Mark B. Garman
and
Michael J. Klass[1]
University of California, Berkeley

On the Estimation of Security Price Volatility from Historical Data[2]

Abstract: *Improved estimators of security price volatility are formulated. These estimators employ data of the type commonly found in the financial pages of a newspaper, namely the high, low, opening and closing prices, and the transaction volume. The new estimators have much higher relative efficiency than the standard estimators.*

Keywords: *Volatility, high-low volatility estimator, high-low-open-close volatility estimator, Garman-Klass volatility estimator*

I. Introduction

This paper examines the problem of estimating capital asset price volatility parameters from the most available forms of public data. While many varieties of such data are possible, we shall consider here only those which are truly universal in their accessibility to investors, that is, data appearing in the financial pages of the newspaper. In particular, we shall consider volatility estimators that are based upon the historical opening, closing, high, and low prices and transaction volume. Since high and low prices require continuous monitoring to obtain, they correspondingly contain superior information content, exploited herein.

Any parameter-estimation procedure must begin with a maintained hypothesis regarding the structural model within which estimation is to be made. Our structural model is described in Section II. Section III discusses the "classical" estimation approach. In Section IV we introduce some more efficient estimators based upon the high and low prices. "Best" analytic estimators, which simultaneously use the high, low, opening, and closing prices, are formulated in Section V. Section VI considers the complications raised by trading volume. Section VII provides a summary.

II. **The Structural Model**

Our maintained model assumes that a diffusion process governs security prices:

$$P(t) = f(B(t)) \tag{1}$$

Here P is the security price, t is time, f is a monotonic, time-independent[3] transformation, and $B(t)$ is a diffusion process with the differential representation

$$dB = s\ dz \tag{2}$$

where dz is the standard Gauss-Wiener process and s is an unknown constant to be estimated. This formulation is sufficiently general to cover the usual hypothesis of the geometric~Brownian motion of stock prices, as well as some of the proposed alternatives to the geometric hypothesis (e.g. Cox and Ross 1975). Throughout this paper, it shall always be implicit that we are dealing with the *transformed* price series $B = f^{-1}P$. Thus in the geometric case, "price" would mean "logarithm of original price," and "volatility" would mean "variance of the logarithm of original prices," and similarly for other transformations.

Of course, there are limitations to our maintained model. First, we are essentially considering each security in isolation, ignoring the covariance thought to exist among securities in specific asset pricing models (e.g., Sharpe 1970). Second, only one parameter is to be estimated: the simultaneous estimation of other unknown parameters, for example, the "drift," is not treated here. Third, the above form of f rules out a significant number of diffusion processes, including those having arbitrary nonzero drift, even when this drift is known.[4] Fortunately, the effect of the foregoing limitation tends to vanish as we shorten the interval over which estimation is made. Finally, dividends and other discrete capital payments are neglected, since these violate the otherwise continuous nature of the assumed diffusion sample paths.

Moreover, the current paper is not concerned with the question of whether the maintained model is the "correct" model of asset price fluctuations. This has been an ongoing subject with many authors over many years, and we certainly could not aspire to settle this complex issue here. Rather, our purposes are to develop the estimation consequences of the model, given the data restrictions described earlier.

III. **"Classical" Estimation**

Under the maintained model, (transformed) price changes over any time interval are normally distributed with mean zero and variance proportional to the length of the interval. Moreover, the prices will always exhibit continuous sample paths. Yet we do not assume that these paths may be everywhere observed. There are at least two factors that interfere with our abilities to continuously observe prices. First, transactions often occur only at discrete points in time.[5] Second, stock exchanges are normally closed during certain periods of time. Our maintained model assumes that the continuous Brownian motion of (2) is followed *during periods between transactions and during periods of exchange closure*, even though prices cannot be observed in such intervals.

As a matter of choice, we shall concentrate herein on estimators of the variance parameter s^2 of $B(t)$. Any such choice of estimation parameter will have disadvantages in some contexts.[6] Since such bias typically vanishes with increasing sample size and is usually small relative to other sources of error, we shall ignore this issue to concentrate solely upon the estimation of s^2.

