

A maximum likelihood approach to volatility estimation for a Brownian motion using the high, low, and close

Malik Magdon-Ismail¹ and Amir F Atiya²

¹ CS Department, Lally 207, RPI, 110 8th Street, Troy, NY 12180, USA

² Department of Computer Engineering, Cairo University, Giza, Egypt

E-mail: magdon@rpi.edu and amir@alumni.caltech.edu

Received 4 March 2003, in final form 20 June 2003

Published

Online at stacks.iop.org/Quant/1

Abstract

Volatility plays an important role in derivatives pricing, asset allocation, and risk management, to name but a few areas. It is therefore crucial to make the utmost use of the scant information typically available in short time windows when estimating the volatility. We propose a volatility estimator using the high and the low information in addition to the close price, all of which are typically available to investors. The proposed estimator is based on a maximum likelihood approach. We present explicit formulae for the likelihood of the drift and volatility parameters when the underlying asset is assumed to follow a Brownian motion with *constant* drift and volatility. Our approach is to then maximize this likelihood to obtain the estimator of the volatility. While we present the method in the context of a Brownian motion, the general methodology is applicable whenever one can obtain the likelihood of the volatility parameter given the high, low, and close information. We present simulations which indicate that our estimator achieves consistently better performance than existing estimators (that use the same information and assumptions) for simulated data. In addition, our simulations using real price data demonstrate that our method produces more stable estimates. We also consider the effects of quantized prices and discretized time.

1. Introduction

The volatility of a financial instrument is a crucial parameter for a number of reasons. It enters as a parameter in pricing formulae for derivative instruments, and plays key roles in asset allocation, and risk management. As a result, considerable attention has been devoted to the accurate estimation of volatility. Because it is recognized that volatility shifts may occur, it is imperative to use only the most recent price observations to construct an estimate of the volatility. To obtain a good estimator, one should thus attempt to make the utmost use of the small number of observations available. In this paper, we propose an estimator that uses the high and

low price information in addition to the closing price used by conventional estimators. In practice, this would be of great interest because most historical data are quoted with both the high and low in addition to the close. **Q.1**

In this paper we propose new estimates for both the volatility and the drift of a Brownian motion using a maximum likelihood approach, under the assumption that the drift and volatility are constant. Our primary focus is on the volatility. The inputs to the algorithm are the close, high, and low prices for some specified period in the past, and our approach is to derive an expression for the joint density of the high and the low given the previous close. We then construct a likelihood

function, which we maximize using a two-dimensional search technique to obtain the maximum likelihood estimates of the drift and volatility parameters. The fidelity of the high and low observations becomes questionable when the price can be quantized and the Brownian is only observed at discrete time intervals, and so we discuss how to incorporate such discretization biases.

The specific formulae that we present assume that the drift and volatility parameters of the Brownian motion are constant during some specified time period. Such an assumption is generally considered to be only an approximate model for real asset dynamics—for example, the distribution function of log-returns is usually leptokurtic [3, 6, 7], and intra-day volatilities may show some periodicities that are related to human behaviour [11]. Despite such evidence that the volatility may not be constant, the constant-volatility geometric Brownian motion model is widely used, partially on account of its analytical simplicity, and we continue with this tradition. In principle, our methodology can be extended to more complex stochastic models, at the expense of further analytical complications. The key requirement is the ability to obtain the joint density of the open, high, and low.

Our methods can be combined with explicitly predictive techniques such as ARMA/GARCH-type models [17], to address changing volatility. We do not address such issues. Our present goal is to present a useful and effective framework for incorporating the high, low, and close information that is usually available in financial data, under the assumption that the drift and volatility are constant over the period for which the high, low, and close have been obtained.

Our simulations indicate that the RMS prediction error of our estimator is about 2.6–2.7 times less than that of the estimator obtained using the closing price alone. In practical terms, using the closing price alone would require about 35–40 days' worth of data to obtain a comparable accuracy to our method based on 5 days. Other methods that use the high–low information also obtain reductions in the RMS prediction error when compared to methods based on the close, but not as substantial as those achieved using our method. For comparison, Parkinson's method obtains a reduction by a factor of about 2.2. With real data, we demonstrate that our method obtains more stable volatility estimates (and hence ones more consistent with the assumptions of the model). By using a sliding window, one can obtain a time-varying volatility estimate. It is not our explicit intent to obtain time-varying volatility estimates; however, we suggest this approach as one possibility for obtaining time-varying volatility estimates using estimators built upon the assumption of a constant volatility over short time periods. The detailed comparison of this approach with existing approaches tailored to time-varying volatility, such as GARCH [17] and stochastic volatility models [19], is not within the scope of this work—we focus on developing a better estimator for the case of constant drift and volatility.

Volatility estimates obtained using high and low prices have been considered to some extent in the literature. All previous studies have considered securities characterized by geometric Brownian motion or Brownian motion.

Parkinson [10] shows that expectation of the high minus the low squared is proportional to σ^2 , and thus constructs an estimate based on the high minus the low. Garman and Klass [5] define a quadratic function of the high, low, and close, and derive the parameters of such a function that result in the estimate being unbiased with minimum variance (their estimate is unbiased only in the case of zero drift). Rogers *et al* [12–14] propose another formula, and show that it is an unbiased estimate even for non-zero drift. The problem with these approaches is that they are not necessarily optimal estimates. In addition, they consider only one period (one day for example). By taking the average of the estimates over the days considered in the data set, unbiasedness of the estimates will prevail, but optimality will generally not be valid.

One of the advantages of the maximum likelihood approach lies in the fact that it produces estimates that are asymptotically efficient. Further, if one assumes independence among the time periods, as is customarily the case, then multiple time periods can be incorporated by using the product likelihood function. In addition, within this probabilistic framework, it is straightforward to employ a fully Bayesian, decision theoretic approach, whereby one enforces certain priors that one might have on the drift and the volatility (for example, in a risk-averse world, the drift should be higher than the risk-free rate [1]).

