Submitted to Econometrica

Most Efficient Homogeneous Volatility Estimators

A. Saichev, D. Sornette, V. Filimonov

August 12, 2009

Department of Management, Technology and Economics, ETH Zurich, Kreuzplatz 5, CH-8032 Zurich, Switzerland

Submitted to Econometrica

MOST EFFICIENT HOMOGENEOUS VOLATILITY ESTIMATORS

A. Saichev, D. Sornette, V. Filimonov

We present a new theory of homogeneous volatility (and variance) estimators for arbitrary stochastic processes. The main tool of our theory is the parsimonious encoding of all the information contained in the OHLC prices for a given time interval by the joint distributions of the high-minus-open, low-minus-open and close-minus-open values, whose analytical expression is derived exactly for Wiener processes with drift. The efficiency of the new proposed estimators is favorably compared with that of the Garman-Klass, Roger-Satchell and maximum likelihood estimators.

Keywords: Variance and volatility estimators, efficiency, homogeneous functions, Schwarz inequality, extremes of Wiener processes.

1. INTRODUCTION

Volatility, defined as the standard deviation of the increments of the log-price over a specific time interval, is a universally used risk indicator. While the growing availability of high-frequency tick-by-tick price time series has permitted the development of new efficient volatility estimators (see, for instance, Yang and Zhang (2000), Corsi et al. (2001), Andersen et al. (2003), Aït-Sahalia (2005), Zhang et al. (2005)), most historical time series as well as databases of price time series, for the many tens of thousands of assets (stocks, commodities, bonds, currencies, derivatives and so on) that exit worldwide, only record price in time steps coarse-grained for convenience (which is often daily). However, it is common practice that not just one (close) price is recorded for a given time step, but four of them, called the open-high-low-close (OHLC) of the price for that given interval. It is natural to exploit these four recorded values per time step to develop better volatility estimators.

Rather than just using the time series of close-prices, here, we present a

comprehensive theory of homogeneous volatility (and variance) estimators of arbitrary stochastic processes that fully exploit the OHLC prices. For this, we develop the theory of most efficient point-wise homogeneous OHLC volatility estimators, valid for any price processes. We introduce the "quasi-unbiased estimators", that can address any type of desirable constraints. The main tool of our theory is the parsimonious encoding of all the information contained in the OHLC prices for a given time interval in the form of general "diagrams" associated with the joint distributions of the high-minus-open, low-minus-open and close-minus-open values. The diagrams can be tailored to yield the most efficient estimators associated to any statistical properties of the underlying log-price stochastic process. Applied to Wiener processes for log-prices with drift, we provide explicit analytical expressions for the most efficient point-wise volatility and variance estimators, based on the analytical expression of the joint distribution of the high-minus-open, low-minus-open and close-minus-open values.

Our work improves on the following papers. Garman and Klass (G&K) (1980) introduced a quadratic estimator for the variance of the Wiener process with drift for the log-price, which has rather low variance but which is biased from non-zero drifts. Parkinson (1980) proposed a simple quadratic variance estimator proportional to $(H-L)^2$, which is using only a part of the information available from OHLC prices. Rogers and Satchell (R&S) (1991,1994) introduced another quadratic estimator for the variance of the Wiener process with drift, which is unbiased for all drifts and has a larger fixed variance for all drifts equal to the variance of the process. Both G&K and R&S estimators are focused on the variance, and do not present estimators for the volatility, which is of obvious interest for financial applications. Furthermore, the variance of their estimators is not provided for non-zero drifts. Magdon-Ismail and Atiya (2003) obtain a maximum likelihood (ML) volatility estimator based on the joint distribution of the high and low,

previously obtained by Dominé (1996). Their estimator does not use the close price and is thus less efficient than the ML estimator using the full information of the OHLC, as shown here. In addition, we will show that the ML estimator is not the most efficient. Yang and Zhang (2000) produced an unbiased and efficient quadratic variance estimator, taking into account the OHLC of log-prices for $n > 1$ consecutive days. Their main novelty is to take into account the possible existence of jumps (or gaps) of prices from yesterday's close till today's open prices. Their minimization of the variance of their estimators requires the estimation of expectations of a quadratic form of the OHLC which they only partly achieve due to the lack of knowledge of the full joint distribution, which we offer in the Appendix. Chan and Lien (2003) compared the empirical effectiveness of four estimators, the Parkinson, the G&K and R&S ones, and the naive excursion range $H - L$ estimator. In sum, the present paper can be viewed as providing the full underpinning theory of all these previous works, since we are able to express efficient estimators in the presence of arbitrary constraints from the explicit knowledge of the joint distribution of the OHLC log-prices.

The paper is organized as follows. Section 2 describes the properties of the stochastic processes for which our theory of most efficient homogeneous estimators is developed. Section 3 derives the general expressions for the most efficient volatility and variance OHLC estimators. Section 4 provides detailed analytical results on the statistical properties of the most efficient homogeneous estimators, for the case of Wiener process with drift. Section 5 compares the exact analytical results with those obtained using numerical simulations of millions of realizations of the Wiener process with drift. Section 6 concludes. The Appendix presents the joint probability density function of the high-minus-open, low-minus-open and close-minus-open values for the Wiener process with drift.

2. VOLATILITY OF STOCHASTIC PROCESSES: MODELS, DEFINITIONS AND PROPERTIES

2.1. *Volatility of simple stochastic log-price process*

The goal of this paper is to construct efficient estimators for the volatility of log-price processes. First, we specify the general properties of the stochastic processes to which our estimators will be applied.

Let us consider the stochastic process $X(t)$, which is interpreted as the log-price of some asset at time t. Its volatility over the time interval of duration T_0 is by definition the standard deviation of the increment $X(t_0+T_0)-X(t_0)$. We assume that $X(t)$ has stationary increments. Accordingly, for simplicity but without loss of generality, we can take $t_0 = 0$ and $X(0) = 0$ and choose the time scale such that $T_0 = 1$. All the rest of the paper is based on the following definition of volatility:

DEFINITION 2.1 The volatility σ of the stochastic process $X(t)$ is equal to the square-root of the variance of its increment per unit time

$$\sigma = \sqrt{D} \ , \qquad D = \mathsf{Var}[X(1)] \ .$$

The estimators of the volatility σ and of the variance $D \equiv \sigma^2$ will be denoted respectively $\hat{\sigma}$ and $\hat{D}$.

We consider the following class of stochastic processes

$$(1) \qquad X(t) = \sigma A(t,\gamma) \ ,$$

where $A(t,\gamma)$ is an auxiliary stochastic process, whose statistical properties are assumed to be known for any given value of the parameter γ. We assume additionally that the expectation and the variance of the stochastic process $A(t,\gamma)$ are finite:

$$\mathrm{E}[A(t,\gamma)] < \infty \ , \qquad \sigma_0^2(t,\gamma) = \mathrm{Var}[A(t,\gamma)] < \infty \ .$$

It follows from (1) and from the definition of volatility that the stochastic process $A(t,\gamma)$ has a unity volatility: $\sigma_0(1,\gamma)=1$.

Let us introduce the following auxiliary stochastic process

$$(2)\qquad Y(t)=\frac{X(t)}{\sigma_0(T,\gamma)}=\sigma B(t,\gamma)\ ,\qquad B(t,\gamma)=\frac{A(t,\gamma)}{\sigma_0(T,\gamma)}\ .$$

By construction, the variance of the increments of the "normalized" stochastic process $Y(t)$ over a time interval of arbitrary duration T coincides with the variance of the increments of the original process $X(t)$ over the unit time interval of duration $T_0=1$:

$$\mathrm{Var}[B(T,\gamma)]=1\qquad\Rightarrow\qquad\mathrm{Var}[Y(T,\gamma)]=\sigma^2\ .$$

Let us consider particular examples of the stochastic processes $X(t)$ given by (1) and of the corresponding $Y(t)$ defined by (2):

EXAMPLE 2.1 The simplest and most common log-price process is the Wiener process

$$(3)\qquad X(t)=\mu t+\sigma W(t)\ ,$$

where $W(t)$ is the standard Wiener process, such that $\mathrm{E}[W(t)]=0$, $\mathrm{Var}[W(t)]=t$, while μ is the drift parameter. In this case, $\sigma_0(T,\gamma)=\sqrt{T}$, so that the auxiliary stochastic process $Y(t)$ (1) takes the form

$$Y(t)=X(t)\Big/\sqrt{T}=\sigma B(t,\gamma)\ ,$$

where

$$(4)\qquad B(t,\gamma)=v(\tau,\gamma)\ ,\qquad v(\tau,\gamma)=\gamma\tau+W(\tau)\ .$$

Here, we introduced the "normalized" time τ and the parameter γ:

$$(5)\qquad \tau=\frac{t}{T}\ ,\qquad \gamma=\frac{\mu}{\sigma}\sqrt{T}\ .$$

REMARK 2.1 In practical applications, the parameter γ, which is is proportional to the drift of the stochastic process $X(t)$ (3), is generally unknown. Our strategy is to proceed in two steps: (i) determine the most efficient volatility and variance estimators for a fixed value of γ, say γ_0; (ii) explore in details the efficiency of the estimators for values of γ that deviate from γ_0.

EXAMPLE 2.2 Let $X(t)$ be defined at discrete times $t = 0, 1, 2, \ldots$ and let it satisfy to recurrent relation

$$X(n+1) = X(n) + \mu + \sigma\epsilon_n \ , \quad X(0) = 0 \ , \quad n = 0, 1, 2, \ldots$$

where $\{\epsilon_n\}$ is a sequence of iid random variables with zero expectation and unit variance $\mathrm{Var}[\epsilon_n] = 1$. In order to estimate the volatility σ from recorded values of $X(n)$ over a discrete time interval of duration N, it is convenient to introduce the "normalized" discrete-time process

$$Y(n) = \frac{X(n)}{\sqrt{N}} = \sigma v(n, \gamma) \ ,$$

where

$$\begin{aligned} v(n,\gamma) = \gamma n + \omega(n) \ , \qquad n = 1, 2, \ldots \ , \qquad v(0,\gamma) = 0 \ , \\ \gamma = \frac{\mu}{\sigma\sqrt{N}} \ , \qquad \omega(n) = \frac{1}{\sqrt{N}} \sum_{k=1}^{n} \epsilon_k \ . \end{aligned} \tag{6}$$

REMARK 2.2 If the random variables $\{\epsilon_k\}$ are Gaussian, the stochastic process $X(n)$ can be interpreted as the discrete-time version of the process $X(t)$ defined by (3). More interesting is the case where $\{\epsilon_k\}$ are non-Gaussian random variables, with a fat tail probability density distribution $f(x) \sim |x|^{-1-p}$ for large $|x|$, with $p > 2$ ensuring that the variance exists (see McKenzie, (2006) for an excellent review of the history of fat tails in financial returns).

2.2. *OHLC volatility and variance estimators*

Given the observed realization of the stochastic process $X(t)$ within some time interval $t \in (0,T)$ over m points, $(t_1, t_2, \ldots, t_{m-1}) \in (0,T), \quad t_m = T$, the most general expression of the estimator of the volatility of $X(t)$ is the function

$$\hat{\sigma}_m = \hat{\sigma}(X_1, X_2, \ldots, X_m) \ ,$$

of the recorded values

$$X_1 = X(t_1) \ , \quad X_2 = X(t_2) \ , \quad \ldots, \quad X_m = X(T) \ .$$

Of particular interest for its widespread use and parsimonious representation of a given realization of the process $X(t)$ over a finite time interval is the case $m = 3$ that corresponds, in particular, to OHLC estimators. The four letters OHLC stand respectively for Open, High, Low and Close. In the following, we focus on this case due to its special significance, while it is understood that one can generalize the theory developed here to higher-order multipoint estimators corresponding to any value $m > 3$.

Without loss of generality, we pose $X(0) = 0$ (in practice, the relevant quantities are simply decreased by the opening value at time = 0). Then, the high, low and close values of a given realization of the stochastic process $X(t)$ within the time interval $t \in (0,T)$ are

$$H = \sup_{t\in(0,T)} X(t) \ , \qquad L = \inf_{t\in(0,T)} X(t) \ , \qquad C = X(T) \ . \tag{7}$$

DEFINITION 2.2 Among all three-points volatility and variance estimators, the *OHLC estimators* are defined as functions of only the three measures (high, low and close) of the realization of the stochastic process $X(t)$ within the time interval $t \in (0,T)$ defined by (7). Specifically, OHLC volatility and variance estimators are functions which can be written as follows:

$$\hat{\sigma} = \hat{\sigma}(H, L, C) \ , \qquad \hat{D} = \hat{D}(H, L, C) \ . \tag{8}$$

Such OHLC estimators are well-known to be more efficient than the equidistant three-points estimators corresponding to $t_k = kT/3$ $(k = 1, 2, 3)$.

