The Annals of Applied Probability
1991, Vol. 1, No. 4, 504–512

ESTIMATING VARIANCE FROM HIGH, LOW AND CLOSING PRICES

BY L. C. G. ROGERS AND S. E. SATCHELL

University of Cambridge

The log of the price of a share is commonly modelled as a Brownian motion with drift, $\sigma B_t + ct$, where the constants c and σ are unknown. In order to use the Black–Scholes option pricing formula, one needs an estimate of σ, though not of c. In this paper, we propose a new estimator of σ based on the high, low, and closing prices in a day's trading. This estimator has the merit of being unbiased whatever the drift c. In common with other estimators of σ, the approximation of the true high and low values of the drifting Brownian motion by the high and low values of a random walk introduces error, often quite a serious error. We shall show how a simple correction can overcome this error almost completely.

1. Introduction. A common assumption in modelling financial markets is that the price of a share at time t may be expressed as $\exp(\sigma B_t + ct)$, where B is a standard Brownian motion on $\mathbb{R}$ and $c \in \mathbb{R}$, $\sigma \geq 0$ are fixed unknowns. In the Black–Scholes option pricing formula, the exact value of the drift c in the log-price process $X_t \equiv \sigma B_t + ct$ is not important, but the variance σ^2 enters explicitly into the formula, and, in practice, must be estimated in order to use the formula. If one saw the whole of the sample path of X, one could deduce σ^2 from the quadratic variation, but this is not helpful to a real observer who will see at best the price at a sequence of closely spaced times. A practically useful estimator of σ^2 will use only a little readily available information; and the most readily available information from a day's trading of a share is the opening and closing prices, together with the day's highest and lowest prices. Usually the number of shares traded is also easily available.

Parkinson [4] first raised this question, then Garman and Klass [2] considered the problem of building an estimator based on the high, low and closing prices in the special case $c = 0$ (so that X is just a multiple of Brownian motion). They found the minimum-variance estimator among a class of quadratic estimators, but this estimator has two drawbacks:

(1a) The estimator will be biased if used in the case of nonzero c.

(1b) In simulations, the numerical value obtained is not as close to the true value as it should be.

The first drawback is to be expected: The estimator was built on the assumption $c = 0$. The second drawback arises because in simulation, one models

Received November 1990; revised January 1991.
AMS 1980 *subject classifications*. Primary 62M05, 62P20, 90A12; secondary 60J60, 60J65.
Key words and phrases. Black–Scholes formula, option pricing, Brownian motion, Wiener–Hopf factorisation.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Applied Probability*. www.jstor.org

Brownian motion by a random walk with Gaussian steps (say). Now the maximum of the random walk will in general be *smaller* than the maximum of the Brownian motion; one is, after all, only viewing the Brownian motion at a discrete set of times. Even taking observations at closer and closer intervals of time corrects this only gradually; between any two times, the Brownian motion wriggles very wildly. This snag was identified by Garman and Klass and also by Beckers [1].

Let $S_t \equiv \sup\{X_u : u \le t\}$, $I_t \equiv \inf\{X_u : u \le t\}$ and define the estimator

$$\hat{\sigma}^2 \equiv S_1(S_1 - X_1) + I_1(I_1 - X_1). \tag{2}$$

In this paper, we shall prove the following (see Section 2):

$$E[S_t(S_t - X_t) + I_t(I_t - X_t)] = \sigma^2 t. \tag{3}$$

The important thing is that *the right-hand side is independent of* c. Thus the estimator works just as well for nonzero c, getting round (1a). Have we had to pay much for this in the case $c = 0$, where the Garman–Klass estimator is optimal? Not really: If $\hat{\sigma}^2_{GK}$ is the Garman–Klass estimator,

$$\hat{\sigma}^2_{GK} \equiv k_1(S_1 - I_1)^2 - k_2(X_1(S_1 + I_1) - 2I_1S_1) - k_3X_1^2,$$

where

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

then

$$\operatorname{var}(\hat{\sigma}^2) = 0.331\sigma^4, \qquad \operatorname{var}(\hat{\sigma}^2_{GK}) = 0.27\sigma^4.$$

Now this gets around (1a), but what of (1b)? This problem remains; but we shall propose an intelligent correction to the estimators which performs well in simulations. The analysis of this corrected estimator is carried out in Section 3, and in Section 4 we present some numerical results.

