This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Bridge Homogeneous Volatility Estimators

A. SAICHEV†‡ D. SORNETTE*†§, V. FILIMONOV† and F. CORSI¶

†ETH Zurich – Department of Management, Technology and Economics, Switzerland
‡Nizhni Novgorod State University – Department of Mathematics, Russia
§Swiss Finance Institute, 40, Boulevard du Pont-d' Arve, Case Postale 3, 1211 Geneva 4,
Switzerland
¶Scuola Normale Superiore, Piazza dei Cavalieri 7, Pisa, 56126, Italy
(18 March 2013)

We present a theory of bridge homogeneous volatility estimators for log-price stochastic processes. Starting with the standard definition of a Brownian bridge as the conditional Wiener process with two endpoints fixed, we introduce the concept of an incomplete bridge by breaking the symmetry between the two endpoints. For any given time interval, this allows us to encode the information contained in the OHLC (open, high, low and close prices) into an incomplete bridge. The efficiency of the new proposed estimators is favorably compared with that of the classical Garman-Klass and Parkinson estimators.

Keywords: volatility, variance, estimators, efficiency, homogeneous functions

1. Introduction

In the theory of financial markets, risks are associated with price variability. Among different measures of variability of financial prices modeled as random walks, the volatility, defined as the square root of the expected square of the increments of the log-price over a specific time interval, plays an important role. With the growing availability of tick-by-tick price time series, a number of high-frequency realized volatility estimators have been developed (see, for instance, Aït-Sahalia *et al.* (2005), Andersen *et al.* (2003), Corsi *et al.* (2001), Yang and Zhang (2000), Zhang *et al.* (2005)). In this article, we suggest new elements to develop better high frequency estimators that exploit four prices, called the open-high-low-close (OHLC).

We present the general theory of volatility estimation for arbitrary stochastic process. Our analysis is based on the parsimonious encoding of the information contained in the mentioned OHLC prices for a given time interval in the form of diagram, associated with the joint probability density (pdf) of the OHLC values. The diagrams can be tailored to yield the most efficient estimators associated to any statistical properties of the underlying log-price stochastic process. When the needed joint pdf cannot be expressed analytically, the diagram can be easily constructed by numerical simulations. We find that our suggested OHLC spot volatility estimators are essentially more efficient than other OHLC estimators that have been previously proposed.

*Corresponding author. Email: dsornette@ethz.ch

Our work also improves on the following results: Garman and Klass (1980) introduced a quadratic variance G&K estimator, which has rather low variance in the case of the Wiener log-price process; Parkinson (1980) proposed a simple quadratic PARK variance estimator proportional to $(H - L)^2$, which is using only a part of the information available from OHLC prices; Rogers and Satchell (1991) and Rogers *et al.* (1994) introduced a quadratic R&S variance estimator, unbiased for any drifts. Yang and Zhang (2000) produced an efficient quadratic variance estimator, based on the OHLC of log-prices for $n > 1$ consecutive days, and taking into account the possible gaps of prices. Chan and Lien (2003) compared empirically the efficiency of the PARK, G&K and R&S estimators. From the perspective offered by these previous works, the present article can be viewed as providing the underpinning theory of, more general than quadratic, most efficient and unbiased bridge homogeneous volatility and variance estimators.

The article is organized as follows. Section 2 describes the bridge homogeneous volatility estimators. Section 3 derives the general expressions for the most efficient bridge homogeneous volatility OHLC estimators. Section 4 provides an exact expression of the joint distribution of extremes of Wiener process with drift. Section 5 compares the efficiency of mentioned most efficient bridge homogeneous estimators and the efficiency of the bridge PARK and G&K estimators. Section 6 tests our results using time series generated by numerical simulations, which mimic some aspects of the tick-by-tick nature of real log-price processes. Section 7 concludes.

2.   Bridge homogeneous volatility estimators

The main goal of this article is to construct efficient volatility estimators using the OHLC of some asset log-price process $A(t)$. The conventional definition of the volatility $V(t_0, T_0)$ of a stochastic process $A(t)$ at time t_0 and time scale T_0 is the square root of the expected square of its increment

$$\Delta(t_0, T_0) = A(t_0 + T_0) - A(t_0)$$

within the time interval $t \in (t_0, t_0 + T_0)$: $V(t_0, T_0) = \sqrt{\mathrm{Var}\left[\Delta(t_0, T_0)\right]}$.

Since different measures of the variability of log-price processes are used in the literature, it is convenient to define a generalized volatility of order λ as follows:

Definition 2.1: The volatility of order λ of the stochastic process $A(t)$ is the power λ of the conventional volatility

$$V_\lambda(t_0, T_0) := V^\lambda(t_0, T_0) = \left(\mathrm{Var}\left[\Delta(t_0, T_0)\right]\right)^{\lambda/2}.$$

Remark 1: For $\lambda = 1$, the above volatility coincides with the conventional volatility, while, for $\lambda = 2$, $V_2(t_0, T_0)$ is the variance of the increment $\Delta(t_0, T_0)$. Most known estimators, for instance the R&S, G&K and PARK ones, are variance estimators. Introducing the volatility of order λ gives us the possibility later on to compare the differences and relations between the volatility and variance estimators.

2.1.   *Wiener process model of log-price increments*

The most efficient estimators, derived in section 3, are general, and valid for a wide range of log-price processes. However, we will mostly analyze the properties of the volatility estimators by applying them to the Wiener process with drift. This particular case is important because

the well-known OHLC estimators, for instance the G&K and R&S ones, are adjusted specifically to the Wiener processes. Our study of the Wiener process with drift allows us to make detailed comparison of the efficiency of our most efficient estimators compared with the well-known G&K, R&S and PARK estimators.

Let us pose, without loss of generality, $t_0 = 0$ and $A(0) = 0$. This implies that log-price process $A(t)$, modeled by the Wiener process with drift, is equal to

$$A(t) := \mu t + \sigma W(t), \tag{1}$$

where μ is the drift of the stochastic process $A(t)$, σ is its standard deviation at $t = 1$, while $W(t)$ is the Wiener process $\sim \mathcal{N}(0,t)$. Its self-similar properties allow us to choose the time scale by $T_0 = 1$ without loss of generality, so that the volatility of order λ is simply equal to the standard deviation σ raised to the power λ:

$$V_\lambda := V_\lambda(t_0 = 0, T_0 = 1) = \sigma^\lambda.$$

Definition 2.2: The stochastic process

$$B(t, \kappa, T) := A(t) - \kappa \frac{t}{T} A(T) = \mu(1 - \kappa)t + \sigma \left[W(t) - \kappa \frac{t}{T} W(T) \right], \tag{2}$$

where κ is an arbitrary constant, is called the incomplete bridge of the original stochastic process $A(t)$. For $\kappa = 1$, the incomplete bridge becomes the (complete) bridge:

$$B(t, T) := \sigma \left[W(t) - \frac{t}{T} W(T) \right].$$

Using the self-similarity of the Wiener process, one can rewrite (1), (2) as:

$$A(t) = \sigma \sqrt{T} X \left(\frac{t}{T}, \gamma \right), \qquad B(t, \kappa, T) = \sigma \sqrt{T} Y \left(\frac{t}{T}, \kappa, \gamma \right), \tag{3}$$

where $Y(t, \kappa, \gamma)$ is the incomplete bridge

$$Y(t, \kappa, \gamma) := X(t, \gamma) - \kappa t X(1, \gamma) \tag{4}$$

of the Wiener process with drift

$$X(t, \gamma) := \gamma t + W(t), \qquad t \in (0, 1), \tag{5}$$

and the auxiliary parameter

$$\gamma = \mu \sqrt{T} / \sigma. \tag{6}$$

Figure 1 shows a realization of the Wiener process $W(t)$ and its bridge. The high and low of the Wiener process and of its bridge are drastically different.

Remark 2: In financial applications, both the drift μ and the standard deviation σ are unknown. Thus, the value of the parameter γ is unknown as well. Nevertheless, for the convenience of analysis, we suppose in the derivations that parameter γ is given.
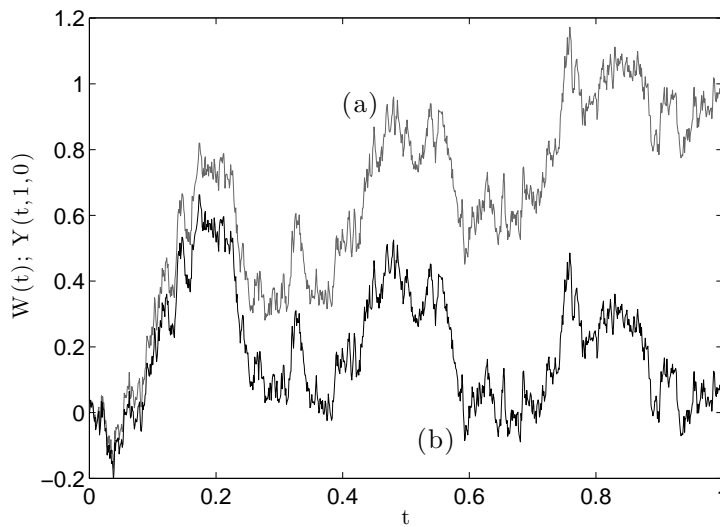
Figure 1. A realization of the (a) Wiener process $W(t)$ and (b) bridge $W(t) - tW(1)$.

2.2.   Discussion of the bridge OHLC and realized tick-by-tick volatility estimators

Below, we present the construction of the incomplete bridge OHLC estimator and discuss its efficiency, showing that the most efficient bridge estimator proposed here outperforms classical tools such as G&K, R&S and PARK estimators.

