Journal Code: S T A N | Article ID: 1 2 0 1 7 | Dispatch: 17.07.13 | CE: Patrimonio, Pat
No. of Pages: 27 | ME:

Statistica Neerlandica (2013)
doi:10.1111/stan.12017

A new look at variance estimation based on low, high and closing prices taking into account the drift

Piotr Fiszeder* and Grzegorz Perczak

Department of Econometrics and Statistics, Faculty of Economic Sciences and Management, Nicolaus Copernicus University, ul. Gagarina 13a, 87-100 Torun, Poland

The joint distribution of low, high and closing prices of the arithmetic Brownian motion is used to evaluate the properties of the most popular estimators of the variance constructed on the basis of high, low and closing prices. The expected values and mean square errors of the Parkinson, Garman–Klass and Rogers–Satchell estimators for the process with a zero drift and a non-zero drift are derived. Moreover, new volatility estimators, more efficient in the majority of financial applications than the Rogers–Satchell estimator, are proposed. The considered estimators are applied to the estimation of the volatility of the Polish stock index WIG20.

Keywords and Phrases: volatility estimators, low, high, closing prices, Brownian motion.

1 Introduction Q1

Volatility plays a key role in many financial applications like portfolio analysis, valuation of assets or risk management. The issue of variance estimation has been a subject of plenty of studies, and many estimators of variance have been proposed in the literature. In numerous analyses, daily returns of financial assets constructed on basis of closing prices are used to estimate variance. However, ANDERSEN and BOLLERSLEV (1998) have shown that although the daily squared return is an unbiased estimator of variance, it is also generally very noisy, which makes it an inefficient estimator. For this reason, more effective estimators of variance based on the use of additional information about prices during the day have been applied. One such approach is the application of information about low and high prices during the day.

Among the best-known volatility measures constructed on the basis of open, low, Q2 high and closing prices estimators of PARKINSON (1980), those of GARMAN and KLASS (1980), ROGERS and SATCHELL (1991) and YANG and ZHANG (2000) can be included. Such estimators, sometimes called price range estimators, are commonly used to estimate volatility by practitioners of a financial market, because they are

*piotr.fiszeder@umk.pl

Published by Wiley Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

more than two to even more than seven times more efficient than the estimator calculated as the daily squared return of closing prices. Over the last few years, a renewed interest in these estimators has been observed also within the scientific community. The usage of low and high prices belongs to the area in which extensive both theoretical and empirical research is now conducted. An overview of such studies can be found in the paper of Chou, CHOU and LIU (2009).

Application of range estimators is an alternative for the usage of data with higher than a daily frequency, so-called intraday data or high frequency data (see, e.g. the papers of MCALEER and MEDEIROS (2008), PATTON (2011), VISSER (2011), HAUTSCH (2012) and PIGORSCH, PIGORSCH and POPOV (2012)). The usage of data with low and high daily prices has many advantages in comparison to intraday data such as the wider availability, the lower acquisition costs, considerably lower databases requirements and a greater robustness to microstructure effects. Furthermore, direct application of intraday data means some problems such as the existence of daily cyclical fluctuations, the existence of strong autocorrelation or a significant impact of the publication of macroeconomic information on quotations. Moreover, the accuracy of the simplest and the least efficient range estimator, namely Parkinson, is similar to the accuracy of realized volatility estimator constructed using four, five or six observations during 1 day (PARKINSON, 1980, ANDERSEN, BOLLERSLEV, 1998). In light of the aforementioned observation, it is worth conducting studies on estimators of variance constructed on the basis of high, low and closing prices; in particular, it is worth examining the properties of such estimators for different model assumptions and searching for new more efficient estimators.

The main contributions of the paper are as follows: (i) to estimate the bias and analytically assess effectiveness of various, popular estimators of variance formulated on the basis of low, high and closing prices for the process with a non-zero drift. Those properties were determined for the process with a zero drift so far; (ii) to propose new estimators of variance which are more efficient in the vast majority of financial applications than commonly used the Rogers–Satchell estimator; (iii) to analytically assess properties of popular and proposed range estimators based on the same mathematical tool. According to our knowledge, it is the first attempt to use the joint density of the random vector of low, high and closing values of the arithmetic Brownian motion with drift to this goal.

The considered estimators are applied in this paper to data from the capital market. We have shown that volatility estimates based on low, high and closing prices are more accurate than the ones formulated on the basis of the GARCH models. CHOU (2005) and LI and HONG (2011), in turn, have demonstrated that range-based volatility models that are formulated on the basis of such data give more accurate forecasts of volatility than the ones based on the GARCH model.

The plan for the rest of the paper is as follows. In section 2, the joint density of the random vector of low, high and closing values of the arithmetic Brownian motion and its characteristic function are presented. Raw moments of random variables low, high and closing values of the arithmetic Brownian motion with both a zero drift and a non-zero drift were also calculated. Section 3 describes the well-known estimators of the variance based on the high, low and closing prices. The analytical evaluation of their efficiency is performed, and also the values of the bias are derived for situations when assumptions under which the estimators were constructed are not met. Afterwards, new variance estimators are proposed. Finally, the considered estimators are applied to the estimation of the volatility of the Polish stock index WIG20. It is worth noting the relatively small number of studies on Polish financial time series in comparison with other emerging or developed markets. Summary is given in section 4. This paper is the extension of the results presented by PERCZAK and FISZEDER (2013).

2 The three-dimensional random vector of low, high and closing values of the arithmetic Brownian motion

2.1 Symbols and assumptions

It is assumed that t is a fixed positive real number, the time unit t is 1 day (i.e. $t = 1$ stands for 1 day equivalent to 24 h), the variable τ satisfies the condition $0 \leq \tau \leq t$, $\tau = 0$ means the moment of the commencement of trading in the current period, B_τ is the Wiener process and S_τ is the price of a financial instrument at time τ. A return defined as $x_t = \ln(S_t/S_0)$ is the realization of the arithmetic Brownian motion $X_\tau = \mu\tau + \sigma B_\tau$ at the point $\tau = t$. Furthermore, it is assumed that $A_t := \min_{0 \leq \tau \leq t} X_\tau$ and $C_t := \max_{0 \leq \tau \leq t} X_\tau$.

Usually, it is assumed that the considered period t is equal to 24 h. The issue of determining the low and high daily returns is more complex in practice (it is presented in Figure 1). F1

Only four values of quotations during the day are usually commonly available: today's open price (O_1), observed today's low price (L_1), observed today's high price (H_1) and today's closing price (S_1). The application of information about prices observed only from the commencement to the closure of market quotations does not allow for the proper calculation of the volatility during the whole day. In this case, one can estimate at most the volatility that takes place during the functioning of the market. Values of L_1 and H_1 are determined for the period when the market is open, but not for the whole day, as shown in Figure 1. Therefore, it is necessary to redefine low and high daily returns so as to cover the period between the two subsequent closures of quotations (i.e. 24 h). Finally, the following definitions of daily low and high returns will be adopted $A_t := \min(\ln L_1, \ln S_0) \ln S_0$, $C_t := \max(\ln H_1, \ln S_0) - \ln S_0$, respectively, which will be used further in the paper. The variable $X_\tau := \ln(S_t/S_0)$ will be defined as the closing return. Terms of low, high and closing returns used in the paper result from the modification of definitions for financial applications; namely, rates of returns are used, and non-trading periods are taken into account. Moreover, the constancy of the variance of the process is assumed during the day, while the variability is permitted on consecutive days. For this reason, the following condition has been established: $0 < t \leq 1$.

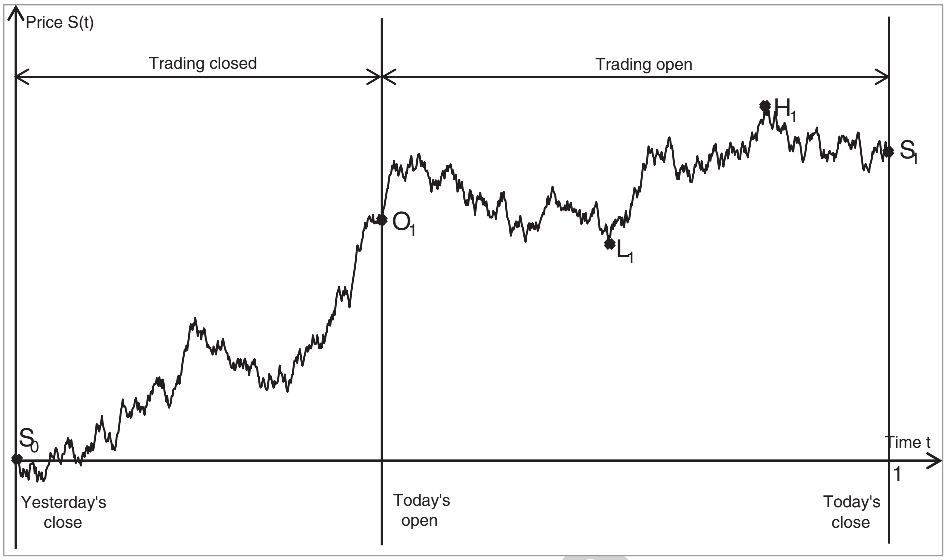

Fig. 1. Example of the price realization of a financial instrument during the day. Source: Own elaboration based on the work of 1980 GARMAN and KLASS (1980, p.70).

2.2 The joint probability density function of random variables low, high and closing returns of the arithmetic Brownian motion

In this section, we describe the density and characteristic functions, which are used later in subsequent parts of the paper to evaluate the properties of well-known and also new proposed estimators of variance.

The knowledge of the joint density of the vector of three random variables (A_t, C_t, X_t) allows us to determine the expected values of selected functions of such variables. The form of this density is relatively complex; that is why, joint densities of random vectors (C_t, X_t) and (A_t, X_t) will be presented firstly. There are many studies in which the issue of the joint density of the random vector (C_t, X_t) is considered (see, e.g. COX and MILLER (1965) and HARRISSON (1985)). This problem is also very often discussed in the valuation of barrier options (see, e. g. LI (1998) and JAKUBOWSKI *et al.* (2006)).

Let us consider the random event $\{C_t \le c, X_t \le x\}$ for $x \le c$ and $0 \le c$. The cumulative distribution function of the joint density of the random vector (C_t, X_t) is described by the following formula (see HARRISSON (1985), p. 13, Equation 8):

$$\mathrm{P}\left(C_t \le c, X_t \le x; \mu, \sigma^2, t\right) = \Phi\left(\frac{x - \mu t}{\sigma\sqrt{t}}\right) - e^{\frac{2\mu c}{\sigma^2}}\Phi\left(\frac{x - 2\mathrm{c} - \mu t}{\sigma\sqrt{t}}\right), \tag{1}$$

where Φ stands for the cumulative standard normal distribution function.