2. The unbiased estimator of σ^2. Let T be an exponential random variable with mean λ^{-1}, independent of B. Then the law of S_T is again exponential, with parameter $\alpha \equiv (\sqrt{c^2 + 2\lambda\sigma^2} - c)/\sigma^2$, and the law of $-I_T$ is exponential, with parameter $\beta \equiv (\sqrt{c^2 + 2\lambda\sigma^2} + c)/\sigma^2$. Now it is one of the features of the classical Wiener–Hopf factorisation of the Lévy process X that S_T and $S_T - X_T$ are independent and $S_T - X_T$ has the same law as $-I_T$ (see, for example, Greenwood and Pitman [3]). Hence immediately

$$\begin{aligned} ES_T(S_T - X_T) &= ES_T E(S_T - X_T) \\ &= -ES_T EI_T \\ &= 1/(\alpha\beta) \\ &= \sigma^2/2\lambda. \end{aligned}$$

But $ES_T(S_T - X_T) = \int_0^\infty \lambda e^{-\lambda t}\, dt\, ES_t(S_t - X_t)$ so inversion of the Laplace transform gives

$$ES_t(S_t - X_t) = \sigma^2 t/2$$

and a symmetric argument gives $EI_t(I_t - X_t) = \sigma^2 t/2$, from which (3) follows.

To compute the variance of this estimator, we must calculate

$$\begin{aligned} E(S_T(S_T - X_T) + I_T(I_T - X_T))^2 &= ES_T^2(S_T - X_T)^2 + 2ES_TI_T(S_T - X_T)(I_T - X_T) + EI_T^2(I_T - X_T)^2 \\ &= 2ES_T^2E(S_T - X_T)^2 + 2ES_TI_T(S_T - X_T)(I - X_T) \\ &= 8/\alpha^2\beta^2 + 2ES_TI_T(S_T - X_T)(I_T - X_T) \end{aligned}$$

since S_T and $S_T - X_T$ are independent exponentials of parameters α, β, respectively;

$$= 2\sigma^4/\lambda^2 + 2ES_TI_T(S_T - X_T)(I_T - X_T).$$

The cross-moments are hard to compute in the case of nonzero drift, but in the case $c = 0$, Garman and Klass have computed enough of the moments to evaluate these; the answer $\text{var}(\hat{\sigma}^2) = 0.331\sigma^4$ comes from applying this.

Notice that the L^2 triangle inequality gives a quick estimate for EY^2, where $Y = Y_1 + Y_2 = S_1(S_1 - X_1) + I_1(I_1 - X_1)$, because, as follows easily from the joint law of S_T and $(S_T - X_T)$, $EY_1^2 = EY_2^2 = \sigma^4/2$; thus $EY^2 \le 2\sigma^4$ and $\text{var}(\hat{\sigma}^2) \le \sigma^4$. Moreover, this bound is, of course, valid for *any* drift c, and not just $c = 0$, which was assumed in the exact computation above.

3. Correcting the estimator. We shall now propose a correction to the estimators of Section 2 which largely overcomes (1b). The amount by which the random walk simulation underestimates S_1 will depend on the fineness of the mesh chosen; the more steps taken by the random walk in the time interval [0, 1], the better the approximation to S_1 we shall obtain. So we shall assume that we know the number N of steps taken by the random walk during [0, 1], and shall take $h \equiv 1/N$. Thus if X is the log-price process, we shall assume that we see the maximum and minimum of $\{X_{kh}: 0 \le k \le N\}$ and X_1, and must estimate S_1 from this information.

