 \sigma_{t-1}^2 , \\
z_t &\sim N(0,1).
\end{aligned} \tag{19}
$$

The estimates of the conditional variance are denoted: $\hat{\sigma}_t^2 = a_0^{(T)} + a_1^{(T)} y_{t-1}^2 + b_1^{(T)} \hat{\sigma}_{t-1}^2$. The realized volatility is computed for sampling frequencies of $m = 250^{-1}$, 200^{-1}, 125^{-1}, 100^{-1}, 50^{-1}, 40^{-1}, 25^{-1}, 20^{-1}, 10^{-1}, 8^{-1}, 5^{-1}, 4^{-1}, 2^{-1}, 1, or equivalently for $\tau = 250$, 200, 125, 100, 50, 40, 25, 20, 10, 8, 5, 4, 2, 1 points in time. The price range measures are computed according to equations (12) and (15). Figure 1 presents a visual inspection of the construction of the realized variance, for the day t, for $\tau = 100$ points in time, or equivalently for a sampling frequency of $m = 1/100$.

[Insert Figure 1 About here]

Table 1 presents the values of the mean squared distance between conditional variance estimate and realized variance. The mean squared distance is usually referred as MSE loss function:

$$
MSE^{(\tau)} = T^{-1} \sum_{t=1}^{T} \left(\hat{\sigma}_t^2 - RV_t^{(\tau)} \right)^2 . \tag{20}
$$

Hansen and Lunde (2006) derived conditions which ensure that the ranking of any two variance forecasts by a loss function is the same whether the ranking is done via the true variance, $\sigma_{[0,1]}^{2(IV)}$, or via a conditionally unbiased variance proxy, i.e. $RV_t^{(\tau)}$. The MSE loss function ensures the equivalence of the ranking of volatility models that is induced by the true volatility and its proxy.

[Insert Table 1 about here]

Naturally, the MSE loss function minimises as $\tau \to \infty$. According to Table 1, both price range proxies are superior to the realized variance measure for moderate values of τ, and inferior to the realized variance for larger values of τ.

We repeat the simulation of the 1.000.000 observations several times in order to investigate the robustness of the findings. Specifically, the simulation is repeated 2.000 times. Table 2 presents the average and the median values of the MSE loss functions corresponding to the 2.000 simulations.

[Insert Table 2 about here]

According to Table 2, the MSE loss function decreases monotonically with τ. The average value of $10^4 MSE^{(1)}$ is 6,506, whereas the average value of $10^4 MSE^{(250)}$ is 0,660. Hence, the volatility measure based on the daily log-returns has on average 10 times higher MSE value compared to the volatility measure which is based on 250 intra-day log-returns. The median values of the $10^4 MSE^{(\tau)}$ provide similar evidence. Both $Range_{[2],t}$ and $Range_{[4],t}$ are superior to the $RV_t^{(\tau)}$ when the realized variance measure is constructed on the basis of a small number of intraday log-returns, i.e $\tau < 8$. More specifically, for the $Range_{[2],t}$, an average value of the MSE loss function of 1,881 indicates that the two-data-points price range volatility estimator is more accurate than the realized volatility estimator which is based on $\tau \leq 5$ intra-day log-returns. In the case of the four-data-points price range volatility estimator, the average of the MSE loss function of 1,491 provides evidence that the $Range_{[4],t}$ is more accurate than the realized volatility estimator when it is based on $\tau \leq 8$ intra-day log-returns. In the case the median value of the MSE loss function is under examination, the results remain qualitatively similar.

Hence, under the ideal situations of a simulated framework, the highest the sampling frequency, the lowest the value of the MSE loss function. However, if intra-day data are not available, or they are available for less than 8 equidistance points in time, then the price range estimators are more accurate volatility estimators than the realized volatility.

5. *Conclusion*

Modern applied financial literature concludes that volatility estimates based on intra-day asset prices are the most accurate estimates of volatility in time series. However, in the cases that either intra-day datasets are unavailable or they require a high cost of data collection, the price range volatility estimator is still an adequate proxy for estimating volatility. The price range estimates can be constructed with data that are available in business newspapers and Japanese candlestick charting techniques.

Two versions of the price range were investigated. The two-data-points price range estimator requires the highest and the lowest prices within the day. The four-data-points price range is based on the highest and the lowest prices as well as on the first and the last prices of the asset. The simulations provide evidence that the price range measures are superior to the realized volatility constructed at low sampling frequency. The two-data-points price range volatility estimator is more accurate than the realized volatility estimator based on $\tau \leq 5$ intra-day log-returns. The four-data-points price range volatility estimator is more accurate than the realized volatility estimator that is based on $\tau \leq 8$ intra-day log-returns.

The comparison of the realized volatility and price range measures under a diffusion process with jumps or the existence of a long memory volatility process would be an interesting issue for future study.

Acknowledgement

The research is supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme (PERG08-GA-276904).