2.3. *Quadratic OHLC variance and volatility estimators*

Almost all known variance OHLC estimators are quadratic forms of H, L and C. Let us introduce the vector $\boldsymbol{X}^T = (H, L, C)$, where ^{T} denotes the transpose operation. Let us denote by $\boldsymbol{Q}$ any positive-definite 3×3 matrix.

DEFINITION 2.3 The variance OHLC estimator $\hat{D}$ is called *quadratic* if it can be expressed as a quadratic form

$$\hat{D} = \frac{1}{\sigma_0^2(T,\gamma)}\, \boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X} = \boldsymbol{Y}^T\boldsymbol{Q}\boldsymbol{Y}\ , \qquad \text{where} \qquad \boldsymbol{Y} = \frac{\boldsymbol{X}}{\sigma_0(T,\gamma)}\ . \tag{9}$$

In turn, the volatility OHLC estimator $\hat{\sigma}$ is called *quadratic* if it can be represented in the form

$$\hat{\sigma} = \frac{1}{\sigma_0(T,\gamma)}\, \sqrt{\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}} = \sqrt{\boldsymbol{Y}^T\boldsymbol{Q}\boldsymbol{Y}}\ . \tag{10}$$

Two well-known OHLC estimators are quadratic, as shown in the two examples 2.3 and 2.4.

EXAMPLE 2.3 Rogers and Satchell (1991) have suggested the following quadratic OHLC variance estimator

$$\hat{D}_{\mathrm{RS}} = \frac{1}{T}\,[H(H-C) + L(L-C)]\ . \tag{11}$$

We will refer to this estimator as the *R&S variance estimator*. The corresponding expression

$$\hat{\sigma}_{\mathrm{RS}} = \frac{1}{\sqrt{T}}\,\sqrt{H(H-C) + L(L-C)} \tag{12}$$

will be called the *R&S volatility estimator*.

The R&S variance estimator (11) has the nice property of being unbiased. Namely, for the Wiener process defined by (3), and for any σ and μ (i.e. for any values of the parameter γ), the expected value of the R&S variance estimator (11) is equal to the variance of the original process over the time interval $[0,1]$: $\mathrm{E}[\hat{D}_{\mathrm{RS}}] = \sigma^2$.

EXAMPLE 2.4 Another quadratic OHLC variance estimator was suggested by Garman and Klass (1980). This *G&K variance estimator* is defined by

$$\hat{D}_{\mathrm{GK}} = \frac{1}{T}\left[k_1\,(H-L)^2 - k_2\,(C(H+L) - 2HL) - k_3\,C^2\right] \ , \tag{13}$$

where $k_1 = 0.511$, $k_2 = 0.019$, $k_3 = 0.383$. The square root of expression (13) will be referred to as the *G&K volatility estimator*.

For the Wiener process (3), the G&K variance estimator is unbiased only if the drift is equal to zero. In general, $\mathrm{E}[\hat{D}_{\mathrm{GK}}] \neq \sigma^2$ if $\mu \neq 0$ ($\gamma \neq 0$). This bias is a shortcoming of the G&K variance estimator. Its advantage is that, for zero drift $\mu = 0$ ($\gamma = 0$), its variance is significantly smaller than the variance of the R&S variance estimator.

2.4. *Homogenous variance and volatility estimators*

In order to more clearly understand the key properties of any quadratic estimators, it is instructive to introduce "generalized" R&S and G&K estimators for the general stochastic process $X(t)$ defined by (1). For definiteness, we will focus here on the "generalized"R&S variance estimator obtained by replacing T by $\sigma_0^2(T,\gamma)$ in (11):

$$\hat{D}_{\mathrm{RS}} = \frac{1}{\sigma_0^2(T,\gamma)}\,\left[H(H-C) + L(L-C)\right] \ .$$

Using relations (2), it can be written in the form

$$\hat{D}_{\mathrm{RS}} = \sigma^2 \hat{d}_{\mathrm{RS}} \ , \qquad \hat{d}_{\mathrm{RS}} = \bar{H}(\bar{H}-\bar{C}) + \bar{L}(\bar{L}-\bar{C}) \ , \tag{14}$$

where $\hat{d}_{\rm RS}$ is function of the high, low and close values of the auxiliary stochastic process $B(t,\gamma)$ defined by (2) within the interval $t \in (0,T)$:

$$\bar{H} = \sup_{t\in(0,T)} B(t,\gamma) \ , \qquad \bar{L} = \inf_{t\in(0,T)} B(t,\gamma) \ , \qquad \bar{C} = B(T,\gamma) \ . \tag{15}$$

Accordingly, the R&S volatility estimator is equal to

$$\hat{\sigma}_{\rm RS} = \sigma \hat{s}_{\rm RS} \ , \qquad \hat{s}_{\rm RS} = \sqrt{\bar{H}(\bar{H}-\bar{C}) + \bar{L}(\bar{L}-\bar{C})} \ . \tag{16}$$

The R&S variance estimator $\hat{D}_{\rm RS}$ given by (14) has the following important property: It is equal to the product of the (unknown) variance σ^2 of the original process $X(t)$ defined by (1) and of the random factor $\hat{d}_{\rm RS}$. The statistical properties of $\hat{d}_{\rm RS}$ are expressed via the statistical properties of auxiliary process $B(t,\gamma)$, which are known by definition. Therefore, for a given γ, the statistical properties of $\hat{d}_{\rm RS}$ do not depend on the variance σ^2 of the original process $X(t)$ defined in (1). Moreover, since the R&S variance estimator is unbiased, the expectation of $\hat{d}_{\rm RS}$ is equal to unity: $\mathrm{E}[\hat{d}_{\rm RS}] \equiv 1$. Correspondingly, one can quantitatively characterize the relative error of the R&S variance estimator by the variance of the factor $\hat{d}_{\rm RS}$,

$$\mathrm{Var}[\hat{d}_{\rm RS}] = \mathrm{E}[\hat{d}^2_{\rm RS}] - 1 \ ,$$

which does not depend (for a given γ) on the sought variance σ^2. Figuratively speaking, one can interpret relation (14) as if the sought variance σ^2 was known while its random factor $\hat{d}_{\rm RS}$ was unknown. Thus, the R&S variance estimator is all the more efficient, the smaller is the variance of factor $\hat{d}_{\rm RS}$.

DEFINITION 2.4 If the OHLC variance estimator is represented in the form

$$\hat{D} = \sigma^2 \hat{d} \tag{17}$$

where the factor

$$\hat{d} = \hat{d}(\bar{H}, \bar{L}, \bar{C}) \tag{18}$$

depends only on the high, low and close values (15) of the auxiliary stochastic process $B(t,\gamma)$ and does not depend (for any given γ) on the variance σ^2 of the original stochastic process $X(t)$, then we refer to the factor $\hat{d}$ (18) as the *canonical variance estimator*. Similarly, if the volatility OHLC estimator is represented in the following form, analogous to (16),

$$\hat{\sigma} = \sigma\hat{s}\ , \qquad \hat{s} = \hat{s}(\bar{H},\bar{L},\bar{C})\ , \tag{19}$$

then the factor $\hat{s}$ is the *canonical volatility estimator*.

Obviously, the estimators (17) and (19) are *unbiased*, for a given $\gamma = \gamma_0$, if

$$\mathrm{E}[\hat{d}|\gamma_0] = 1\ , \qquad \mathrm{E}[\hat{s}|\gamma_0] = 1\ . \tag{20}$$

Here and below, we use the notations $\mathrm{E}[\ldots|\gamma_0]$, $\mathrm{Var}[\ldots|\gamma_0]$ for conditional expectations and variances, under the condition that the parameter γ is equal to γ_0.

REMARK 2.3 In general, all volatility $\hat{\sigma}$ and variance $\hat{D}$ estimators (8) are functions of H, L and C. However, it is not true that all of them accept canonical estimators $\hat{s}$ and $\hat{d}$ (18), (19), depending on the variables $\bar{H}$, $\bar{L}$, $\bar{C}$ (15). In the present paper, we explore only *homogeneous* estimators, defined below, which are expressed via canonical estimators.

DEFINITION 2.5 The OHLC variance estimator is called homogeneous if it can be represented in the form

$$\hat{D}(H,L,C) = h_2(H,L,C)\Big/\sigma_0^2(T,\gamma)\ , \tag{21}$$

where $h_2(H,L,C)$ is a second-order homogeneous function. Analogously, the volatility estimator is called homogeneous if it can be represented in the form

$$\hat{\sigma}(H,L,C) = h_1(H,L,C)\Big/\sigma_0(T,\gamma)\ , \tag{22}$$

where h_1 is a first-order homogeneous function.

THEOREM 2.1 *The homogeneous OHLC variance estimators* $\hat{D}(H,L,C)$ (21) *accept the representation form* (17), (18).

Proof. It follows from (7), (15) and definition (2) of the auxiliary stochastic process $B(t,\gamma)$, that the following equalities are true

$$H = \alpha\bar{H}\ , \quad L = \alpha\bar{L}\ , \quad C = \alpha\bar{C}\ , \quad \alpha = \sigma\sigma_0(T,\gamma)\ . \tag{23}$$

Thus, one can rewrite relation (21) in the form

$$\hat{D}(H,L,C) = h_2(\alpha\bar{H},\alpha\bar{L},\alpha\bar{C})\Big/\sigma_0^2(T,\gamma)\ .$$

From the homogeneity property of the second order homogeneous function h_2,

$$h_2(\alpha\bar{H},\alpha\bar{L},\alpha\bar{C}) \equiv \alpha^2 h_2(\bar{H},\bar{L},\bar{C})\ ,$$

we obtain

$$\hat{D}(H,L,C) = \sigma^2 h_2(\bar{H},\bar{L},\bar{C})\ .$$

Thus, the homogeneous estimators (21) are reduced to (17) and (18), where $\hat{d} = h_2(\bar{H},\bar{L},\bar{C})$. □

REMARK 2.4 One can prove in a similar way that homogenous volatility estimators (22) are reduced to the form (19).

DEFINITION 2.6 The variance (21) and volatility (22) estimators are the *most efficient homogeneous estimators*, for a given γ_0, if the corresponding canonical variance and volatility estimators satisfy relations (20), while their variances

$$\mathsf{Var}[\hat{d}|\gamma_0] = \mathsf{E}[\hat{d}^2|\gamma_0] - 1\ , \quad \mathsf{Var}[\hat{s}|\gamma_0] = \mathsf{E}[\hat{s}^2|\gamma_0] - 1\ ,$$

are the smallest among the variances of all possible canonical homogeneous estimators, which are unbiased at γ_0,

REMARK 2.5 All quadratic estimators are homogeneous. This results from their definition (9), since the quadratic form $\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}$ is a second order homogeneous function of its argument $\boldsymbol{X}$. Analogously, the quadratic volatility estimator (10) is homogeneous, because $\sqrt{\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}}$ is a homogeneous function of order one. In particular, the quadratic R&S (11) and G&K (13) variance estimators are homogeneous.

More insight in homogeneous OHLC estimators can be obtained by representing (H, L, C) in the following "spherical" (or geographic) coordinates which embody parsimoniously the homogeneity property:

$$
\begin{gathered}
H = R\cos\Theta\cos\Phi\ , \quad L = R\cos\Theta\sin\Phi\ , \quad C = R\sin\Theta\ , \\
R = \sqrt{H^2 + L^2 + C^2}\ , \\
\Theta = \arctan\left(\frac{C}{\sqrt{H^2+L^2}}\right)\ , \quad \Phi = \arctan\left(\frac{L}{H}\right)\ .
\end{gathered} \tag{24}
$$

THEOREM 2.2 *Any variance estimator of the form*

$$\hat{D} = \frac{R^2}{\sigma_0^2(T,\gamma)}\varphi(\Theta,\Phi)\ , \tag{25}$$

where R, Θ and Φ are given by (24) and $\varphi(\theta,\phi)$ is an arbitrary function, is a homogeneous variance estimator. Reciprocally, any homogenous variance estimator (21) *can be expressed in the form (25). Similarly,*

$$\hat{\sigma} = \frac{R}{\sigma_0(T,\gamma)}\psi(\Theta,\Phi)\ , \tag{26}$$

where $\psi(\theta,\phi)$ is arbitrary function, is a homogeneous volatility estimator and reciprocally.