Let us just review the situation. If S_1 denotes the maximum of X by time 1, and S denotes the maximum of the embedded random walk, then we have that

$$S_1 = S + \Delta,$$

say, so that

$$S_1(S_1 - X_1) = \Delta^2 + (2S - X_1)\Delta + S(S - X_1).$$

Similarly, if I denotes the minimum of the random walk by time 1, then $I_1 = I - \tilde{\Delta}$ where $\tilde{\Delta}$ has the same law as Δ, and so

$$I_1(I_1 - X_1) = \tilde{\Delta}^2 - \tilde{\Delta}(2I - X_1) + I(I - X_1).$$

Thus the estimator $\hat{\sigma}^2$ is

$$\begin{aligned} \hat{\sigma}^2 = (\Delta^2 + \tilde{\Delta}^2) + \Delta(2S - X_1) - \tilde{\Delta}(2I - X_1) \\ + S(S - X_1) + I(I - X_1). \end{aligned} \tag{4}$$

Now we shall show that $E\Delta \doteqdot a\sigma\sqrt{h}$ for some a (in fact, $a \equiv \sqrt{2\pi}[\frac{1}{4} - (\sqrt{2} - 1)/6]$), and that $E\Delta^2 \doteqdot b\sigma^2 h$ for some other constant b (in fact, $b \equiv (1 + (3\pi/4))/12$). Thus we propose to use the estimator $\hat{\sigma}_h$, where $\hat{\sigma}_h$ is the positive root of the equation

$$(5) \qquad \hat{\sigma}_h^2 = 2b\hat{\sigma}_h^2 h + 2(S - I)a\hat{\sigma}_h\sqrt{h} + S(S - X_1) + I(I - X_1),$$

obtained from (4) by replacing Δ, $\tilde{\Delta}$, Δ^2 and $\tilde{\Delta}^2$ by their expected values. All that remains is to compute the values of a and b.

Suppose that the maximum value S taken by the random walk $(X_{kh})_{k=1}^N$ is x, achieved at time t, a multiple of h. We shall ignore the possibility that X *may take a value greater than x outside of the interval* $[t - h, t + h]$, so that the underestimate $\Delta \equiv S_1 - S = S_1 - x$ is assumed to be well approximated by $\sup\{X_s: |t - s| \le h\} - x$. If $H_x \equiv \inf\{u: X_u = x\}$, we have

$$(6) \qquad \begin{aligned} &P[H_x \in ds | t - h < H_x < t, X_t = x]/ds \\ &\quad = h_x(s)p_{t-s}(0) \Big/ \int_{t-h}^{t} h_x(u)p_{t-u}(0)\,du, \end{aligned}$$

where $p_t(\cdot)$ is the transition density for X, and $h_x(\cdot)$ is the density of H_x. Since the interval $[t - h, t]$ is typically small, we shall assume that $h_x(s)$ is *effectively constant through this interval* and since $p_{t-s}(0) = \exp(-c^2(t - s)/2\sigma^2)/\sqrt{2\pi\sigma^2(t - s)}$ is approximately $(2\pi\sigma^2(t - s))^{-1/2}$, we shall assume that the conditional density (6) can be approximated by

$$(7) \qquad P(t - H_x \in du | t - h < H_x < t, X_t = x)/du \doteqdot (4uh)^{-1/2}.$$

Now we consider the piece of the path of X between $H_x = t - u$ and t. This is a Brownian bridge of duration u, and as such may be represented as

$$(8) \qquad Y_s = \sigma\left(1 - \frac{2}{u}\right)W\left(\frac{us}{u - s}\right), \qquad 0 \le s \le u,$$

where W is a standard Brownian motion (see, e.g., Rogers and Williams [5], Theorem IV.40.3). The distribution of the maximum of Y is easy to compute:

$$\begin{aligned} P[Y_s > y \text{ for some } 0 \le s \le u] &= P[\sigma W_v > y(u + v)/u \text{ for some } v \ge 0] \\ &= P[W_v - yv/u\sigma > y/\sigma \text{ for some } v \ge 0] \\ &= \exp(-2y^2/u\sigma^2) \end{aligned}$$

by the well-known result on the law of the maximum of a downward-drifting Brownian motion. Hence if $Z \equiv \sup\{X_s: t - h \le s \le t\} - x$, we have