Moreover, it is convenient to think of the interval $t \in [0, 1]$ as representing one trading interval or "day," since this will prove to be a satisfactory paradigm for the problems of weekly and monthly data also. Our "day" will be divided into two portions, an initial period when the market is closed, followed immediately by a trading period where the market is active. Figure 1 shows this diagrammatically.

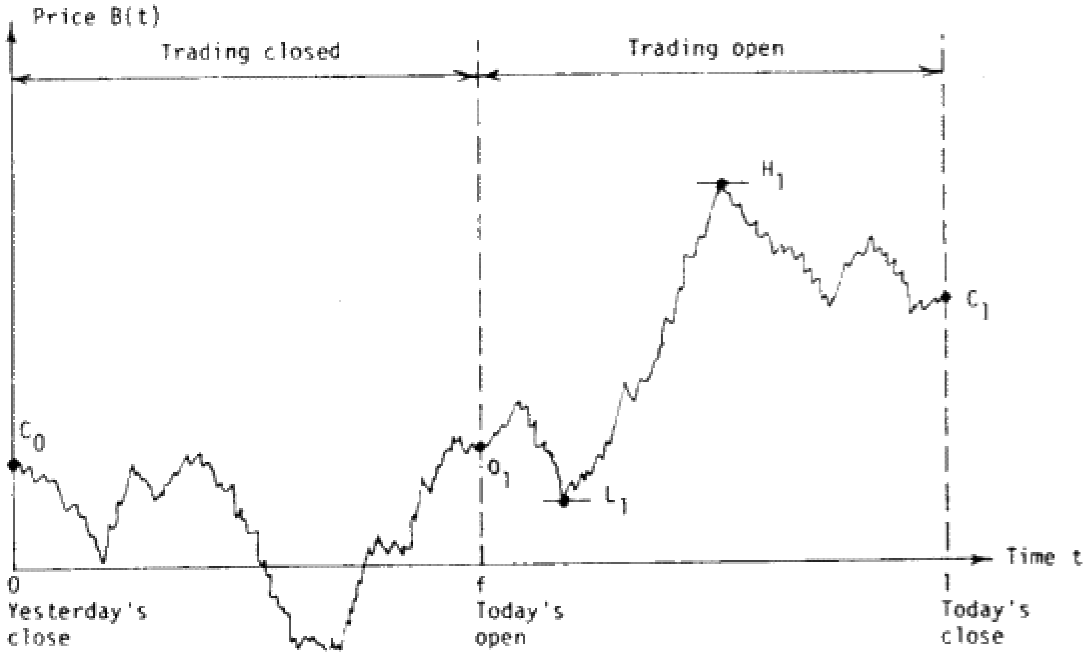

FIG. 1.—Price versus time

In figure 1 trading is closed initially, starting with yesterday's closing when the price was C_0. The price sample path is then *unobservable* until trading opens, at time f when the price is O_1. In the interval [f, 1] we assume (ignoring transaction volume for the moment) that the entire sample path is continuously monitored, having a high value of H_1, a low value of L_1 , and a closing value of C_1. (The effects of monitoring at discrete transactions will be considered later.) We then adopt notation as follows:

s^2 = unknown constant variance (volatility) of price change;
f = fraction of the day (interval [0, 1]) that trading is closed;
$C_0 = B(0)$, previous closing price;
$O_1 = B(f)$, today's opening price;
$H_1 = \max B(t), f \leq t \leq 1$, today's high;
$L_1 = \min B(t), f \leq t \leq 1$, today's low;
$C_1 = B(1)$, today's close;
$u = H_1 - O_1$, the normalized high;
$d = L_1 - O_1$, the normalized low;
$c = C_1 - O_1$, the normalized close;
$g(u, d, c;\ s^2)$ = the joint density of (u, d, c) given s^2 and $f = 0$.

The classical estimation procedure employs

$$\hat{s}_0^2 \equiv (C_1 - C_0)^2$$

as an unbiased estimator of s^2. The advantages of this classical estimator are its simplicity of usage and its freedom from obvious sources of error or bias. Closing prices are measured in a consistent fashion from period to period, and there is little question about the time interval being spanned by the estimator. Its principal disadvantage is the fact that it ignores other readily available information which may contribute to estimator efficiency. As we shall see, even minor additions to the utilized information can have remarkable impact.