Proof. It follows from (24) that R^2 is a second order homogeneous function of its arguments (H, L, C), while Θ and Φ are zero order homogeneous functions of the same arguments. Accordingly, $\varphi(\Theta,\Phi)$ is a zero order homogenous function of (H, L, C), while $R^2\varphi(\Theta,\Phi)$ is a second order homogeneous function of (H, L, C). Thus, due to theorem 2.1, the estimator (25)

is represented in homogeneous form as

$$\hat{D} = \sigma^2 \hat{d}(\bar{H}, \bar{L}, \bar{C}) \ , \qquad \text{where} \qquad \hat{d}(\bar{H}, \bar{L}, \bar{C}) = \bar{R}^2 \varphi(\Theta, \Phi) \ , \tag{27}$$
$$\bar{R} = \sqrt{\bar{H}^2 + \bar{L}^2 + \bar{C}^2} \ .$$

In turn, it is obvious that any homogeneous estimator (21), is represented in the form (25), where

$$\varphi(\Theta, \Phi) = h_2(\cos\Theta\cos\Phi, \cos\Theta\sin\Phi, \sin\Theta) \ .$$

Using a similar derivation, it is easy to prove that $\hat{\sigma}$ given by (26) is a homogeneous volatility estimator, i.e.,

$$\hat{\sigma} = \sigma \hat{s}((\bar{H}, \bar{L}, \bar{C}) \ , \quad \text{where} \quad \hat{s} = \bar{R}\psi(\Theta, \Phi) \ . \qquad \square \tag{28}$$

REMARK 2.6 The inequalities

$$\bar{L} \leqslant \bar{C} \leqslant \bar{H} \ , \qquad \bar{H} \geqslant 0 \ , \qquad \bar{L} \leqslant 0 \ ,$$

resulting from the definition of H, L, C, impose that $\bar{R}$, Θ and Φ should satisfy

$$0 \leqslant \bar{R} < \infty \ , \quad -\frac{\pi}{2} \leqslant \Phi \leqslant 0 \ , \quad s(\Phi) \leqslant \Theta \leqslant c(\Phi) \ ,$$
$$s(\phi) = \arctan(\sin\phi) \ , \qquad c(\phi) = \arctan(\cos\phi) \ .$$

DEFINITION 2.7 We will refer to the functions $\varphi(\theta, \phi)$ and $\psi(\theta, \phi)$, defined respectively by (27) and (28), as the *diagrams* of the homogeneous OHLC variance and volatility estimators.

EXAMPLE 2.5 From the definitions (11) and (13), the diagrams of the R&S and G&K variance estimators are

$$\varphi_{\rm RS}(\theta, \phi) = \cos^2\theta - \frac{1}{2}\sin 2\theta(\cos\phi + \sin\phi) \ ,$$
$$\varphi_{\rm GK}(\theta, \phi) = k_1 \cos^2\theta(\cos\phi - \sin\phi)^2 \tag{29}$$
$$+k_2\left[\cos^2\theta\sin 2\phi - \frac{1}{2}\sin 2\theta(\cos\phi + \sin\phi)\right] - k_3\sin^2\theta \ .$$

It is probable that R&S and G&K estimators are not the most efficient quadratic estimators for any given value γ_0 of the parameter γ. It is therefore natural to search for the most efficient quadratic estimators, at a given value γ_0, which might be more efficient than R&S and G&K estimators. We will determine below the most efficient homogeneous volatility and variance OHLC estimators for any given γ_0. The following theorem summarizes the relations between the most efficient quadratic and the most efficient homogeneous OHLC estimators.

THEOREM 2.3 *Let $Var_q[\hat{d}|\gamma_0]$ and $Var_q[\hat{s}|\gamma_0]$ be the variances of the most efficient quadratic canonical OHLC estimators for a given γ_0. Let $Var_h[\hat{d}|\gamma_0]$ $Var_h[\hat{s}|\gamma_0]$ be the variances of the most efficient homogeneous canonical OHLC estimators for the same given γ_0. Then, the following inequalities hold true*

$$Var_q[\hat{d}|\gamma_0] \geqslant Var_h[\hat{d}|\gamma_0] \ , \qquad Var_q[\hat{s}|\gamma_0] \geqslant Var_h[\hat{s}|\gamma_0] \ . \tag{30}$$

In another words, at a given value γ_0, the most efficient homogeneous OHLC estimator is no less efficient than the most efficient quadratic OHLC estimator.

Proof. Denoting as Ω_q the set of quadratic OHLC estimators, and as Ω_h the set of homogeneous OHLC estimators, we have $\Omega_q \subset \Omega_h$. The inequalities (30) derive from this inclusion. □

3. DIAGRAMS OF MOST EFFICIENT OHLC HOMOGENEOUS ESTIMATORS

In this section, we derive the expressions for the most efficient (at $\gamma = \gamma_0$) homogeneous variance and volatility OHLC estimators , whose canonical estimators are given by expressions (27) and (28). To make clear that these estimators depend on γ_0, we will use the following notations for the diagrams of the most efficient homogeneous estimators: $\varphi(\theta, \phi; \gamma_0)$ and $\psi(\theta, \phi; \gamma_0)$.

We assume the existence of the joint probability density function (pdf) $\bar{\mathcal{Q}}(h,l,c;\gamma)$ of the random variables $(\bar{H},\bar{L},\bar{C})$ given by equalities (15). The pdf $\bar{\mathcal{Q}}(h,l,c;\gamma)$ depends on the parameter γ. The pdf $\bar{\mathcal{Q}}(h,l,c;\gamma)$ is defined by

$$\bar{\mathcal{Q}}(h,l,c;\gamma)dhdldc = \Pr\{\bar{H}\in(h,h+dh),\bar{L}\in(l,l+dl),\bar{C}\in(c,c+dc)\}\ , \tag{31}$$

which expresses the probability that $(\bar{H},\bar{L},\bar{C})$ take specific values to within infinitesimal intervals. The Appendix gives the explicit expression of the pdf $\bar{\mathcal{Q}}(h,l,c;\gamma)$ for the special case of the Wiener process $v(\tau,\gamma)$ defined in (4). Let us consider first the canonical variance estimator

$$\hat{d} = \bar{R}^2\varphi(\Theta,\Phi;\gamma_0)\ . \tag{32}$$

The diagram of this estimator can be written as

$$\varphi(\theta,\phi;\gamma_0) = \frac{G(\theta,\phi;\gamma_0)}{\mathrm{E}[\bar{R}^2G(\Theta,\Phi;\gamma_0)|\gamma_0]}\ , \tag{33}$$

where the function $G(\theta,\phi;\gamma_0)$ will be defined below. The expectation term in the denominator of expression(33) is equal to

$$\mathrm{E}[\bar{R}^2G(\Theta,\Phi;\gamma_0)|\gamma_0] = \int_{-\pi/2}^{0}d\phi\int_{s(\phi)}^{c(\phi)}\cos\theta d\theta G(\theta,\phi;\gamma_0)g_2(\theta,\phi;\gamma_0)\ ,$$

where

$$g_n(\theta,\phi;\gamma) = \int_0^\infty \rho^{2+n}\bar{\mathcal{Q}}(\rho\cos\theta\cos\phi,\rho\cos\theta\sin\phi,\rho\sin\theta;\gamma)d\rho\ . \tag{34}$$

We stress the important property that the canonical OHLC variance estimator given by (32) with (33) is unbiased at $\gamma=\gamma_0$, since its expectation is

$$\mathrm{E}[\hat{d}|\gamma_0] = \frac{\mathrm{E}[\bar{R}^2G(\Theta,\Phi;\gamma_0)|\gamma_0]}{\mathrm{E}[\bar{R}^2G(\Theta,\Phi;\gamma_0)|\gamma_0]} = 1\ .$$

Thus, we look for the function $G(\Theta,\Phi;\gamma_0)$ that makes the unbiased canonical variance estimator (32) with (33) the most efficient for a given γ_0.

THEOREM 3.1 *The diagram of the unbiased most efficient homogeneous canonical variance estimator for a given γ_0 is equal to*

$$\varphi(\theta,\phi;\gamma_0)=\frac{1}{\mathcal{E}(\gamma_0)}\ \frac{g_2(\theta,\phi;\gamma_0)}{g_4(\theta,\phi;\gamma_0)}\ , \tag{35}$$

where $g_n(\theta,\phi;\gamma)$ is defined by expression (34) and

$$\mathcal{E}(\gamma)=\int_{-\pi/2}^{0}d\phi\int_{s(\phi)}^{c(\phi)}\cos\theta d\theta\ \frac{g_2^2(\theta,\phi;\gamma)}{g_4(\theta,\phi;\gamma)}\ . \tag{36}$$

Proof. The variance of the unbiased homogeneous canonical estimator (32) with (33) is equal to

$$\mathrm{Var}\left[\hat{d}|\gamma_0\right]=\frac{\displaystyle\int_{-\pi/2}^{0}d\phi\int_{s(\phi)}^{c(\phi)}\cos\theta d\theta\ G^2(\theta,\phi;\gamma_0)g_4(\theta,\phi;\gamma_0)}{\left(\displaystyle\int_{-\pi/2}^{0}d\phi\int_{s(\phi)}^{c(\phi)}\cos\theta d\theta\ G(\theta,\phi;\gamma_0)g_2(\theta,\phi;\gamma_0)\right)^2}-1\ . \tag{37}$$

We use the Schwarz inequality to determine the minimal value of the variance given by (37) of the canonical estimator. Omitting for the sake of conciseness the limits in the integrals, we represent the Schwarz inequality in the form

$$\left(\iint A(\theta,\phi)B(\theta,\phi)d\theta d\phi\right)^2\leqslant\iint A^2(\theta,\phi)d\theta d\phi\iint B^2(\theta,\phi)d\theta d\phi\ ,$$

where $A(\theta,\phi)$ and $B(\theta,\phi)$ are arbitrary real-valued functions. Taking here

$$A(\theta,\phi)=G(\theta,\phi;\gamma_0)\sqrt{g_4(\theta,\phi;\gamma_0)\cos\theta}\ ,$$

$$B(\theta,\phi)=g_2(\theta,\phi;\gamma_0)\sqrt{\frac{\cos\theta}{g_4(\theta,\phi;\gamma_0)}}\ ,$$

we obtain

$$\left(\iint G(\theta,\phi;\gamma_0)g_2(\theta,\phi;\gamma)\cos\theta d\theta d\phi\right)^2\leqslant$$

$$\iint G^2(\theta,\phi;\gamma_0)g_4(\theta,\phi;\gamma_0)\cos\theta d\theta d\phi\iint\frac{g_2^2(\theta,\phi;\gamma_0)}{g_4(\theta,\phi;\gamma_0)}\cos\theta d\theta d\phi\ .$$

It follows from (37) and from the last inequality that the variance of any canonical variance estimator satisfies the inequality

$$\text{(38)} \qquad \text{Var}\left[\hat{d}(\Theta,\Phi;\gamma_0)|\gamma_0\right] \geqslant V(\gamma_0)\ , \qquad V(\gamma) = \frac{1}{\mathcal{E}(\gamma)} - 1\ ,$$

where $\mathcal{E}(\gamma)$ is defined by expression (36). Taking into account (36), (37) and (38), the variance of the canonical variance estimator reaches its minimal value $V(\gamma_0)$ for the following choice of the function $G(\theta,\phi;\gamma_0)$:

$$G(\theta,\phi;\gamma_0) = \frac{g_2(\theta,\phi;\gamma_0)}{g_4(\theta,\phi;\gamma_0)}\ .$$

This corresponds to the diagram $\varphi(\theta,\phi;\gamma_0)$ given by expression (35). □

An analogous derivation provides the unbiased most efficient canonical volatility estimator, for a given γ_0. The main corresponding results are summarized in the following theorem.

THEOREM 3.2 *The diagram $\psi(\theta,\phi;\gamma_0)$ of the unbiased most efficient homogeneous canonical OHLC volatility estimator, defined by*

$$\text{(39)} \qquad \hat{s} = \bar{R}\psi(\Theta,\Phi;\gamma_0)\ ,$$

is equal to

$$\text{(40)} \qquad \begin{aligned} \psi(\theta,\phi;\gamma_0) &= \frac{1}{\mathcal{F}(\gamma_0)}\ \frac{g_1(\theta,\phi;\gamma_0)}{g_2(\theta,\phi;\gamma_0)}\ , \\ \mathcal{F}(\gamma) &= \int_{-\pi/2}^{0} d\phi \int_{s(\phi)}^{c(\phi)} \cos\theta d\theta\ \frac{g_1^2(\theta,\phi;\gamma)}{g_2(\theta,\phi;\gamma)}\ . \end{aligned}$$

The variance of the most efficient canonical OHLC volatility estimator is equal to

$$\text{(41)} \qquad W(\gamma_0) = \frac{1}{\mathcal{F}(\gamma_0)} - 1\ .$$

DEFINITION 3.1 $V(\gamma)$ defined in (38) is called the *lowest bound* of the variance of the canonical variance estimator, for a given value of the parameter γ. Analogously, $W(\gamma)$ given by (41) is called the *lowest bound* of the variance of the canonical volatility estimator, for the given value of the parameter γ.

4. PROPERTIES OF MOST EFFICIENT OHLC VARIANCE ESTIMATORS FOR THE WIENER PROCESS

The Appendix derives the explicit expression of the pdf $\bar{\mathcal{Q}}(h, l, c; \gamma)$ of the high, low and close values of the Wiener process $v(\tau, \gamma)$ defined in (4). This section uses this explicit knowledge to explore the quantitative properties of the most efficient canonical estimators for this particular case and compare them with those of the R&S and G&K canonical variance estimators.