References

1. **Alexander, C.O. (2008).** Market Risk Analysis: Quantitative Methods in Finance, Volume 1. John Wiley and Sons, New York.
2. **Alizadeh, S., Brandt, M.W. and Diebold, F.X. (2002).** Range-Based Estimation of Stochastic Volatility Models. *Journal of Finance*, LV11, 1047-1091.
3. **Andersen, T. and Bollerslev, T. (1998).** Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*, 39, 885-905.
4. **Andersen, T. and Benzoni, L. (2009).** Realized Volatility. In (eds.) Andersen, T.G, Richard A.D., J.P. Kreiss and T. Mikosch, in *Handbook of Financial Time Series*, Springer-Verlag, 555-575.
5. **Andersen, T., Bollerslev, T., Diebold, F.X. and Ebens, H. (2001a).** The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics*, 61, 43-76.
6. **Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. (2001b).** The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, 96, 42-55.
7. **Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. (2003).** Modeling and Forecasting Realized Volatility. *Econometrica*, 71, 529-626.
8. **Andersen, T., Bollerslev, T. and Diebold, F.X. (2010).** Parametric and Nonparametric Volatility Measurement. In (eds.) Yacine Ait-Sahalia and Lars Peter Hansen, *Handbook of Financial Econometrics*, Amsterdam, Elsevier Science B.V, 67-128.
9. **Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996).** Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 74, 3-30.

10. **Barndorff-Nielsen, O.E. and Shephard, N. (2001).** Non-Gaussian Ornstein-Uhlenbeck based Models and Some of their Uses in Financial Economics. *Journal of the Royal Statistical Society, Series B*, 63, 197-241.

11. **Barndorff-Nielsen, O.E. and Shephard, N. (2002).** Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society, Series B*, 64, 253–280.

12. **Barndorff-Nielsen, O.E. and Shephard, N. (2004).** Power and Bipower Variation with Stochastic Volatility and Jumps, *Journal of Financial Econometrics*, 2, 1–37.

13. **Barndorff-Nielsen, O.E. and Shephard, N. (2005).** How Accurate is the Asymptotic Approximation to the Distribution of Realised Volatility? In (eds.) Andrews, D., Powell, J., Ruud, P., and Stock, J., *Identification and Inference for Econometric Models*. Cambridge University Press, Cambridge.

14. **Drost, F.C. and Werker, B.J.M. (1996).** Closing the GARCH Gap: Continuous Time GARCH Modeling. *Journal of Econometrics*, 74, 31-57.

15. **French, K.R., Schwert, G.W. and Stambaugh, R.F. (1987).** Expected Stock Returns and Volatility. *Journal of Financial Economics*, 19, 3-29.

16. **Garman, M. and Klass, M. (1980).** On the Estimation of Security Price Volatilities from Historical Data. *Journal of Business*, 53, 67-78.

17. **Hansen, P.R. and Lunde, A. (2005).** A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873-889.

18. **Hansen, P.R. and Lunde, A. (2006).** Consistent Ranking of Volatility Models. *Journal of Econometrics*, 131, 97-121.

19. **Madhavan, A. (2000).** Market Microstructure: A Survey. *Journal of Financial Markets*, 3, 205-258.

20. **McAleer, M. and Medeiros, M.C. (2008).** Realized Volatility: A Review, *Econometric Reviews*, 27(1), 10-45.

21. **Merton, R.C. (1980).** On Estimating the Expected Return on the Market: An Explanatory Investigation. *Journal of Financial Economics*, 8, 323-361.

22. **Nison, S. (2001).** Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Investment Techniques of the Far East. New York Institute of Finance, New York.

23. **Nelson, D. (1990).** ARCH Models as Diffusion Approximations. *Journal of Econometrics*, 45, 7-38.

24. **Oomen, R. (2001).** Using High Frequency Stock Market Index Data to Calculate, Model and Forecast Realized Volatility. *Department of Economics, European University Institute,* Manuscript.

25. **Parkinson, M. (1980).** The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Business*, 53(1), 61-65.

26. **Schwert, G.W. (1989).** Why Does Stock Market Volatility Changes Over Time. *Journal of Finance*, 44, 1115-1153.

27. **Schwert, G.W. (1990).** Stock Volatility and the Crash of ‘87. *Review of Financial Studies*, 3, 77-102.

28. **Schwert, G.W. and Seguin, P. (1990).** Heteroskedasticity in Stock Returns. *Journal of Finance*, 45, 1129–1155.

Figures and Tables

Table 1. Values of the MSE loss functions. The data generating process is the continuous time diffusion

$$\log(p(t+dt)) = \log(p(t)) + \sigma(t)\sqrt{dt}W_1(t),$$

$$\sigma^2(t+dt) = 0{,}00108dt + \sigma^2(t)\left(1 - 0{,}083dt + \sqrt{0{,}084dt}W_2(t)\right).$$

The conditional variance, $\hat{\sigma}_t^2 = a_0^{(T)} + a_1^{(T)} y_{t-1}^2 + b_1^{(T)} \hat{\sigma}_{t-1}^2$, is estimated from the GARCH(1,1) model, represented by $y_t = \sigma_t z_t$, $\sigma_t^2 = a_0 + a_1 y_{t-1}^2 + b_1 \sigma_{t-1}^2$ and $z_t \sim N(0,1)$ for the dependent variable $y_t = \left(\log P_{t_{1000}} - \log P_{t-1_{1000}}\right)$, or $y_t = \log(p(1.000t)) - \log(p(1.000(t-1)))$.