This paper is organized as follows. First, we develop the maximum likelihood formulation of the problem and then extensions to discrete time and price. We then present extensive simulations to compare our method to existing methods (the close estimator, Parkinson's estimator, Rogers and Satchell's estimator, and the Garman–Klass estimator). Finally we demonstrate the use of our method with real data. We compare our method for estimating the volatility with Parkinson's [10], Rogers and Satchell's [13], and with the estimate based on the close price alone.

2. Maximum likelihood approach

We consider an instrument that follows a standard Brownian motion:

$$dx(t) = \mu dt + \sigma dW(t) \quad (1)$$

where μ and σ denote respectively the drift and the volatility of the instrument, which are constants, and $W(t)$ is a standard Wiener process. Usually, financial instruments are assumed to show geometric Brownian motion, which can be converted to a standard Brownian motion using a transformation of variables. According to Ito's calculus, if $x(t)$ follows the dynamics given in (1), then $dS(t) = e^{x(t)}$ follows a geometric Brownian motion given by

$$S(t) = (\mu + \frac{1}{2}\sigma^2)S(t) dt + \sigma S(t) dW(t) \quad (2)$$

and thus we can estimate the drift and volatility of $\log S(t)$ and construct the drift and volatility of $S(t)$ as specified in (2).

Let the instrument value at time 0 be $x_0 = x$. Consider for the time being the single-period case, for example one day. We will derive a volatility estimator using the high and the low information. Later in the section, we will show how to extend

Q.2

Q.3

this estimator in a straightforward manner to the multi-period case (several days in the example that we consider).

Denote the high and the low for the period by h and l respectively, i.e.,

$$h = \sup_{0 \leq t \leq T} x(t), \quad (3)$$

$$l = \inf_{0 \leq t \leq T} x(t), \quad (4)$$

where T is the length of the period. The idea behind the proposed method is to evaluate the conditional density

$$p(h, l|x, \mu, \sigma, T) \quad (5)$$

and then obtain the μ and the σ that maximize this likelihood function. To obtain such a probability density, we revisit a classical result for the problem of first passage time of a Brownian motion with drift and with two absorbing boundaries [4]. In the first-passage problem we have a Brownian motion and two boundaries h_1 and h_2 with $h_1 < x(0) < h_2$. The first passage time is the time until $x(t)$ first crosses either of the boundaries. The density function and the distribution function of the first passage time have been derived (see [4]), and they are in the form of a series. Let the distribution function be

$$F(T|h_1, h_2, x, \mu, \sigma) \quad (6)$$

which represents the probability that first passage occurs in the interval $[0, T]$. One can see that the distribution function corresponding to the required density $p(h, l|x, \mu, \sigma, T)$ with respect to the high and low random variables is equal to $1 - F$. Hence we can obtain the density p by differentiation as follows:

$$p(h, l|x, \mu, \sigma, T) = -\frac{\partial^2}{\partial h_1 \partial h_2} F(T|h_1, h_2, x, \mu, \sigma) \Big|_{h_1=l, h_2=h} \quad (7)$$

where the right-hand side represents the second-order partial derivative with respect to the two barrier levels evaluated at the low and high. Dominé [4] has computed a series expansion for exactly this first-passage-time distribution F ; hence what remains is to compute the necessary derivatives. The formulae are tedious and their explicit form is given in the appendix—see equation (33) in the appendix, which is a series representation for the density $p(h, l|x, \mu, \sigma, T)$, and can be computed to any desired accuracy by taking sufficiently many terms. As a practical point, it is found that as the volatility decreases, more terms in the series should be computed to maintain the accuracy of the estimate.

Consider now the multi-day case. Assume that one observes a set of prices, which consists of the open x_0 of the first day and the triples $\{h_i, l_i, c_i\}_{i=1}^N$ where i indexes the consecutive days for which one has data, and h, l, c represent the high, low, close respectively. We assume that the close of any given day is the open of the next day; hence we can define the series of opens by $o_1 = x_0$ and $o_i = c_{i-1}$, for $i > 1$. (If there are gaps in the data between a close and the next open, then the contribution of this period to the likelihood function can be obtained from a Gaussian distribution, since there is no high and low information to exploit.) Because of the

Markov property, the likelihood for the set of N days becomes the product of the likelihoods of each day. The log-likelihood then becomes a sum and is given by the formula

$$L(\mu, \sigma) = \sum_{i=1}^N \log p(l_i, h_i|o_i, \mu, \sigma, T) \quad (8)$$

where $p(l_i, h_i|o_i, \mu, \sigma, T)$ is given by (7) and an explicit series representation is given in (33). $L(\mu, \sigma)$ is the function that we wish to maximize with respect to μ and σ . The values of μ and σ that maximize (8) are our estimates $\hat{\mu}$ and $\hat{\sigma}$.

3. Discrete time and price

While a maximum likelihood (or Bayesian) approach may be optimal, the volatility estimates that result in practical situations may contain errors that arise from two possible sources. The first is that the price changes are quantized—the unit of change is denoted by a *tick*; for example, in many financial markets, quotes are usually to the nearest 1/16 (US) dollar. Further, the price changes only occur when the trades are made, and the waiting time between consecutive trades is itself a random variable [3, 15]. Thus, the price is quoted at discrete time intervals in integral units of the tick. From these price data the high and low are obtained. To facilitate analysis, we will proceed like Rogers and Satchell [13] and approximate these discretization phenomena by assuming that the price process is actually a Brownian motion that is sampled at discrete times $t = 0, \tau, 2\tau, \dots$, and this price is then rounded to the nearest tick. Thus, the Brownian motion is not directly observed, the observed process being a random walk that samples the Brownian motion at regular (possibly large) finite intervals. While this model is probably only a first approximation to tick data, it is close to reality for prices quoted every 5 min (for example).

One does not expect the price quantization to significantly bias the high (or the low) one way or the other, since intuitively one might argue that sometimes the rounding will result in a positive error and sometimes a negative. On the other hand, the time discretization produces an error that is systematic and can be significant. Since the actual high (low) could occur between the times when the price is observed, the observed high (low) will always be below (above) the true high (low). This can lead to significant underestimation of the volatility. Here, we present a discussion of these discretization effects. We do not describe all the details since this is not the main theme of the paper; however, we do present the relevant formulae.