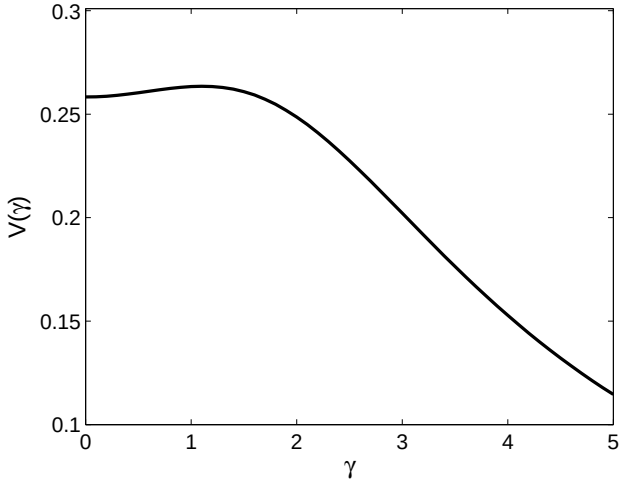

Fig. 1: Dependence of the lowest bound $V(\gamma)$ given by (38) of the variance of the homogeneous canonical variance estimator as a function of γ. $V(0) = 0.2583$.

4.1. *Variance of canonical variance estimators*

Let us first consider the lowest bound $V(\gamma)$ given by (38) of the variance of the homogeneous canonical variance estimator. For the Wiener process model, it is easy to calculate numerically the function $V(\gamma)$, which is represented in figure 1. The variance of the most efficient canonical variance estimator at $\gamma_0 = 0$ is equal to $V(0) \approx 0.258$, which can be compared with the corresponding variances for the G&K and R&S canonical variance estimators: $\mathrm{Var}[\hat{d}_{\mathrm{GK}}|0] \approx 0.27$, $\mathrm{Var}[\hat{d}_{\mathrm{RS}}|0] \approx 0.331$ (Rogers and Satchell, 1991).

Thus, at $\gamma = 0$, the G&K variance estimator has almost the same efficiency as the most efficient (for $\gamma_0 = 0$) homogeneous variance estimator, while the efficiency of the R&S estimator is significantly worse. These results are reflecting the closeness of the diagrams of the G&K and most efficient estimators, while the diagram of the R&S estimator drastically differs from the diagram of most efficient estimator, as shown in figure 2.

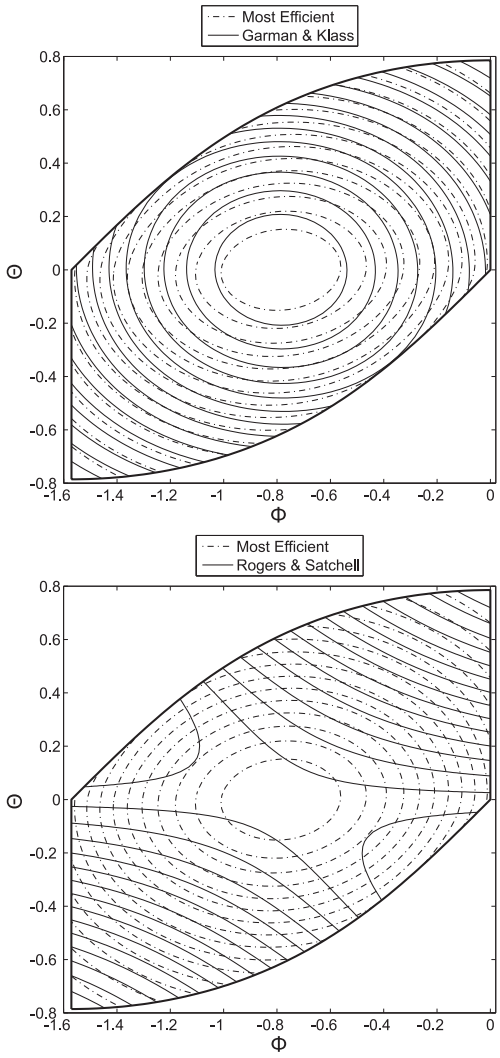

Fig. 2: Diagrams of the R&S, G&K and most efficient (for $\gamma_0 = 0$) variance estimators. See definition 2.7 for the meaning and construction of the diagrams.

4.2. *Bias and efficiency of the most efficient (γ_0) variance estimator*

The homogeneous variance estimator with diagram (35) is unbiased and most efficient only for a given value γ_0. In general, the value of γ is unknown. It is thus necessary to quantify the bias and efficiency of the homogeneous estimator for different values $\gamma \neq \gamma_0$, and compare it with the biases and efficiencies of the G&K and R&S variance estimators.

For this, we first determine the expected value and the variance of an arbitrary canonical variance estimator given by (32). Calculations similar to those performed in the previous section yield

$$\begin{aligned} &\mathrm{E}\left[\hat{d}(\Theta,\Phi)|\gamma\right] = \mathcal{K}_1(\gamma)\ , \quad \mathrm{Var}\left[\hat{d}(\Theta,\Phi)|\gamma\right] = \mathcal{K}_2(\gamma) - \mathcal{K}_1^2(\gamma)\ , \\ &\mathcal{K}_n(\gamma) = \int_{-\pi/2}^{0} d\phi \int_{s(\phi)}^{c(\phi)} \cos\theta d\theta\ g_{2n}(\theta,\phi;\gamma)\varphi^n(\theta,\phi)\ . \end{aligned} \tag{42}$$

Substituting the expression (35) for the diagram of the most efficient estimator into equation (42) yields

$$\begin{aligned} \mathrm{E}\left[\hat{d}(\Theta,\Phi;\gamma_0)|\gamma\right] &= \frac{\mathcal{E}(\gamma,\gamma_0)}{\mathcal{E}(\gamma_0)}\ , \\ \mathcal{E}(\gamma,\gamma_0) &= \int_{-\pi/2}^{0} d\phi \int_{s(\phi)}^{c(\phi)} \cos\theta d\theta\ \frac{g_2(\theta,\phi;\gamma)g_2(\theta,\phi;\gamma_0)}{g_4(\theta,\phi;\gamma_0)}\ . \end{aligned}$$

Figure 3 presents the dependence as a function of γ of the expected value of the most efficient canonical variance estimators given by (32) with (35). The expectations of the R&S and G&K canonical variance estimators, whose diagrams are given by (29), are also shown for comparison. While the R&S variance estimator is unbiased for all γ's, the most efficient estimators at γ_0 are unbiased only in the neighborhood of $\gamma = 0$ and of $\gamma = \gamma_0$. Comparing the G&K and the most efficient estimators, the homogeneous estimator, which is the most efficient for $\gamma_0 = 1$ for instance, is not significantly biased over the whole range $0 \leqslant \gamma \lesssim 1.5$ and remains much less biased than the G&K estimator over the range $0 \leqslant \gamma \leqslant 2$.

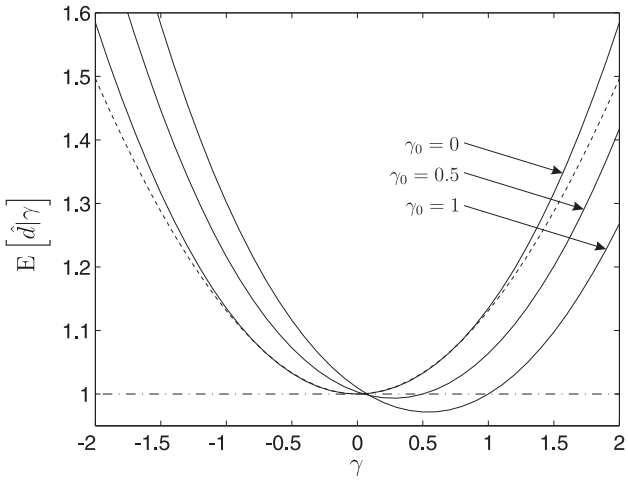

Fig. 3: Dependence as a function of γ of the expected values of the R&S (dash-dot line) and G&K (dashed line) canonical variance estimators and of the most efficient variance estimators for $\gamma_0 = 0; 0.5; 1$ (solid lines, top-down)

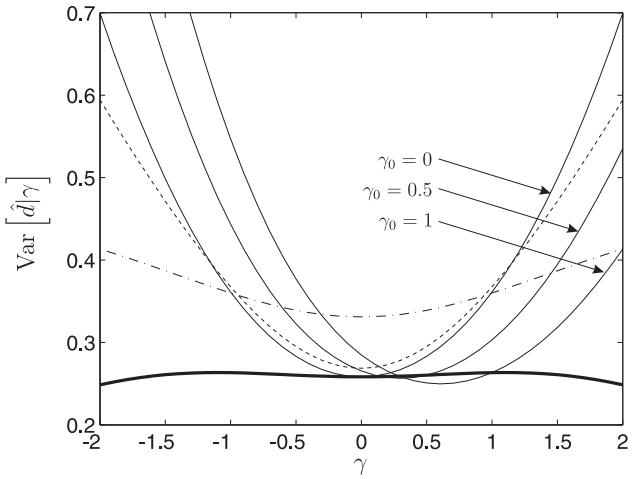

Fig. 4: Dependence as a function of γ of the variances of the R&S (dash-dot line), G&K (dashed line) and most efficient variance estimators for $\gamma_0 = 0; 0.5; 1$ (solid lines). The heavy solid line is the lowest bound variance given by (38).

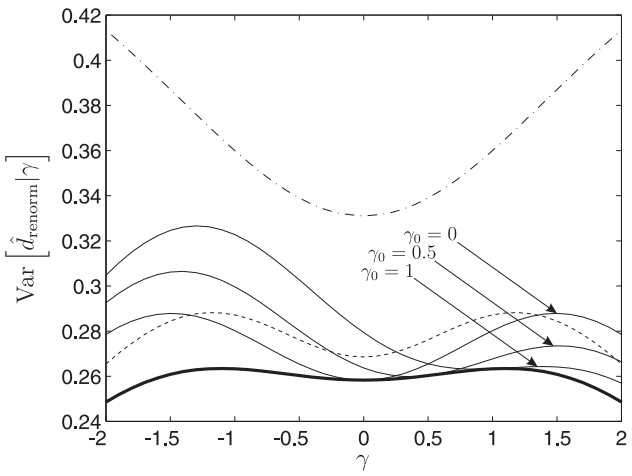

Fig. 5: Dependence as a function of γ of the variances of the *renormalized* R&S (dash-dot line) and G&K (dashed line) canonical variance estimators, and the most efficient estimators (solid lines), as defined by expression (43).

Calculation of the variance (for any γ) of the canonical variance estimator, which is most efficient at γ_0, gives

$$\mathrm{Var}\left[\hat{d}(\Theta,\Phi;\gamma_0)|\gamma\right] = \frac{\mathcal{M}(\gamma,\gamma_0) - \mathcal{E}^2(\gamma,\gamma_0)}{\mathcal{E}^2(\gamma_0)} \ ,$$

$$\mathcal{M}(\gamma,\gamma_0) = \int_{-\pi/2}^{0} d\Phi \int_{s(\phi)}^{c(\phi)} \cos\theta d\theta \, \frac{g_4(\theta,\phi;\gamma) g_2^2(\theta,\phi;\gamma_0)}{g_4^2(\theta,\phi;\gamma_0)} \ .$$

Figure 4 shows the dependence as a function of γ of the variances of the R&S and G&K canonical variance estimators and of the most efficient homogeneous variance estimators for different γ_0. One can observe that the homogeneous variance estimator, which is the most efficient at $\gamma_0 = 1$, is both less biased and significantly more efficient than the G&K estimator over the interval $0 \lesssim \gamma \lesssim 2$.

One should not be surprised to observe in figure 4 several intervals along the γ axis in which the variances of the estimators are smaller than the lower bound $V(\gamma)$ given by (38). Indeed, the lower bound for the variance given by (38) is suitable only for unbiased estimators. Therefore, the "strange"

behavior of the variance plots does not mean that these estimators are more efficient than the most efficient estimator at the given γ, but rather that they are biased at this point. With the proper renormalization

$$\hat{d}_{\rm renorm} = \hat{d} \Big/ \mathrm{E}\left[\hat{d}|\gamma\right] \ , \tag{43}$$

one can see that, for any γ values, the estimators have variances which are indeed bounded from below by the lower bound $V(\gamma)$, as shown in figure 5.

4.3. *Probabilistic properties of homogeneous estimators*

Knowing the exact explicit expression of the pdf $\bar{\mathcal{Q}}(h,l,c;\gamma)$ of the high, low and close values of the Wiener process $v(\tau,\gamma)$ defined in (4) given in the Appendix, we can go beyond the calculations of the expectations and variances of the estimators described in previous subsections and derive their full distribution. In particular, the knowledge of the full distribution of the estimators allows one to determine the confidence intervals of the quasi-unbiased estimators introduced in section 4.4.