From a practical point of view, the construction of the bridge requires the full tick-by-tick price dataset. This might lead to question the usefulness of considering only OHLC values compared with exploiting the realized volatility estimators based on the full tick-by-tick time series. Indeed, in the case of the geometrical Brownian motion whose increments are statistically independent, the increase of information (via the number of samples) does improve significantly the efficiency of the estimator. However, in the case of real financial time series, returns are not statistically independent, exhibiting for instance long-term correlations in their absolute values. Moreover, with the increase of the sampling frequency, additional strong negative serial correlation due to the bid-ask bounce effect starts playing an important role, resulting in what is called the *microstructure noise* and *microstructure effect* (see for instance, Zhang *et al.* (2005) and Aït-Sahalia *et al.* (2005)). In particular, the microstructure noise leads to the divergence of the realized volatility with increasing sampling frequency (the so-called *volatility signature plot*). For quantitative descriptions of the microstructure noise and for models explaining it, see for instance Bacry *et al.* (2013) and Saichev and Sornette (2012) and references therein.

In comparison, characteristics of the stochastic price movement that are more coarse-grained or involve cumulative or integral properties, such as its extreme values (high and low values in the notation of this article), are much less sensitive to the existence of microstructure effects and short-term anti-persistency. Considering microstructure noise as an additive zero mean stochastic process (as in Zhang *et al.* (2005)), taking the integral of returns over time (or their cumulative sum in the case of the discrete version of the process) amounts to an effective low-pass filter that has the property of reducing the impact of the high-frequency microstructure noise. This is vividly illustrated by the vanishing of negative correlations when considering aggregated time intervals (see the empirical analysis in (Cont 2001) and the theoretical background in Saichev and Sornette (2012)). Thus, estimating volatility with OHLC values mitigates negative effects of

the microstructure noise while volatility measures using realized tick-by-tick data are polluted by it.

At much longer time scales, the volatility $\sigma(t)$ cannot be assumed constant anymore as assumed in expression (1). In such cases, the time varying volatility is often modeled by conditional or stochastic volatility models. The so-called *integrated variance (IV)* then provides a useful characteristic of price movements. It is defined as the integral over time ($IV = \int_0^T \sigma^2(t)dt$) of the *instantaneous variance* (or *spot variance*), and is typically estimated using realized volatility measures. Specifically, the interval $[0, T)$ is split into N sub-intervals $[t_i, t_{i+1})$, $i = 0 \dots N-1$. In each of these N sub-intervals, equation (1) is assumed to hold with constant drift μ and volatility σ terms. The coefficients μ and σ can be different from a sub-interval to the next. In each of these sub-intervals, the instantaneous variance is estimated using the realized measure such as the *quadratic variation* ($\hat{\sigma}_i^2 = (A(t_{i+1}) - A(t_i))^2$). The integrated variance is then constructed at the sum of these local estimates ($\hat{IV} = \sum_{i=0}^{N-1} \hat{\sigma}_i^2$). Theoretically, it can be shown that such realized estimation of the integrated variance is a consistent estimator in the sense that it converges to the integrated variance as $N \to \infty$ (see e.g. Protter (2005)). However in the presence of the microstructure effects discussed above, the division of the interval $[0, T]$ into a growing number N of sub-intervals becomes limited by the discreteness of the order arrivals and the bid-ask bounce effect. As N must therefore remain finite and not too high to remove these deleterious effects, the realized volatility estimator then suffers from a well-known bias problem and is generally inconsistent with the integrated variance (see e.g. Zhang *et al.* (2005)). A known solution to this problem is to replace the quadratic variation with more consistent measures.

In order to address the biases of standard realized variance estimators for ultra-high frequency data (Zhang *et al.* 2005), we should mention that other theories on alternative realized variance estimators have been developed (Zhang *et al.* 2005, Zhang 2006). Based on a multi-scale approach for ultra-high frequency data, they combine multiple frequencies in a manner which corrects the bias of the volatility estimate. Alternatively, Barndor-Nielsen et al. (2008, 2011) introduced a class of realized sampled and sub-sampled kernel estimators of the increments of quadratic variation in the presence of noise, which are very efficient to obtain similar unbiased volatility estimates.

Here, we propose another perspective in terms of the most efficient bridge volatility estimator, which becomes a very useful input to the integrated variance, as it is less subjected to the microstructure noise than the realized volatility measures. Moreover, the method proposed in this article allows one to construct volatility estimator tailored to the specific properties of the real price time series such as heavy-tails, long-memory and microstructure as well as its discrete-time nature (this idea will be illustrated with the toy model of the discrete random walk in the section 6).

2.3. *Homogeneous bridge volatility estimators*

Definition 2.3: A volatility estimator V_λ is called an *homogeneous* bridge OHLC volatility estimator of order λ if it has the form

$$V_\lambda = h_\lambda(\bar{H}, \bar{L}, \bar{C})/T^{\lambda/2}, \tag{7}$$

where h_λ is a homogeneous function of order λ, $\bar{H}$ and $\bar{L}$, are the high and low of the incomplete bridge $B(t, \kappa, T)$ (2) within the observation interval $(0, T)$:

$$\bar{H} := \sup_{t \in (0,T)} B(t, \kappa, T), \qquad \bar{L} := \inf_{t \in (0,T)} B(t, \kappa, T),$$

while $\bar{C} := A(T) = \mu T + \sigma W(T)$ is the close value of the stochastic process $A(t)$.

A remarkable property of homogeneous estimators defined by (7) is that, for a given γ, their statistical properties do not depend on the duration T of the observation interval. Mathematically, this fact is expressed by the theorem:

Theorem 2.4:   *The estimator defined by* (7) *is equal to*

$$V_\lambda = \sigma^\lambda h_\lambda(H, L, C), \tag{8}$$

where H and L are the high and low of the incomplete bridge $Y(t, \kappa, \gamma)$ (4):

$$H := \sup_{t \in (0,1)} Y(t, \kappa, \gamma), \quad L := \inf_{t \in (0,1)} Y(t, \kappa, \gamma), \tag{9}$$

while $C := X(1, \gamma)$ is the close value of the process $X(t, \gamma)$, defined by (5).

Proof: Relation (8) follows from self-similarity of the Wiener process and from homogeneity of the function h_λ.   □

Definition 2.5: We refer to the function

$$\hat{e}_\lambda = h_\lambda(H, L, C) \tag{10}$$

as the *canonical bridge* OHLC volatility estimator of order λ. Using this definition, one can rewrite expression (8) in the form

$$V_\lambda = \sigma^\lambda \hat{e}_\lambda. \tag{11}$$

3.   Most efficient bridge homogeneous estimators

3.1.   *Diagrams of bridge homogeneous estimators*

It results from expressions (11), (10) that the homogeneous estimator (7) is unbiased, if the expected value of the corresponding canonical estimator given by (10) is equal to unity:

$$\mathrm{E}[\hat{e}_\lambda] = \mathrm{E}[h_\lambda(H, L, C)] = 1. \tag{12}$$

Definition 3.1: The bridge homogeneous volatility estimator of order λ given by (7) is called the most efficient one, for a given value γ_0 of the parameter γ and for a fixed κ, if, for $\gamma = \gamma_0$ and fixed κ and λ, the equality (12) holds while the variance of the corresponding canonical estimator achieves the minimal value among the variances of all canonical estimators of order λ and for the same parameters $\gamma = \gamma_0$ and κ.

Below we provide the explicit expressions of the most efficient bridge homogeneous volatility estimators. For this, it is convenient to use a change of variables from the random variables $\{H, L, C\}$ to their spherical coordinates $\{R, \Theta, \Phi\}$:

$$H = R\cos\Theta\cos\Phi, \quad L = R\cos\Theta\sin\Phi, \quad C = R\sin\Theta. \tag{13}$$

Inversely, we have

$$R = \sqrt{H^2 + L^2 + C^2}, \quad \Theta = \arctan\left(\frac{C}{\sqrt{H^2 + L^2}}\right), \quad \Phi = \arctan\left(\frac{L}{H}\right). \tag{14}$$

Putting (13) into (10) and bearing in mind the homogeneity function h_λ, we obtain

$$\hat{e}_\lambda = R^\lambda \, \psi_\lambda(\Theta, \Phi), \tag{15}$$

where

$$\psi_\lambda(\theta, \phi) = h_\lambda(\cos\theta\cos\phi, \cos\theta\sin\phi, \sin\theta). \tag{16}$$

Definition 3.2: The function $\psi_\lambda(\theta, \phi)$ (16) is called the *canonical estimator diagram.*

Remark 3: The spherical coordinates are intrinsic to homogeneous estimators, allowing to split them into a power function R^λ and an arbitrary function of two variables Θ and Φ (see Eq. (15)). The spherical coordinates reduce the search of efficient OHLC estimators from 3D functions to the appropriate 2D function $\psi_\lambda(\theta, \phi)$.