Based on Equation 1, the following expression for the joint density of the random vector (C_t, X_t) was derived:

$$f_{C_t,X_t}(c,x;\mu,\sigma^2,t) = \frac{\partial^2 P(C_t \le c, X_t \le x;\mu,\sigma^2,t)}{\partial c \partial x} = \frac{\sqrt{2}(2c-x)e^{-\frac{(x-2c-\mu t)^2-4\mu ct}{2\sigma^2 t}}}{\sqrt{\pi}\sigma^3 t^{\frac{3}{2}}}. \quad (2)$$

The characteristic function of the joint density of the random vector (C_t, X_t)

$$\varphi_{C_t,X_t}(q,s;\mu,\sigma^2,t) = \int_0^{\infty}\int_{-\infty}^{c} e^{iqc+isx} f_{C_t,X_t}(c,x;\mu,\sigma^2,t)\,dxdc \quad (3)$$

is described in detail in Appendix A.

Let us analyse another random event $\{A_t > a, X_t > x\}$ for $a \le 0$ and $a \le x$. By analogy with the previously considered event, one can formulate the probability

$$\mathrm{P}(A_t > a, X_t > x;\mu,\sigma^2,t) = \Phi\left(\frac{-x+\mu t}{\sigma\sqrt{t}}\right) - e^{\frac{2\mu a}{\sigma^2}}\Phi\left(\frac{-x+2a+\mu t}{\sigma\sqrt{t}}\right). \quad (4)$$

The density and its characteristic function were also derived in this case. The joint density of the random vector (A_t, X_t) is given by the formula:

$$f_{A_t,X_t}(a,x;\mu,\sigma^2,t) = \frac{\partial^2 P(A_t > a, X_t > x;\mu,\sigma^2,t)}{\partial a \partial x} = \frac{\sqrt{2}(x-2a)e^{-\frac{(x-2a-\mu t)^2-4\mu at}{2\sigma^2 t}}}{\sqrt{\pi}\sigma^3 t^{\frac{3}{2}}}. \quad (5)$$

The characteristic function of the joint density of the random vector (A_t, X_t)

$$\varphi_{A_t,X_t}(p,s;\mu,\sigma^2,t) = \int_{-\infty}^{0}\int_{a}^{\infty} e^{ipa+isx} f_{A_t,X_t}(a,x;\mu,\sigma^2,t)\,dxda \quad (6)$$

is also given in detail in Appendix A.

The issue of finding the density of the random vector (A_t, C_t, X_t) was considered, for instance, in the papers of Cox and Miller (1965) and Li (1999). The density of X_t Q6 with upper and lower absorbing barriers equal to c and a, respectively, is given by the following formula (see Cox and Miller (1965), p. 222, Equation 78): Q7

$$f_{X_t}(x;A_t > a, C_t \le c;\mu,\sigma^2,t) = \frac{1}{\sqrt{2\pi t}\sigma}\sum_{k=-\infty}^{\infty} e^{\frac{2k(c-a)\mu}{\sigma^2}}\left(e^{-\frac{(x-2k(c-a)-\mu t)^2}{2\sigma^2 t}} - e^{\frac{2c\mu t-(x-2c-2k(c-a)-\mu t)^2}{2\sigma^2 t}}\right), \quad (7)$$

where $a \le 0 \le c, a \le x \le c$.

Using the expression (7), the joint density of random vector (A_t, C_t, X_t) was derived:

$$\begin{aligned} f_{A_t,C_t,X_t}(a,\ c,\ x;\mu,\sigma^2,t) &= -\frac{\partial^2 f_{X_t}(x;C_t \le c, A_t > a;\mu,\sigma^2,t)}{\partial a \partial c} \\ &= \frac{1}{\sqrt{2\pi}\sigma^5 t^{5/2}}\sum_{k=-\infty}^{\infty}(g(a,c,x;k,k,\mu,\sigma,t) \\ &\quad -g(a,c,x;k,k+1,\mu,\sigma,t)), \end{aligned} \quad (8)$$

where the function g is described as follows:

$$g(a,c,x;m,n,\mu,\sigma,t) = 4mn\left[(x-2(nc-ma))^2-\sigma^2 t\right]e^{\frac{2(nc-ma)\mu t-(x-2(nc-ma)-\mu t)^2}{2\sigma^2 t}}. \tag{9}$$

The characteristic function of the joint density of the random vector (A_t,C_t,X_t) is given by the following formula:

$$\varphi_{A_t,C_t,X_t}(p,q,s;\mu,\sigma^2,t) = \int_{-\infty}^{0}\int_{0}^{\infty}\int_{a}^{c} e^{ipa+iqc+isx} f_{A_t,C_t,X_t}(a,\ c,\ x;\mu,\sigma^2,t)\,dx\,dc\,da. \tag{10}$$

The calculation of the aforementioned integral is presented in Appendix A.

The form of the characteristic functions allows for the calculation of the expected values of selected random variables. For instance, for integers u, v, w, we obtain

$$E[A_t{}^u C_t{}^v X_t{}^w] = \frac{1}{i^n}\frac{\partial^n \varphi_{A_t,C_t,X_t}(p,q,s;\mu,\sigma^2,t)}{\partial p^u \partial q^v \partial s^w}, \tag{11}$$

where $n = u + v + w$.

When $u = 0$ or $v = 0$, it is convenient to apply the following formulae:

$$E[C_t{}^v X_t{}^w] = \frac{1}{i^n}\frac{\partial^n \varphi_{C_t,X_t}(q,s;\mu,\sigma^2,t)}{\partial q^v \partial s^w}, \tag{12}$$

$$E[A_t{}^u X_t{}^w] = \frac{1}{i^n}\frac{\partial^n \varphi_{A_t,X_t}(p,s;\mu,\sigma^2,t)}{\partial p^u \partial s^w}, \tag{13}$$

respectively.

General expressions for the expected values of products of any powers of variables (A_t,C_t,X_t) given in Equations 11–13 are applied to derive all first, second and fourth raw moments of the considered variables (details are presented in Appendices B and C, respectively, for the processes with a non-zero drift and with a zero drift).

3 Volatility estimators with the usage of low, high and closing prices

3.1 Review of variance estimators based on high, low and closing prices

The derived expressions for raw moments are applied in this section to discuss the properties of well-known estimators of variance of the arithmetic Brownian motion. According to our knowledge, there is no study in which a comprehensive comparison of variance estimators based on the same mathematical tool would have been analytically performed. In section 3.1, biases of selected popular estimators of the variance for a non-zero drift are formulated; furthermore, variances of those estimators are analytically derived. Such characteristics have not been published yet.

The most popular, unbiased estimator of the variance of the process is constructed based on only closing prices, but the value of the drift is required. It can be presented by the following formula:

$$\widehat{s^2}_0(\mathrm{t}) = (X_t - \mu t)^2 \tag{14}$$

and its variance is equal to

$$\mathrm{Var}\left[\widehat{s^2}_0(\mathrm{t})\right] = E\left[\left((X_t - \mu t)^2 - \sigma^2 t\right)^2\right] = E\left[(X_t - \mu t)^4 - \sigma^4 t^2\right] = 2\sigma^4 t^2 \tag{15}$$

The high variance of this estimator obviously results from the only fragmentary usage of the data set A_t, C_t, X_t.

The intermediate results for the derivation of formulae for the expected value, variance and the mean square error (MSE) of estimators presented in this section are given in Appendix D.

The Parkinson (1980) estimator can be described by the following formula:

$$\widehat{s^2}_1(\mathrm{t}) = \frac{(C_t - A_t)^2}{4\ln 2}. \tag{16}$$

Note that the values of X_t are not used in the valuation. Moreover, it is an unbiased estimator of the variance of the arithmetic Brownian motion only if $\mu = 0$. It follows from Equations C3 and C4 in Appendix C. In practice, the estimator is often applied when the value of the drift is unknown or when it is different from zero. A formula for the expected value of the Parkinson estimator for the process with a non-zero drift was derived:

$$\begin{aligned} E\left[\widehat{s^2}_1(t)\right] &= \left(1 + \frac{1}{4\ln 2}\sum_{m=1}^{\infty} \upsilon^{2m} \sum_{j=1}^{m} \frac{(-1)^{m+j}\left(2^{2j+1}-1\right)\zeta(2\mathrm{j}+1)}{2^{m+2j-1} m(\mathrm{m}+1)(\mathrm{m}-\mathrm{j})!(\mathrm{j}-1)!}\right)\sigma^2 t \\ &= \left[1 + \frac{7\zeta(3)}{32\ln 2}\upsilon^2 + \left(-\frac{7\zeta(3)}{192\ln 2} + \frac{31\zeta(5)}{768\ln 2}\right)\upsilon^4 \right. \\ &\quad + \left(\frac{7\zeta(3)}{1536\ln 2} - \frac{31\zeta(5)}{3072\ln 2} + \frac{127\zeta(7)}{24576\ln 2}\right)\upsilon^6 \\ &\quad \left. + \left(-\frac{7\zeta(3)}{15360\ln 2} + \frac{31\zeta(5)}{20480\ln 2} - \frac{127\zeta(7)}{81920\ln 2} + \frac{511\zeta(9)}{983040\ln 2}\right)\upsilon^8 + \cdots\right]\sigma^2 t, \end{aligned} \tag{17}$$

where $\upsilon := \frac{\mu\sqrt{t}}{\sigma}$, and $\zeta(n) := \sum_{k=1}^{\infty}\frac{1}{k^n}$ is the so-called Riemann Zeta function defined for $n \in \mathbb{Z}, n > 1$.

Figure 2 gives an overview on the size of the bias of the estimator depending on the F2
value of the v parameter.