Let us consider the pdf of the canonical variance estimator (27). For a given γ, it is defined by the following expression

$$f(u;\gamma) = \mathrm{E}\left[\delta(u - \bar{R}^2\varphi(\Theta,\Phi))|\gamma\right] \ .$$

Using the standard properties of the delta-function of a composite argument, we can rewrite the previous definition (4.3) in the form

$$f(u;\gamma) = \mathrm{E}\left[\frac{1}{\sqrt{u\varphi(\Theta,\Phi)}}\delta\left(\bar{R} - \sqrt{\frac{u}{\varphi(\Theta,\Phi)}}\right)\Big|\gamma\right] \ ,$$

or more explicitly

$$f(u;\gamma) = \frac{\sqrt{u}}{2}\int_{-\pi/2}^{0} d\phi \int_{s(\phi)}^{c(\phi)} \frac{\cos\theta d\theta}{\varphi^{3/2}(\theta,\phi)} \times \tag{44}$$
$$\bar{\mathcal{Q}}\left(\sqrt{\frac{u}{\varphi(\theta,\phi)}}\cos\theta\cos\phi, \sqrt{\frac{u}{\varphi(\theta,\phi)}}\cos\theta\sin\phi, \sqrt{\frac{u}{\varphi(\theta,\phi)}}\sin\theta;\gamma\right) \ .$$

We use expression (44) to obtain, by numerical integration, the pdf's of R&S, G&K and of the most efficient ($\gamma_0 = 0$) canonical variance estimators, calculated for $\gamma = 0$. These three pdf's are represented in figure 6.

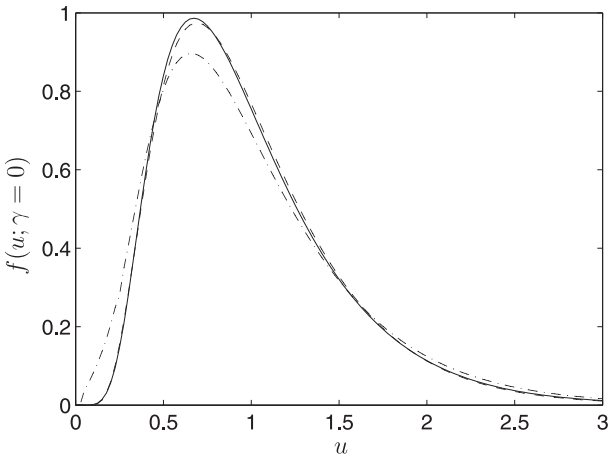

Fig. 6: Pdfs of the R&S (dash-dot line), G&K (dashed line) and of the most efficient ($\gamma_0 = 0$) canonical variance estimators (solid line), at $\gamma = 0$.

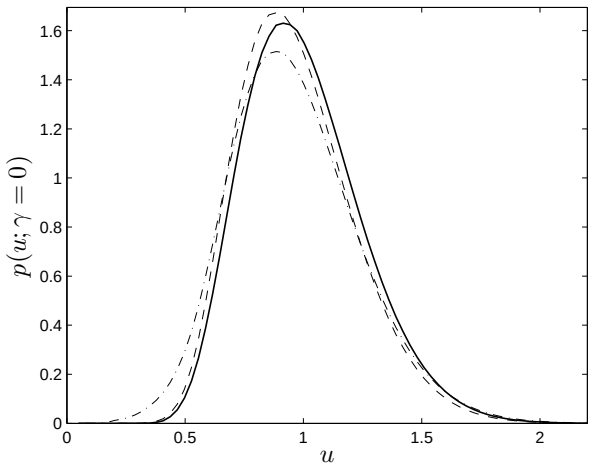

Fig. 7: Pdfs of the R&S (dash-dot line), G&K (dashed line) and of the most efficient ($\gamma_0 = 0$) canonical volatility estimators (solid line) at $\gamma = 0$.

Similarly, the pdf of the canonical volatility estimator is defined by

$$p(u;\gamma) = \mathrm{E}[\delta(u - \bar{R}\psi(\Theta,\Phi))|\gamma] \ ,$$

and its explicit expression, analogous to (44) formula, reads

$$p(u;\gamma) =$$

$$(45) \qquad u^2 \int_{-\pi/2}^{0} d\phi \int_{s(\phi)}^{c(\phi)} \frac{\cos\theta d\theta}{\psi^3(\theta,\phi)} \bar{\mathcal{Q}}\left(\frac{u\cos\theta\cos\phi}{\psi(\theta,\phi)}, \frac{u\cos\theta\sin\phi}{\psi(\theta,\phi)}, \frac{u\sin\theta}{\psi(\theta,\phi)}; \gamma\right) \ .$$

Figure 7 shows the pdf's given by (45) of the R&S, G&K and of the most efficient ($\gamma_0 = 0$) canonical volatility estimators, for $\gamma = 0$.

4.4. *Quasi-unbiased quasi-optimal estimators*

The previous subsections have made it clear that the most efficient unbiased (γ_0) estimators are not the most efficient for $\gamma \neq \gamma_0$, nor are they unbiased. Since varying γ_0 corresponds to scanning these most efficient estimators, which remain efficient in a neighborhood of their γ_0, this suggests to introduce new reasonably efficient and approximately unbiased estimators, obtained as linear superpositions of the most efficient canonical homogeneous variance estimators:

$$(46) \qquad \hat{d}(\Theta,\Phi) = \bar{R}^2 \int_{-\infty}^{\infty} \frac{h(\gamma_0)}{\mathcal{E}(\gamma_0)} \frac{g_2(\Theta,\Phi;\gamma_0)}{g_4(\Theta,\Phi;\gamma_0)} d\gamma_0 \ .$$

Here, $h(\gamma_0)$ is some weighting function, whose explicit expression must be determined from some optimization criterion. A possible requirement is that $h(\gamma_0)$ be such as to both minimize the bias of the estimator (46) and maximize its efficiency within some given γ interval, according to some criterion. To demonstrate the principle of this approach, we search for the function $h_0(\gamma_0)$ that ensures that the estimator (46) is unbiased. The corresponding condition is that the expected value of the composed estimator (46) given by

$$\mathrm{E}\left[\hat{d}(\Theta,\Phi)|\gamma\right] = \int_{-\infty}^{\infty} \frac{h(\gamma_0)\mathcal{E}(\gamma,\gamma_0)}{\mathcal{E}(\gamma_0)} d\gamma_0 \ ,$$

be equal to 1. Condition $\mathrm{E}\left[\hat{d}(\Theta,\Phi)|\gamma\right]=1$ then provides an integral equation for the function $h_0(\gamma_0)$. In practice, it is more convenient to look for *quasi-unbiased* estimators, which are exactly unbiased at $2K+1$ values of the parameter γ, for instance at

$$\gamma_i = i\,\frac{\Gamma}{K}\ , \qquad i=-K,-K+1,\cdots-1,0,1,\ldots,K-1,K\ . \tag{47}$$

Given these $2K+2$ constraints, it is natural to search for a solution constructed as the sum of $2K+1$ most efficient (γ_i) canonical variance estimators:

$$\hat{d}(\Theta,\Phi)=\bar{R}^2\sum_{i=-K}^{K} h_i\varphi_i(\Theta,\Phi)\ , \quad \varphi_i(\theta,\phi)=\frac{1}{\mathcal{E}(\gamma_i)}\,\frac{g_2(\theta,\phi;\gamma_i)}{g_4(\theta,\phi;\gamma_i)}\ . \tag{48}$$

The $2K+1$ unknown coefficients $\{h_i, i=-K,...,+K\}$ are to be determined from the $2K+2$ constraints of an absence of bias at the discrete γ values (47). We refer to Γ as the *band width* of the quasi-unbiased estimator (48), while K is its *order*.

In particular, the quasi-unbiased estimator of zero order corresponds to the previously studied most efficient $(\gamma_0=0)$ canonical variance estimator. The first order quasi-unbiased estimator is equal to

$$\hat{d}(\Theta,\Phi)=\bar{R}^2\left[h_{-1}\varphi_{-1}(\Theta,\Phi)+h_0\varphi_0(\Theta,\Phi)+h_1\varphi_1(\Theta,\Phi)\right]\ , \tag{49}$$

and so on.

The expected value of the quasi-unbiased estimator (48) is equal to

$$\mathrm{E}\left[\hat{d}(\Theta,\Phi)|\gamma\right]=\sum_{i=-K}^{K} h_i\frac{\mathcal{E}(\gamma,\gamma_i)}{\mathcal{E}(\gamma_i)}\ . \tag{50}$$

Equating expression (50) to 1 for the $2K+1$ values (47) yields the set of $2K+1$ linear equations:

$$\mathrm{E}\left[\hat{d}(\Theta,\Phi)|\gamma_j\right]=1 \quad\Rightarrow\quad \sum_{i=-K}^{K}\varepsilon_{i,j}\,h_i=1\ , \quad \varepsilon_{i,j}=\frac{\mathcal{E}(\gamma_j,\gamma_i)}{\mathcal{E}(\gamma_i)}\ . \tag{51}$$

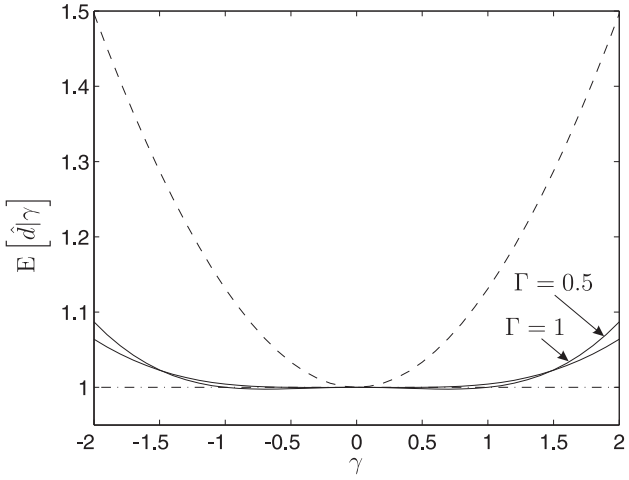

Fig. 8: Dependence as a function of γ of the expected values of the R&S (dash-dot line), G&K (dashed line) and of the quasi-unbiased first-order variance estimators for the band widths $\Gamma = 0.5; 1$ (solid lines).

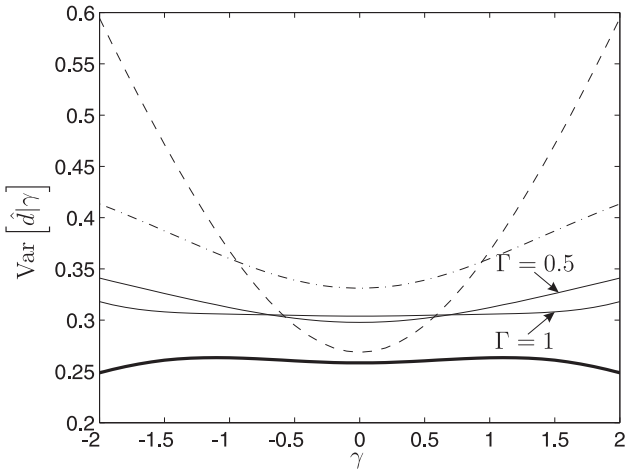

Fig. 9: Dependence as a function of γ of the variances of the R&S (dash-dot line), G&K (dashed line) and of the quasi-unbiased first-order variance estimators for the band widths $\Gamma = 0.5; 1$ (solid lines).

The statistical symmetry of the Wiener process (4) implies that the solution of equations (51) satisfies the following symmetry conditions: $\varepsilon_{i,j} = \varepsilon_{-i,-j}$

$\Rightarrow h_i = h_{-i}$.

Exploiting this symmetry for the first order case $K = 1$ yields to two equations for $h_1 = h_{-1}$ and h_0:

$$h_0\varepsilon_{0,1} + h_1(\varepsilon_{-1,1} + \varepsilon_{1,1}) = 1 \ , \quad h_0\varepsilon_{0,0} + 2h_1\varepsilon_{1,0} = 1 \ ,$$

whose solution reads

$$h_0 = \frac{2\varepsilon_{1,0} - \varepsilon_{-1,1} - \varepsilon_{1,1}}{2\varepsilon_{0,1}\varepsilon_{1,0} - \varepsilon_{0,0}(\varepsilon_{-1,1} + \varepsilon_{1,1})} \ , \quad h_{\pm 1} = \frac{\varepsilon_{0,0} - \varepsilon_{0,1}}{\varepsilon_{0,0}(\varepsilon_{-1,1} + \varepsilon_{1,1}) - 2\varepsilon_{0,1}\varepsilon_{1,0}} \ .$$

Figure 8 shows the dependence as a function of γ of the expected values of the first-order quasi-unbiased canonical variance estimators for band widths $\Gamma = 0.5; 1$. Figure 9 presents the dependence as a function of γ of the variances of these estimators. For comparison, the expected values and variances of the R&S and G&K estimators are also shown. We can state that the quasi-unbiased canonical variance estimators constructed here provide the best of both world: (i) they exhibit a very weak bias up to rather large values of γ, thus competing reasonably well with the R&S estimator; (ii) their variance is very weakly dependent on γ and significantly smaller than that of the R&S estimator for all γ's and than that of the G&K estimator, except for a central zone around $\gamma = 0$.