Volatility Proxy	$10^4 MSE^{(\tau)} = 10^4 T^{-1} \sum_{t=1}^{T} \left(\hat{\sigma}_t^2 - RV_t^{(\tau)}\right)^2$
$RV_t^{(250)}$	1,071
$RV_t^{(200)}$	1,083
$RV_t^{(125)}$	1,126
$RV_t^{(100)}$	1,168
$RV_t^{(50)}$	1,295
$RV_t^{(40)}$	1,434
$RV_t^{(25)}$	1,457
$RV_t^{(20)}$	1,640
$RV_t^{(10)}$	2,057
$RV_t^{(8)}$	2,431
$RV_t^{(5)}$	2,975
$RV_t^{(4)}$	4,227
$RV_t^{(2)}$	5,630
$RV_t^{(1)}$	10,640
$Range_{[2],t}$	2,812
$Range_{[4],t}$	2,087

Table 2. Average and median values of the MSE loss functions of the 2.000 simulations. The data generating process is the continuous time diffusion

$$\log(p(t+dt)) = \log(p(t)) + \sigma(t)\sqrt{dt}W_1(t),$$
$$\sigma^2(t+dt) = 0{,}00108dt + \sigma^2(t)\left(1 - 0{,}083dt + \sqrt{0{,}084dt}W_2(t)\right).$$

The conditional variance, $\hat{\sigma}_t^2 = a_0^{(T)} + a_1^{(T)} y_{t-1}^2 + b_1^{(T)} \hat{\sigma}_{t-1}^2$, is estimated from the GARCH(1,1) model, represented by $y_t = \sigma_t z_t$, $\sigma_t^2 = a_0 + a_1 y_{t-1}^2 + b_1 \sigma_{t-1}^2$ and $z_t \sim N(0,1)$ for the dependent variable $y_t = \left(\log P_{t_{1000}} - \log P_{t-1_{1000}}\right)$, or $y_t = \log(p(1.000t)) - \log(p(1.000(t-1)))$.

Volatility Proxy	Average of $10^4 MSE^{(\tau)} = 10^4 T^{-1} \sum_{t=1}^{T} \left(\hat{\sigma}_t^2 - RV_t^{(\tau)}\right)^2$	Median of $10^4 MSE^{(\tau)} = 10^4 T^{-1} \sum_{t=1}^{T} \left(\hat{\sigma}_t^2 - RV_t^{(\tau)}\right)^2$
$RV_t^{(250)}$	0,660	0,471
$RV_t^{(200)}$	0,677	0,481
$RV_t^{(125)}$	0,740	0,515
$RV_t^{(100)}$	0,776	0,539
$RV_t^{(50)}$	0,887	0,626
$RV_t^{(40)}$	0,926	0,655
$RV_t^{(25)}$	1,030	0,735
$RV_t^{(20)}$	1,111	0,795
$RV_t^{(10)}$	1,411	1,033
$RV_t^{(8)}$	1,565	1,150
$RV_t^{(5)}$	1,994	1,483
$RV_t^{(4)}$	2,318	1,714
$RV_t^{(2)}$	3,748	2,791
$RV_t^{(1)}$	6,506	4,965
$Range_{[2],t}$	1,881	1,402
$Range_{[4],t}$	1,491	1,106

Figure 1. Determination of realized variance for day t, $RV_t^{(\tau)} = \sum_{j=1}^{\tau}\left(\log P_{t_j} - \log P_{t_{j-1}}\right)^2$, when 1000 intraday observations are available and $\tau = 100$ equidistance points in time are considered.

$$
\left.\begin{array}{l}
\left.\begin{array}{c} \log P_{t_1} \\ . \\ . \\ . \\ \log P_{t_{10}} \end{array}\right\} \quad \log P_{t_{10}} - \log P_{t_1} = y_{t_{10}} \\
\left.\begin{array}{c} \log P_{t_{11}} \\ . \\ . \\ . \\ \log P_{t_{20}} \end{array}\right\} \quad \log P_{t_{20}} - \log P_{t_{11}} = y_{t_{20}} \\
\begin{array}{ccc} . & \qquad & . \\ . & & . \\ . & & . \end{array} \\
\left.\begin{array}{c} \log P_{t_{j-9}} \\ . \\ . \\ . \\ \log P_{t_j} \end{array}\right\} \quad \log P_{t_j} - \log P_{t_{j-9}} = y_{t_j} \\
\begin{array}{ccc} . & \qquad & . \\ . & & . \\ . & & . \end{array} \\
\left.\begin{array}{c} \log P_{t_{991}} \\ . \\ . \\ . \\ \log P_{t_{1000}} \end{array}\right\} \quad \log P_{t_{1000}} - \log P_{t_{991}} = y_{t_{1000}}
\end{array}\right\} \quad \sum_{j=1}^{100} y_{t_j}^2
$$