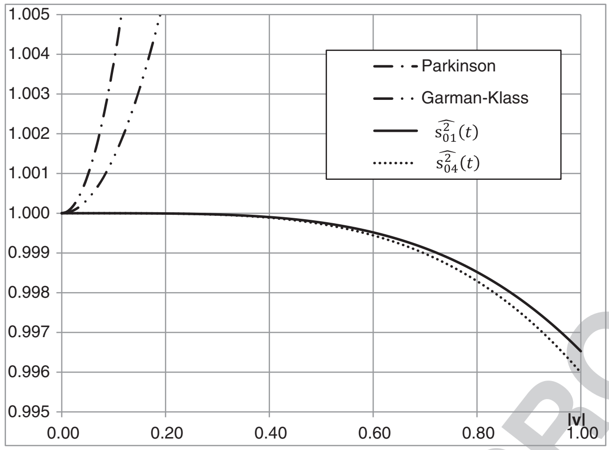

Fig. 2. The bias of the Parkinson, Garman–Klass, $\widehat{s^2}_2(t)$ and $\widehat{s^2}_3(t)$ estimators for a non-zero drift. The quotient of the estimator's expected value and the variance of the arithmetic Brownian motion as a function of $|v|$. Q12

Because the Parkinson estimator is biased for $\mu\neq 0$, it is better to present the MSE of this estimator, instead of its variance. It can be expressed as follows:

$$MSE\left[\widehat{s^2}_1(t)\right] = \left[-1 + \frac{9\zeta(3)}{16\ln^2 2} + \left(-\frac{3\zeta(3)}{32\ln^2 2} - \frac{7\zeta(3)}{16\ln 2} + \frac{93\zeta(5)}{128\ln^2 2}\right)v^2 + \left(\frac{3\zeta(3)}{256\ln^2 2} + \frac{7\zeta(3)}{96\ln 2} - \frac{93\zeta(5)}{512\ln^2 2} - \frac{31\zeta(5)}{384\ln 2} + \frac{1905\zeta(7)}{8192\ln^2 2}\right)v^4 + \left(-\frac{3\zeta(3)}{2560\ln^2 2} - \frac{7\zeta(3)}{768\ln 2} + \frac{279\zeta(5)}{10240\ln^2 2} + \frac{31\zeta(5)}{1536\ln 2} - \frac{1143\zeta(7)}{16384\ln^2 2} - \frac{127\zeta(7)}{12288\ln 2} + \frac{3577\zeta(9)}{81920\ln^2 2}\right)v^6 + \cdots\right]\sigma^4 t^2. \quad (18)$$

The efficiency of the estimator can be assessed for the different values of $|v|$ on the basis of Figure 3. **F3**

For $\mu = 0$, the MSE of the Parkinson estimator is equal to its variance and can be described as follows:

$$Var\left[\widehat{s^2}_1(t)\right] = \left(-1 + \frac{9\zeta(3)}{16\ln^2 2}\right)\sigma^4 t^2 \approx 0.407332\sigma^4 t^2. \quad (19)$$

This result does not differ from the outcome of Parkinson (1980, p. 63), but it diverges from the estimate of Garman and Klass (1980), p. 71), according to which the estimator $\widehat{s^2}_1$(t) is 5.2 times more effective than $\widehat{s^2}_0$ (t). In none of the discussed papers, exact formulae for the derivation of those values have been given.

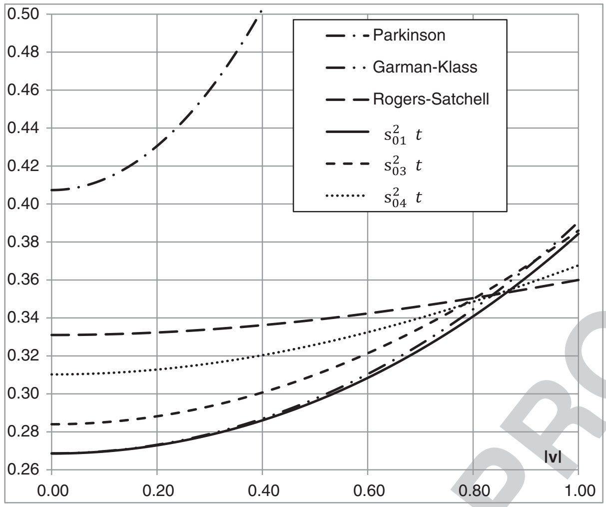

Fig. 3. The efficiency of the analysed estimators. The quotient of the estimator's mean square error and the variance of the arithmetic Brownian motion as a function of $|v|$.

GARMAN and KLASS (1980) considered also only the case when $\mu = 0$. They constructed the estimator of the variance of the arithmetic Brownian motion given by the expression (hereinafter referred to as the G-K estimator):

$$\widehat{s^2}_2(\mathrm{t}) = \frac{(C_t - A_t)^2}{2} - (2\ln 2 - 1)X_t{}^2. \tag{20}$$

Under the adopted assumptions, its unbiasedness can be proved by applying Equations C3 and C4 in Appendix C. The expected value of the G-K estimator for $v \neq 0$ ($\mu \neq 0$) can be constructed based on the following equation:

$$\begin{aligned} E\left[\widehat{s^2}_2(t)\right] &= \left[1 + (1 - 2\ln 2)v^2 + \sum_{n=1}^{\infty} v^{2n} \sum_{j=1}^{n} (-1)^{n+j} \frac{\left(2^{2j+1} - 1\right)\zeta(2j+1)}{2^{n+2j} n(n+1)(n-j)!(j-1)!}\right]\sigma^2 t \\ &= \left[1 + \left(1 - 2\ln 2 + \frac{7}{16}\zeta(3)\right)v^2 + \left(-\frac{7}{96}\zeta(3) + \frac{31}{384}\zeta(5)\right)v^4 \right. \\ &\quad + \left(\frac{7\zeta(3)}{768} - \frac{31\zeta(5)}{1536} + \frac{127\zeta(7)}{12288}\right)v^6 \\ &\quad \left. + \left(-\frac{7\zeta(3)}{7680} + \frac{31\zeta(5)}{10240} - \frac{127\zeta(7)}{40960} + \frac{511\zeta(9)}{491520}\right)v^8 + \cdots\right]\sigma^2 t. \end{aligned} \tag{21}$$

The size of the bias of the estimator depending on the value of the v parameter can be assessed on the basis of Figure 2.

The MSE of the G-K estimator is equal to:

$$
\begin{aligned}
MSE\left[\widehat{s^2}_2(t)\right] &= E\left[\left(\frac{(C_t-A_t)^2}{2}-(2\ln 2-1)X_t{}^2-\sigma^2 t\right)^2\right] \\
&= \left[2-8\ln 2+4\ln^2 2+\left(4-\frac{7}{2}\ln 2\right)\zeta(3)\right. \\
&\quad +\left(4(1-2\ln 2)^2+\frac{11\zeta(3)}{8}-\frac{21\zeta(3)\ln 2}{4}+\frac{155\zeta(5)}{32}-\frac{31\zeta(5)\ln 2}{8}\right)\upsilon^2 \\
&\quad +\left((1-2\ln 2)^2+\frac{29\zeta(3)}{96}-\frac{7\zeta(3)\ln 2}{32}-\frac{31\zeta(5)}{48}-\frac{31\zeta(5)\ln 2}{64}\right. \\
&\quad \left.\left.+\frac{3175\zeta(7)}{2048}-\frac{635\zeta(7)\ln 2}{512}\right)\upsilon^4+\cdots\right]\sigma^4 t^2.
\end{aligned}
\tag{22}
$$

The efficiency of the estimator can be assessed for the different values of $|\upsilon|$ on the basis of Figure 3. For $\mu=0$, the MSE of the G-K estimator is equal to the variance and can be presented as follows:

$$
Var\left[\widehat{s^2}_2(t)\right] = \left(2-8\ln 2+4\ln^2 2+\left(4-\frac{7}{2}\ln 2\right)\zeta(3)\right)\sigma^4 t^2 \approx 0.268654\sigma^4 t^2.
\tag{23}
$$

The obtained result is in line with the estimate of Garman and Klass (1980, p. 74), according to which the estimator $\widehat{s^2}_2(\mathrm{t})$ is 7.4 times more effective than $\widehat{s^2}_0(\mathrm{t})$, and with the outcome of Rogers and Satchel (1991, p. 505), according to which the variance of the estimator $\widehat{s^2}_2(\mathrm{t})$ is equal to $0.27\sigma^4 t^2$.

Rogers and Satchell (1991) conducted research for the process with a non-zero drift. They proposed the following variance estimator of X_t (further referred to as the R-S estimator):

$$
\widehat{s^2}_3(\mathrm{t}) = C_t(C_t-X_t)+A_t(A_t-X_t),
\tag{24}
$$

which is unbiased for all values of $\mu \neq 0$ (it follows from equations (B16, B17, B19, B20) in Appendix B), and the value of this parameter is not necessary for variance estimation. The variance of this estimator can be expressed as follows:

$$
\begin{aligned}
Var\left[\widehat{s^2}_3(t)\right] &= E\left[\left(C_t(C_t - X_t) + A_t(A_t - X_t) - \sigma^2 t\right)^2\right] \\
&= \left[1 - 4\ln 2 + \frac{7\zeta(3)}{4} + \left(-\frac{4\ln 2}{3} - \frac{7\zeta(3)}{8} + \frac{31\zeta(5)}{16}\right)\upsilon^2 \right. \\
&\quad + \left(-\frac{7\zeta(3)}{192} - \frac{217\zeta(5)}{384} + \frac{635\zeta(7)}{1024}\right)\upsilon^4 \\
&\quad + \left(\frac{7\zeta(3)}{1920} + \frac{93\zeta(5)}{1280} - \frac{5969\zeta(7)}{30720} + \frac{3577\zeta(9)}{30720}\right)\upsilon^6 \\
&\quad \left. + \left(-\frac{7\zeta(3)}{23040} - \frac{341\zeta(5)}{46080} + \frac{2921\zeta(7)}{92160} - \frac{9709\zeta(9)}{245760} + \frac{2047\zeta(11)}{131072}\right)\upsilon^8 + \cdots\right]\sigma^4 t^2.
\end{aligned} \tag{25}
$$

The efficiency of the estimator can be evaluated for different values of $|\upsilon|$ on the basis of Figure 3. For $\mu = 0$, the variance of the estimator is equal to

$$Var\left[\widehat{s^2}_3(t)\right] = \left(1 - 4\ln 2 + \frac{7}{4}\zeta(3)\right)\sigma^4 t^2 \approx 0.331011\sigma^4 t^2. \tag{26}$$

The value of $Var\left[\widehat{s^2}_3(\mathrm{t})\right]$ is in line with the results of ROGERS and SATCHELL (1991, p. 505). Also in this case, expressions for the derivation of this value were not given.

The estimators constructed on the basis of only the single vector of random variables (A_t, C_t, X_t) are considered in this paper. However, in financial studies, a number of other volatility estimators, which are built on the basis of a wider set of information, are also used. One of them is the estimator proposed by KUNITOMO (1992), which is significantly more effective than the earlier-mentioned estimators. The so-called corrected low and high values are used in the construction of this estimator; however, the quotations of intraday returns are necessary. Therefore, the estimator was not considered in this paper.

As it has been already mentioned, the presented estimators are the most often applied for the calculation of the variance for 1 day of quotations. Their properties are derived with the assumption that the log returns are realizations of the arithmetic Brownian motion with given values of the drift μ and the variance σ^2. Sometimes, however, the estimators are used to calculate the variance for the period longer than 1 day. Then, it is assumed that the variance of returns is constant in the considered period and its estimate for 1 day is the arithmetic mean of daily variances calculated for consecutive days belonging to the adopted period. In such a case, a significant strengthening of assumptions of the adopted model takes place. It should be noted, however, that such strengthened assumption is not valid in the light of the results of empirical studies for most financial series (see, e.g. the papers of CORSI, MITTNIK, PIGORSCH and PIGORSCH (2008) and TERÄSVIRTA (2009)). For this reason, this method of use of the presented estimators is not of interest in this study.