5. TESTS OF THEORETICAL RESULTS OF VARIANCE AND VOLATILITY ESTIMATORS USING SYNTHETIC TIME SERIES OF THE WIENER PROCESS

The present section implements the variance and volatility estimators discussed above for synthetic time series of the Wiener process (3). Because our results are mathematically exact, these tests on synthetic time series offer the opportunity to study the impact of finite size and discreteness effects, and give the opportunity to study additional properties of the estimators. We will also determine the Maximum Likelihood estimator for the variance and the volatility and will compare them to the other estimators. The homogeneity of the estimators under study allow us to restrict σ to the value

1 and to construct time series on the unit time interval ($T = 1$), without losing generality. With these parameter values, we have $\mu = \gamma$, and $X(t)$ are replaced by $v(\tau, \gamma)$ given by (4).

5.1. *Test on numerical convergence of the discrete to the continuous Wiener process*

It is interesting to illustrate and test the theoretical results of previous sections by numerical simulation of the Wiener process $X(t)$ given by (3). Numerical simulations require replacing the continuous time stochastic process $v(\tau, \gamma)$ given by (4) by its discrete counterpart $v(n, \gamma)$ given by (6), where $\{\epsilon_k\}$ are Gaussian.

The Gaussian discrete process $v(n, \gamma)$ represents rather accurately the continuous time process $v(\tau, \gamma)$ only for sufficiently large N. On the other hand, the discrete process (6) obtained for not too large N might describe the stochastic behavior of some financial markets more adequately than the continuous time process $v(\tau, \gamma)$. From a practitioner point of view, N could be interpreted as the typical number of transactions within the time interval of interest. From a theoretical point of view, N should be chosen large enough to simulate the variables $\bar{H}$, $\bar{L}$ and $\bar{C}$ defined by (15), which are known to be distributed according to the analytically derived pdf (A.16) with (A.17). To determine the appropriate value for N, we repeated $M = 10^6$ simulations of the discrete process $v(n, \gamma)$ (6), and calculated for each of these M samples the corresponding G&K and R&S variance estimators at $\gamma = 0$. Averaging over the M realizations, we found the dependence of the expected value of the G&K and R&S variance estimators as a function of N, which is shown in figure 10. In particular, the statistical average value of the canonical R&S estimator, for $N = 10^6$, is found to be $\mathrm{E}\left[\hat{d}_{\mathrm{RS}} \middle| \gamma = 0\right] = 0.9987$, which is close enough to the theoretical one ($\mathrm{E}\left[\hat{d}_{\mathrm{RS}} \middle| \gamma = 0\right] = 1$).

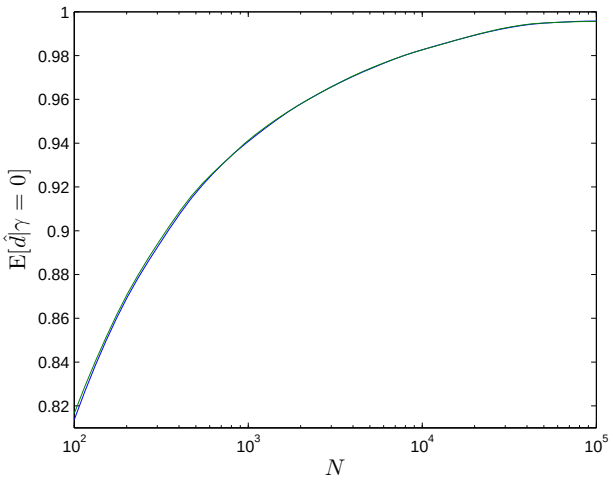

Fig. 10: Dependence as a function of N of the statistical average value of the G&K and R&S variance estimators for $\gamma = 0$, where the statistical average is performed over $M = 10^6$ realizations of the discrete time Wiener process $v(\tau, \gamma)$ given by (4). Note that the two curves are almost undistinguishable, but not exactly the same.

5.2. *Variance estimators*

Figures 11 and 12 show the expected values and variances of the G&K, R&S and of the most efficient variance estimators, obtained theoretically and by numerical simulations with $M = 10^5$ realizations of $v(n, \gamma)$, each of length $N = 10^6$. One can observe an excellent agreement between the simulations and the theory.

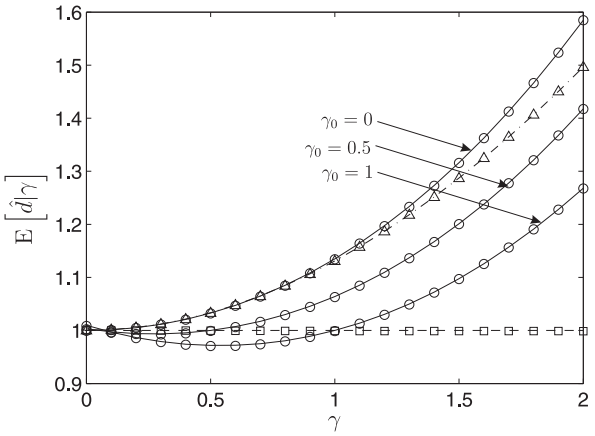

Fig. 11: Dependence of the expected values of the R&S (squares and dashed line), G&K (triangles and dash-dot line) and of the most efficient (at $\gamma_0 = 0; 0.5; 1$) (circles and solid lines) variance estimators as a function of γ. The markers show the values obtained by numerical simulations described in the text; the continuous lines correspond to the theoretical results presented in sections 3 and 4.

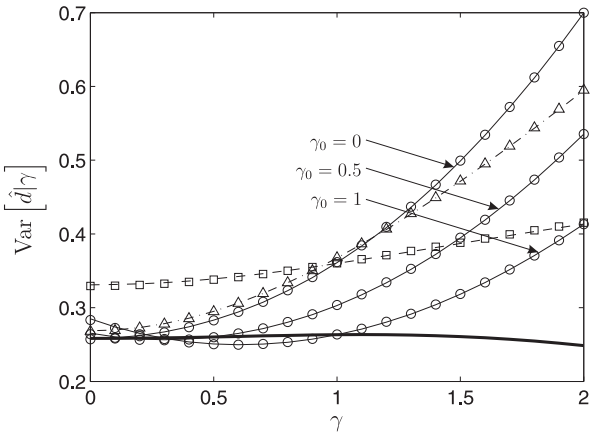

Fig. 12: Dependence of the variances of the R&S (squares and dashed line), G&K (triangles and dash-dot line) and of the most efficient estimators (at $\gamma_0 = 0; 0.5; 1$) (circles and solid lines) of the variance estimators as a function of γ. The markers show the values obtained by numerical simulations described in the text; the continuous lines correspond to the theoretical results presented in sections 3 and 4.

5.3. *Volatility estimators*

We now compare the efficiency and bias of the R&S, G&K and of the most efficient canonical volatility estimators. Recall that, while the R&S canonical variance estimator (14) is unbiased for all γ's, the R&S canonical volatility estimator (16) is biased for all γ's. The same holds true for the G&K volatility estimator, which is biased even for $\gamma = 0$. Figure 13 shows the dependence of the expected values of these estimators as a function of γ. In particular, the G&K and R&S volatility estimators have the following biases at $\gamma = 0$:

$$1 - \mathrm{E}[\hat{s}_{\mathrm{GK}}|\gamma = 0] = 0.0309 \ , \qquad 1 - \mathrm{E}[\hat{s}_{\mathrm{RS}}|\gamma = 0] = 0.0386 \ .$$

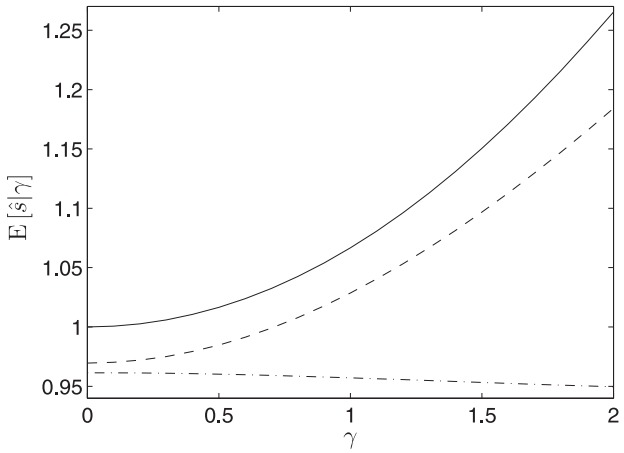

Fig. 13: Dependence as a function of γ of the expected values of R&S (dash-dot), G&K (dashed) and of the most efficient at $\gamma_0 = 0$ volatility estimator (solid line).

In order to provide an appropriate comparison between the efficiency of the R&S, G&K and of the most efficient volatility estimators, we normalize them by their values reached at $\gamma = 0$:

$$\hat{s}_{\mathrm{norm}}(\Theta, \Phi) = \bar{R}\psi(\Theta, \Phi) \Big/ \mathrm{E}[\hat{s}|\gamma = 0] \ . \tag{52}$$

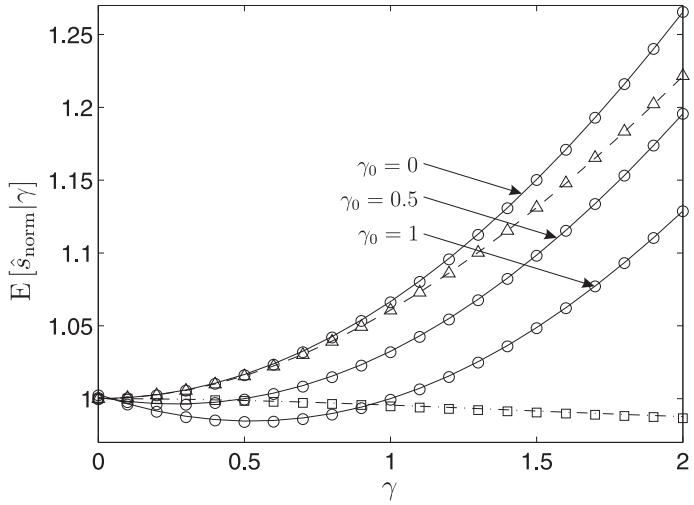

Fig. 14: Expected values of the normalized (according to (52)) R&S, G&K and of the most efficient unbiased homogeneous canonical volatility estimators at $\gamma_0 = 0; 0.5; 1$.

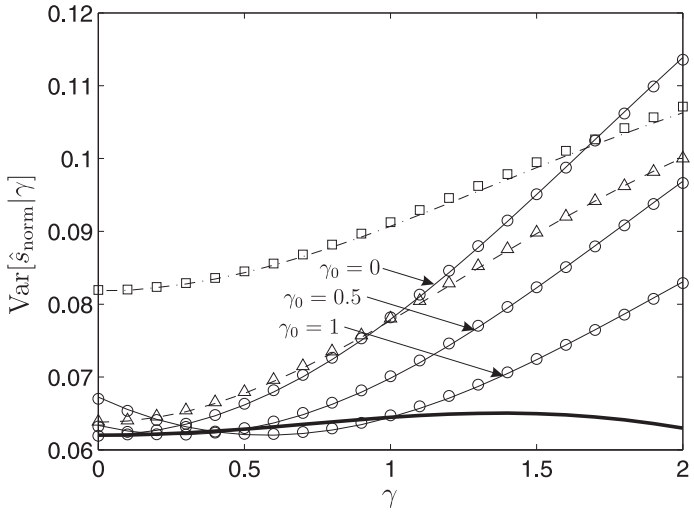

Fig. 15: Variances of the normalized (according to (52)) R&S, G&K and of the most efficient unbiased homogeneous canonical volatility estimators at $\gamma_0 = 0; 0.5; 1$. The heavy solid line corresponds to the lowest bound variance $W(\gamma)$ given by (41).

Figures 14 and 15 show the expected values and variances of the normalized (according to (52)) R&S, G&K and of the most efficient unbiased homogeneous canonical volatility estimators at $\gamma_0 = 0; 0.5; 1$. In particular, the variances at $\gamma = 0$ of the normalized G&K, R&S and of the most efficient (at $\gamma_0 = 0$) volatility estimators are equal to

$$\begin{gathered} \mathrm{Var}[\hat{s}_{\mathrm{GK}}|\gamma = 0] = 0.06379 \ , \quad \mathrm{Var}[\hat{s}_{\mathrm{RS}}|\gamma = 0] = 0.08186 \ , \\ \mathrm{Var}[\hat{s}(\gamma_0 = 0)|\gamma = 0] = 0.06201 \ . \end{gathered} \tag{53}$$

The theoretical results shown in figures 14 and 15 are also compared with the numerical calculations performed using $M = 10^5$ different realizations of the discrete Wiener process (6) with length $N = 10^6$..