3.2. *Domain of possible $\{\Theta, \Phi\}$ values*

Below, we will need the domain of existence of the random variables $\{R, \Theta, \Phi\}$ (14). Note that $R \in (0, \infty)$. The domain $\mathcal{S}_\kappa$ of the possible values of random variables $\{\Theta, \Phi\}$ depends on the parameter κ. It is the same for the arguments $\{\theta, \phi\}$ of the diagram $\psi_\lambda(\theta, \phi)$ (16): $\{\theta, \phi\} \in \mathcal{S}_\kappa$. Since $H \geqslant 0$ and $L \leqslant 0$, in view of (13), we have

$$\tan\Phi = \frac{L}{H} \in (-\infty, 0] \qquad \Rightarrow \qquad -\frac{\pi}{2} \leqslant \Phi < 0.$$

In turn, as seen from (4) and (5), the values $\{H, L, C\}$ satisfy the inequalities $L \leqslant (1-\kappa)C \leqslant H$ or, using (13),

$$\sin\Phi \leqslant (1 - \kappa)\tan\Theta \leqslant \cos\Phi \Rightarrow \arctan\left(\frac{\sin\Phi}{1 - \kappa}\right) \leqslant \Theta \leqslant \arctan\left(\frac{\cos\Phi}{1 - \kappa}\right).$$

Thus

$$\mathcal{S}_\kappa = \left\{\arctan\left(\frac{\sin\Phi}{1 - \kappa}\right) \leqslant \Theta \leqslant \arctan\left(\frac{\cos\Phi}{1 - \kappa}\right), -\frac{\pi}{2} \leqslant \Phi < 0\right\}. \tag{17}$$

3.3. *Most efficient bridge OHLC homogeneous estimators*

Let us denote the joint pdf of the variables $\{H, L, C\}$ by $Q(h, \ell, c; \kappa, \gamma)$. Then, the expectation of the canonical estimator defined by (15) is equal to

$$\mathrm{E}[\hat{e}_\lambda | \kappa, \gamma] = \mathcal{M}_\lambda(\kappa, \gamma) := \iint_{\mathcal{S}_\kappa} \psi_\lambda(\theta, \phi) \, g_\lambda(\theta, \phi; \kappa, \gamma) \cos\theta d\theta d\phi, \tag{18}$$

where

$$g_\lambda(\theta, \phi; \kappa, \gamma) = \int_0^\infty \rho^{\lambda+2} Q(\rho \cos \theta \cos \phi, \rho \cos \theta \sin \phi, \rho \sin \theta; \kappa, \gamma) d\rho. \tag{19}$$

Accordingly, at $\gamma = \gamma_0$ and given κ, one can represent the diagram of any unbiased, homogeneous estimator in the form

$$\psi_\lambda(\theta, \phi; \kappa, \gamma_0) = \frac{G(\theta, \phi)}{\iint\limits_{\mathcal{S}_\kappa} G(\theta, \phi) \; g_\lambda(\theta, \phi; \kappa, \gamma_0) \cos \theta d\theta d\phi}, \tag{20}$$

where $G(\theta, \phi)$ is an arbitrary function. The following theorem gives the expression for the diagram (20) of the most efficient (at any given κ and $\gamma = \gamma_0$) homogeneous estimator.

Theorem 3.3: *The diagram of the most efficient (for a given κ and $\gamma = \gamma_0$) bridge homogeneous estimator of order λ is equal to*

$$\psi_{me,\lambda}(\theta, \phi; \kappa, \gamma_0) = \frac{G_\lambda(\theta, \phi; \kappa, \gamma_0)}{\mathcal{E}_\lambda(\kappa, \gamma_0)}, \qquad \{\theta, \phi\} \in \mathcal{S}_\kappa, \tag{21}$$

where

$$G_\lambda(\theta, \phi; \kappa, \gamma) = \frac{g_\lambda(\theta, \phi; \kappa, \gamma)}{g_{2\lambda}(\theta, \phi; \kappa, \gamma)}, \quad \mathcal{E}_\lambda(\kappa, \gamma) = \iint\limits_{\mathcal{S}_\kappa} \frac{g_\lambda^2(\theta, \phi; \kappa, \gamma)}{g_{2\lambda}(\theta, \phi; \kappa, \gamma)} \cos \theta d\theta d\phi. \tag{22}$$

The proof is given in Appendix A.

3.4. *Discussion of theorem 3.3*

As seen from the proof of theorem 3.3, the variances of the homogeneous estimators given by relations (15), (21), (22) reach the lower bound among the variances of any unbiased homogeneous estimators $\hat{e}_\lambda$, for any statistics of OHLC values of the underlying log-price stochastic process. Since quadratic estimators constitute special cases of homogeneous ones, the most efficient homogeneous estimators should be more efficient than known quadratic ones. Thus, theorem 3.3 provides the tools to develop novel estimators, which are most efficient for given statistics of log-price processes associated with particular financial markets, taking into account, for instance, the chaotic jumps and the heavy tail characteristics of the probability density functions (pdf) of returns.

Discussing the most efficient estimators requires a detailed study of their statistics for each financial market behavior. The present article opens the path towards such studies targeting different processes that are now more fashionable, such as GARCH and their siblings. But before embarking on this general program, the present article has the goal of developing the basis of comparisons between estimators already known and our most efficient ones, so as to motivate the future development on more realistic stochastic processes. Until now, most known volatility estimators have been adjusted to meet some properties in the frame of Wiener processes with drift. For instance, the seminal G&K estimator is the most efficient quadratic estimator in the particular case of the Wiener log-price process (with zero drift). Similarly, the R&S's quadratic estimator is constructed in such a way that, in the frame of Wiener processes with drift, it is unbiased for any drift (nonzero γ). It thus seems important, as a first step, to compare the

efficiency of our most efficient estimators and the well-known G&K, R&S, and PARK estimators in the frame of Wiener processes for log-price with drift, for which they are so-to-speak "native".

The introduction of quadratic estimators have been in part due to the lack of sufficient knowledge concerning the statistics of the extremes of the Wiener process with drift. In the present article, we provide a significant advance in the form of an analytically exact explicit expression for the joint pdf of the incomplete bridge OHLC values. We offer this novel mathematical result in the hope that it fosters new developments in econometrics.

4.   Statistical description of incomplete bridges

4.1.   Wiener process identical in law

In order to get the most efficient bridge homogeneous OHLC estimator, we need the pdf $Q(h, \ell, c; \kappa, \gamma)$ of the OHLC values of the incomplete bridge $Y(t, \kappa, \gamma)$ defined by (4) and the close value C of the underlying process $X(t, \gamma)$ (5). Before giving the explicit pdf $Q(h, \ell, c; \kappa, \gamma)$, it is useful to discuss some properties of the stochastic process $Y(t, \kappa, \gamma)$.

Theorem 4.1:   *The incomplete bridge $Y(t, \kappa, \gamma)$ given by (4) is identical in law to the diffusion process*

$$\mathcal{Y}(t, \kappa, \gamma) := \gamma(1 - \kappa)t + \mathcal{W}(t, \kappa), \tag{23}$$

where

$$\mathcal{W}(t, \kappa) := (1 - t + (1 - \kappa)^2 t)W\left(\frac{t}{1 - t + (1 - \kappa)^2 t}\right). \tag{24}$$

Proof: After substituting (5) into (4), we obtain

$$Y(t, \kappa, \gamma) = \gamma(1 - \kappa)t + \Omega(t, \kappa), \qquad \Omega(t, \kappa) := W(t) - \kappa t W(1). \tag{25}$$

One can verify that $\Omega(t, \kappa)$ is a Gaussian process with zero mean and the covariance:

$$\mathrm{E}[\Omega(t_1, \kappa)\Omega(t_2, \kappa)] = (t_1 \wedge t_2) - [1 - (1 - \kappa)^2]t_1 t_2, \quad 0 \leqslant t_1, t_2 \leqslant 1. \tag{26}$$

Direct calculations show that the Gaussian process $\mathcal{W}(t, \kappa)$ defined by (24) is also of a zero mean and the same covariance (26). Thus, the incomplete bridge $Y(t, \kappa, \gamma)$ (25) is identical in law to the diffusion process $\mathcal{Y}(t, \kappa, \gamma)$ defined in (23). □

Henceforth, for the analysis of the incomplete bridge $Y(t, \kappa, \gamma)$ (4) statistics, we use the equivalence in law stated in theorem 4.1, which allows us to replace the incomplete bridge by the diffusion process $\mathcal{Y}(t, \kappa, \gamma)$ (23). In turn, it is convenient to explore the properties of the diffusion process $\mathcal{Y}(t, \kappa, \gamma)$ using the change of time

$$\tau = \tau(t, \kappa) := \frac{(1 - \kappa)^2 t}{1 - t + (1 - \kappa)^2 t} \quad \Leftrightarrow \quad t = t(\tau, \kappa) := \frac{\tau}{\tau + (1 - \kappa)^2(1 - \tau)}.$$

The function $\tau(t, \kappa)$ maps the interval $t \in (0, 1)$ onto the same interval $\tau \in (0, 1)$.

Let us introduce the auxiliary stochastic process

$$\mathcal{Z}(\tau, \kappa, \gamma) := \mathcal{Y}(t(\tau, \kappa), \kappa, \gamma). \tag{27}$$

Using relations (23), (24), and self-similar properties of the Wiener process, we rewrite $\mathcal{Z}(\tau, \kappa, \gamma)$ in the form

$$\mathcal{Z}(\tau, \kappa, \gamma) = \frac{1 - \kappa}{\tau + (1 - \kappa)^2 (1 - \tau)} \left[\gamma \tau + W(\tau) \right]. \tag{28}$$

Below we assume, for definiteness, $\kappa < 1$.