The constancy of the variance for the period longer than 1 day is also assumed in the formulation of the estimator proposed by YANG and ZHANG (2000). It is constructed on the basis of the sequence of random vectors (A_t, C_t, X_t). Additionally the so-called night returns, that is, returns between the closing price the day before and the next day's opening price, are taken (Figure 1). In view of a high efficiency and the common availability of data, which are used in its construction, it is one of the most commonly used volatility estimators based on open, high, low and closing prices. However, it is not possible to use this estimator to calculate the variance for only 1 day. For this reason and because of the very strong assumptions, the estimator has also not been widely discussed in the paper.

3.2 Modifications of the existing volatility estimators based on the low and high prices

The variance estimators considered so far are constructed for different assumptions. The estimator $\widehat{s^2}_0(t)$ based on the closing prices requires the value of the drift μ, and the Parkinson estimator $\widehat{s^2}_1(t)$ and the G-K estimator $\widehat{s^2}_2(t)$ were constructed for the assumption that the value of the drift is not only known but is also equal to zero. Only for the R-S estimator $\widehat{s^2}_3(t)$, the information of the drift is not used. The attempt to find more efficient variance estimators for different assumptions is undertaken in this section of the paper.

Let us consider a very general formulation of the variance estimator:

$$\widehat{s^2}\ (a,b,c,d,u;t)a\left(A_t^{\,2}+C_t^{\,2}\right)+bA_tC_t+cX_t^{\,2}+d(A_tX_t+C_tX_t)+u\mu^2t^2, \quad (27)$$

where $a,b,c,d,u\in\mathbb{R}$. Applying the formulae presented in Appendix B, one can show that the expected value and the MSE of this estimator can always be written as the sequences:

$$E\left[\widehat{s^2}\ (a,b,c,d,u;t)\right]=\left(r_0+r_1v^2+r_2v^4+r_3v^6+\cdots\right)\sigma^2t, \quad (28)$$

$$MSE\left[\widehat{s^2}\ (a,b,c,d,u;t)\right]=\left(s_0+s_1v^2+s_2v^4+s_3v^6+\cdots\right)\sigma^4t^2, \quad (29)$$

where $r_0, r_1, r_2, \cdots\in\mathbb{R}, s_0, s_1, s_2, \cdots\in\mathbb{R}$.

Because the form of the minimum variance estimator depends on the value of the estimated parameter, it is not possible to construct the most efficient estimator. However, in practice, for daily financial time series, the value of the drift is relatively low, that is, $|\mu t|\ll\sigma\sqrt{t}\Rightarrow|v|\ll 1$, and such assumption permits to formulate some propositions. Let us assume at the beginning that the value of the drift μ is known.

Scenario 1

Find the parameters a, b, c, d, u which minimize s_0 in the formula (29), subject to $r_0=1,\ r_1=0$.

The solution is (due to the complicated form, only decimal approximations are given) $a \approx 0.510995$, $b \approx -0.984239$, $c \approx -0.383321$, $d \approx -0.018875$, and $u \approx -0.134291$, and the value of s_0 is approximately equal to 0.268581. This estimator can be written as follows:

$$\widehat{s_{01}^2}(t) := \widehat{s^2}(0.510995, -0.984239, -0.383321, -0.018875, -0.134291; t) \quad (30)$$

For $\mu = 0$ (i.e. $\upsilon = 0$), it is, to the accuracy of the applied approximation, an unbiased estimator of the variance of the arithmetic Brownian motion. The MSE (in this case also the variance) of this estimator is approximately equal to $0.268581\sigma^4 t^2$, and it is the lowest among all of the estimators described by the expression (B11); in particular, it is lower than the variance of the PARKINSON and G-K estimators. However, the estimator $\widehat{s_{01}^2}(t)$ is only slightly more effective than the G-K estimator (see formula B3). Moreover, when the value of the drift is known and $\mu \neq 0$, then the estimator $\widehat{s_{01}^2}(t)$ is biased. Notwithstanding, the bias of the estimator does not exceed 0.4% of the value of the variance of the arithmetic Brownian motion ($r_1 = 0$, $r_2 \approx -0.003878$ and $r_3 \approx 0.000443$ and the subsequent values of r_i in the sequence (28) are very low).

Because the increase of the efficiency of the estimator $\widehat{s_{01}^2}(t)$ in comparison to the G-K estimator is very low, the latter estimator served as a basis for the construction of a more parsimonious formulation. Assuming, in addition, that in scenario 1 $a = \frac{1}{2}$, $b = -1$, $c = -(2\ln 2 - 1) \approx -0.386294$, $d = 0$ and based on the equation (21), the following value of the parameter u was obtained $u = -\left(1 - 2\ln 2 + \frac{7\zeta(3)}{16}\right) \approx -0.139606$, for which $s_0 = 2 - 8\ln 2 + 4\ln^2 2 + \left(4 - \frac{7}{2}\ln 2\right)\zeta(3) \approx 0.268654$, $r_2 \approx -0.003940$ and $r_3 \approx 0.000450$. This estimator can be expressed as follows:

$$\widehat{s_{02}^2}(t) := \frac{(C_t - A_t)^2}{2} - (2\ln 2 - 1)X_t^2 - \left(1 - 2\ln 2 + \frac{7\zeta(3)}{16}\right)\mu^2 t^2. \quad (31)$$

The bias of the estimator $\widehat{s_{02}^2}(t)$ of the variance of the arithmetic Brownian motion, for the known value of the drift satisfying the condition $|\mu t| \ll \sigma\sqrt{t}$, is not significant in practical applications. Whereas the MSE of the estimator is only slightly higher than the MSE of the estimator $\widehat{s_{01}^2}(t)$, and for $\mu = 0$, it is equal to $0.268654\sigma^4 t^2$

Scenario 2

Find the parameters a, b, c, d, u that minimize s_0 in the formula (29), subject to $r_0 = 1$ and $r_1 = 0$, for i>0.

In comparison to scenario 1, an explicit requirement of the unbiasedness of the estimator $\widehat{s^2}(a, b, c, d, u; t)$ is included now. The condition $r_i = 0$ for i > 0 means that

$b = 0$, and the value of the parameter u is responsible for the condition $r_1 = 0$ The solution is the following set of parameters: $a = -d = \frac{8}{12-16\ln 2+7\zeta(3)} \approx 0.857997$ and $c = u = 1 - a \approx 0.142003$. This estimator can be formulated as follows:

$$\widehat{s_{03}^2}(t) := \frac{8}{12 - 16\ln 2 + 7\zeta(3)}\left(C_t(C_t - X_t) + A_t(A_t - X_t)\right) + \frac{4 - 16\ln 2 + 7\zeta(3)}{12 - 16\ln 2 + 7\zeta(3)}\left(X_t^2 - \mu^2 t^2\right). \tag{32}$$

It is an unbiased estimator of the variance of the arithmetic Brownian motion for the known μ and for $\mu = 0$, its MSE is the lowest among all the unbiased estimators described by the expression (27). For $\mu = 0$, the MSE of the estimator $\widehat{s_{03}^2}(t)$ is equal to $2 - 2a \approx 0.284006\sigma^4 t^2$

Let us assume now that the value of the drift μ is not known.

Scenario 3

Find the parameters a, b, c, d that minimize s_0 in the formula (29), subject to $u = 0$, $r_0 = 1$ and $r_1 = 0$.

The solution is (due to the complicated form, only decimal approximations are given) $a \approx 0.590262$, $b \approx -1.136916$, $c \approx -0.597904$ and $d \approx -0.021803$, for which $s_0 \approx 0.310244$. This estimator can be given as follows:

$$\widehat{s_{04}^2}(t) := \widehat{s^2}(0.590262, -1.136916, -0.597904, -0.021803, 0; t). \tag{33}$$

The estimator $\widehat{s_{04}^2}(t)$ is biased; however, its bias does not exceed 0.5% of the value of the variance of the arithmetic Brownian motion ($r_1 = 0$, $r_2 \approx -0.004564$ and $r_3 \approx 0.000522$ and the subsequent values of r_i in the sequence (28) are very low) and is negligible from a practical point of view. More importantly, the estimator can be used when the value of the drift is unknown. For $\mu = 0$, the MSE of the estimator $\widehat{s_{04}^2}(t)$ is equal to 0.310244.

The additional restriction $d = 0$ imposed in scenario 3 simplifies the form of the estimator, without causing a significant increase in the value of s_0. The obtained estimator has the following formula:

$$\widehat{s_{05}^2}(t) = \widehat{s^2}(0.582491, -1.158478, -0.612495, 0, 0, t). \tag{34}$$

For these parameters, s_0 is approximately equal to 0.310253, and the values of r_2 and r_3 are the same as for the estimator $\widehat{s_{04}^2}(t)$. The MSE of the estimator is only slightly higher than the MSE of the estimator $\widehat{s_{04}^2}(t)$, and for $\mu = 0$, it is equal to 0.310253.

Scenario 4

Find the parameters a, b, c, d that minimize s_0 in the formula (29), subject to $u = 0$, $r_0 = 1$ and $r_1 = 0$ for $i > 0$ It failed to discover a more effective estimator than the R-S one for the defined problem.

The size of bias of the proposed estimators $\widehat{s^2_{01}}\,(t)$ and $\widehat{s^2_{04}}\,(t)$ depending on the value of the parameter υ can be assessed on the basis of Figure 2. The biases of the estimators $\widehat{s^2_{02}}\,(t)$ and $\widehat{s^2_{05}}\,(t)$ are almost the same as the ones of the estimators $\widehat{s^2_{01}}\,(t)$ and $\widehat{s^2_{04}}\,(t)$, respectively (the estimators would be indistinguishable in Figure 2; that is why, they are omitted). Moreover the quotients of the estimator's MSE and the square of the variance of the arithmetic Brownian motion for the estimators $\widehat{s^2_{01}}\,(t)$, $\widehat{s^2_{03}}\,(t)$ and $\widehat{s^2_{04}}\,(t)$. are presented in Figure 3. The efficiency of the estimators $\widehat{s^2_{02}}\,(t)$ and $\widehat{s^2_{05}}\,(t)$ is almost the same as the one of the estimators $\widehat{s^2_{01}}\,(t)$ and $\widehat{s^2_{04}}\,(t)$, respectively (the estimators would be indistinguishable in Figure 3; that is why, they are omitted).