5.4. *Maximum likelihood estimators*

The Appendix derives the exact expression for the joint distribution of the H, L, C of a Wiener process. Being a function of the volatility σ, this joint distribution allows us to obtain the maximum likelihood (ML) estimator of σ, as we now describe. It turns out that the MLE is less efficient than the most efficient homogenous estimators described above.

Let us start from $\bar{\mathcal{Q}}(h, l, c; \gamma)$ given by (A.16) in the Appendix, which is the pdf of the high, low and close values $(\bar{H}, \bar{L}, \bar{C})$ defined by (15) of the Wiener process $v(\tau, \gamma)$ (4) with unit volatility. Knowing $\bar{\mathcal{Q}}(h, l, c; \gamma)$, one can recover the pdf $\mathcal{Q}(\eta, \lambda, \xi; \mu, \sigma)$ of the high, low and close values (H, L, C) defined by (7) of the original Wiener process $X(t)$ (3) for $t \in (0, T)$, by using the relation

$$\begin{gathered} \mathcal{Q}(\eta, \lambda, \xi; \mu, \sigma) = \frac{1}{\sigma^3 T\sqrt{T}} \bar{\mathcal{Q}}\left(\frac{\eta}{\sigma\sqrt{T}}, \frac{\lambda}{\sigma\sqrt{T}}, \frac{\xi}{\sigma\sqrt{T}}; \frac{\mu}{\sigma}\sqrt{T}\right) = \\ \frac{1}{\sigma^3 T\sqrt{2\pi T}} \exp\left(-\frac{(\xi - \mu T)^2}{2\sigma^2 T}\right) \mathcal{R}\left(\frac{\eta}{\sigma\sqrt{T}}, \frac{\lambda}{\sigma\sqrt{T}}\middle| \frac{\xi}{\sigma\sqrt{T}}\right) \ . \end{gathered} \tag{54}$$

This expression for the pdf of (H, L, C) allows us to construct the maximum likelihood OHLC estimators $\hat{\mu}_{\mathrm{ML}}$ and $\hat{\sigma}_{\mathrm{ML}}$ of the drift and the volatility of

the Wiener process $X(t)$ defined by (3). The MLE are obtained by replacing the arguments (η, λ, ξ) in (54) by the realized samples (H, L, C), and by searching for the values $\hat{\mu}_{\rm ML}$ and $\hat{\sigma}_{\rm ML}$ that maximize the likelihood function

$$\mathcal{L}(H, L, C; \mu, \sigma) = \ln \mathcal{Q}(H, L, C; \hat{\mu}, \hat{\sigma}) =$$

$$-\frac{(C - \hat{\mu}_{\rm ML} T)^2}{2\hat{\sigma}^2_{\rm ML} T} + \ln \mathcal{R}\left(\frac{H}{\hat{\sigma}_{\rm ML}\sqrt{T}}, \frac{L}{\hat{\sigma}_{\rm ML}\sqrt{T}} \middle| \frac{C}{\hat{\sigma}_{\rm ML}\sqrt{T}}\right) - 3 \ln \hat{\sigma}_{\rm ML} \ .$$

We obtain the ML drift estimator,

$$\hat{\mu}_{\rm ML} = \frac{C}{T} \ . \tag{55}$$

We recall that this drift estimator (55) has the minimal possible variance among all estimators, since it realizes the lower bound given by the Cramer-Rao inequality.

The ML volatility estimator $\hat{\sigma}_{\rm ML}$ maximizes the function

$$\ln \mathcal{R}\left(\frac{H}{\hat{\sigma}_{\rm ML}\sqrt{T}}, \frac{L}{\hat{\sigma}_{\rm ML}\sqrt{T}} \middle| \frac{C}{\hat{\sigma}_{\rm ML}\sqrt{T}}\right) - 3 \ln \hat{\sigma}_{\rm ML} \ . \tag{56}$$

The following theorem then derives.

THEOREM 5.1 *The ML volatility estimator $\hat{\sigma}_{ML}$ is homogeneous, i.e., analogously to* (28)*, it can written in the form*

$$\hat{\sigma}_{ML} = \sigma \hat{s}_{ML}(\bar{H}, \bar{L}, \bar{C}) \ ,$$

where $\hat{s}_{ML}(h, l, c)$ is a first order homogeneous function.

Proof. Replacing $\hat{\sigma}_{\rm ML}$ by $\hat{\sigma}_{\rm ML} = \sigma \hat{s}_{\rm ML}$ in expression (56), using the equalities

$$\bar{H} = \frac{H}{\sigma\sqrt{T}} \ , \quad \bar{L} = \frac{L}{\sigma\sqrt{T}} \ , \quad \bar{C} = \frac{C}{\sigma\sqrt{T}} \ ,$$

and omitting the nonessential constant $3 \ln \sigma$, we obtain that $\hat{s}_{\rm ML}$ should maximize the function

$$\mathcal{N}(\bar{H}, \bar{L}, \bar{C}, \hat{s}_{\rm ML}) = \ln \mathcal{R}\left(\frac{\bar{H}}{\hat{s}_{\rm ML}}, \frac{\bar{L}}{\hat{s}_{\rm ML}} \middle| \frac{\bar{C}}{\hat{s}_{\rm ML}}\right) - 3 \ln \hat{s}_{\rm ML} \ . \tag{57}$$

Here, $\mathcal{R}(h,l|c)$ is a deterministic function given by (A.17). Accordingly, the value $\hat{s}_{\rm ML}$, which maximizes the function $\mathcal{N}(\bar{H},\bar{L},\bar{C},\hat{s}_{\rm ML})$, is a deterministic function $\hat{s}_{\rm ML}=\hat{s}_{\rm ML}(\bar{H},\bar{L},\bar{C})$ of the variables $(\bar{H},\bar{L},\bar{C})$. Its homogeneity is obvious. □

REMARK 5.1 From general properties of maximum likelihood estimators, the ML variance estimator is also homogeneous and it is equal to the square of the volatility estimator:

$$\hat{D}=\sigma^2\hat{d}_{\rm ML}(\bar{H},\bar{L},\bar{C})\ , \qquad \hat{d}_{\rm ML}(\bar{H},\bar{L},\bar{C})=\hat{s}^2_{\rm ML}(\bar{H},\bar{L},\bar{C})\ . \tag{58}$$

In general, ML estimators are biased. It is therefore convenient to normalize it by its value as some given $\gamma=\gamma_0$ to obtain

$$\hat{s}_{\rm norm}=\frac{\hat{s}_{\rm ML}(\bar{H},\bar{L},\bar{C})}{{\rm E}[\hat{s}_{\rm ML}(\bar{H},\bar{L},\bar{C})|\gamma_0]}\ .$$

Since ML estimators are homogeneous, they may not be more efficient than the most efficient estimators at the same γ_0 value. In practice, unbiased ML estimators are significantly less efficient than the most efficient one. Let illustrate this fact using the normalized ML volatility estimator at $\gamma_0=0$. For this case, the numerical calculation with ($N=10^6$, $M=10^6$) of the expected value and variance, at $\gamma=0$, of the canonical ML estimator yields

$$ {\rm E}[\hat{s}_{\rm ML}|0]\approx 0.9202\ , \quad {\rm Var}[\hat{s}_{\rm ML}|0]\approx 0.0712 \quad \Rightarrow \quad {\rm Var}[\hat{s}_{\rm norm}|0]\approx 0.0840\ . \tag{59}$$

Comparing these values with those reported in (53), one can see that the efficiency of the ML volatility estimator is significantly worse than for the most efficient one, and even worse than that of the R&S volatility estimator. The corresponding values for the ML canonical variance estimator are

(60)

$$\mathrm{E}[\hat{d}_{\mathrm{ML}}|0] \approx 0.9179\ , \quad \mathrm{Var}[\hat{d}_{\mathrm{ML}}|0] \approx 0.2756 \quad \Rightarrow \quad \mathrm{Var}[\hat{d}_{\mathrm{norm}}|0] \approx 0.3271\ .$$

While smaller than the variance of the R&S canonical variance estimator ($\mathrm{Var}_{\mathrm{RS}}[\hat{d}|0] \approx 0.331$), the variance $\mathrm{Var}[\hat{d}_{\mathrm{norm}}|0]$ is 27% larger than the variance of the most efficient one ($V(0) \approx 0.258$).

6. CONCLUSIONS

We have laid the first stones for a comprehensive theory of homogeneous volatility (and variance) estimators of arbitrary stochastic processes. Our focus has been to exploit the universally quoted OHLC (open-high-low-close) prices, which can span time intervals extending from seconds to years, in order to develop new efficient estimators. Our theory opens many possibilities to design new efficient estimators, such as the "quasi-unbiased estimators", that address any type of desirable constraints. The main tool of our theory is the parsimonious encoding of all the information contained in the OHLC in the form of general "diagrams" associated with the joint distributions of the high minus open, low minus open and close minus open values. The diagrams can be tailored to yield the most efficient estimators associated to any statistical properties of the underlying log-price stochastic process.

Our theory opens several interesting developments. First, the accurate determination of the key functions $g_n(\theta, \phi; \gamma)$, defining the above diagrams, gives the tools to develop efficient estimators of the variance and volatility (as well as any other quantities of interest) for arbitrary non-Gaussian log-price processes, including the presence of micro-structure as in tick-by-tick price series. Our methods should lead to the development of fast and effective algorithms for low- and high-frequency OHLC variance and volatility estimators, that can be applied in practice to any kind of financial markets.

APPENDIX A: EXTREMES OF WIENER PROCESSES

In the main text, we lay out the basic stones for a comprehensive theory of homogenous OHLC volatility and variance estimators, which are most efficient for any specific value of the normalized drift parameter γ of the underlying price stochastic process. This theory uses the OHLC (open-high-low-close) prices in the given time interval or scale of interest.

All expressions depend on a fundamental quantity, which is the joint probability density function (pdf) $\bar{\mathcal{Q}}(h,l,c;\gamma)$ defined by (31) of the high, low and close values given by (15) of the auxiliary stochastic process $B(t,\gamma)$ (2). In general, it is only possible to construct the sought pdf $\bar{\mathcal{Q}}(h,l,c;\gamma)$ by numerical simulations generating a huge number of realizations of the underlying stochastic process $B(t,\gamma)$. For certain stochastic process $X(t)$ (1), the pdf $\bar{\mathcal{Q}}(h,l,c;\gamma)$ can be calculated analytically. In this Appendix, we obtain the explicit analytical expression for $\bar{\mathcal{Q}}(h,l,c;\gamma)$ in the case of the Wiener process, $B(t,\gamma) \equiv v(\tau,\gamma)$ given by expression (4).

As shown below, the sought pdf $\bar{\mathcal{Q}}(h,l,c;\gamma)$ will be derived from the solution of the diffusion equation

$$\text{(A.1)} \quad \frac{\partial f(c;\tau,\gamma)}{\partial \tau} + \gamma \frac{\partial f(c;\tau,\gamma)}{\partial c} = \frac{1}{2}\frac{\partial^2 f(c;\tau,\gamma)}{\partial c^2} \ ,$$

where the reduced time τ and parameter γ are defined in (5). The well-known solution of the diffusion equation (A.1), satisfying the initial condition

$$\text{(A.2)} \quad f(c;\tau=0,\gamma) = \delta(c) \ ,$$

is

$$\text{(A.3)} \quad f(c;\tau,\gamma) = g(c-\gamma\tau,\tau) \ , \qquad g(x,\tau) = \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{x^2}{2\tau}\right) \ .$$

A.1. *Distribution of the maximal value*

The full derivation of the pdf $\bar{\mathcal{Q}}(h,l,c;\gamma)$ for the Wiener process $v(\tau,\gamma)$ (4) involves rather extensive calculations. In order to present the intuition behind these calculations, it is useful to consider the reduced problem of determining the joint pdf of the high (maximum) and close values of the Wiener process $v(\tau';\gamma)$ within a given time interval $\tau' \in (0,\tau)$. This reduced problem is tightly connected with the so-called "absorption" of the process $v(\tau;\gamma)$ at the given level h. The existence of absorption amounts to supplement the diffusion equation (A.1) by the absorption condition

$$f(c=h;\tau,\gamma)=0\ , \qquad h>0\ . \tag{A.4}$$

We denote the solution of the initial-boundary value problem (A.1), (A.2), (A.4) by $f(c,h;\tau,\gamma)$. This function is the pdf of the values, at time τ, of the realizations of the stochastic process $v(\tau';\gamma)$, that has not reached the level h for all times $\tau' \in (0,\tau)$, i.e.,

$$f(c,h;\tau,\gamma)dx=\Pr\{v(\tau;\gamma)\in(x,x+dx)\cap\bar{H}<h\}\ , \quad x<h\ , h>0\ , \tag{A.5}$$

where

$$\bar{H}=\sup_{\tau'\in(0,\tau)} v(\tau',\gamma)\ .$$

Correspondingly, expression (A.5) implies that the joint pdf of the random variables $\bar{C}=v(\tau,\gamma)$ and maximum $\bar{H}$ is equal to

$$\bar{\mathcal{Q}}(h,c;\gamma,\tau)=\frac{\partial f(c,h;\tau,\gamma)}{\partial h}\ , \qquad h>0\ , \quad c<h\ . \tag{A.6}$$

Then, the joint pdf of the high and close values of the stochastic process $v(\tau',\gamma)$ within the interval $\tau' \in (0,1)$ is obtained by taking $\tau=1$ in expression (A.6), which reads

$$\bar{\mathcal{Q}}(h,c;\gamma)=\frac{\partial f(c,h;\tau=1,\gamma)}{\partial h}\ , \qquad h>0\ , \quad c<h\ . \tag{A.7}$$

The joint pdf of the high and close values of Brownian motions was derived by Paul Lévy (1948).