Suppose that the quoted (observed) high is h_o , and the quoted low is l_o . Let the true high and low be h and l respectively, and define the differences $\delta_h = h - h_o$ and $\delta_l = l - l_o$. δ_h and δ_l will have some joint distribution. Ideally one would like to obtain this distribution, and incorporate it into the maximum likelihood estimator according to the formula

$$p(l_o, h_o|o, \mu, \sigma, T) = \int dh dl p(l_o, h_o|l, h, o, \mu, \sigma, T) p(l, h|o, \mu, \sigma, T), \quad (9)$$

where the first term in the integral is the distribution of the observed high and low given the true high and low, and the

second term is the distribution of the true high and low obtained in equation (33) in the appendix. This is a tedious approach, and an alternative approximate one was suggested in [13]. In [13], Rogers and Satchel replace h and h^2 in the expression for their volatility estimators by the expected values of h and h^2 (and similarly for l and l^2). Since by the definitions of h_o and l_o we have that

$$\begin{aligned} E[h] &= h_o + E[\delta_h|h_o, l_o] \quad \text{and} \\ E[h^2] &= h_o^2 + 2h_o E[\delta_h|h_o, l_o] + E[\delta_h^2|h_o, l_o], \\ E[l] &= l_o + E[\delta_l|h_o, l_o] \quad \text{and} \\ E[l^2] &= l_o^2 + 2l_o E[\delta_l|h_o, l_o] + E[\delta_l^2|h_o, l_o], \end{aligned}$$

it suffices to compute $E[\delta_h|h_o, l_o]$, $E[\delta_h^2|h_o, l_o]$, and similarly for δ_l . This is the approach that we suggest for the maximum likelihood estimator: namely, replace h, l, h^2, l^2 in (33) by their expected values in order to compute the likelihood.

3.1. Discretized price

We make the approximation that in analysing the discretization effect, we can treat the maximum and minimum independently. The joint density of the maximum and the close of a Wiener process (assuming the open is zero) has been computed in [16]:

$$\begin{aligned} f_T(h, c) &= \frac{2(2h - c)}{(2\pi)^{1/2}(\sigma^2 T)^{3/2}} \\ &\times \exp[-((2h - c)^2 - 2\mu c T + \mu^2 T^2)/2\sigma^2 T]. \end{aligned} \quad (10)$$

Using this density, and assuming that the price quantization is 2ϵ , one can compute the conditional density $f_T(h|c, h \in [h_o - \epsilon, h_o + \epsilon])$, conditioning on the fact that the true high must be within $\pm\epsilon$ of the observed high. This density is given by

$$\begin{aligned} f_T(h|c, h \in [h_o - \epsilon, h_o + \epsilon]) \\ = \begin{cases} \frac{1}{Z}(2h - c) \exp\left(-\frac{4h(h - c)}{2\sigma^2 T}\right) & h \in [h_o - \epsilon, h_o + \epsilon], \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (11)$$

where $Z = \int_{h_o - \epsilon}^{h_o + \epsilon} dx (2x - c) \exp(-4x(x - c)/(2\sigma^2 T))$. It is now a routine but tedious calculation to compute the expected values, $E[\delta_h|h_o]$ and $E[\delta_h^2|h_o]$, using this conditional density. In the limit $\epsilon \rightarrow 0$, and assuming that $h_o - c > \gg \epsilon$, the result is

$$\begin{aligned} E[\delta_h|h_o] &= \frac{2\epsilon^2}{3(2h_o - c)} \left[1 - \frac{(2h_o - c)^2}{\sigma^2 T} \right] + o(\epsilon^2), \\ E[\delta_h^2|h_o] &= \frac{\epsilon^2}{3} + o(\epsilon^2). \end{aligned} \quad (12)$$

As already mentioned, we expect the effect of the rounding to be minor as can be observed from the fact that the first order correction is $o(\epsilon^2)$ which is small if ϵ is small.

The corrections for the minimum, $E[\delta_l|l_o]$ and $E[\delta_l^2|l_o]$, can be obtained in a similar manner. The joint density for the magnitude of the minimum and the close can be obtained by making the changes $\mu \rightarrow -\mu$, $c \rightarrow -c$ in the density for the

maximum. All the calculations are then analogous, and we finally arrive at

$$\begin{aligned} E[\delta_l|l_o] &= -\frac{2\epsilon^2}{3(2l_o + c)} \left[1 - \frac{(2l_o + c)^2}{\sigma^2 T} \right] + o(\epsilon^2), \\ E[\delta_l^2|l_o] &= \frac{\epsilon^2}{3} + o(\epsilon^2). \end{aligned} \quad (13)$$

One could also develop a similar correction for the close, but since the argument is similar, we do not pursue it further here. An alternative approach to estimating volatility with price quanta was given in [12], where it is assumed that rather than observing at fixed time intervals and then rounding, the price is only observed after it has moved by ϵ .

3.2. Discretized time

The effect of observing the Brownian motion at discrete time intervals is expected to have the systematic effect of underestimating the maximum and overestimating the minimum. We follow the approximation used in [13]. Once again, we write $h = h_o + \delta_h$, where now δ_h is necessarily positive. A non-zero δ_h arises from the fact that the Brownian motion can fluctuate above h_o in the intervals where the Brownian is not observed. As a first approximation, we assume that the largest $\delta_h > 0$ occurs in one of the two intervals adjacent to the time at which h_o is observed, once again following [13]. Let τ be the size of the time interval. Then, it is necessary to obtain the distribution of the maximum of a Brownian process given that it is h_o at $t = 0$ and less than h_o at $t = \pm\tau$. In [13], the expected values, $E[\delta_h|h_o]$ and $E[\delta_h^2|h_o]$, are computed. In the limit when $\tau \rightarrow 0$, the result is given by

$$\begin{aligned} E[\delta_h|h_o] &= \sqrt{2\pi} \left(\frac{1}{4} - \frac{\sqrt{2} - 1}{6} \right) \sigma \sqrt{\tau} + o(\tau^{1/2}), \\ E[\delta_h^2|h_o] &= \left(\frac{1}{12} + \frac{\pi}{16} \right) \sigma^2 \tau + o(\tau). \end{aligned} \quad (14)$$

A similar calculation can be performed for the distribution of the minimum with analogous results, namely, $E[\delta_l|l_o] = -E[\delta_h|h_o]$ and $E[\delta_l^2|l_o] = E[\delta_h^2|h_o]$.