It is seen from (27) and (28) that the following inequalities are equivalent

$$L \leqslant \mathcal{Y}(t, \kappa, \gamma) \leqslant H, \quad \Leftrightarrow \quad a + \alpha \tau \leqslant W(\tau) \leqslant b + \beta \tau, \quad t, \tau \in (0, 1), \tag{29}$$

where

$$a = (1 - \kappa)L, \ b = (1 - \kappa)H, \ \alpha = \frac{1 - (1 - \kappa)^2}{1 - \kappa} L - \gamma, \ \beta = \frac{1 - (1 - \kappa)^2}{1 - \kappa} H - \gamma. \tag{30}$$

Note also, the close value $C = X(1, \gamma)$ of the process $X(t, \gamma)$ (5) is tied to the close value of the incomplete bridge $Y(t, \kappa, \gamma)$ (4) by the equality $Y(1, \kappa, \gamma) = (1 - \kappa)C$. In turn, it follows from the identity in law of the two stochastic processes $Y(t, \kappa, \gamma)$ and $\mathcal{Y}(t, \kappa, \gamma)$ and from relations (27), (28) that one may replace $Y(1, \kappa, \gamma)$ by

$$\mathcal{Z}(1, \kappa, \gamma) = (1 - \kappa)[\gamma + W(1)].$$

Thus, one obtains

$$W(1) = C - \gamma \quad \Leftrightarrow \quad C = W(1) + \gamma. \tag{31}$$

4.2. *Diffusion equation*

Let us define the probability

$$f(h, \ell, c; \kappa, \gamma)dc := \Pr\{C \in (c, c + dc) \cap \ell \leqslant Y(t, \kappa, \gamma) \leqslant h; t \in (0, 1)\}.$$

Then, the joint pdf of the high and low values $\{H, L\}$ (9) of the incomplete bridge $Y(t, \kappa, \gamma)$, and of the close value C of the original process $X(t, \gamma)$ (5), is equal to

$$Q(h, \ell, c; \kappa, \gamma) = -\frac{\partial^2 f(h, \ell, c; \kappa, \gamma)}{\partial h \partial \ell},$$

$$h > h_-, \ \ell < \ell_+, \ \frac{\ell}{1 - \kappa} \leqslant c \leqslant \frac{h}{1 - \kappa}, \ h_- = 0 \vee (1 - \kappa)c, \ \ell_+ = 0 \wedge (1 - \kappa)c. \tag{32}$$

From the relations of the previous subsection, one can express the function $f(h, \ell, c; \kappa, \gamma)$ via the auxiliary function $\varphi(\omega; \tau)$

$$\varphi(\omega; \tau)d\omega := \Pr\{W(\tau) \in (\omega, \omega + d\omega) \cap a + \alpha \tau' \leqslant W(\tau') \leqslant b + \beta \tau'; \tau' \in (0, \tau)\},$$

so that

$$f(h, \ell, c; \kappa, \gamma) = \varphi(c - \gamma; 1, a, b, \alpha, \beta). \tag{33}$$

Note that $\varphi(\omega; \tau)$ is the solution of the diffusion equation

$$\frac{\partial \varphi}{\partial \tau} = \frac{1}{2} \frac{\partial^2 \varphi}{\partial \omega^2}, \tag{34}$$

with initial condition

$$\varphi(\omega; \tau = 0) = \delta(\omega) \tag{35}$$

and absorbing boundary conditions

$$\varphi(\omega = a + \alpha\tau; \tau) = 0, \quad \varphi(\omega = b + \beta\tau; \tau) = 0, \quad \tau > 0, \tag{36}$$

which account for the inequalities (29).

Below, we solve this initial-boundary problem (34)–(36) and determine the joint pdf of the high and low values of the incomplete bridge $Y(t, \kappa, \gamma)$ and of the close value of the Wiener process $X(t, \gamma)$ with drift, using the following relation that derives from (32) and (33):

$$Q(h, \ell, c; \kappa, \gamma) = -\frac{\partial^2 \varphi(c - \gamma; 1, a, b, \alpha, \beta)}{\partial h \partial \ell}. \tag{37}$$

4.3.   Solution of the initial-boundary problem

The solution of the initial-boundary problem (34), (35), (36) is stated in the theorem:

Theorem 4.2:   *The solution of the diffusion equation* (34), *satisfying the initial-boundary conditions* (35), (36), *is given by*

$$\varphi(\omega; \tau) = \sum_{m=-\infty}^{\infty} e^{2(\alpha-\beta)(b-a)m^2 + 2(\alpha b - \beta a)m} \times$$
$$\left[g(\omega + 2m(b-a); \tau) - e^{2a(2(\beta-\alpha)m-\alpha)} g(\omega + 2m(b-a) - 2a; \tau) \right], \tag{38}$$
$$g(\omega; \tau) = \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{\omega^2}{2\tau}\right).$$

The proof is given in Appendixes B and C. Putting (38), (30) into (33), we get

$$f(h, \ell, c; \kappa, \gamma) = g(c - \gamma) \sum_{m=-\infty}^{\infty} e^{-2(h-\ell)^2 m^2 - 2m(h-\ell)(1-\kappa)c} \left[1 - e^{4(h-\ell)\ell m - 2\ell(\ell-(1-\kappa)c)} \right], \tag{39}$$

where $g(c) = g(c; 1)$.

4.4. Joint pdf of high, low and close values

Using relations (33), (37) and (39), we obtain the sought pdf $Q(h, \ell, c; \kappa, \gamma)$ (37):

$$Q(h, \ell, c; \kappa, \gamma) = g(c - \gamma)\mathcal{R}(h, \ell; \kappa | c), \tag{40}$$

where

$$\mathcal{R}(h, \ell; \kappa | c) = \sum_{m=-\infty}^{\infty} m\big[m\mathcal{D}(m(h - \ell), (1 - \kappa)c) + (1 - m)\mathcal{D}(m(h - \ell) + \ell, (1 - \kappa)c)\big], \tag{41}$$

$$\mathcal{D}(h, c) = 4[(c - 2h)^2 - 1]e^{2h(c-h)}.$$

Let us thus define the joint pdf limit

$$Q(h, \ell, c; \gamma) := \lim_{\kappa \to 1} Q(h, \ell, c; \kappa, \gamma).$$

Expressions (5) and (41) show that it is equal to

$$Q(h, \ell, c; \gamma) = g(c - \gamma)\mathcal{R}(h, \ell), \qquad -\infty < c < \infty, \quad h > 0, \qquad \ell < 0,$$

$$\mathcal{R}(h, \ell) = \sum_{m=-\infty}^{\infty} m\left[m\mathcal{D}(m(h - \ell)) + (1 - m)\mathcal{D}(m(h - \ell) + \ell)\right], \tag{42}$$

$$\mathcal{D}(h) = 4(4h^2 - 1)\, e^{-2h^2}.$$

Expression (42) has a clear probabilistic sense. It means that the high and low values of the bridge $Y(t, 1, \gamma)$ are statistically independent from the close value $C = X(1, \gamma)$ of the original Wiener process with drift. Accordingly, $\mathcal{R}(h, \ell)$ in (42) is the unconditional joint pdf of the bridge high and low values.

4.5. Diagrams of the most efficient bridge homogeneous OHLC estimators

Knowing the joint pdf of the random variables $\{H, L, C\}$, one can calculate the functions $g_\lambda(\theta, \phi; \kappa, \gamma)$ (19) containing at the diagrams (21), (22) of the most efficient volatility estimators. It is seen from (19) and (40), (41), that

$$g_\lambda(\theta, \phi; \kappa, \gamma) = \frac{1}{\sqrt{2\pi}}e^{-\gamma^2/2} \times$$

$$\sum_{m=-\infty}^{\infty} m\big[mI_\lambda(m(\tilde{h} - \tilde{l}), \tilde{c}; \kappa, \gamma) + (1 - m)I_\lambda(m(\tilde{h} - \tilde{l}) + \tilde{l}, \tilde{c}; \kappa, \gamma)\big], \tag{43}$$

$$I_\lambda(h, c; \kappa, \gamma) = \int_0^\infty \rho^{2+\lambda} \exp\left(\gamma c\rho - \frac{c^2}{2}\rho^2\right) \mathcal{D}(h\rho, (1 - \kappa)c\rho)d\rho,$$

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

All calculations done, the explicit expression of $I_\lambda(h, c; \kappa, \gamma)$ reads

$$I_\lambda(h, c; \kappa, \gamma) = \left(\frac{2}{a}\right)^{3+\frac{\lambda}{2}} \times$$

$$\left[b\sqrt{2a}\,\Gamma\left(\frac{5+\lambda}{2}\right)M\left(\frac{5+\lambda}{2}, \frac{1}{2}, \frac{d^2}{2a}\right) - a\sqrt{\frac{a}{2}}\,\Gamma\left(\frac{3+\lambda}{2}\right)M\left(\frac{3+\lambda}{2}, \frac{1}{2}, \frac{d^2}{2a}\right)\right.$$

$$\left. +2db\,\Gamma\left(3+\frac{\lambda}{2}\right)M\left(3+\frac{\lambda}{2}, \frac{3}{2}, \frac{d^2}{2a}\right) - da\,\Gamma\left(2+\frac{\lambda}{2}\right)M\left(2+\frac{\lambda}{2}, \frac{3}{2}, \frac{d^2}{2a}\right)\right].$$

Here,

$$M(a, b, z) := \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)}\int_0^1 du\; e^{zu}u^{a-1}(1-u)^{b-a-1}\;, \qquad \mathrm{Re}\{b\} > \mathrm{Re}\{a\} > 0$$

is the Kummer function (see Abramowitz and Stegun (1964)). We have used the following notations

$$a = 4h(h - (1-\kappa)c) + c^2, \quad b = (2h - (1-\kappa)c)^2, \qquad d = \gamma c.$$

In the particular case $\gamma = 0$, we obtain

$$I_\lambda(h, c; \kappa) := I_\lambda(h, c; \kappa, \gamma = 0) = \tag{44}$$

$$2^{\frac{5+\lambda}{2}}\Gamma\left(\frac{3+\lambda}{2}\right)\frac{(3+\lambda)[2h - (1-\kappa)c]^2 - (2h - c)^2 - 4\kappa ch}{[(2h-c)^2 + 4h\kappa c]^{\frac{5+\lambda}{2}}}.$$

Figure 2 shows a 3D plot of the diagram obtained from (21), (22), (43), (44) of the most efficient bridge variance estimator, for $\kappa = 0.95$ and zero drift ($\gamma = 0$). On the plane (θ, ϕ), the boundary of the domain $\mathcal{S}_\kappa$ (17) is shown.