$$(9) \qquad \begin{aligned} &P(Z > \alpha | t - h < H_x < t, X_t = x) \\ &\quad = \int_0^h e^{-2\alpha^2/u\sigma^2} P[t - H_x \in du | t - h < H_x < t, X_t = x] \\ &\quad = \int_0^h e^{-2\alpha^2/u\sigma^2} \frac{du}{(4uh)^{1/2}} \\ &\quad = \int_{h^{-1}}^{\infty} e^{-2\alpha^2 s/\sigma^2} \frac{ds}{(4hs^3)^{1/2}}. \end{aligned}$$

Hence we deduce

$$E(Z|t - h < H_x < t, X_t = x) = \sigma(2\pi h)^{1/2}/8. \tag{10}$$

If now we set $Z' = \sup\{X_s: t \leq s \leq t + h\} - x$, then the situation for Z' is essentially the same as for Z, and Z and Z' are independent conditional on $X_t = x$ and $X_s < x$ for $|s - t| > h$. The same approximation to the distribution of Z' can be used, so we shall assume that Z and Z' are independent, with distribution given by (9). The amount by which the random walk maximum underestimates the maximum of X is thus (approximately) $Z \vee Z'$, whose mean value we now compute, by calculating

$$\begin{aligned}
E(Z \wedge Z') &= \int_0^\infty d\alpha \int_{h^{-1}}^\infty e^{-2\alpha^2 s/\sigma^2} \frac{ds}{(4hs^3)^{1/2}} \int_{h^{-1}}^\infty e^{-2\alpha^2 u/\sigma^2} \frac{du}{(4hu^3)^{1/2}} \\
&= \frac{\sqrt{2\pi}}{16h}\sigma \int_{h^{-1}}^\infty ds \int_{h^{-1}}^\infty du\,(s^3u^3(s+u))^{-1/2} \\
&= \frac{\sqrt{2\pi h}}{16}\sigma \int_1^\infty dv \int_1^\infty dy\,(v^3y^3(v+y))^{-1/2} \\
&= \frac{\sqrt{2\pi h}}{16}\sigma \int_0^1 ds \int_0^1 dt\,(s+t)^{-1/2} \\
&= \frac{\sqrt{2\pi h}\,(\sqrt{2}-1)\sigma}{6},
\end{aligned}$$

after some calculation. Hence we have

$$\begin{aligned}
E(Z \vee Z') &= E(Z + Z' - Z \wedge Z') \\
&= \sigma\sqrt{2\pi h}\left[\frac{1}{4} - \frac{\sqrt{2}-1}{6}\right].
\end{aligned} \tag{11}$$

Thus we have shown that $E\Delta \doteqdot E(Z \vee Z') = a\sigma\sqrt{h}$, where a is the constant stated above. It remains to prove that $E\Delta^2 \doteqdot E(Z \vee Z')^2 = b\sigma^2 h$.

Recalling the distribution (7) of Z, we first compute

$$\begin{aligned}
EZ^2 &= \int_0^\infty 2\alpha\, d\alpha \int_{h^{-1}}^\infty e^{-2\alpha^2 s/\sigma^2} \frac{ds}{(4hs^3)^{1/2}} \\
&= \sigma^2 \int_{h^{-1}}^\infty \frac{1}{2s} \frac{ds}{(4hs^3)^{1/2}} \\
&= \frac{\sigma^2 h}{6},
\end{aligned} \tag{12}$$

and then

$$
\begin{aligned}
E(Z \wedge Z')^2 &= \int_0^\infty 2\alpha\, d\alpha \int_{h^{-1}}^\infty e^{-2\alpha^2 s/\sigma^2} \frac{ds}{(4hs^3)^{1/2}} \int_{h^{-1}}^\infty e^{-2\alpha^2 u/\sigma^2} \frac{du}{(4hu^3)^{1/2}} \\
&= \sigma^2 \int_{h^{-1}}^\infty \frac{ds}{(4hs^3)^{1/2}} \int_{h^{-1}}^\infty \frac{du}{(4hu^3)^{1/2}} \frac{1}{2(s+u)} \\
&= \frac{\sigma^2 h}{8} \int_0^1 ds \int_0^1 dt \frac{\sqrt{st}}{s+t} \\
&= \frac{\sigma^2 h}{4}\left(1 - \frac{\pi}{4}\right),
\end{aligned}
\tag{13}
$$