For example, suppose that opening prices are available in addition to the closing prices. In this case, we can form the estimator

$$\hat{s}_1^2 \equiv \frac{(O_1 - C_0)^2}{2f} + \frac{(C_1 - O_1)^2}{2(1-f)}, \quad 0 < f < 1 \tag{3}$$

The latter is a "better" estimator, in the sense discussed next.

The classical estimator above will provide the benchmark by which we shall judge all other estimators. Therefore, define the relative efficiency of an arbitrary estimator $\hat{y}$ via the ratio

$$Eff(\hat{y}) \equiv \frac{\mathrm{var}(\hat{s}_0^2)}{\mathrm{var}(\hat{y})} \tag{4}$$

Since $\mathrm{var}(\hat{s}_0^2) = 2s^4$ and $\mathrm{var}(\hat{s}_1^2) = s^4$, it follows[7] that $Eff(\hat{s}_1^2) = 2$, independent of *f*. Thus we see that, by simply including the opening price in our estimation procedure, we may halve the variance of our volatility estimates.

The importance of high relative efficiency in a volatility estimator is obvious, inasmuch as improved confidence may be ascribed to the estimation. Alternatively, investigators may adopt the tactic of purposely restricting data usage to combat nonstationarity; For example, suppose a researcher possesses a data base spanning 10 months. If she discovers an estimator having a high relative efficiency, say 10, this permits her to reduce the estimator confidence regions by a factor of $\sqrt{10}$. Alternatively, she might choose to use only one month's data and retain the old confidence regions; the reason for doing this might be to use the most recent month's data, since it presumably has more predictive content in the presence of unknown nonstationarity.

IV. High/Low Estimators

High and low prices during the trading interval require continuous monitoring to establish their values. The opening and closing prices, on the other hand, are merely "snapshots" of the process. Intuition would then tell us that high/low prices contain more information regarding volatility than do open/close prices. This intuition is correct, as Parkinson (1976) has demonstrated.[8] He assumes $f = 0$ and constructs the estimator

$$\hat{s}_2^2 \equiv \frac{(H_1 - L_1)^2}{4(\log_e 2)} = \frac{(u-d)^2}{4(\log_e 2)} \tag{5}$$

Here, $Eff(\hat{s}_2^2) \cong 5.2$. When the high, low, open. and close prices are simultaneously available, we can also form the composite estimator

$$\hat{s}_3^2 \equiv a\frac{(O_1 - C_0)^2}{f} + (1-a)\frac{(u-d)^2}{(1-f)4(\log_e 2)}, \quad 0 < f < 1 \tag{6}$$

which has minimum variance when $a = 0.17$, independent of the value of *f*. In this case, $Eff(\hat{s}_3^2) \cong 6.2$.

One criticism of the estimators which are based solely on the quantity (u-d) is that they ignore the joint effects between the quantities u, d, and c, which may be utilized to further increase efficiency. In the following section we therefore characterize the best analytic estimators of σ^2.

V. "Best" Analytic Scale-invariant Estimators

For our purposes herein, an estimator is "best" when it has minimum variance and is unbiased. We shall also impose the requirements that the estimators be analytic and scale-invariant, as explained later.
To simplify the initial analysis, we suppose $f = 0$, that is, trading is open throughout the interval [0, 1]. Next, consider estimators of the form $D(u, d, c)$, that is, decision rules which are functions only of the quantities u, d, and c. We restrict attention to these normalized values because the process $B(t)$ renews itself everywhere, including at $t =$

0, and so only the increments from the level O_1 (= C_0) are relevant. According to the lemma established in Appendix A, any minimum-squared-error estimator D (u, d, c) should inherit the invariance properties of the joint density of (u, d, c). Two such invariance properties may be quickly recounted: For all $\sigma^2 > 0$ and all $d \le c \le u$, $d \le 0 \le u$, we have

$$g(u,d,u;s^2) = g(-d,-u,-c;s^2) \tag{7}$$

and

$$g(u,d,u;s^2) = g(u-c,d-c,-c;s^2) \tag{8}$$

The first condition represents price symmetry: for the Brownian motion of form (2), $B(t)$ and $-B(t)$ have the same distribution. Whenever $B(t)$ generates the realization (u, d, c), - $B(t)$ generates (u, d, c). The second condition represents time symmetry: $B(t)$ and $B(1 - t)$ - $B(1)$ have identical distributions. Whenever $B(t)$ produces (u, d, c), $B(1-t)$ - $B(1)$ produces (u - c, d - c, -c). By the lemma of Appendix A, then, any decision rule D (u, d, c) may be replaced by an alternative decision rule which preserves the invariance properties (7) and (8) without increasing the expected (convex) loss associated with the estimator. Therefore, we seek decision rules which satisfy