5.   Comparison of the most efficient bridge estimators with the G&K and PARK estimators

5.1.   *Expectation and variance of arbitrary canonical bridge estimators*

Below we compare the efficiency of the most efficient bridge homogeneous estimators with that of the G&K and PARK estimators. We thus give the formulas for the expected value and the variance of arbitrary canonical bridge homogeneous OHLC estimators defined by (15). First, their expected values are $\mathcal{M}_\lambda(\kappa, \gamma)$ given by (18).

In general, estimators (15) are biased. Thus, one needs a normalization procedure for a comparison. For definiteness, we normalize the estimators at zero drift ($\gamma = 0$). Thus, for each estimator (15), we consider its normalized version

$$\tilde{e}_\lambda = R^\lambda \frac{\psi_\lambda(\Theta, \Phi)}{\mathcal{M}_\lambda(\kappa)},$$
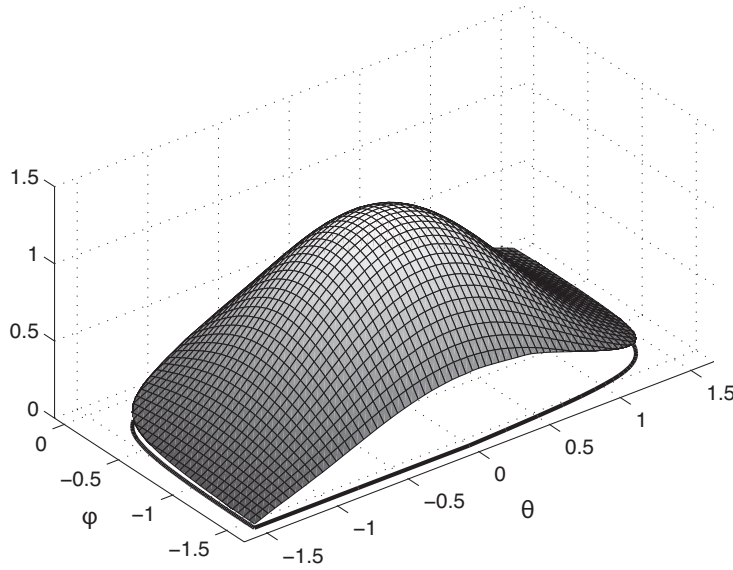
Figure 2. Diagram of the most efficient variance estimator, for $\kappa = 0.95$ and $\gamma = 0$.

where $\mathcal{M}_\lambda(\kappa) = \mathcal{M}(\kappa, \gamma = 0)$. The expected value of the normalized estimator is

$$\mathrm{E}[\tilde{e}_\lambda|\kappa, \gamma] = \frac{\mathcal{M}_\lambda(\kappa, \gamma)}{\mathcal{M}_\lambda(\kappa)}.$$

The variance of the normalized (at $\gamma = 0$) estimator is

$$\mathrm{Var}[\tilde{e}_\lambda|\kappa, \gamma] = \frac{\mathcal{N}_\lambda(\kappa, \gamma) - \mathcal{M}_\lambda^2(\kappa, \gamma)}{\mathcal{M}_\lambda^2(\kappa)},$$

$$\mathcal{N}_\lambda(\kappa, \gamma) = \iint\limits_{\mathcal{S}_\kappa} \psi_\lambda^2(\theta, \phi) g_{2\lambda}(\theta, \phi; \kappa, \gamma) \cos\theta \, d\theta d\phi.$$

5.2.   *Modified G&K bridge estimator*

We recall that the G&K canonical variance estimator is given by

$$\hat{e}_{\mathrm{GK}} = k_1(H_0 - L_0)^2 - k_2(C(H_0 + L_0) - 2H_0 L_0) - k_3 C^2, \tag{45}$$

$$k_1 = 0.511, \qquad k_2 = 0.019, \qquad k_3 = 0.383.$$

The random variables $\{H_0, L_0, C\}$ are the high, low and close values of the Wiener process $X(t, \gamma)$ with drift (5). In order to compare the efficiencies of the G&K and of most efficient bridge estimators, we modify the G&K estimator (45) by replacing the high, low and close values of the process $X(t, \gamma)$ by the high, low $\{H, L\}$ (9) and close values of the incomplete bridge $Y(t, \kappa, \gamma)$ defined by (4). This yields

$$\hat{e}_{\mathrm{GK}}(\kappa) = k_1(H - L)^2 - k_2((1 - \kappa)C(H + L) - 2HL) - k_3(1 - \kappa)^2 C^2. \tag{46}$$

The estimator (46) can be expressed in a form analogous to (15),

$$\hat{e}_{\mathrm{GK}} = R^2 \psi_{\mathrm{GK}}(\Theta, \Phi, \kappa), \tag{47}$$

with

$$\psi_{\mathrm{GK}}(\theta, \phi, \kappa) = k_1 \cos^2 \theta (\cos \phi - \sin \phi)^2$$
$$+ k_2 \left[\cos^2 \theta \sin 2\phi - \frac{1 - \kappa}{2} \sin 2\theta (\cos \phi + \sin \phi) \right] - k_3 (1 - \kappa)^2 \sin^2 \theta.$$

To compare efficiencies of the modified G&K estimator and of the most efficient bridge estimators of any order λ, we introduce the G&K estimator of order λ:

$$\tilde{e}_{\mathrm{GK},\lambda} = \frac{R^\lambda}{\mathcal{M}_{\mathrm{GK},\lambda}(\kappa)} \, \psi_{\mathrm{GK}}^{\lambda/2}(\Theta, \Phi, \kappa), \tag{48}$$

where $\mathcal{M}_{\mathrm{GK},\lambda}(\kappa)$ is given by the following expression

$$\mathcal{M}_{\mathrm{GK},\lambda}(\kappa) = \iint\limits_{\mathcal{S}_\kappa} \psi_{\mathrm{GK}}^{\lambda/2}(\theta, \phi, \kappa) g_\lambda(\theta, \phi; \kappa) \cos \theta d\theta d\phi.$$

For $\kappa = 0$, $\lambda = 2$, the estimator (48) reduces to the regular G&K estimator (45).

5.3. *Modified PARK bridge estimator*

The canonical PARK variance estimator is given by

$$\tilde{s}_{\mathrm{P}} = \frac{(H_0 - L_0)^2}{4 \ln 2}. \tag{49}$$

We modify it, replacing it by the bridge estimator of order λ,

$$\tilde{e}_{\mathrm{P},\lambda} = \frac{R^\lambda}{\mathcal{M}_{\mathrm{P},\lambda}(\kappa)} \, \psi_{\mathrm{P}}^{\lambda/2}(\Theta, \Phi),$$
$$\psi_{\mathrm{P}}(\theta, \phi) = \frac{\cos^2 \theta (1 - \sin 2\phi)}{4 \ln 2}, \quad \mathcal{M}_{\mathrm{P},\lambda}(\kappa) = \iint\limits_{\mathcal{S}_\kappa} \psi_{\mathrm{P}}^{\lambda/2}(\theta, \phi) g_\lambda(\theta, \phi; \kappa) \cos \theta d\theta d\phi. \tag{50}$$

Remark 4 : Below, we compare the efficiencies of the G&K, PARK and of the most efficient estimators, and do not discuss the efficiency of the R&S estimator. Indeed, it follows from our preliminary calculations for $\kappa \simeq 1$ that the R&S bridge estimator is much less efficient than even the PARK estimator.

Remark 5 : For convenience and based on the similarity with their homonyms, we are referring to the estimators (48) and (50) as the *modified* G&K and PARK ones. However, these estimators, based on the incomplete bridge, are novel. One can recover the regular G&K and PARK estimators by taking $\kappa = 0$ in (48), (50). It will be seen below that our modified bridge G&K and PARK estimators are more efficient than their regular counterparts, as quantified by their biases and variances.
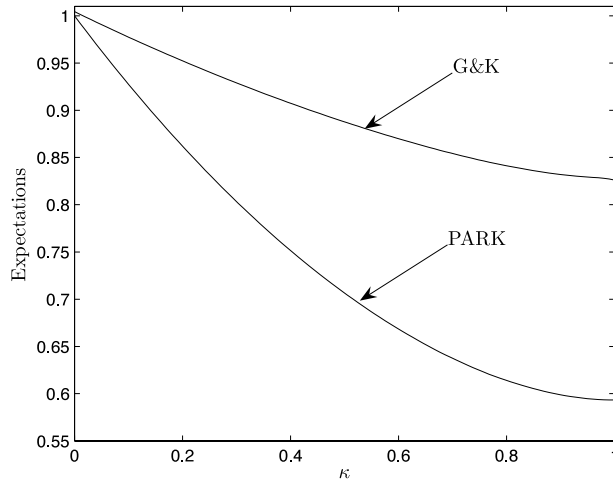
Figure 3. Dependence of the expected values of the G&K (46) and PARK (49) canonical bridge estimators as a function of the parameter κ, in the zero drift ($\gamma = 0$) case.