4. Simulation results

The proposed approach involves the maximization of the likelihood function given in (8). The results may depend on the specific algorithm that is used. For our simulations, we used a standard multi-dimensional unconstrained Nelder-Mead optimization algorithm that is available in Matlab [8], as the `fminsearch` function (all our simulations were run in Matlab). In order to enforce the positivity constraint on σ , we used the trick of defining $\sigma = u^2$ and then optimizing with respect to u which is now unconstrained. As initial values for the optimization, we used the values obtained by the close estimator. The iterative optimization was continued until convergence to within error tolerances of 10^{-4} was reached for the likelihood function and the parameters.

4.1. Artificially generated data

To compare various volatility estimators, we generated artificial data for window sizes ranging from 5 to 50 days. For each day (of length $T = 1$), we simulated the Brownian motion (B) using 100 000 time steps to obtain the high, low, and close data for the window. The changes in the Brownian motion are computed according to $\Delta B = \mu \Delta t + \epsilon \sigma \sqrt{\Delta t}$ where $\Delta t = T/100\,000$ and ϵ is an independent, normally distributed random variable at each time step. Note that since we model the parameters of the Brownian motion as constants, we did not use stochastic volatility models to compare the predictors—all the predictors are developed under the assumption that the drift and volatility are constant. In principle we could also investigate how much each predictor would be affected by a stochastic volatility, including for comparison GARCH and stochastic volatility models; however, this is not within the scope of the present discussion. In our simulations, we compare different estimators by looking at their RMS prediction error ($\sqrt{E[(\hat{\sigma} - \sigma)^2]}$) using 2000 realizations for each window size.

In our first simulation, we will assume that μ is known and does not need to be estimated. It is frequently the case that this assumption is made by equating the drift to the risk-free rate (this can be done provided that $\mu \gg \sigma^2$). For our simulations, we set $T = 1$, $\mu = 0.02$, $\sigma = 0.5$, and $x_0 = 0$. We assume that the drift μ is known and only the volatility σ needs to be estimated. Shown in the first four columns of table 1 is a comparison of four methods. The first method uses the close prices only, and the estimate is given by

$$\hat{\sigma}_{\text{close}} = \sqrt{\frac{1}{NT} \sum_{i=1}^N (c_i - o_i - \mu T)^2}. \quad (15)$$

The second method (Parkinson's estimator) uses only the high and low values [10], and the estimate is given by

$$\hat{\sigma}_{\text{Park}} = \sqrt{\frac{1}{4NT \ln 2} \sum_{i=1}^N (h_i - l_i)^2}. \quad (16)$$

The third method (the Rogers–Satchell estimator) uses the high low and close prices [13], and the estimate is given by

$$\sigma_{\text{RS}} = \sqrt{\frac{1}{NT} \sum_{i=1}^N (h_i - o_i)(h_i - c_i) + (l_i - o_i)(l_i - c_i)}. \quad (17)$$

The fourth method is our method based maximum likelihood, and the estimate is given by

$$\hat{\sigma}_{\text{ML}} = \underset{v}{\operatorname{argmax}} L(\mu, v), \quad (18)$$

where $L(\mu, v)$ is given in (8).

We show the results of these simulations in the first four columns of table 1 (μ known). The results are also summarized in figure 1(a). From these results it is clear that our method produces a superior estimate, in terms of accuracy. In terms of CPU time, the competing methods are essentially

instantaneous, requiring, for example, about 100 floating point operations (flops) to estimate the volatility on a window of size 10. Our method requires roughly 10^7 flops or is about 100 000 times slower. This is on account of the need to perform an optimization as well as the time needed to compute the complex likelihood function. On a modern 2 GHz machine which requires roughly 200 clock cycles per flop, this takes roughly 1 s. Since the goal is to compute the volatility as accurately as possible, extra CPU time is not an issue, as long as it is within reasonable limits.

One might note that the Parkinson and Rogers–Satchell estimates do not rely on knowledge of μ . Thus, to fairly compare these estimates to the maximum likelihood estimator, one should not assume that μ is known. We therefore repeat the previous simulation without assuming that μ is known. In this case, both μ and σ need to be estimated—this additional concern is only relevant to the close estimator and our maximum likelihood estimator. The results of this simulation are shown in table 1 in the ' μ unknown' columns. The numbers shown in parentheses in these two columns are the resulting RMS errors in the estimates of μ for the methods. These results are also summarized in figure 1(b). Our method not only remains superior to all the other methods, but the fact that μ needs to be estimated as well has not significantly worsened the performance. Note, however, that the RMS error in μ for the close estimate is slightly better than the ML estimate. The intuition for this is that the high and low seem to convey more information about the volatility than the drift (as can also be seen by the μ -independence of the Parkinson estimate).