3.3 The estimation of the volatility of the Polish stock index

In order to illustrate the usefulness of the considered estimators, they are applied to the estimation of the variance of the Polish stock index WIG20. The index is chosen deliberately, due to the fact that it includes the 20 largest and most liquid companies listed on the Warsaw Stock Exchange. This is important, because intraday data are used in the study and the quality of such data depends to a large extent on the market liquidity. It is worth noting the relatively small number of studies on Polish financial time series in comparison with other emerging or developed markets.

The following estimators of variance are applied: $\widehat{s^2_0}\,(t)$ constructed on the basis of closing prices, Parkinson $\widehat{s^2_{01}}\,(t)$, G-K $\widehat{s^2_{02}}\,(t)$, R-S $\widehat{s^2_{03}}\,(t)$ and the ones proposed in the paper , $\widehat{s^2_{01}}\,(t)$, $\widehat{s^2_{03}}\,(t)$ and $\widehat{s^2_{04}}\,(t)$. They are described by the expressions 14, 16, 20, 24, 30, 32 and 33, respectively. The analysis is performed for the 10-year period from 30 September 2002 to 28 September 2012, that is, for the period that covers both the bull and bear markets and, what is important, the financial crises. A total of 2513 vectors of logarithmic returns multiplied by 100: (a_1, c_1, x_1), (a_2, c_2, x_2), … , $(a_{2513}, c_{2513}, x_{2513})$ are calculated, and the drift is estimated by the following formula: $\hat{\mu} = \frac{1}{2513}\sum_{i=1}^{i=2513} x_i \approx 0.0326135$. The constancy of the drift in time is assumed. The simplest form of the estimator of the drift is adopted. If a more effective estimator of the drift is applied, then one would expect that estimates of the volatility based on the proposed estimators would not be less accurate from those presented in the paper. The aforementioned seven estimators are evaluated for each day in the sample, that is, for $1 \le i \le 2513$.

For comparison, estimates of variances are also calculated on the basis of univariate GARCH model (Bollerslev, 1986). Due to the presence of a negative correlation between lagged returns and the conditional variance, the following models are also applied: EGARCH model (Nelson, 1991), GJR (Glosten, Q8

JAGANNATHAN, RUNKLE, 1993), TGARCH (RABEMANANJARA, ZAKOIAN, 1993) and AGARCH (ENGLE, NG, 1993). Models with the Student-t conditional innovation distribution are used to better describe the fat tails of the distribution of the WIG20 returns. There was no statistically significant autocorrelation; that is why, the following specification of the conditional mean is adopted: $x_{\mathrm{i}} = \gamma_0 + \epsilon_i$. Parameters of these models are estimated using a maximum likelihood method.

As a measure of ex-post realized variance, for the evaluation of estimates accuracy, the sum of squared intraday returns is applied. A significant problem with the use of such data is the choice of the appropriate frequency of observations (see, e.g. PIGORSCH *et al.* (2012)). For that reason, the realized volatility was estimated in various variants by using returns in 5, 10, 15, 20, 30 and 60 min. The evaluation of estimates accuracy of the considered methods was performed on the basis of the following measures (a detailed description of the methods can be found, e.g. in POON and GRANGER (2003)): the mean error (ME), Q9 the root mean squared error, the heteroskedasticity adjusted root mean squared error, the LINEX loss function (for $a = -0.01$ and $a = 0.01$) and the coefficient of determination (R^2) Q10 for the regression of ex-post realized variances on estimates of variances.

The results of the performed study for the case when ex-post realized variances are estimated as the sum of squared 10-min returns are presented in Table 1. The rankings **T1** of the considered methods are very similar also when ex-post realized variances are estimated on the basis of returns in 5, 15, 20, 30 and 60 min.

The outcomes of the ME indicate that estimates of variances constructed on the basis of low and high prices are significantly undervalued. However, it does not result from the poor quality of those estimators but from the adopted measure of ex-post realized variance; in this case, the sum of squared intraday returns (see, e.g. BARNDORFF-NIELSEN and SHEPHARD (2004) and ANDERSEN, DOBREV and SCHAUMBURG (2012)). Moreover, similarly as for daily data, the squared intraday return is an inefficient estimator of variance of the intraday return.

The results of other measures presented in Table 1 indicate the visible superiority of the variance estimators based on low and high prices over the estimators formulated on the basis of the GARCH models. The asymmetric GARCH models perform better than the simple GARCH model, but the advantage is small in comparison to the one resulting from the application of low and high prices.

It seems that the relatively weak performance of the estimators R-S and $\widehat{s_{04}^2}(t)$, which can be used when the value of the drift is unknown, may be due to the excessive existence, in comparison to the arithmetic Brownian motion, of the phenomenon: both C_t and $A_t - X_t$ or both (C_t, X_t) and A_t are close to zero. As a consequence, estimates of the variance based on the R-S estimator are lower than the ones based on the G-K and particularly the Parkinson. This may be related to the unstable, during consecutive days, drift or the existence of autocorrelation in intraday returns; however, it needs further comprehensive studies.

All the measures presented in Table 1 (with the exception of the ME) favour the estimator $\widehat{s_{03}^2}(t)$. Obviously, the superiority of this estimator is not always meaningful, and in many cases, it is probably not statistically significant.

Table 1. The evaluation of the accuracy of variance estimates for the WIG20 index – realized volatility estimated on the basis of squared 10-min returns

Estimators	ME	RMSE	HRMSE	LINEX ($a = 0.01$)	LINEX ($a = -0.01$)	R^2
$\widehat{s_0^2}(t)$	0.0811	4.8953	1.2882	1.19E-03	1.23E-03	0.2472
$\widehat{s_1^2}(t)$ – Parkinson	0.5376	2.5397	0.5202	3.34E-04	3.14E-04	0.6437
$\widehat{s_2^2}(t)$ – G-K	0.7144	2.5562	0.4704	3.36E-04	3.21E-04	0.6532
$\widehat{s_3^2}(t)$ – R-S	0.6946	2.7346	0.5502	3.85E-04	3.66E-04	0.6230
$\widehat{s_{01}^2}(t)$	0.7161	2.5609	0.4720	3.37E-04	3.22E-04	0.6523
$\widehat{s_{03}^2}(t)$	0.6074	2.4183	0.4586	3.02E-04	2.85E-04	0.6853
$\widehat{s_{04}^2}(t)$	0.7990	2.8458	0.5541	4.15E-04	3.99E-04	0.5937
GARCH	0.0596	3.3217	1.0577	6.05E-04	5.09E-04	0.3718
EGARCH	0.0843	3.2492	1.0550	5.79E-04	4.86E-04	0.4152
GJR	0.0538	3.1991	1.0424	5.59E-04	4.73E-04	0.4210
TGARCH	0.0735	3.2476	1.0533	5.78E-04	4.86E-04	0.4070
AGARCH	0.0738	3.2634	1.0467	5.83E-04	4.91E-04	0.4015

Note: ME, mean error; RMSE, root mean squared error; HRMSE, heteroskedasticity adjusted root mean squared error.

4 Conclusions

The joint probability density function of random variables low, high and closing returns of the arithmetic Brownian motion is presented in the paper. The application of this density allowed to analytically evaluate the main properties of the most popular estimators of the variance constructed on the basis of low, high and closing prices of the arithmetic Brownian motion. In particular, the unbiasedness of the Parkinson and Garman–Klass estimators for the process with a zero drift and of the Rogers–Satchell estimator for any drift is proved. Thus, the main results from the papers of the mentioned authors are confirmed.

Moreover, the expected values of the Parkinson and Garman–Klass estimators for the arithmetic Brownian motion with a non-zero drift are derived. The MSEs of the Parkinson, Garman–Klass and Rogers–Satchell estimators for the process with a non-zero drift are also formulated. According to our knowledge, those characteristics have not been published yet.

Furthermore, new volatility estimators, more efficient in the majority of financial applications than the Rogers–Satchell estimator, are proposed. The considered estimators are applied to the estimation of the volatility of the Polish stock index WIG20. It is shown that volatility estimates based on low, high and closing prices are more accurate than the ones formulated on the basis of the GARCH models. The estimators based on low, high and closing prices can be applied in the future to the construction of the GARCH models, so that it will be possible to obtain even more accurate volatility estimates.

Acknowledgements Q11

The paper is financed by the National Science Centre project number 2012/05/B/HS4/00675 entitled ‘Modelling and forecasting volatility – usage of additional information contained in low and high prices’. The authors would like to thank an anonymous reviewer for constructive comments and useful recommendations.

Appendix A

Characteristic functions of the joint density of the random vectors of low, high and closing returns of the arithmetic Brownian motion

Here, we describe in detail the characteristic functions, which are used in section 3 to evaluate properties of well-known and also new proposed estimators of variance. The characteristic function of the joint density of the random vector (C_t, X_t) is given by the following formula:

$$\begin{aligned}\varphi_{C_t,X_t}(q,s;\mu,\sigma^2,t) &= \int_0^{\infty}\int_{-\infty}^{c} e^{iqc+isx} f_{C_t,X_t}(c,x;\mu,\sigma^2,t)\,dx\,dc \\ &= \frac{((q+s)\sigma^2 - i\mu)e^{\frac{1}{2}(q+s)\left(-(q+s)\sigma^2+2i\mu\right)t}\operatorname{erfc}\left(\frac{\left(-\mu-i(q+s)\sigma^2\right)\sqrt{t}}{\sqrt{2}\sigma}\right)}{(q+2s)\sigma^2-2i\mu} \\ &\quad + \frac{(s\sigma^2 - i\mu)e^{\frac{1}{2}s\left(-s\sigma^2+2i\mu\right)t}\operatorname{erfc}\left(\frac{\left(\mu+is\sigma^2\right)\sqrt{t}}{\sqrt{2}\sigma}\right)}{(q+2s)\sigma^2-2i\mu},\end{aligned} \tag{A1}$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ and $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$.