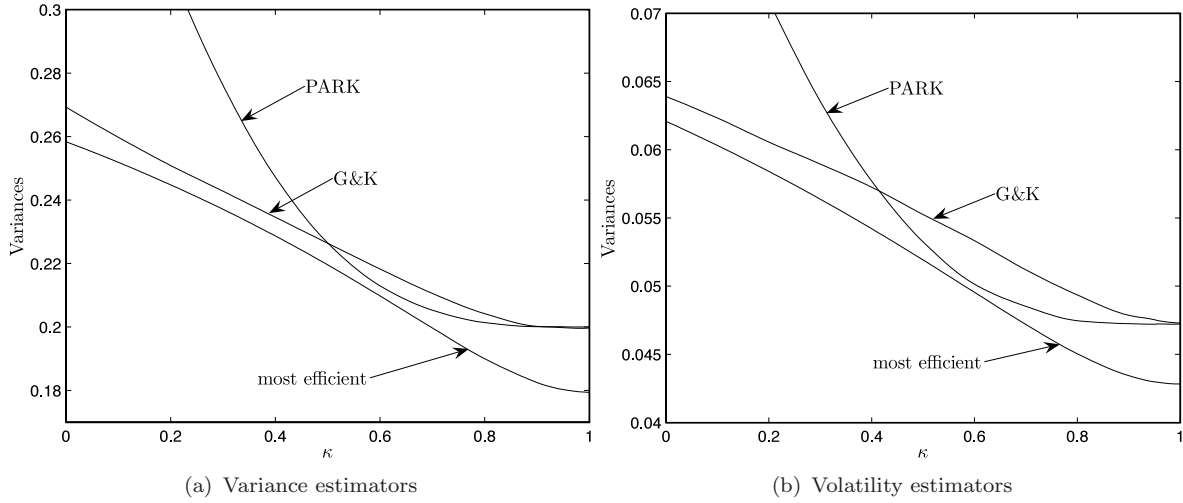(a) Variance estimators   (b) Volatility estimators

Figure 4. Variances of the most efficient, normalized G&K (48) and PARK (50) (a) variance and (b) volatility bridge estimators, as functions of the parameter κ, in the case of zero drift ($\gamma = 0$). It is seen that the closer parameter κ to unity, the more efficient are all discussed estimators.

5.4. *Comparison of the volatility estimators*

Figure 3 shows the dependence of the expected values of the G&K (46) and PARK (49) variance estimators as a function of the bridge parameter κ. One can see that, if $\kappa \neq 0$, the regular G&K and PARK estimators are biased, so one needs to compare their normalized versions (48), (50).

Figure 4(a) plots the dependencies on κ of the variances of the most efficient canonical bridge variance estimator, and of the modified G&K and PARK canonical variance estimators (48), (50). For $\kappa = 0$, i.e. in the case of "regular" OHLC estimators, the variances of the most efficient and of the G&K estimators are rather close, while the PARK estimator variance is much larger than the former ones:

$$\mathrm{Var}[\hat{e}_{\mathrm{me},2}|\kappa = 0] = 0.2584, \quad \mathrm{Var}[\tilde{e}_{\mathrm{GK},2}|\kappa = 0] = 0.2693, \quad \mathrm{Var}[\tilde{e}_{\mathrm{P},2}|\kappa = 0] = 0.4073.$$

In contrast, in the case of an almost complete bridge $\kappa \in (0.9, 1)$, the variance of the most efficient variance estimator is much smaller than the variances of the modified G&K and PARK estimators:

$$\text{Var}[\hat{e}_{\text{me},2}|\kappa = 1] = 0.1794, \quad \text{Var}[\tilde{e}_{\text{GK},2}|\kappa = 1] \simeq \text{Var}[\tilde{e}_{\text{P},2}|\kappa = 1] \simeq 0.2.$$

Similarly, figure 4(b) shows the dependencies on κ of the variances of the most efficient canonical bridge volatility ($\lambda = 1$) estimator, and the modified G&K and PARK volatility estimators (48), (50). In the case of almost complete bridges $\kappa \in (0.9, 1)$, the variance of the most efficient volatility estimator is much smaller than the variances of the modified G&K and PARK estimators:

$$\text{Var}[\hat{e}_{\text{me},1}|\kappa = 1] = 0.0428, \quad \text{Var}[\tilde{e}_{\text{GK},1}|\kappa = 1] = 0.0473, \quad \text{Var}[\tilde{e}_{\text{P},1}|\kappa = 1] = 0.0472.$$

Remark 6 : The fact that the OHLC bridge estimators, with $\kappa \simeq 1$, are more efficient than the "regular" ones, with $\kappa = 0$, can be intuitively explained as follows. It is known that the high and low values of the Wiener process are most probably found at the edges vicinity, while, by construction of the bridge, its high and low values are mostly distant from the edges of the observation interval. This is illustrated by the figure 5 (borrowed from the Lapinova *et al.* (2013)), where we present the pdf $\phi(\theta, \gamma)$ of the random instant of the high (low) value of the Wiener process with drift $X(t, \gamma)$ and the uniform pdf $\varphi(\theta)$ of the instant of the high (low) value of the corresponding complete bridge process ($\kappa = 1$). Thus, the bridge's high and low incorporate essentially more information about the behavior of the original stochastic process than its own high and low values.

6.   Synthetic most efficient estimators

In section 3, we have derived the most efficient bridge estimator diagrams (21), (22), defined in terms of the function $g_\lambda(\theta, \phi; \kappa, \gamma)$ (19). It depends in turn on the pdf $Q(h, \ell, c; \kappa, \gamma)$ (40). Notice that it is possible to determine the function $g_\lambda(\theta, \phi; \kappa, \gamma)$ even if the pdf is unknown. The function $g_\lambda(\theta, \phi; \kappa, \gamma)$ can indeed be determined by simulating $M \gg 1$ times the stochastic process $X(t)$ and then estimating $g_\lambda(\theta, \phi; \kappa, \gamma)$ by statistical averaging. This is a convenient approach when analytical formulas are not available, as occurs when considering stochastic log-price processes more complex than the Wiener process with drift.

To illustrate this possibility of simulating the diagrams associated with the most efficient estimators, consider the discrete random walk

$$X(k) = \frac{1}{\sqrt{K}} \sum_{i=1}^{k} \epsilon_i, \qquad k = 1, \ldots, K, \qquad X(0) = 0, \tag{51}$$

where $\{\epsilon_i\}$ is a sequence of iid random variables with zero expectation and unit variance. The discrete random walk (51) mimics the discrete, tick-by-tick, nature of the log-price stochastic process at the micro time scale.

In the limit $K \to \infty$, the random walk $X(k)$ (51) converges (even if $\{\epsilon_i\}$ are non-Gaussian but the tails of their pdf are not heavy) to the Wiener process $W(t)$, so that the joint pdf of the variables $\{H, L, C\}$ is known. In contrast, if the number of ticks is finite ($K < \infty$), the joint pdf is unknown. Then one can obtain an approximate expression for the diagram of the most efficient estimator by numerical simulation of paths of the discrete time process (51).
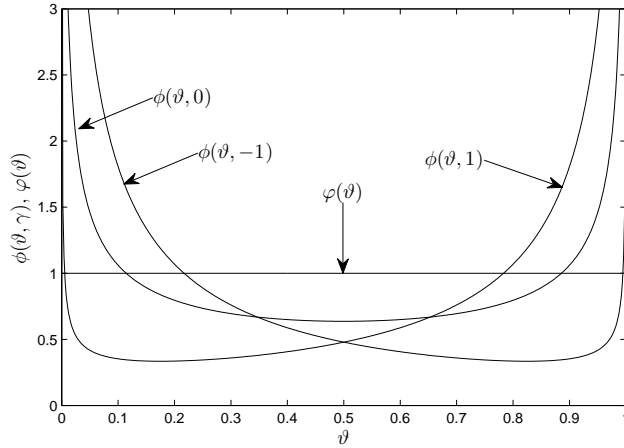
Figure 5. The pdf $\phi(\theta, \gamma)$ of the high (low) value of the Wiener process with drift $X(t, \gamma)$ for $\gamma = -1, 0, 1$ and the uniform pdf $\varphi(\theta)$ of the instant of the high (low) value of the corresponding complete bridge. The fiigure is borrowed from the Lapinova *et al.* (2013).
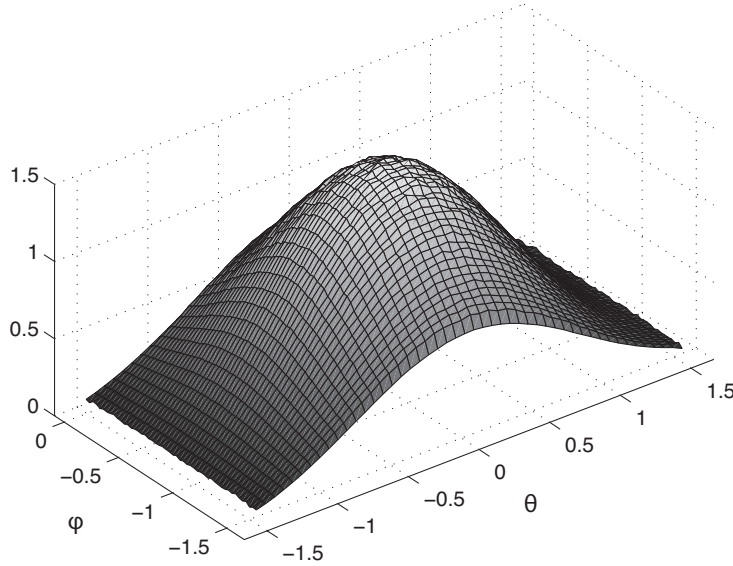

Figure 6. Synthetic diagram of the most efficient variance canonical bridge estimator for $K = 10$ and $\kappa = 1$, $\gamma = 0$.