The solution of the initial-boundary value problem (A.1), (A.2), (A.4) can be obtained by the *reflection method* as follows. The reflection method consists in replacing the initial-boundary value problem by the following auxiliary initial-value problem

$$\frac{\partial f(c,h;\tau,\gamma)}{\partial \tau} + \gamma \frac{\partial f(c,h;\tau,\gamma)}{\partial c} = \frac{1}{2}\frac{\partial^2 f(c,h;\tau,\gamma)}{\partial c^2} \ , \tag{A.8}$$
$$f(c,h;\tau=0,\gamma) = \delta(c) - A\delta(c-2h) \ ,$$

where the constant A has to be chosen such that the solution of the initial-value problem (A.8) satisfies the absorbtion boundary condition (A.4).

The solution of the initial value problem (A.8) is nothing but

$$f(c,h;\tau,\gamma) = g(c-\gamma\tau,\tau) - Ag(c-2h-\gamma\tau,\tau) \ ,$$

where $g(x,\tau)$ is given in (A.3). Substituting this expression into the boundary condition (A.4) yields $A = e^{2h\gamma}$. Thus, the solution of the initial-boundary value problem is

$$f(c,h;\tau,\gamma) = g(c-\gamma\tau,\tau) - e^{2h\gamma} g(c-2h-\gamma\tau,\tau) \ . \tag{A.9}$$

Substituting it into expression (A.7) yields the joint pdf of the high and close variables,

$$\bar{\mathcal{Q}}(h,c;\gamma) = f(c;\gamma)\mathcal{R}(h|c) \ , \qquad c < h \ , \qquad h > 0 \ , \tag{A.10}$$

where

$$f(c;\gamma) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(c-\gamma)^2}{2}\right) \tag{A.11}$$

is the pdf of the close value $c = v(1,\gamma)$, while

$$\mathcal{R}(h|c) = 2(2h-c)e^{2h(c-h)} \ , \qquad h \geqslant \max\{0,c\},$$

is the pdf of the high value $\bar{H}$, under the condition that the close value is equal to c.

A.2. *Wiener process between two absorbing boundaries*

The joint pdf $Q(h,l,c;\gamma)$ defined by (31) of the high, low and close values of the Wiener process can be expressed similarly to relation (A.7) via the solution of the diffusion equation (A.1) in the presence of two absorbing boundaries. We thus the new initial-boundary problem

$$\begin{gathered}\frac{\partial f(c,h,l;\tau,\gamma)}{\partial \tau}+\gamma\frac{\partial f(c,h,l;\tau,\gamma)}{\partial c}=\frac{1}{2}\frac{\partial^2 f(c,h,l;\tau,\gamma)}{\partial c^2}\ ,\\ f(c,h,l;\tau=0,\gamma)=\delta(c)\ ,\\ f(c=h+u\tau,h,l;\tau,\gamma)=0\ ,\qquad f(c=l+v\tau,h,l;\tau,\gamma)=0\ .\end{gathered}\tag{A.12}$$

Using the reflection method and a derivation similar to that leading to expression (A.9), we obtain

$$\begin{gathered}f(c,h,l;\tau,\gamma)=\sum_{m=-\infty}^{\infty}e^{2(v-u)(m(h-l)+l)}\times\\ \Big[e^{2(v-\gamma)(h-l)m}g(c-\gamma\tau+2(h-l)m,\tau)-\\ e^{2(\gamma-v)((h-l)m+l)}g(c-\gamma\tau-2l-2(h-l)m,\tau)\Big]\ .\end{gathered}\tag{A.13}$$

Figure A1 plots the function

$$f(c,h,l;\tau,\gamma)+0.05\cdot\tau\tag{A.14}$$

as a function of the close value c. The $0.05\cdot\tau$ is added in order to show clearly that $f(c,h,l;\tau,\gamma)$ indeed satisfies the moving absorption conditions (A.12).

We need the particular case corresponding to static boundaries ($u=v=0$) to transform the general solution (A.13) into

$$\begin{gathered}f(c,h,l;\tau,\gamma)=\sum_{m=-\infty}^{\infty}\Big[e^{2\gamma(l-h)m}g(c-\gamma\tau+2(h-l)m,\tau)-\\ e^{2\gamma((h-l)m+l)}g(c-\gamma\tau-2l-2(h-l)m,\tau)\Big]\ ,\\ l<c<h\ ,\quad h>0\ ,\quad l<0\ .\end{gathered}\tag{A.15}$$

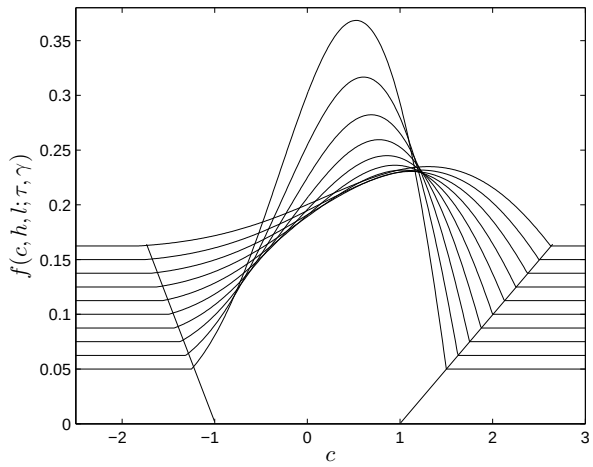

Fig. A1: Plots of the function (A.14) as a function of the close value c for $h = 1$, $l = -1$, $u = 0.5$, $v = -0.25$, $\gamma = 0.8$ and for $\tau = 1 + 0.25 \cdot k$, where $k = 0, 1, \ldots, 9$.

A.3. *Distribution of high, low, close values*

The joint pdf $\bar{\mathcal{Q}}(h, l, c; \gamma)$ corresponding to the diffusion process $v(\tau', \gamma)$ within the time interval $\tau' \in (0, 1)$ is obtained via the pdf $f(c, h, l; \tau, \gamma)$ given by (A.15) by the following relation, which is analogous to (A.7):

$$\bar{\mathcal{Q}}(h, l, c; \gamma) = -\frac{\partial f(c, h, l; \tau = 1, \gamma)}{\partial h \partial l} \ .$$

Analogously to expression (A.10), we obtain

$$\bar{\mathcal{Q}}(h, l, c; \gamma) = f(c; \gamma)\mathcal{R}(h, l|c) \ , \quad h > 0 \ , \quad l < 0 \ , \quad l < c < h \ , \tag{A.16}$$

where $f(c; \gamma)$ is given by (A.11), while $\mathcal{R}(h, l|c)$ is the joint pdf of the high and low values under the condition that the close value is equal to c:

$$\mathcal{R}(h, l|c) = 4 \sum_{m=-\infty}^{\infty} m\Big[m\mathcal{D}(m(h-l), c) + (1-m)\mathcal{D}(m(h-l)+l, c)\Big] \ , \tag{A.17}$$

$$\mathcal{D}(h, c) = [(c-2h)^2 - 1]e^{2h(c-h)} \ .$$

Figure A2 shows the contour lines of the conditional pdf $\mathcal{R}(h,l|c)$ for $c=0$ in the plane (h,l). Skorohod (1964) reported the joint distribution of the high-low-close for random walks with zero drift ($\gamma=0$).

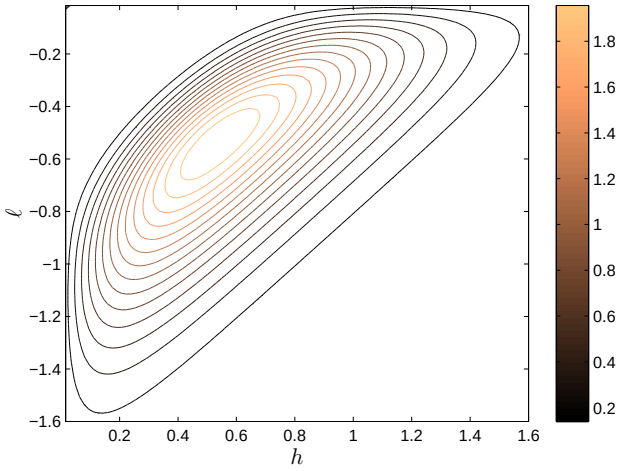

Fig. A2: Contour lines of the conditional pdf $\mathcal{R}(h,l|c)$ given by (A.17) for $c=0$ in the plane (h,l).

A.4. *Function g_n defined in expression* (34)

As seen from expressions (35) and (40), the diagrams (see definition 2.7) of the most efficient estimators are expressed via the function $g_n(\theta,\phi;\gamma)$ defined by the equation (34). The above calculations valid for the Wiener process show that it is equal to

$$g_n(\theta,\phi;\gamma)=\frac{4}{\sqrt{2\pi}}e^{-\gamma^2/2}\times$$
$$\sum_{m=-\infty}^{\infty} m\Big[mI_n(m(\tilde{h}-\tilde{l}),\tilde{c};\gamma)+(1-m)I_n(m(\tilde{h}-\tilde{l})+\tilde{l},\tilde{c};\gamma)\Big]\ ,$$

where

$$I_n(h,c,\gamma)=\int_0^\infty \rho^{2+n}\exp\left(\gamma c\rho-\frac{c^2}{2}\rho^2\right)\mathcal{D}(h\rho,c\rho)d\rho$$

and

$$\tilde{h}=\cos\theta\cos\phi\ ,\qquad \tilde{l}=\cos\theta\sin\phi\ ,\qquad \tilde{c}=\sin\theta\ .$$

In particular,

$$I_n(h, c, \gamma = 0) = \frac{F(n)}{|2h - c|^{3+n}} , \qquad F(n) = 2^{\frac{1+n}{2}}(2 + n)\Gamma\left(\frac{3 + n}{2}\right) .$$

REFERENCES

Aït-Sahalia, Y., Mykland, P.A. and Zhang, L. (2005) How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Fin. Stud. 18, 351-416.

Andersen T., G., T. Bollershev, F. X. Diebolt and P. Labys (2003): Modeling and Forecasting Realized Volatility, *Econometrica*, 71, 529-626.

Chan L. and Lien D. (2003) Using high, low, open, and closing prices to estimate the effects of cash settlement on futures prices. International Review of Financial Analysis, 12, 35-47.

Corsi, F., Zumbach, G., Müller, U., and Dacorogna, M. (2001) Consistent high-precision volatility from high-frequency data. Economic Notes, 30, 183-204.

Dominé (1996) First passage time distribution of a wiener process with drift concerning two elastic barriers, Journal of Applied Probability, 33, 164-175.

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

Lévy, P. (1948) Processus Stochastiques et Mouvement Brownien, Paris, Chapitre 6, section 42, p.193.

Magdon-Ismail, M. and A.A. Atiya (2003) A maximum likelihood approach to volatility estimation for a Brownian motion using the high, low and close, Quantitative Finance, 3 (5), 376-384.

McKenzie, D. (2006) An engine, not a camera (how financial models shape markets), MIT Press, Cambridge, MA.

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

Rogers, L. C. G., and S. E. Satchell (1991) Estimating Variance From High, Low and Closing Prices, *The Annals of Applied Probability*, 4, 504-512.

Rogers, L. C. G., and S. E. Satchell, and Y. Yoon (1994) Estimating the Volatility of Stock Prices: A Comparison of Methods that use High and Low Prices, *Applied Financial Economics*, 4, 241-247.

Skorohod, A.V. (1964) Stochastic processes with independent increments, M.:, Nauka, Chapter 6, section 27, p. 169 (in Russian).

Yang, D., and Q. Zhang (2000) Drift-independent Volatility Estimation Based on High, Low, Open and Close Prices, Journal of Business, 73 (3), 477-491.

Zhang, L., Mykland, P.A. and At-Sahalia, Y. (2005) A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100, 1394-1411.