after some calculation. Hence immediately from (12) and (13),

$$
E\left[(Z \vee Z')^2\right] = \sigma^2 h\left\{\frac{1}{3} - \frac{1}{4}\left(1 - \frac{\pi}{4}\right)\right\} = \frac{\sigma^2 h}{12}\left\{1 + \frac{3\pi}{4}\right\},
$$

which is to say

$$
E\Delta^2 = \sigma^2 \rho_h := \sigma^2 \frac{h}{12}\left\{1 + \frac{3\pi}{4}\right\}.
$$

Note that we may equally well apply this correction to the Garman–Klass estimator and in this case obtain the estimator $\hat{\sigma}^2_{GK,h}$, where $\hat{\sigma}_{GK,h}$ solves

$$
\begin{aligned}
\sigma^2 = k_1\big(S + I + 4(S - I)a\sigma\sqrt{h} + 2\sigma^2 h(b + a^2)\big) - k_2 X_1(S + I) \\
+ 2k_2\big(IS - (S - I)a\sigma\sqrt{h} - a^2\sigma^2 h\big) - k_3 X_1^2.
\end{aligned}
$$

TABLE 1

Simulation performance of the four estimators, variance 1, 400 *simulations*

c	N	$\hat{\sigma}^2$	$\hat{\sigma}^2_h$	$\hat{\sigma}^2_{GK}$	$\hat{\sigma}^2_{GK,h}$
0.000	20	0.689 ± 0.045	0.992 ± 0.061	0.678 ± 0.039	1.023 ± 0.058
0.000	100	0.856 ± 0.051	0.999 ± 0.058	0.856 ± 0.047	1.020 ± 0.057
0.000	500	0.968 ± 0.059	1.035 ± 0.062	0.962 ± 0.051	1.038 ± 0.055
0.000	2500	0.921 ± 0.051	0.948 ± 0.052	0.911 ± 0.044	0.942 ± 0.046
1.000	20	0.649 ± 0.049	1.016 ± 0.069	0.795 ± 0.050	1.274 ± 0.083
1.000	100	0.805 ± 0.054	0.967 ± 0.062	0.948 ± 0.055	1.152 ± 0.068
1.000	500	0.914 ± 0.059	0.986 ± 0.062	1.035 ± 0.055	1.124 ± 0.060
1.000	2500	0.904 ± 0.056	0.936 ± 0.057	1.049 ± 0.058	1.089 ± 0.060
2.000	20	0.540 ± 0.052	1.026 ± 0.075	1.026 ± 0.057	1.790 ± 0.099
2.000	100	0.784 ± 0.063	1.006 ± 0.074	1.331 ± 0.073	1.678 ± 0.092
2.000	500	0.948 ± 0.066	1.047 ± 0.070	1.424 ± 0.072	1.568 ± 0.080
2.000	2500	0.964 ± 0.058	1.008 ± 0.060	1.435 ± 0.074	1.497 ± 0.077
3.000	20	0.390 ± 0.046	1.070 ± 0.073	1.462 ± 0.073	2.748 ± 0.138
3.000	100	0.720 ± 0.066	1.007 ± 0.079	1.807 ± 0.085	2.339 ± 0.109
3.000	500	0.840 ± 0.065	0.963 ± 0.070	1.870 ± 0.088	2.089 ± 0.098
3.000	2500	0.922 ± 0.066	0.979 ± 0.069	1.980 ± 0.096	2.079 ± 0.101