In the present case, in order to construct such a synthetic diagram for $K = 10$; 10^2 and 10^3, for each K, we generated $M = 10^8$ realizations of the random walk $X(k)$ (51) with Gaussian summands $\{\epsilon_i\}$. We have divided the domain S_κ defined in (17) in 50×50 rectangle bins and, then calculated the function $g_\lambda(\theta, \phi; \kappa)$ (for $\gamma = 0$) using the approximate statistical relation

$$g_\lambda(\theta, \phi; \kappa) \cos\theta d\theta d\phi \simeq \frac{1}{M} \sum_{m=1}^{M} R_m^\lambda I_\delta(\Theta_m, \Phi_m) \ . \tag{52}$$

Here the set $\{\Theta_m, \Phi_m, R_m\}$ of the random variables (14) obtained from the m-th realization; I_δ
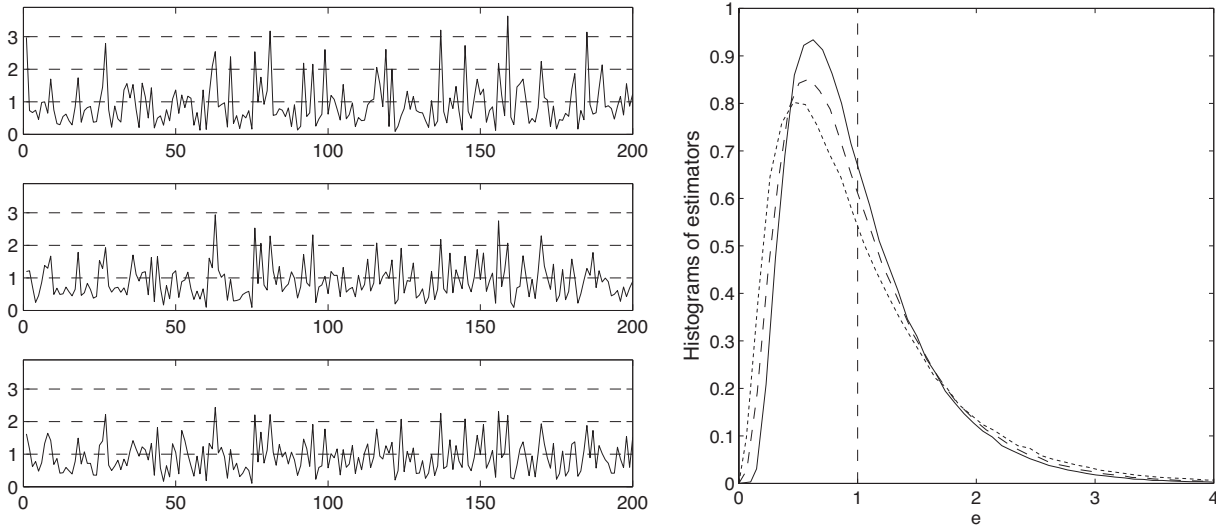
Figure 7. Left panel: G&K estimators for $\kappa = 0$ (top axes) and $\kappa = 1$ (middle axes), and most efficient canonical bridge variance estimator for $\kappa = 1$ (bottom axes), obtained for 200 realizations of a discrete random walk $X(k)$ (51) with $K = 10$.

Right panel: Empirical histograms of G&K estimators for $\kappa = 0$ (dotted line) and $\kappa = 1$ (dashed line), and most efficient canonical bridge variance estimator for $\kappa = 1$ (solid line), obtained using $2 \cdot 10^5$ realizations of a discrete random walk $X(k)$ (51) with $K = 10$.

Table 1. Variances of modified G&K and simulated most efficient variance estimators. The variances of the G&K estimators and of the synthetic most efficient variance estimators, both for $\kappa = 0$ and $\kappa = 1$, are obtained by averaging over $M = 10^6$ simulations of the discrete random walk $X(k)$ (51).

	$K = 10$	100	1000	∞
$\mathrm{Var}[\hat{e}_{\mathrm{GK},2}](\kappa = 0)$	0.5103	0.3272	0.2858	0.2693
$\mathrm{Var}[\hat{e}_{\mathrm{me},2}](\kappa = 0)$	0.4759	0.3130	0.2755	0.2584
$\mathrm{Var}[\hat{e}_{\mathrm{GK},2}](\kappa = 1)$	0.4062	0.2434	0.2125	0.1996
$\mathrm{Var}[\hat{e}_{\mathrm{me},2}](\kappa = 1)$	0.3373	0.2151	0.1896	0.1794

is the indicator of the set $\delta = (\theta, \theta + d\theta) \times (\phi, \phi + d\phi)$, which is equal to 1 within for $(\Theta_m, \Phi_m) \in \delta$ and equal to 0 outside; and the summation is performed over all M realizations of the random walk (51). The histograms of functions $g_\lambda(\theta, \phi; \kappa)$, that we obtain here, we substituted into the expression for the diagram function (21), (22).

Figure 6 presents the 3D plot of the synthetic diagram of the most efficient bridge variance estimator, obtained by statistical averaging for $K = 10$ and $\kappa = 1$, $\gamma = 0$. Notwithstanding the visible fluctuations, table 1 shows that this level of numerical approximation is sufficient to obtain much better efficient estimators than, for instance, the G&K estimator. Table 1 gives the variances of the canonical bridge variance estimators. The values shown in Table 1 have been obtained by statistical averaging over $M = 10^6$ simulations of the random walk (51).

As an illustration, we present results of numerical simulations. We have simulated 200 paths of discrete Wiener process (51) for $K = 10$ and recorded their related bridge OHLC values. Figure 7 (left) shows the samples of the synthetic most efficient canonical bridge variance estimator, for $\kappa = 1$, and the samples of the modified G&K estimators, for $\kappa = 0$ and $\kappa = 1$. The figure

shows that the synthetic most efficient bridge variance estimator is much more efficient than the G&K estimators, having much less extreme estimates. This is vividly illustrated by figure 7 (right), which presents histograms of these 3 estimators obtained from numerical simulations of $2 \cdot 10^5$ paths of the discrete Wiener process (51) for $K = 10$. Although all three estimators after correction for bias have mean value 1, the mode (most probable value) of the most efficient canonical bridge estimator is much closer to the value 1 than the modes of the G&K estimators for both $\kappa = 0$ and $\kappa = 1$. Moreover, the right tail of the canonical bridge estimator is lower than that of the two others, indicating a smaller probability of obtaining an "outlier" in the estimations.

7. Conclusions

In this article, we have pursued the development of a comprehensive theory of homogeneous volatility estimators. 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 pdfs of the high-minus-open, low-minus-open and close-minus-open values of the original log-price process and its bridge. The diagrams can be tailored to yield the most efficient estimators associated to any statistical properties of the underlying log-price stochastic process.

Previous works have developed variance estimators which are quadratic functions of the OHLC. Our main contribution is to stress that quadratic estimators are only particular cases of general homogeneous estimators. Our theory gives the tools to find most efficient homogenous estimators which, by construction, are always more efficient than the quadratic estimators. Another advantage of homogeneous estimators is that they give the possibility to develop efficient volatility in addition to variance estimators, while quadratic estimators are specialized to variance estimators.

Our theory opens several interesting developments. First, the determination of the key functions $g_\lambda(\theta, \phi; \gamma)$ provides the most efficient bridge volatility estimators for non-Gaussian log-price processes. For more realistic price processes for which the analytic distribution of the OHLC is not known, the most efficient estimators can be found via the diagram obtained from general relations (21) and (22) together with the statistical relation (52). An example of this procedure applied to a non-Wiener process, the discrete random walk $X(k)$ defined by (51) mimicking the discrete nature of real tick-by-tick trading process, is presented in section 6. Our methods should lead to the development of effective algorithms for high-frequency OHLC bridge volatility estimators, which can be applied in practice to any kind of financial markets.

References

Abramowitz, M., and Stegun, A., Handbook of Mathematical Functions, 1964 (National Bureau of Standards Applied Mathematics Series - 55).

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

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

Bacry, E., Delattre S., Hoffmann M., and Muzy J.-F., Modeling Microstructure Noise with Mutually Exciting Point Processes. *Quantitative Finance*, 2013, **13(1)**, 65-77.

Barndor-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica, 2008, 76 (6), 1481-1536.

Barndor-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, Subsampling realised kernels. Journal of Econometrics, 2011, 160 (1), 204-219.

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

Cont, R. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance*, 2001, **1**: 223236.

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

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

Lapinova S., Saichev A. and Tarakanova M. Efficiency and Probabilistic Properties of Bridge Volatility Estimator. *Physica A*, 2013, **392 (6)**: 14391451.

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

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

Protter, P., Stochastic Integration and Differential Equations, 2005 (Springer Verlag, NY)

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

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

Saichev A., Sornette D. A simple microstructure return model explaining microstructure noise and Epps effects [online] Swiss Finance Institute Research Paper No. 12-08. 2012. Available online at: http://ssrn.com/abstract=2009392.

Saichev, A., Sornette, D. and Filimonov, V., Most Efficient Homogeneous Volatility Estimators [online]. ETH Zurich working paper, 2009. Available online at: http://ssrn.com/abstract=1470004.

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

Zhang, L., Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli, 2006, 12 (6), 1019-1043.