The characteristic function of the joint density of the random vector (A_t,X_t) is described as follows:

$$\begin{aligned}\varphi_{A_t,X_t}(p,s;\mu,\sigma^2,t) &= \int_{-\infty}^{0}\int_{a}^{\infty} e^{ipa+isx} f_{A_t,X_t}(a,x;\mu,\sigma^2,t)\,dxda\\ &= \frac{((p+s)\sigma^2-i\mu)e^{\frac{1}{2}(p+s)\left(-(p+s)\sigma^2+2i\mu\right)t}\operatorname{erfc}\left(\frac{\left(\mu+i(p+s)\sigma^2\right)\sqrt{t}}{\sqrt{2}\sigma}\right)}{(p+2s)\sigma^2-2i\mu}\\ &\quad+\frac{(s\sigma^2-i\mu)e^{\frac{1}{2}s\left(-s\sigma^2+2i\mu\right)t}\operatorname{erfc}\left(\frac{\left(-\mu-is\sigma^2\right)\sqrt{t}}{\sqrt{2}\sigma}\right)}{(p+2s)\sigma^2-2i\mu}.\end{aligned} \tag{A2}$$

The characteristic function of the joint density of the random vector (A_t, C_t, X_t) is given by the following formula:

$$\begin{aligned}\varphi_{A_t,C_t,X_t}(p,q,s;\mu,\sigma^2,t) &= \int_{-\infty}^{0}\int_{0}^{\infty}\int_{a}^{c} e^{ipa+iqc+isx} f_{A_t,C_t,X_t}(a,\ \ c,\ \ x;\mu,\sigma^2,t)\,dxdcda\\ &= \sum_{k=-\infty}^{-2}(\psi(2k,2k,\ \ p,\ \ q,\ \ s;\mu,\sigma t)\\ &\quad-\psi(2k+2,2k,\ \ p,\ \ q,\ \ s;\mu,\sigma,t))\\ &\quad+\psi(-2,-2,\ \ p,\ \ q,\ \ s;\mu,\sigma,t)\\ &\quad+\sum_{k=1}^{\infty}(\psi(2k,2k,\ \ p,\ \ q,\ \ s;\mu,\sigma,t)\\ &\quad-\psi(2k+2,2k,\ \ p,\ \ q,\ \ s;\mu,\sigma,t)),\end{aligned} \tag{A3}$$

where

$$\begin{aligned}\psi(m,n,\ \ p,\ \ q,\ \ s;\mu,\sigma,t) &= \frac{mn(p\sigma^2+s\sigma^2-i\mu)^2\left(\operatorname{sgn}(m-1)+\operatorname{erf}\left(\frac{\left(ip\sigma^2+is\sigma^2+\mu\right)\sqrt{t}}{\sqrt{2}(m-1)\sigma}\right)\right)}{2(m-1)(p\sigma^2+ms\sigma^2-im\mu)((m-1)q\sigma^2+n(p\sigma^2+s\sigma^2-i\mu))}\\ &\quad\times\exp\left(-\frac{(p\sigma^2+s\sigma^2-i\mu)^2t}{2(m-1)^2\sigma^2}-\frac{\mu^2t}{2\sigma^2}\right)\\ &\quad+\frac{mn(q\sigma^2+s\sigma^2-i\mu)^2\left(\operatorname{sgn}(n+1)-\operatorname{erf}\left(\frac{\left(iq\sigma^2+is\sigma^2+\mu\right)\sqrt{t}}{\sqrt{2}(n+1)\sigma}\right)\right)}{2(n+1)(-q\sigma^2+ns\sigma^2-in\mu)((n+1)p\sigma^2+m(q\sigma^2+s\sigma^2-i\mu))}\\ &\quad\times\quad\exp\left(-\frac{(q\sigma^2+s\sigma^2-i\mu)^2t}{2(n+1)^2\sigma^2}-\frac{\mu^2t}{2\sigma^2}\right)\\ &\quad+\frac{np^2\sigma^4\left(-\operatorname{sgn}(m)-\operatorname{erf}\left(\frac{ip\sigma\sqrt{t}}{\sqrt{2}m}\right)\right)\exp\left(-\frac{p^2\sigma^2t}{2m^2}-\frac{\mu^2t}{2\sigma^2}\right)}{2(p\sigma^2+ms\sigma^2-im\mu)((n+1)p\sigma^2+m(q\sigma^2+s\sigma^2-i\mu))}\\ &\quad+\frac{mq^2\sigma^4\left(-\operatorname{sgn}(n)+\operatorname{erf}\left(\frac{iq\sigma\sqrt{t}}{\sqrt{2}n}\right)\right)\exp\left(-\frac{q^2\sigma^2t}{2n^2}-\frac{\mu^2t}{2\sigma^2}\right)}{2(-q\sigma^2+ns\sigma^2-in\mu)((m-1)q\sigma^2+n(p\sigma^2+s\sigma^2-i\mu))},\end{aligned} \tag{A4}$$

where $\operatorname{sgn}(x)$ is the sign function.

Appendix B

Raw moments of random variables low, high and closing returns of the arithmetic Brownian motion with a non-zero drift

Relatively complicated formulae arise in the calculation of specified moments. That is why it is reasonable to introduce first additional symbols and explanatory definitions, which are presented in expressions B1–B11:

The following symbols and definitions are adopted:

$$v := \frac{\mu\sqrt{t}}{\sigma} \tag{B1}$$

$$Z(n, y, \ v) := \sum_{k=1}^{\infty} \frac{1}{(k+y)^n} e^{-\frac{2k(k+1)v^2}{(2k+1)^2}}, \tag{B2}$$

for $n \in \mathbb{Z}, \ \ n > 0, y > 0$.

$$\zeta(n) := Z(n, 0, 0) = \sum_{k=1}^{\infty} \frac{1}{k^n}, \tag{B3}$$

for $n \in \mathbb{Z}, \mathrm{n} > 1 \zeta(n)$ is the so-called Riemann Zeta function,

The explanatory definitions are as follows:

$$\begin{aligned} Z_1(v) :&= Z(1, 0, v) - 2Z\left(1, \frac{1}{2}, v\right) + Z(1, 1, v) \\ &= 3 - 4\ln 2 + 2v^2 + \sum_{m=1}^{\infty} v^{2m} \sum_{j=1}^{m} \frac{(-1)^{m+j+1}\left(2^{2j+1} - 1\right)\zeta(2j+1)}{2^{m+2j-1} m(m-j)!(j-1)!} \\ &= 3 - 4\ln 2 + \left(2 - \frac{7\zeta(3)}{4}\right)v^2 + \left(\frac{7\zeta(3)}{16} - \frac{31\zeta(5)}{64}\right)v^4 \\ &\quad + \left(-\frac{7\zeta(3)}{96} + \frac{31\zeta(5)}{192} - \frac{127\zeta(7)}{1536}\right)v^6 \\ &\quad + \left(\frac{7\zeta(3)}{768} - \frac{31\zeta(5)}{1024} + \frac{127\zeta(7)}{4096} - \frac{511\zeta(9)}{49152}\right)v^8 + \cdots, \end{aligned} \tag{B4}$$

$$\begin{aligned} Z_2(v) :&= 1 - Z(2, 0, v) + Z(2, 1, v) \\ &= 2(3 - 4\ln 2)v^2 + 2v^4 + \sum_{m=2}^{\infty} v^{2m} \sum_{j=1}^{m-1} \frac{(-1)^{m+j}\left(2^{2j+1} - 1\right)\zeta(2j+1)}{2^{m+2j-3} m(m-1)(m-j-1)!(j-1)!} \\ &= 2(3 - 4\ln 2)v^2 + \left(2 - \frac{7\zeta(3)}{4}\right)v^4 + \left(\frac{7\zeta(3)}{24} - \frac{31\zeta(5)}{96}\right)v^6 \\ &\quad + \left(-\frac{7\zeta(3)}{192} + \frac{31\zeta(5)}{384} - \frac{127\zeta(7)}{3072}\right)v^8 \\ &\quad + \left(\frac{7\zeta(3)}{1920} - \frac{31\zeta(5)}{2560} + \frac{127\zeta(7)}{10240} - \frac{511\zeta(9)}{122880}\right)v^{10} + \cdots, \end{aligned} \tag{B5}$$

$$
\begin{aligned}
\mathrm{Z}_3(v) :&= \mathrm{Z}(3,0,v) + \mathrm{Z}(3,1,v) \\
&= 2\zeta(3) - 1 + 8(2 - 3\ln 2)v^2 + \left(12 - 8\ln 2 - \frac{21}{4}\zeta(3)\right)v^4 + \frac{4}{3}v^6 \\
&\quad + \sum_{m=3}^{\infty} v^{2m} \sum_{j=1}^{m-1} \frac{(-1)^{m+j}(4j-m-2)\left(2^{2j+1}-1\right)\zeta(2j+1)}{2^{m+2j-3}m(m-1)(m-2)(m-j-1)!(j-1)!} \\
&= 2\zeta(3) - 1 + 8(2 - 3\ln 2)v^2 + \left(12 - 8\ln 2 - \frac{21}{4}\zeta(3)\right)v^4 + \frac{4}{3}v^6 \\
&\quad + \left(\frac{4}{3} - \frac{7\zeta(3)}{24} - \frac{31\zeta(5)}{32}\right)v^6 + \left(\frac{7\zeta(3)}{192} + \frac{31\zeta(5)}{384} - \frac{127\zeta(7)}{1024}\right)v^8 \\
&\quad + \left(-\frac{7\zeta(3)}{1920} - \frac{31\zeta(5)}{7680} + \frac{127\zeta(7)}{6144} - \frac{511\zeta(9)}{40960}\right)v^{10} + \cdots,
\end{aligned}
\tag{B6}
$$

$$
\begin{aligned}
\mathrm{Z}\left(3, \frac{1}{2}, v\right) &= -8 + \sum_{m=0}^{\infty} v^{2m} \sum_{j=1}^{m+1} \frac{(-1)^{m+j+1}\left(2^{2j+1}-1\right)\zeta(2j+1)}{2^{m+2j-2}(m-j+1)!(j-1)!} \\
&= -8 + 7\zeta(3) + \left(-\frac{7\zeta(3)}{2} + \frac{31\zeta(5)}{8}\right)v^2 \\
&\quad + \left(\frac{7\zeta(3)}{8} - \frac{31\zeta(5)}{16} + \frac{127\zeta(7)}{128}\right)v^4 \\
&\quad + \left(-\frac{7\zeta(3)}{48} + \frac{31\zeta(5)}{64} - \frac{127\zeta(7)}{256} + \frac{511\zeta(9)}{3072}\right)v^6 + \cdots \\
&\quad + \left(\frac{7\zeta(3)}{384} - \frac{31\zeta(5)}{384} + \frac{127\zeta(7)}{1024} - \frac{511\zeta(9)}{6144} + \frac{2047\zeta(11)}{98304}\right)v^8 + \cdots,
\end{aligned}
\tag{B7}
$$