As one further comparison, we consider the special case of $\mu = 0$. For this particular situation, Garman–Klass [5] have constructed the optimal (in the least-squares sense) quadratic estimator as

$$\sigma_{\text{GK}} = \left(\frac{1}{N} \sum_{i=1}^N 0.511(\tilde{h}_i - \tilde{l}_i)^2 - 0.019(\tilde{c}_i(\tilde{h}_i + \tilde{l}_i) - 2\tilde{l}_i\tilde{h}_i) - 0.383\tilde{c}_i^2 \right)^{1/2}. \quad (19)$$

The tilde over the symbols indicates that one 'normalizes' the quantities by subtracting the open price. For example, $\tilde{h}_i = h_i - o_i$. A comparison between this estimator and our maximum likelihood estimator for the case of $\mu = 0$ is given in the last two columns of table 1. In this special case of zero drift, our estimator does not beat the Garman–Klass estimator for small window sizes, but it approaches the optimal estimator as the window size increases. This is to be expected since the Garman–Klass estimator is the optimal quadratic estimator. These results are also summarized in figure 2(a), along with the performance of the other estimators for this special case of $\mu = 0$. From figure 2, it is seen that the Garman–Klass estimator is slightly better than the maximum likelihood estimator; however, one can also note that the maximum likelihood estimator is asymptotically approaching the Garman–Klass estimator as one might expect due to the asymptotic efficiency of maximum likelihood estimators. This convergence to optimality is more evident from figure 2(b), where we show

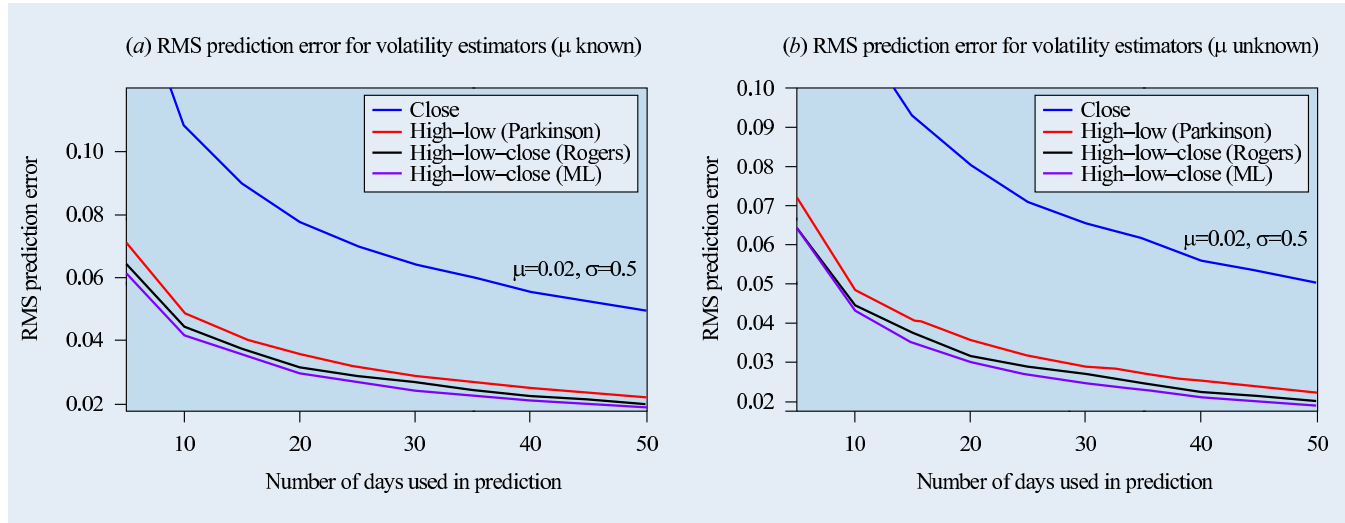


Figure 1. Comparison of volatility prediction methods using the RMS prediction error (a) when the drift is known, and (b) when the drift is unknown and has to be estimated as well.

Table 1. Comparison of different volatility estimation methods.

Days used	RMS volatility prediction error							
	μ known				μ unknown		$\mu = 0$	
	Close	Park	R-S	ML	Close	ML	G-K	ML
5	0.1597	0.0713	0.0641	0.0619	0.1752 (0.2254)	0.0639 (0.2471)	0.0591	0.0640
10	0.1090	0.0489	0.0448	0.0424	0.1152 (0.1618)	0.0434 (0.1747)	0.0399	0.0417
15	0.0900	0.0410	0.0375	0.0351	0.0930 (0.1289)	0.0354 (0.1415)	0.0337	0.0346
20	0.0781	0.0360	0.0317	0.0303	0.0808 (0.1098)	0.0307 (0.1208)	0.0292	0.0304
25	0.0702	0.0317	0.0289	0.0270	0.0709 (0.0996)	0.0270 (0.1080)	0.0260	0.0271
30	0.0645	0.0292	0.0270	0.0245	0.0654 (0.0910)	0.0248 (0.1007)	0.0233	0.0245
35	0.0605	0.0272	0.0245	0.0227	0.0615 (0.0847)	0.0229 (0.0919)	0.0224	0.0232
40	0.0556	0.0252	0.0227	0.0213	0.0559 (0.0797)	0.0215 (0.0849)	0.0205	0.0215
45	0.0526	0.0238	0.0215	0.0199	0.0534 (0.0738)	0.0202 (0.0816)	0.0192	0.0196
50	0.0499	0.0222	0.0204	0.0189	0.0505 (0.0689)	0.0191 (0.0753)	0.0186	0.0191

the difference in RMS prediction error between the various estimators and the optimal Garman–Klass estimator. The main drawback of the Garman–Klass estimator is that it only applies to the case of zero drift.

4.2. Real data

We have also applied our method to real data. We tested the methods on many different stocks, and obtained similar results each time. We only show the results from IBM. We used high, low, and close price data from 2 January 1987 to 29 October 1999, a total of 3241 trading days. These data are commercially available, and we obtained them from [9]. We set the open price for a day as the close price for the previous day, and we discard the first day (as we have no open price for this day). We report the results of estimating the volatility from a 10-day window of high–low–close prices. This window is shifted a day at a time to generate volatility estimates based on different 10-day windows. As the window is moved from left to right, the volatility estimates for the close estimator, Parkinson’s high–low estimator, Rogers and Satchell’s high–low–close estimator, and our maximum likelihood estimator are shown in figure 3, column (a). Notice the large peak

in the close estimate, around day 200. This peak is present in all the estimators, but shrinks as the estimator gets better. This peak corresponds to all the 10-day windows that contain Black Monday, 19 October 1987, corresponding to the stock market crash of October 1987. The other peaks in the close estimate centred around days 1500, 2700, 3200 have largely disappeared in the better estimates, and can be considered spurious peaks, not real outliers like the 19 October 1987 crash.