TABLE 2
Simulation of the four estimators on daily stock prices, 400 simulations

Return	N	$10^4\hat{\sigma}^2$	$10^4\hat{\sigma}_h^2$	$10^4\hat{\sigma}_{GK}^2$	$10^4\hat{\sigma}_{GK,h}^2$
		Annual variance 0.1 = daily variance 0.0002739			
0.000	20	1.887 ± 0.124	2.717 ± 0.166	1.858 ± 0.106	2.804 ± 0.158
0.000	100	2.346 ± 0.140	2.738 ± 0.160	2.345 ± 0.130	2.795 ± 0.155
0.000	500	2.653 ± 0.160	2.835 ± 0.169	2.636 ± 0.139	2.844 ± 0.149
0.050	20	1.881 ± 0.132	2.729 ± 0.175	1.876 ± 0.109	2.857 ± 0.165
0.050	100	2.280 ± 0.141	2.662 ± 0.159	2.307 ± 0.124	2.745 ± 0.147
0.050	500	2.506 ± 0.140	2.676 ± 0.148	2.455 ± 0.123	2.646 ± 0.132
0.100	20	1.818 ± 0.129	2.652 ± 0.174	1.857 ± 0.113	2.831 ± 0.173
0.100	100	2.166 ± 0.134	2.524 ± 0.151	2.167 ± 0.116	2.578 ± 0.138
0.100	500	2.655 ± 0.146	2.833 ± 0.154	2.580 ± 0.127	2.779 ± 0.137
0.150	20	1.758 ± 0.123	2.549 ± 0.165	1.758 ± 0.103	2.679 ± 0.156
0.150	100	2.262 ± 0.147	2.644 ± 0.167	2.291 ± 0.129	2.732 ± 0.154
0.150	500	2.611 ± 0.154	2.787 ± 0.162	2.527 ± 0.130	2.722 ± 0.140
		Annual variance 0.2 = daily variance 0.0005479			
0.000	20	3.684 ± 0.269	5.335 ± 0.361	3.643 ± 0.221	5.542 ± 0.335
0.000	100	4.721 ± 0.281	5.512 ± 0.319	4.732 ± 0.254	5.638 ± 0.303
0.000	500	5.009 ± 0.302	5.349 ± 0.319	4.917 ± 0.267	5.299 ± 0.287
0.050	20	3.765 ± 0.281	5.512 ± 0.371	3.852 ± 0.227	5.896 ± 0.340
0.050	100	4.513 ± 0.293	5.276 ± 0.332	4.545 ± 0.254	5.423 ± 0.302
0.050	500	5.094 ± 0.303	5.444 ± 0.320	5.033 ± 0.265	5.428 ± 0.286
0.100	20	3.709 ± 0.275	5.392 ± 0.371	3.738 ± 0.234	5.686 ± 0.351
0.100	100	4.474 ± 0.265	5.231 ± 0.300	4.553 ± 0.23	5.429 ± 0.281
0.100	500	4.802 ± 0.273	5.129 ± 0.289	4.758 ± 0.246	5.127 ± 0.265
0.150	20	3.683 ± 0.277	5.299 ± 0.369	3.613 ± 0.225	5.456 ± 0.330
0.150	100	4.679 ± 0.300	5.452 ± 0.338	4.663 ± 0.254	5.548 ± 0.302
0.150	500	5.061 ± 0.303	5.406 ± 0.320	4.989 ± 0.271	5.379 ± 0.293
		Annual variance 0.3 = daily variance 0.0008219			
0.000	20	5.423 ± 0.387	7.876 ± 0.517	5.419 ± 0.325	8.359 ± 0.488
0.000	7100	6.880 ± 0.454	8.054 ± 0.521	7.009 ± 0.425	8.369 ± 0.508
0.000	500	7.888 ± 0.451	8.431 ± 0.477	7.833 ± 0.409	8.447 ± 0.442
0.050	20	5.782 ± 0.455	8.413 ± 0.607	5.874 ± 0.371	8.960 ± 0.558
0.050	100	6.755 ± 0.424	7.878 ± 0.476	6.755 ± 0.340	8.049 ± 0.403
0.050	500	7.426 ± 0.413	7.937 ± 0.436	7.363 ± 0.367	7.940 ± 0.397
0.100	20	5.684 ± 0.403	8.230 ± 0.534	5.631 ± 0.324	8.566 ± 0.489
0.100	100	6.604 ± 0.404	7.729 ± 0.459	6.726 ± 0.361	8.026 ± 0.429
0.100	500	7.812 ± 0.450	8.342 ± 0.473	7.660 ± 0.372	8.257 ± 0.400
0.150	20	5.090 ± 0.384	7.445 ± 0.513	5.209 ± 0.320	7.985 ± 0.483
0.150	100	7.094 ± 0.435	8.256 ± 0.490	6.946 ± 0.364	8.262 ± 0.432
0.150	500	7.917 ± 0.453	8.458 ± 0.477	7.815 ± 0.386	8.426 ± 0.414