$$
\begin{aligned}
\mathrm{Z}_4(v) :&= 1 - \mathrm{Z}(4,0,v) + \mathrm{Z}(4,1,v) \\
&= 4(10 - 16\ln 2 + \zeta(3))v^2 + 2(25 - 24\ln 2 - 7\zeta(3))v^4 \\
&\quad + \left(12 - \frac{16\ln 2}{3} - \frac{14\zeta(3)}{3} - \frac{31\zeta(5)}{12}\right)v^6 + \frac{2}{3}v^8 \\
&\quad + \sum_{m=4}^{\infty} v^{2m} \sum_{j=2}^{m-1} \frac{(-1)^{m+j}(2j-m-1)\left(2^{2j+1}-1\right)\zeta(2j+1)}{2^{m+2j-6}m(m-1)(m-2)(m-3)(m-j-1)!(j-2)!} \\
&= 4(10 - 16\ln 2 + \zeta(3))v^2 + 2(25 - 24\ln 2 - 7\zeta(3))v^4 \\
&\quad + \left(12 - \frac{16\ln 2}{3} - \frac{14\zeta(3)}{3} - \frac{31\zeta(5)}{12}\right)v^6 + \left(\frac{2}{3} - \frac{31\zeta(5)}{96} - \frac{127\zeta(7)}{384}\right)v^8 \\
&\quad + \left(\frac{31\zeta(5)}{960} - \frac{511\zeta(9)}{15360}\right)v^{10} + \cdots,
\end{aligned}
\tag{B8}
$$

$$
\begin{aligned}
\mathrm{H}_1(\upsilon) :&= \frac{1}{2} + \frac{1}{2\upsilon^2}\left(-3 + \mathrm{Z}_2(\upsilon) + \mathrm{Z}_3(\upsilon)\right) + \frac{1}{4\upsilon^4}\left(8\mathrm{Z}_2(\upsilon) - \mathrm{Z}_4(\upsilon)\right) \\
&= 3 - 4\ln 2 + \left(4 - \frac{8\ln 2}{3} - \frac{7\zeta(3)}{4}\right)\upsilon^2 \\
&\quad + \left(\frac{1}{2} - \frac{7\zeta(3)}{96} - \frac{155\zeta(5)}{384}\right)\upsilon^4 + \left(\frac{7\zeta(3)}{960} + \frac{31\zeta(5)}{640} - \frac{889\zeta(7)}{15360}\right)\upsilon^6 \\
&\quad + \left(-\frac{7\zeta(3)}{11520} - \frac{217\zeta(5)}{46080} + \frac{2159\zeta(7)}{184320} - \frac{511\zeta(9)}{81920}\right)\upsilon^8 + \cdots,
\end{aligned} \tag{B9}
$$

$$
\begin{aligned}
\mathrm{H}_2(\upsilon) :&= -\mathrm{Z}\left(3, \frac{1}{2}, \upsilon\right) + 4\mathrm{Z}_1(\upsilon) + \frac{1}{2}\mathrm{Z}_2(\upsilon) \\
&\quad - \frac{3}{2\upsilon^2}\left(4\mathrm{Z}_1(\upsilon) + \mathrm{Z}_2(\upsilon)\right) + \frac{3}{\upsilon^4}\mathrm{Z}_2(\upsilon) \\
&= 5 - 4\ln 2 - \frac{7\zeta(3)}{4} + \left(8 - 4\ln 2 - \frac{21\zeta(3)}{8} - \frac{31\zeta(5)}{16}\right)\upsilon^2 \\
&\quad + \left(1 - \frac{7\zeta(3)}{64} - \frac{31\zeta(5)}{128} - \frac{635\zeta(7)}{1024}\right)\upsilon^4 \\
&\quad + \left(\frac{7\zeta(3)}{640} + \frac{31\zeta(5)}{1280} + \frac{2413\zeta(7)}{30720} - \frac{3577\zeta(9)}{30720}\right)\upsilon^6 \\
&\quad + \left(-\frac{7\zeta(3)}{7680} - \frac{31\zeta(5)}{15360} - \frac{127\zeta(7)}{15360} + \frac{6643\zeta(9)}{245760} - \frac{2047\zeta(11)}{131072}\right)\upsilon^8 + \cdots,
\end{aligned} \tag{B10}
$$

$$
\begin{aligned}
\mathrm{H}_3(\upsilon) :&= -\frac{3}{2} - \frac{\upsilon^2}{2} + 6\mathrm{Z}_1(\upsilon) + \frac{3}{4}\mathrm{Z}_2(\upsilon) - \frac{3}{2}\mathrm{Z}\left(3, \frac{1}{2}, \upsilon\right) \\
&\quad + \frac{1}{\upsilon^2}\left(\frac{9}{4} - 9\mathrm{Z}_1(\upsilon) - 3\mathrm{Z}_2(\upsilon) - \frac{3}{4}\mathrm{Z}_3(\upsilon)\right) \\
&\quad + \frac{1}{\upsilon^4}\left(-\frac{3}{2} + 3\zeta(3)e^{-\frac{\upsilon^2}{2}} + 3\mathrm{Z}_2(\upsilon) - \frac{3}{2}\mathrm{Z}_3(\upsilon) + \frac{3}{4}\mathrm{Z}_4(\upsilon)\right) \\
&= 6 - 6\ln 2 - \frac{9\zeta(3)}{4} + \left(8 - 4\ln 2 - \frac{9\zeta(3)}{4} - \frac{93\zeta(5)}{32}\right)\upsilon^2 \\
&\quad + \left(1 - \frac{5\zeta(3)}{32} + \frac{31\zeta(5)}{256} - \frac{1905\zeta(7)}{2048}\right)\upsilon^4 \\
&\quad + \left(\frac{\zeta(3)}{64} - \frac{93\zeta(5)}{2560} + \frac{3937\zeta(7)}{20480} - \frac{3577\zeta(9)}{20480}\right)\upsilon^6 \\
&\quad + \left(-\frac{\zeta(3)}{768} + \frac{31\zeta(5)}{6144} - \frac{889\zeta(7)}{30720} + \frac{24017\zeta(9)}{491520} - \frac{6141\zeta(11)}{262144}\right)\upsilon^8 + \cdots.
\end{aligned} \tag{B11}
$$

Each of the descriptive functions given in Equations B4–B11 is calculated in two alternative forms: first as a function of exponential series of the variable υ and then as a power series of the variable υ. The first forms are more compact;

however, numerical calculations of their values are more complex. The second forms have more complicated notations, but the convergence of such series is much faster. For the same reasons, expressions that describe the properties of considered estimators given in section 3.1 are also given in two forms; however, the first form is presented in Appendix D.

Based on formulae 11–13, one can calculate the expected values of any powers of variables A_t, C_t, X_t and their products. Derived expressions for selected raw moments till the fourth order for $\upsilon \neq 0$ $(\mu \neq 0)$ are presented next. The first moments are given in the following forms:

$$E[X_t] = \mu t = \upsilon \sigma \sqrt{t}, \tag{B12}$$

$$E[C_t] = \frac{\sigma\sqrt{t}}{2}\left(\left(\frac{1}{\upsilon} + \upsilon\right)\mathrm{erf}\left(\frac{\upsilon}{\sqrt{2}}\right) + \sqrt{\frac{2}{\pi}}e^{-\frac{\upsilon^2}{2}} + \upsilon\right), \tag{B13}$$

$$E[A_t] = \frac{\sigma\sqrt{t}}{2}\left(-\left(\frac{1}{\upsilon} + \upsilon\right)\mathrm{erf}\left(\frac{\upsilon}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}e^{-\frac{\upsilon^2}{2}} + \upsilon\right). \tag{B14}$$

The second moments are described by the following formulae:

$$E\left[X_t^{\,2}\right] = \sigma^2 t + \mu^2 t^2 = \sigma^2 t\left(1 + \upsilon^2\right), \tag{B15}$$

$$E\left[C_t^{\,2}\right] = \sigma^2 t\left[\left(-\frac{1}{2\upsilon^2} + 2\upsilon^2 + 1\right)\mathrm{erf}\left(\frac{\upsilon}{\sqrt{2}}\right) + \frac{(1+\upsilon^2)e^{-\frac{\upsilon^2}{2}}}{\sqrt{2\pi}\upsilon} + 1 + \frac{\upsilon^2}{2}\right], \tag{B16}$$

$$E\left[A_t^{\,2}\right] = \sigma^2 t\left[\left(\frac{1}{2\upsilon^2} - 2\upsilon^2 - 1\right)\mathrm{erf}\left(\frac{\upsilon}{\sqrt{2}}\right) - \frac{(1+\upsilon^2)e^{-\frac{\upsilon^2}{2}}}{\sqrt{2\pi}\upsilon} + 1 + \frac{\upsilon^2}{2}\right], \tag{B17}$$

$$E(A_t C_t) = \sigma^2 t\left(-\frac{1}{2} + \frac{1}{4\upsilon^2}\left(1 - \mathrm{Z}(2,0,\upsilon) + \mathrm{Z}(2,1,\upsilon)\right)\right) \tag{B18}$$

$$E[C_t X_t] = \sigma^2 t\left[\left(-\frac{1}{2\upsilon^2} + 2\upsilon^2 + 1\right)\mathrm{erf}\left(\frac{\upsilon}{\sqrt{2}}\right) + \frac{(1+\upsilon^2)e^{-\frac{\upsilon^2}{2}}}{\sqrt{2\pi}\upsilon} + \frac{1}{2} + \frac{\upsilon^2}{2}\right], \tag{B19}$$

$$E[A_t X_t] = \sigma^2 t\left[\left(\frac{1}{2\upsilon^2} - 2\upsilon^2 - 1\right)\mathrm{erf}\left(\frac{\upsilon}{\sqrt{2}}\right) - \frac{(1+\upsilon^2)e^{-\frac{\upsilon^2}{2}}}{\sqrt{2\pi}\upsilon} + \frac{1}{2} + \frac{\upsilon^2}{2}\right]. \tag{B20}$$

The third moment is expressed as follows:

$$E\left[X_t^{\,3}\right] = 3\sigma^2 \mu t^2 + \mu^3 t^3 = \sigma^3 t^{\frac{3}{2}}\left(3\upsilon + \upsilon^3\right). \tag{B21}$$

The fourth moments are described by the following equations:

$$E\left[X_t^{\,4}\right] = 3\sigma^4 t^2 + 6\sigma^2\mu^2 t^3 + \mu^4 t^4, \tag{B22}$$

$$E\left[A_t^{\,2} X_t^{\,2} + C_t^{\,2} X_t^{\,2}\right] = 4\sigma^4 t^2 + 7\sigma^2\mu^2 t^3 + \mu^4 t^4, \tag{B23}$$

$$E\left[A_t^{\,4} + C_t^{\,4}\right] = 6\sigma^4 t^2 + 8\sigma^2\mu^2 t^3 + \mu^4 t^4, \tag{B24}$$

$$E\left[A_t{}^3X_t + C_t{}^3X_t\right] = \frac{9}{2}\sigma^4t^2 + \frac{15}{2}\sigma^2\mu^2t^3 + \mu^4t^4, \tag{B25}$$

$$E\left[A_tX_t{}^3 + C_tX_t{}^3\right] = 3\sigma^4t^2 + 6\sigma^2\mu^2t^3 + \mu^4t^4, \tag{B26}$$

$$E\left[A_t{}^2C_t{}^2\right] = \sigma^4t^2\mathrm{H}_1(\upsilon), \tag{B27}$$

$$E\left[A_t{}^2C_tX_t + A_tC_t{}^2X_t\right] = \sigma^4t^2\left(\mathrm{H}_2(\upsilon) - \frac{\upsilon^2}{2} - \frac{1}{2}\right), \tag{B28}$$

$$E\left[A_tC_tX_t{}^2\right] = \sigma^4t^2\left(\frac{1}{2}\mathrm{H}_2(\upsilon) - \frac{\upsilon^2}{2} - \frac{1}{2}\right), \tag{B29}$$

$$E\left[A_t{}^3C_t + A_tC_t{}^3\right] = \sigma^4t^2\mathrm{H}_3(\upsilon). \tag{B30}$$

Appendix C

Raw moments of random variables low, high and closing returns of the arithmetic Brownian motion with a zero drift

Moments of random variables, presented in Equations B12–B30, can be calculated analogously for $\upsilon = 0$ ($\mu = 0$) using the same formulae 11–13. However, in this case, limits of functions in the right hand have to be evaluated for $\upsilon \to 0$ ($\mu \to 0$). This is relatively easy if the expressions in Equations B4–B11 are applied. The results of such calculations, presented in Equations C1–C13, can be compared with the outcomes of Garman and Klass (1980, p. 74).