Since no ground truth is available against which to measure the various estimates, we compare the methods on the basis of stability measures of the volatility estimate. The volatility should not vary drastically from day to day. At most, it should be slowly varying, given that we model the stock process using a constant volatility. We define a measure of stability of the volatility estimate which we call the *variability* of the volatility estimate. The variability is the standard deviation of the estimate over twenty consecutive 10-day windows. As these twenty windows move from left to right, the variability of the estimate is plotted in figure 3, column (b). The average variability of the estimate is also given in the figure. The standard deviation of the estimates over the entire period is given in table 2.

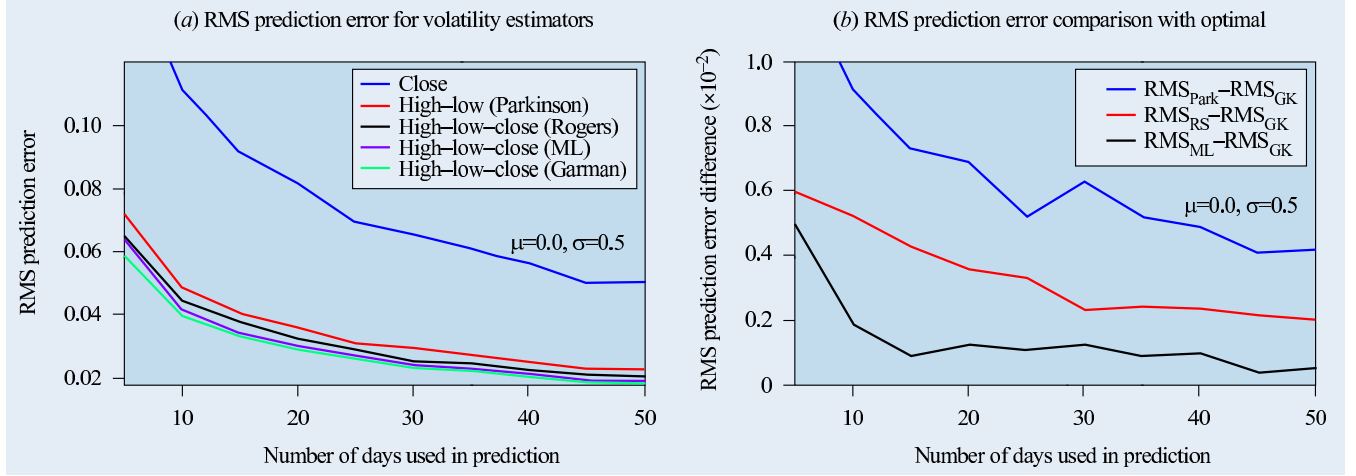


Figure 2. Comparison of volatility prediction methods using the RMS prediction error for the special case of zero drift.

Table 2.

$\hat{\sigma}_{ML}$	$\hat{\sigma}_{RS}$	$\hat{\sigma}_{Park}$	$\hat{\sigma}_{close}$
0.0751	0.0912	0.104	0.134

It is clear both from figure 3(b) and the table above that the maximum likelihood estimate is the most stable estimate.

Figure 3, column (c) shows the histogram of the logarithm of the volatility estimates over the entire time period from 1987 to 1999. Also shown is the normal distribution having the same mean and variance. The log-volatility estimates are remarkably close to a normal distribution, indicating that the actual volatility itself appears to be distributed log-normally, in accordance with previous studies [11, 18, 19]. We have also included in the figure the histogram that should have resulted had the volatility been a constant equal to its mean over the entire period. Notice that even if the volatility is a constant, the estimate will not be constant due to random fluctuations. In all cases, the histogram that would arise due to random fluctuations is narrower than the observed histogram, suggesting that the volatility may not be accurately modelled as a constant. However, note that using the small-window estimate does give a log-normally distributed estimate for the volatility, in accordance with [11, 18, 19]. This suggests that this approach may be applicable to stochastic volatility estimation. In general, when using a constant-volatility method to produce time-varying volatility estimates by using a sliding window, it is desirable to use as small a prediction window as one can afford while still getting a reasonable estimate of the volatility. Under such a paradigm, our method will outperform the existing constant-volatility methods, as we have demonstrated by simulation. However, it is not clear how such an approach would fare against other methods that are specifically tailored to time-varying volatility, such as GARCH and stochastic volatility methods. Such a detailed comparison by simulation is not within the scope of the present discussion, and we leave such issues as avenues for future work.

5. Conclusions

We have presented a formula for obtaining the joint distribution of the high and low given the open and the parameters of the Wiener process. Using this formula, one can construct a likelihood for the observed data given the parameters and hence obtain a maximum likelihood estimate for the parameters. One could also employ a fully Bayesian framework to obtain a Bayes optimal estimator under some risk measure, if one had some prior information on the possible values of the volatility. We have shown that in simulations, our estimator achieves a significant improvement over the conventional close estimator as well as other estimators based on the high, low, and close values. Further, we have demonstrated that our method produces a much more stable estimator when using real data, thus enabling one to make reliable volatility estimates using the few most recent data points. It is expected that a more accurate estimate of the time-varying volatility will lead to more efficient pricing of volatility-based derivative instruments such as options.