4. Numerical results. The simulations simulated a random walk with Gaussian step distribution. If $N = 1/h$ is the number of steps taken by the random walk in $[0, 1]$, then each has mean ch and variance σh. We standardised on 400 simulations for each choice of the parameters (σ^2, c, N) chosen. The tables show the values of the estimators together with a 95% confidence interval in each case.

If one looks at the entries of the table in which the confidence interval contains the true value, one gets a good picture of how the four estimators perform relative to one another. In Table 1, we see that of the 16 readings for $\hat{\sigma}_h^2$, all but one are successful (that is, the confidence interval contains the true value.) The corrected Garman–Klass estimator is successful in three out of four cases when the drift is 0, but not successful in any other cases. This is not surprising, in that we know that $\hat{\sigma}_{GK,h}^2$ is designed to be good in precisely this

TABLE 3
Simulation of the four estimators on annual stock prices, 400 *simulations*

Return	N	$100\hat{\sigma}^2$	$100\hat{\sigma}_h^2$	$100\hat{\sigma}_{GK}^2$	$100\hat{\sigma}_{GK,h}^2$
			Annual variance 0.100		
0.000	500	0.0972 ± 0.0061	0.1039 ± 0.0064	0.0957 ± 0.0050	0.1039 ± 0.0054
0.000	2500	0.0914 ± 0.0048	0.0941 ± 0.0049	0.0933 ± 0.0044	0.0968 ± 0.0045
0.050	500	0.0956 ± 0.0055	0.1020 ± 0.0058	0.0933 ± 0.0049	0.1002 ± 0.0052
0.050	2500	0.0981 ± 0.0058	0.1010 ± 0.0059	0.0967 ± 0.0049	0.1008 ± 0.0050
0.100	500	0.0924 ± 0.0054	0.0987 ± 0.0057	0.0915 ± 0.0049	0.1006 ± 0.0055
0.100	2500	0.0947 ± 0.0054	0.0975 ± 0.0056	0.0952 ± 0.0049	0.0991 ± 0.0051
0.150	500	0.0962 ± 0.0063	0.1031 ± 0.0067	0.0986 ± 0.0060	0.1089 ± 0.0067
0.150	2500	0.0993 ± 0.0058	0.1023 ± 0.0060	0.1020 ± 0.0056	0.1079 ± 0.0060
			Annual variance 0.200		
0.000	500	0.1782 ± 0.0102	0.1907 ± 0.0108	0.1794 ± 0.0090	0.1917 ± 0.0096
0.000	2500	0.2004 ± 0.0118	0.2063 ± 0.0121	0.1977 ± 0.0104	0.2038 ± 0.0109
0.050	500	0.1905 ± 0.0112	0.2035 ± 0.0118	0.1883 ± 0.0098	0.2013 ± 0.0106
0.050	2500	0.1972 ± 0.0124	0.2033 ± 0.0127	0.2010 ± 0.0109	0.2080 ± 0.0113
0.100	500	0.1770 ± 0.0116	0.1893 ± 0.0122	0.1787 ± 0.0100	0.1949 ± 0.0109
0.100	2500	0.1959 ± 0.0109	0.2017 ± 0.0111	0.1955 ± 0.0101	0.2034 ± 0.0108
0.150	500	0.1884 ± 0.0111	0.2013 ± 0.0118	0.1877 ± 0.0099	0.2081 ± 0.0110
0.150	2500	0.2015 ± 0.0111	0.2075 ± 0.0114	0.1986 ± 0.0098	0.2092 ± 0.0104
			Annual variance 0.300		
0.000	500	0.2931 ± 0.0184	0.3132 ± 0.0195	0.2896 ± 0.0158	0.3097 ± 0.0169
0.000	2500	0.2825 ± 0.0152	0.2911 ± 0.0156	0.2862 ± 0.0134	0.2918 ± 0.0137
0.050	500	0.2733 ± 0.0165	0.2922 ± 0.0174	0.2730 ± 0.0146	0.2921 ± 0.0159
0.050	2500	0.2889 ± 0.0170	0.2976 ± 0.0174	0.2880 ± 0.0155	0.2982 ± 0.0159
0.100	500	0.2696 ± 0.0150	0.2883 ± 0.0158	0.2693 ± 0.0130	0.2899 ± 0.0142
0.100	2500	0.2825 ± 0.0164	0.2907 ± 0.0168	0.2757 ± 0.0142	0.2885 ± 0.0147
0.150	500	0.2752 ± 0.0158	0.2941 ± 0.0167	0.2750 ± 0.0140	0.2989 ± 0.0156
0.150	2500	0.2829 ± 0.0158	0.2915 ± 0.0162	0.2850 ± 0.0146	0.2997 ± 0.0152