The first moments have the following forms:

$$E[X_t] = 0, \tag{C1}$$

$$E[C_t] = -E[A_t] = \frac{\sqrt{2}\sigma\sqrt{t}}{\sqrt{\pi}}. \tag{C2}$$

The second moments are described by the following expressions:

$$E\left[X_t{}^2\right] = E\left[C_t{}^2\right] = E\left[A_t{}^2\right] = \sigma^2t, \tag{C3}$$

$$E(A_tC_t) = (1 - 2\ln 2)\sigma^2t, \tag{C4}$$

$$E[C_tX_t] = E[A_tX_t] = \frac{1}{2}\sigma^2t. \tag{C5}$$

The fourth moments are expressed by the following equations:

$$E\left[X_t{}^4\right] = E\left[A_t{}^4\right] = E\left[C_t{}^4\right] = 3\sigma^4t^2, \tag{C6}$$

$$E\left[A_t{}^2X_t{}^2 + C_t{}^2X_t{}^2\right] = 4\sigma^4t^2, \tag{C7}$$

$$E\left[A_t{}^3X_t + C_t{}^3X_t\right] = \frac{9}{2}\sigma^4t^2 \tag{C8}$$

$$E\left[A_t{}^4 + C_t{}^4\right] = 6\sigma^4t^2 + 8\sigma^2\mu^2t^3 + \mu^4t^4, \tag{C9}$$

$$E\left[A_t^2 C_t^2\right] = (3 - 4\ln 2)\sigma^4 t^2, \tag{C10}$$

$$E\left[A_t^2 C_t X_t + A_t C_t^2 X_t\right] = \left(\frac{9}{2} - 4\ln 2 - \frac{7}{4}\zeta(3)\right)\sigma^4 t^2, \tag{C11}$$

$$E\left[A_t C_t X_t^2\right] = \left(2 - 2\ln 2 - \frac{7}{8}\zeta(3)\right)\sigma^4 t^2, \tag{C12}$$

$$E\left[A_t C_t^3 + A_t^3 C_t\right] = \left(6 - 6\ln 2 - \frac{9}{4}\zeta(3)\right)\sigma^4 t^2. \tag{C13}$$

Appendix D

Properties of estimators of PARKINSON, GARMAN–KLASS and ROGERS–SATCHELL

Raw moments of random variables presented in Appendices B and C are applied to derive the expected value and the MSE of estimators of Parkinson $\widehat{s^2}_1(t)$, G-K - $\widehat{s^2}_2(t)$, R-S - $\widehat{s^2}_3(t)$. The following formulae are valid for $\upsilon \neq 0$ ($\mu \neq 0$):

$$E\left[\widehat{s^2}_1(t)\right] = \frac{1}{4\ln 2}\left(3 - \frac{1}{2\upsilon^2}\mathrm{Z}_2(\upsilon) + \upsilon^2\right)\sigma^2 t, \tag{D1}$$

$$\begin{aligned} MSE\left[\widehat{s^2}_1(t)\right] = &\left[1 + \frac{\mathrm{Z}_2(\upsilon)}{4\upsilon^2 \ln 2}\right. \\ &\left. + \frac{6 - 24\ln 2 + (8 - 8\ln 2)\upsilon^2 + \upsilon^4 + 6\mathrm{H}_1(\upsilon) - 4\mathrm{H}_3(\upsilon)}{16\ln^2 2}\right]\sigma^4 t^2, \end{aligned} \tag{D2}$$

$$E\left[\widehat{s^2}_2(t)\right] = \left(\left(\frac{5}{2} - 2\ln 2\right) + \left(\frac{3}{2} - 2\ln 2\right)\upsilon^2 - \frac{1}{4\upsilon^2}\mathrm{Z}_2(\upsilon)\right)\sigma^2 t \tag{D3}$$

$$\begin{aligned} MSE\left[\widehat{s^2}_2(t)\right] = &\left[\frac{11}{2} - 18\ln 2 + 12\ln^2 2 + \left(13 - 36\ln 2 + 24\ln^2 2\right)\upsilon^2\right. \\ &\left. + \left(\frac{3}{2} - 2\ln 2\right)^2 \upsilon^4 + \frac{3\mathrm{H}_1(\upsilon)}{2} + (2\ln 2 - 1)\mathrm{H}_2(\upsilon) - \mathrm{H}_3(\upsilon) + \frac{\mathrm{Z}_2(\upsilon)}{2\upsilon^2}\right]\sigma^4 t^2, \end{aligned} \tag{D4}$$

$$Var\left[\widehat{s^2}_3(t)\right] = (2\mathrm{H}_1(\upsilon) - \mathrm{H}_2(\upsilon))\sigma^4 t^2. \tag{D5}$$

After further transformations of Equations D1–D5, the forms presented, respectively, in (17), (18), (21), (22) and (25) are obtained.

References

ANDERSEN, T. G. and T. BOLLERSLEV (1998), Answering the skeptics: yes, standard volatility models do provide accurate forecasts, *International Economic Review*, **39**, 885–905.

BOLLERSLEV, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.

ANDERSEN, T. G., D. DOBREV and E. SCHAUMBURG (2012), Jump-robust volatility estimation using nearest neighbor truncation, *Journal of Econometrics*, **169**, 75–93.

BARNDORFF-NIELSEN, O. E. and N. SHEPHARD (2004), Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics*, **2**, 1–37.

CHOU, R. Y. (2005), Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model, *Journal of Money, Credit, and Banking*, **37**, 561–582.

CHOU, R. Y., H.- C. CHOU and N. LIU (2009), Range volatility models and their applications in finance, in: C.-F. Lee and A. C. Lee, J. Lee (eds.), *Handbook of quantitative finance and risk management*, 1273–1281.

CORSI, F., S. MITTNIK, C. PIGORSCH and U. PIGORSCH (2008), The volatility of realized volatility, *Econometric Reviews*, **27**, 46–78.

COX, D. R. and M. D. MILLER (1965), *The theory of stochastic processes*, Methuen and Co., London.

ENGLE, R. F. and V. K. NG (1993), Measuring and testing the impact of news on volatility, *Journal of Finance*, **48**, 1749–1778.

GARMAN, M. B. and M. J. KLASS (1980), On the estimation of security price volatilities from historical data, *Journal of Business*, **53**, 67–78.

GLOSTEN, L. R, R. JAGANNATHAN and D. E. RUNKLE (1993), On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, **48**, 1779–1801.

HARRISSON, J. M. (1985), *Brownian motion and stochastic flow systems*, John Wiley & Sons, New York.

HAUTSCH, N. (2012), *Econometrics of financial high-frequency data*, Springer-Verlag Berlin Heidelberg.

JAKUBOWSKI, J., A. PALCZEWSKI, M. RUTKOWSKI and Ł. STETTNER (2006), *Matematyka finansowa, instrumenty pochodne*, WNT, Warszawa.

KUNITOMO, N. (1992), Improving the Parkinson method of estimating security price volatilities, *Journal of Business*, **65**, 295–302.

KUNITOMO, N. and M. IKEDA (1992), Pricing options with curved boundaries, *Mathematical Finance*, **2**, 275–298. Q12

LI, A. (1999), The pricing of double barrier options and their variations, *Advances in Futures and Options Research*, **10**, 17–41.

LI, H. and Y. HONG (2011), Financial volatility forecasting with range-based autoregressive volatility model, *Finance Research Letters*, **8**, 69–76.

MCALEER, M. and M. C. MEDEIROS (2008), Realized volatility: a review, *Econometric Reviews*, **27**, 10–45.

NELSON, D. B. (1991), Conditional heteroskedasticity in asset returns: a New approach, *Econometrica*, **59**, 347–370.

PARKINSON, M. (1980), The extreme value method for estimating the variance of the rate of return, *Journal of Business*, **53**, 61-65.

PATTON, A. J. (2011), Data-based ranking of realised volatility estimators, *Journal of Econometrics*, **161** 284–303.

PERCZAK, G. and P. FISZEDER (2013), Estymacja wariancji arytmetycznego ruchu Browna na podstawie znanych wartości minimum, maksimum, końcowej oraz dryfu, *Przegląd Statystyczny*, **60**, 39–62.

PIGORSCH, C., U. PIGORSCH and I. POPOV (2012), Volatility estimation based on high-frequency data, in: J. C. Duan, W. K. Härdle and J. E. Gentle (eds.), *Handbook of computational finance*, Springer-Verlag Berlin Heidelberg, 335–369.

RABEMANANJARA, R. and J. M. ZAKOIAN (1993), Threshold ARCH models and asymmetries in volatility, *Journal of Applied Econometrics*, **8**, 31–49.

ROGERS, L. C. G. and S. E. SATCHELL (1991) Estimating variance from high, Low and closing prices, *The Annals of Applied Probability*, **1**, 504–512.
TERÄSVIRTA, T. (2009), An introduction to univariate GARCH models, in: T. Mikosch, T. Kreiss, J. P. Davis, R. A. Andersen and T. Gustav (eds.), *Handbook of financial time series*, Springer-Verlag Berlin Heidelberg, 17–42.
VISSER, M. P. (2011), Garch parameter estimation using high-frequency data, *Journal of Financial Econometrics*, **9**, 162–197.
YANG, D. and Q. ZHANG (2000), Drift-independent volatility estimation based on high, Low, open, and closing prices, *Journal of Business*, **73**, 477–491.

Received: 10 April 2013. Revised: 12 June 2013.