Appendix. Joint density of the high and the low

In this appendix, we compute a series expansion for the joint density of the high and the low given the open, and the drift and volatility parameters. This is the expression that is needed for the computation of the likelihood as given in (8). We start with the distribution function (6). A series expansion is derived in [4] which we reproduce here for convenience:

$$F(T|h_1, h_2, x, \mu, \sigma) = \sum_{k=1}^{\infty} 2\sigma^4 k\pi \tilde{C}(k, x, \mu, \sigma, h_1, h_2) \times f(k, \mu, \sigma, T, (h_1 - h_2)^2) \quad (20)$$

where the functions \tilde{C} and f are given by

$$\tilde{C}(k, x, \mu, \sigma, h_1, h_2) = \left[\exp\left(\frac{\mu}{\sigma^2}(h_2 - x)\right) (-1)^{k+1} + \exp\left(\frac{\mu}{\sigma^2}(h_1 - x)\right) \right] \sin\left(k\pi \frac{x - h_1}{h_2 - h_1}\right) \quad (21)$$

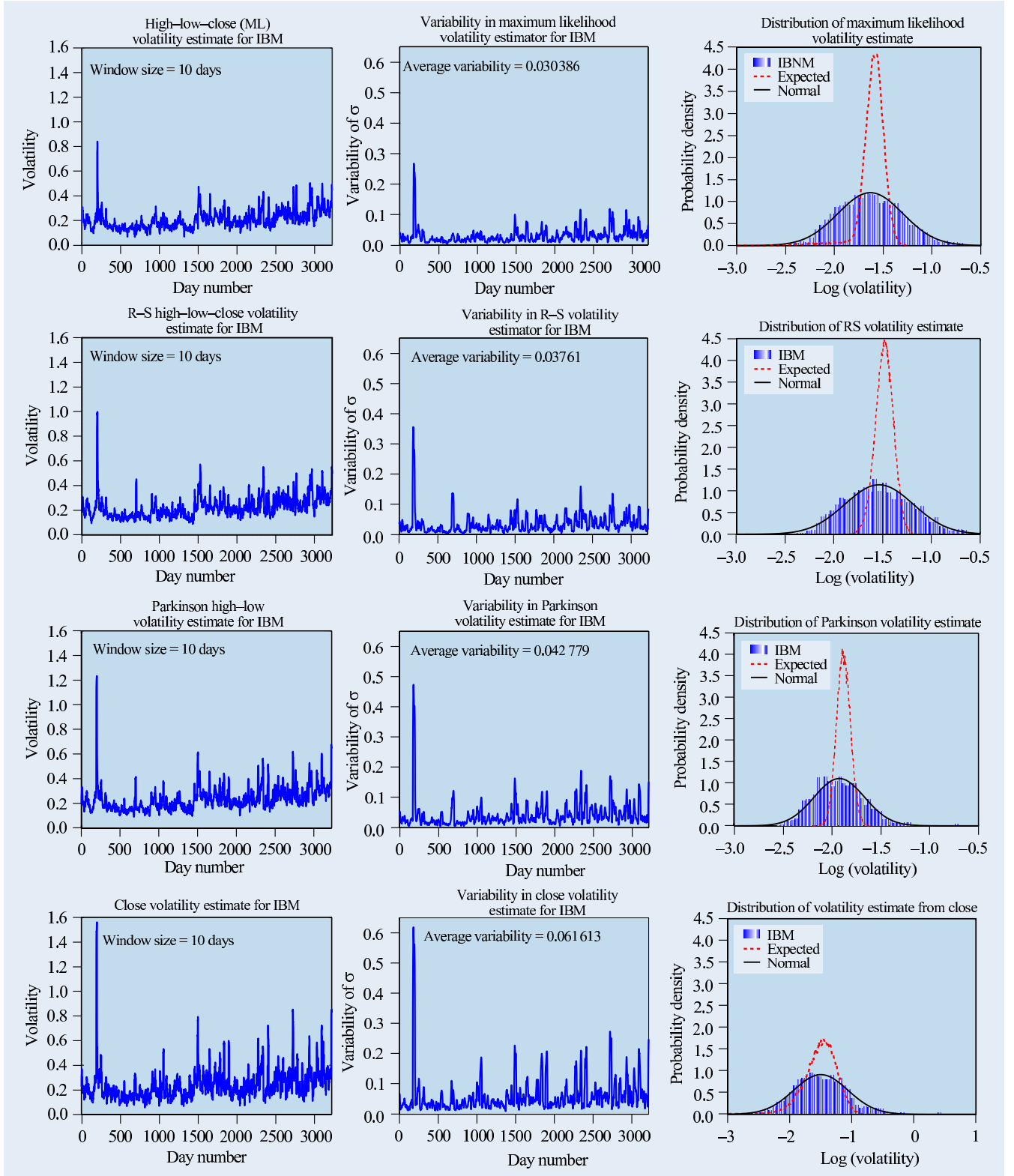


Figure 3. Volatility estimates based on IBM's return series for various predictors. From top to bottom, the predictors are (i) $\hat{\sigma}_{ML}$, (ii) $\hat{\sigma}_{RS}$, (iii) $\hat{\sigma}_{Park}$, (iv) $\hat{\sigma}_{close}$. The columns from left to right are (a) the volatility estimates; (b) the variability of the volatility estimates; (c) a histogram of the volatility estimates showing the normal with the same mean and variance as well as the histogram that would have been expected had the volatility been a constant equal to the mean.

and

where

$$f(k, \mu, \sigma, T, u) = \frac{\exp(-g(k, \mu, \sigma, u)T/(2\sigma^2 u))}{g(k, \mu, \sigma, u)} \quad (22)$$

$$g(k, \mu, \sigma, u) = \mu^2 u + \sigma^4 k^2 \pi^2. \quad (23)$$

In order to make the notation more concise, we will suppress the dependence on k, x, μ, σ, T when referring to the above functions (keeping only the dependence on h_1, h_2); for example, we will write $f(u)$ instead of $f(k, \mu, \sigma, u)$. The derivatives of $\tilde{C}(h_1, h_2)$ and $f(u)$ will be needed. We will use the usual subscript notation to denote the partial derivatives with respect to the arguments; for example, $\tilde{C}_{i,j}(h_1, h_2)$ is the i th partial derivative with respect to the first argument and the j th partial derivative with respect to the second argument. Using this notation, the joint density (7) is given by

$$p(h, l|x, \mu, \sigma, T) = -F_{1,1}(h_1, h_2)|_{h_1=l, h_2=h}. \quad (24)$$

Q.8 We will need the partial derivatives $f_1(u), f_2(u), \tilde{C}_{0,1}(h_1, h_2), \tilde{C}_{1,0}(h_1, h_2)$, and $\tilde{C}_{1,1}(h_1, h_2)$. Tedious but straightforward computations yield the following expressions:

$$f_1(u) = f(u) \left[\frac{T}{2\sigma^2 u} \left(\frac{g(u)}{u} - \mu^2 \right) - \frac{\mu^2}{g(u)} \right], \quad (25)$$

$$f_2(u) = \frac{f_1(u)^2}{f(u)} + f(u) \left[\frac{T}{\sigma^2 u^2} \left(\mu^2 - \frac{g(u)}{u} \right) + \frac{\mu^4}{g(u)^2} \right]. \quad (26)$$