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

Appendix A: Proof of Theorem 3.3

Proof: For given values of the parameters κ and $\gamma = \gamma_0$, the variance of the unbiased canonical estimator, with diagram (20), is equal to

$$\mathrm{Var}[\hat{e}_\lambda; \kappa, \gamma_0] = \frac{\iint\limits_{\mathcal{S}_\kappa} G^2(\theta, \phi)\, g_{2\lambda}(\theta, \phi; \kappa, \gamma_0) \cos\theta d\theta d\phi}{\left(\iint\limits_{\mathcal{S}_\kappa} G(\theta, \phi)\, g_\lambda(\theta, \phi; \kappa, \gamma_0) \cos\theta d\theta d\phi\right)^2} - 1. \tag{A1}$$

Using the Schwarz inequality

$$\left(\iint\limits_{\mathcal{S}_\kappa} A(\theta, \phi) B(\theta, \phi) d\theta d\phi\right)^2 \leqslant \iint\limits_{\mathcal{S}_\kappa} A^2(\theta, \phi) d\theta d\phi \iint\limits_{\mathcal{S}_\kappa} B^2(\theta, \phi) d\theta d\phi,$$

for arbitrary locally integrable real-valued functions $A(\theta, \phi)$ and $B(\theta, \phi)$, we take

$$A(\theta, \phi) = G(\theta, \phi) \sqrt{g_{2\lambda}(\theta, \phi; \kappa, \gamma_0) \cos\theta},$$

$$B(\theta, \phi) = g_\lambda(\theta, \phi; \kappa, \gamma_0) \sqrt{\frac{\cos\theta}{g_{2\lambda}(\theta, \phi; \kappa, \gamma_0)}},$$

and obtain

$$\left(\iint\limits_{\mathcal{S}_\kappa} G(\theta, \phi) g_\lambda(\theta, \phi; \kappa, \gamma) \cos\theta d\theta d\phi\right)^2 \leqslant$$

$$\iint\limits_{\mathcal{S}_\kappa} G^2(\theta, \phi) g_{2\lambda}(\theta, \phi; \kappa, \gamma_0) \cos\theta d\theta d\phi \iint\limits_{\mathcal{S}_\kappa} \frac{g_\lambda^2(\theta, \phi; \kappa, \gamma_0)}{g_{2\lambda}(\theta, \phi; \kappa, \gamma_0)} \cos\theta d\theta d\phi.$$

It follows from (A1) and from the above inequality that the variance of any canonical homogeneous volatility estimator of order λ satisfies the inequality

$$\mathrm{Var}[\hat{e}_\lambda; \kappa, \gamma_0] \geqslant V_\lambda(\kappa, \gamma_0), \qquad V_\lambda(\kappa, \gamma) = \frac{1}{\mathcal{E}_\lambda(\kappa, \gamma)} - 1, \tag{A2}$$

where $\mathcal{E}_\lambda(\kappa, \gamma)$ is defined by expression (22). It follows from (A1), (A2) and (22) that the variance of the canonical volatility estimator of order λ reaches its minimal value for a given $\gamma = \gamma_0$ and κ, if $G(\theta, \phi)$ is given by the left equality of (22). $\qquad\qquad\square$

Appendix B: Useful properties of the solutions of diffusion equations

In order to solve explicitly the initial-boundary problem (34), (35), (36), it is useful to present some general properties of its solutions. Firstly, if $\varphi(\omega; \tau)$ is some solution of the diffusion equation (34), then $A\varphi(\omega + a; \tau)$, where a and A are arbitrary constants, is also a solution. Such

relation tying together different solutions of the same diffusion equation can be written as

$$\varphi(\omega; \tau) \qquad \longleftrightarrow \qquad A\varphi(\omega + a; \tau). \tag{B1}$$

In order to solve the initial-boundary problem (34),–(36), we will need two lemmas.

Lemma B.1:   *If $\varphi(\omega, \tau)$ of the form*

$$\varphi(\omega; \tau) = \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^{\infty} \varphi(y) \exp\left(-\frac{(\omega - y)^2}{2\tau}\right) dy \tag{B2}$$

is a solution of the diffusion equation (34), satisfying the initial condition

$$\varphi(\omega; t) = \varphi(\omega),$$

where $\varphi(\omega)$ is such that $\varphi(\omega, \tau)$ is a continuous function of ω for any $\tau > 0$, then it generates a family of continuous solutions via the transformation

$$\varphi(\omega; \tau) \qquad \longleftrightarrow \qquad A\varphi(2\alpha\tau - \omega; \tau)\, e^{2\alpha(\alpha\tau - \omega)}, \tag{B3}$$

where A and α are arbitrary constants.

Proof: Let us write the function on the right of the relation (B3) in explicit form:

$$\varphi(2\alpha\tau - \omega; \tau)\, e^{2\alpha(\alpha\tau - \omega)} = \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^{\infty} \varphi(y) \exp\left(-\frac{(2\alpha\tau - \omega - y)^2}{2\tau}\right) dy\, e^{2\alpha(\alpha\tau - \omega)}.$$

Since

$$-\frac{(2\alpha\tau - \omega - y)^2}{2\tau} + 2\alpha(\alpha\tau - \omega) = -\frac{(\omega + y)^2}{2\tau} + 2\alpha y,$$

the right-hand side of relation (B3) is a continuous solution of the diffusion equation (34), satisfying the initial condition

$$\tilde{\varphi}(\omega) = \varphi(-\omega)\, e^{2\alpha\omega},$$

analogously to (B2).   □

The second lemma needed to find the solution of the initial-boundary problem (34), (35), (36) can be stated as follows.

Lemma B.2:   *Consider the function $\varphi(\omega)$ which verifies to symmetry relation*

$$\varphi(\omega) = -\varphi(2a - \omega)\, e^{2\alpha(a - \omega)}. \tag{B4}$$

Then, the solution $\varphi(\omega; \tau)$ of the diffusion equation (B2), which is continuous with respect to ω and with initial condition equal to $\varphi(\omega)$, is vanishing on the line $\omega = a + \alpha\tau$:

$$\varphi(a + \alpha\tau; \tau) = 0, \qquad \tau > 0.$$

Proof: Consider the function

$$\tilde{\varphi}(\omega; \tau) = \varphi(2\alpha\tau + 2a - \omega; \tau) \, e^{2\alpha(\alpha\tau + a - \omega)}, \tag{B5}$$

where $\varphi(\omega; \tau)$ is given by expression (B2) and $\varphi(\omega)$ obeys to symmetry relation (B4). It follows from (B1), (B3) and from the lemma B.1 conditions, that $\tilde{\varphi}(\omega; \tau)$ satisfies the diffusion equation (34) and is, for $\tau > 0$, a continuous function of the argument ω. Expressions (B5) and (B4) ensure that the $\tilde{\varphi}(\omega; \tau)$ satisfies the initial condition

$$\tilde{\varphi}(\omega; \tau = 0) = \varphi(2a - \omega) \, e^{2\alpha(a - \omega)} = -\varphi(\omega).$$

This means in turn that

$$\tilde{\varphi}(\omega; \tau) = -\varphi(\omega; \tau) \quad \Rightarrow \quad \varphi(\omega; \tau) = -\varphi(2\alpha\tau + 2a - \omega; \tau) \, e^{2\alpha(\alpha\tau + a - \omega)}.$$

In particular

$$\varphi(a + \alpha\tau; \tau) = -\varphi(a + \alpha\tau; \tau) \quad \Rightarrow \quad \varphi(a + \alpha\tau; \tau) = 0, \quad \tau > 0.$$

$\square$

Appendix C: Proof of Theorem 4.2

Proof: Let us replace the initial condition (35) by the more general one

$$\varphi(\omega; \tau = 0) = \varphi(\omega), \qquad \omega \in (a, b). \tag{C1}$$

At the end of proof, we obtain formula (38) by taking $\varphi(\omega) = \delta(\omega)$.

The idea of the proof consists in redefining the function $\varphi(\omega)$ in (C1) outside the interval $\omega \in (a, b)$ in such a way that the solution of equation (34), supplemented by the initial condition

$$\varphi(\omega; \tau = 0) = \varphi(\omega), \qquad \omega \in (-\infty, \infty), \tag{C2}$$

satisfies the absorbing boundary conditions (36). In other words, it should be equal to zero on the lines $\omega = a + \alpha\tau$, $\omega = b + \beta\tau$, $\tau > 0$. Let us define the auxiliary function

$$\varphi_0(\omega) = \varphi(\omega) \mathrm{I}_{(a,b)}(\omega), \qquad \omega \in (-\infty, \infty), \tag{C3}$$

where $\mathrm{I}_E(x)$ is the indicator of the set E.

It follows from lemma B.2 that the solution of the diffusion equation, supplemented by the initial condition (C2), satisfies the boundary conditions (36) if $\varphi(\omega)$ obeys to symmetry relations

$$\varphi(\omega) = -\varphi(2a - \omega) \, e^{2\alpha(a - \omega)}, \qquad \varphi(\omega) = -\varphi(2b - \omega) \, e^{2\beta(b - \omega)}. \tag{C4}$$

Using the first of these two equalities and definition (C3) of the function $\varphi_0(\omega)$, let us redefine $\varphi(\omega)$ onto the interval $\omega \in (2a - b, b)$ as follows:

$$\varphi(\omega) = \varphi^0(\omega), \qquad \omega \in (2a - b, b), \qquad \varphi^0(\omega) = \varphi_0(\omega) - \varphi_0(2a - \omega) \, e^{2\alpha(a - \omega)}.$$

Then, the equalities (C4) provide the "quasiperiodic" relation

$$\varphi(\omega) = \varphi(\omega + 2(b-a)) \; e^{2(\beta-\alpha)(\omega+b-a)+2(\alpha b-\beta a)},$$

which yields

$$\varphi(\omega) = \sum_{m=-\infty}^{\infty} \varphi^m(\omega),$$

$$\varphi^m(\omega) = \varphi^0(\omega + 2(b-a)m) \; e^{2(\beta-\alpha)(\omega+m(b-a))m+2(\alpha b-\beta a)m}. \tag{C5}$$

Putting $\varphi(\omega)$ given by (C5) into (B2), we obtain the sought solution of the initial-boundary problem (34)–(36). In particular, using $\varphi_0(\omega) = \delta(\omega)$, that is

$$\varphi^0(\omega) \qquad \Rightarrow \qquad \varphi^0(\omega) = \delta(\omega) - e^{-2\alpha a} \; \delta(\omega - 2a),$$

we obtain the solution (38). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$