situation. We see also that for $N = 20$, 100, the uncorrected and corrected estimators are widely different and even for $N = 500$, the uncorrected values lie outside the confidence intervals of the corrected estimators. For $N = 2500$, the uncorrected values lie in the confidence intervals of the corrected estimators. For $c = 3$, only $\hat{\sigma}_h^2$ is any good at all.

In Table 2, we give the performance of the estimators in a situation where the variance and drift are more typical of the financial applications. We take some typical figures for the annual variance of the log-price and for the return [if the return is 5% say, we assume that the price process $X_t \equiv \exp(\sigma B_t + ct)$ satisfies $E(X_1) = 1.05X_0$]. We then consider a single day's trading of that stock. Again, $\hat{\sigma}_h^2$ is doing well, but now we see also that $\hat{\sigma}_{GK,h}^2$ is doing comparably well; the drift is small compared to the variance, and so the Garman–Klass estimator is not really suffering. All the same, it is clear that the corrected estimators are performing far better than the uncorrected ones; indeed, for variance 0.2, *none* of the trials with the uncorrected estimators is successful.

Table 3 takes annual data for the log-price of some stock; it shows what happens if one attempts to estimate the variance of log-price from the annual high, low and closing prices. In this situation, typically the number of shares traded will be large, and so one expects that the effect of using the correction will be less pronounced. This indeed turns out to be the case, but if one looks at the success rates of the uncorrected estimators, then one sees that taking $N = 500$ is generally not enough to save the estimator, and even when $N = 2500$ the estimator may be significantly in error. The Garman–Klass estimator is still doing well here, even though the size of the drift (compared to the variance) is larger.

REFERENCES

[1] BECKERS, S. (1983). Variances of security price returns based on high, low and closing prices. *J. Business* **56** 97–112.

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

[3] GREENWOOD, P. and PITMAN, J. W. (1980). Fluctuation identities for Lévy processes and splitting at the maximum. *Adv. Appl. Probab.* **12** 893–902.

[4] PARKINSON, M. (1980). The extreme value method for estimating the variance of the rate of return. *J. Business* **53** 61–65.

[5] ROGERS, L. C. G. and WILLIAMS, D. (1987). *Diffusions, Markov Processes, and Martingales* **2**. Wiley, Chichester.

SCHOOL OF MATHEMATICAL SCIENCES
QUEEN MARY AND WESTFIELD COLLEGE
UNIVERSITY OF LONDON
MILE END ROAD
LONDON E1 4NS
ENGLAND

DEPARTMENT OF APPLIED ECONOMICS
SIDGWICK AVENUE
CAMBRIDGE CB3 9DE
ENGLAND