In order to write the derivatives of \tilde{C} more compactly, we introduce the function $A(h_1, h_2) = k\pi(x - h_1)/(h_2 - h_1)$. The derivatives of A are then given by

$$A_{0,1}(h_1, h_2) = -\frac{A(h_1, h_2)}{(h_2 - h_1)}, \quad (27)$$

$$A_{1,0}(h_1, h_2) = A(h_1, h_2) \left(\frac{1}{h_2 - h_1} - \frac{1}{x - h_1} \right), \quad (28)$$

$$A_{1,1}(h_1, h_2) = \frac{A(h_1, h_2)}{h_2 - h_1} \left(\frac{1}{x - h_1} - \frac{2}{h_2 - h_1} \right). \quad (29)$$

One then finds for the derivatives of \tilde{C} :

$$\begin{aligned} \tilde{C}_{0,1}(h_1, h_2) &= A_{0,1} \cot(A) \tilde{C} \\ &+ (-1)^{k+1} \frac{\mu}{\sigma^2} \sin(A) \exp\left(\frac{\mu}{\sigma^2}(h_2 - x)\right), \end{aligned} \quad (30)$$

$$\begin{aligned} \tilde{C}_{1,0}(h_1, h_2) &= A_{1,0} \cot(A) \tilde{C} \\ &+ \frac{\mu}{\sigma^2} \sin(A) \exp\left(\frac{\mu}{\sigma^2}(h_1 - x)\right), \end{aligned} \quad (31)$$

$$\begin{aligned} \tilde{C}_{1,1}(h_1, h_2) &= A_{1,1} \cot(A) \tilde{C} + \frac{A_{1,0} A_{0,1}}{\sin^2(A)} \tilde{C} + A_{1,0} \cot(A) \tilde{C}_{0,1} \\ &+ \frac{\mu}{\sigma^2} A_{0,1} \cos(A) \exp\left(\frac{\mu}{\sigma^2}(h_1 - x)\right), \end{aligned} \quad (32)$$

where we have suppressed the dependence on h_1, h_2 of the

functions A and \tilde{C} on the right-hand side. Finally, the function $-F_{1,1}(h_1, h_2)$ can be obtained from the following expression:

$$\begin{aligned} F_{1,1} &= 2\sigma^4 \pi \sum_{k=1}^{\infty} k [\tilde{C}_{1,1} f(u) + 2(h_2 - h_1) f_1(u) (\tilde{C}_{1,0} - \tilde{C}_{0,1}) \\ &- 4\tilde{C} f_2(u) (h_2 - h_1)^2 - 2\tilde{C} f_1(u)]. \end{aligned} \quad (33)$$

References

- [1] Basso A and Pianca P 1999 A more informative estimation procedure for the parameters of a diffusion process *Physica D* **269** 45–53
- [2] Brandimarte P 2002 *Numerical Methods in Finance: a MATLAB-based Introduction (Wiley Series in Probability and Statistics)* (New York: Wiley)
- [3] Clark P K 1973 A subordinate stochastic process model with finite variance for speculative prices *Econometrica* **41** 135–55
- [4] Dominé M 1996 First passage time distribution of a Wiener process with drift concerning two elastic barriers *J. Appl. Probab.* **33** 164–75
- [5] Garman M and Klass M 1980 On the estimation of security price volatilities from historical data *J. Business* **53** 67–78
- [6] Mandelbrot B 1963 The variation of certain speculative prices *J. Business* **36** 394–419
- [7] Mantegna R N and Stanley H E 1995 Scaling behavior in the dynamics of an economic index *Nature* **376** 46–9
- [8] MathWorks *Matlab: The Language of Technical Computing* available through <http://www.mathworks.com/>
- [9] Momentum 1996 *Historical Market Data* (Traders Press, Inc., PO Box 6206, Greenville, SC-29606, USA)
- [10] Parkinson M 1980 The extreme value method for estimating the variance of the rate of return *J. Business* **53** 61–5
- [11] Raberto M, Scalas E, Cuniberti G and Riani M 1999 Volatility in the Italian stock market: an empirical study *Physica A* **269** 148–55
- [12] Rogers L 1998 Volatility estimation with price quanta *Math. Finance* **8** 277–90
- [13] Rogers L and Satchell S 1991 Estimating variance from high, low and closing prices *Ann. Appl. Probab.* **1** 504–12
- [14] Rogers L, Satchell S and Yoon Y 1994 Estimating the volatility of stock prices: a comparison of methods that use high and low prices *Appl. Financial Econom.* **4** 241–7
- [15] Scalas E, Gorenflo R and Mainardi F 2000 Fractional calculus and continuous time finance *Physica A* **284** 376–84
- [16] Shepp L A 1979 The joint density of the maximum and its location for a Wiener process with drift *J. Appl. Probab.* **16** 423–7
- [17] Tsay R and Ruey 2002 *Analysis of Financial Time Series (Wiley Series in Probability and Statistics)* (New York: Wiley)
- [18] Wiggins J B 1987 Option values under stochastic volatility: theory and empirical estimates *J. Financial Econom.* **19** 351–72
- [19] Wilmott P and Oztukel A 1998 Uncertain parameters, an empirical stochastic volatility model and confidence limits *Int. J. Theor. Appl. Finance* **1** 175–89