$$D(u,d,c) = D(-d,-u,-c) \tag{9}$$

and

$$D(u,d,c) = D(u-c,d-c,-c) \tag{10}$$

Next, we observe that a scale-invariance property should hold in the volatility parameter space: for any $\lambda > 0$,

$$g(u,d,c;s^2) = g(lu, ld, lc; l^2 s^2) \tag{11}$$

In consequence of (11), we now restrict our attention to scale-invariant decision rules for which

$$D(lu, ld, lc) = l^2 D(u,d,c), \qquad l > 0 \tag{12}$$

If we now adopt the regularity condition that the decision rules considered must be analytic in a neighborhood of the origin, condition (12) implies that the decision rule D (u, d, c) must be quadratic in its arguments. (Proof of this is given in Appendix B.) Thus we have

$$D(u,d,c) = a_{200}u^2 + a_{020}d^2 + a_{002}c^2 + a_{110}ud + a_{101}uc + a_{011}dc \tag{13}$$

Scale invariance and analyticity are combined to reduce the search for a method of estimating σ^2 from an infinite-dimensional problem to a six-dimensional one. Applying the symmetry property (9) to equation (13), we have the implications $a_{200} = a_{020}$ and $a_{101} = a_{011}$. By virtue of property (10), we have the additional constraint $2a_{200} + a_{110} + 2a_{101} = 0$, hence we have

$$D(u,d,c) = a_{200}(u^2 + d^2) + a_{002}c^2 - 2(a_{200} + a_{101})ud + a_{101}c(u+d) \tag{14}$$

Insisting that D (u, d, c) be unbiased, that is, E[D (u, d, c)] = σ^2, leads to one further reduction. Since[9] $E[u^2] = E[d^2] = E[c^2] = E[c(u+d)] = \sigma^2$ and $E[ud] = (1 - 2 \log_e 2)\sigma^2$, we may restrict attention further to the two-parameter family of decision rules $D(.)$ of the form

$$D(u,d,c) = a_1(u-d)^2 + a_2\{c(u+d) - 2ud\} + \{1 - 4(a_1 + a_2)\log_e 2 + a_2\}c^2 \tag{15}$$

To minimize this quantity, note that, for any random variables X, Y, and Z, the quantity $V(a_1, a_2) \equiv E[(a_1X + a_2Y + Z)^2]$ is minimized by a_1 and a_2 which satisfy the first-order conditions

$$E[(a_1X + a_2Y + Z)X] = E[(a_1X + a_2Y + Z)Y] = 0 \tag{16}$$

Solving the above for a_1 and a_2, we have

$$a_1^* = \frac{E[XY]E[YZ] - E[Y^2]E[XZ]}{E[X^2]E[Y^2] - (E[XY])^2} \tag{17a}$$

and

$$a_2^* = \frac{E[XY]E[XZ] - E[X^2]E[YZ]}{E[X^2]E[Y^2] - (E[XY])^2} \tag{17b}$$

In the problem at hand,

$$\begin{aligned} X &= (u-d)^2 - (4\log_e 2)c^2 \\ Y &= c(u+d) - 2ud + (1 - 4\log_e 2)c^2 \\ Z &= c^2 \end{aligned} \tag{18}$$

Analysis via generating functions (Appendix C) reveals the following fourth moments:

$$\begin{aligned}
E[u^4] = E[d^4] = E[c^4] &= 3s^4 \\
E[u^2c^2] = E[d^2c^2] &= 2s^4 \\
E[u^3c] = E[d^3c] &= 2.25s^4 \\
E[uc^3] = E[dc^3] &= 1.5s^4 \\
E[ud^2c] = E[u^2dc] &= \left\{\frac{9}{4} - 2\log_e 2 - \frac{7}{8}z(3)\right\}s^4 = -0.1881s^4 \\
E[u^2d^2] &= [3 - 4\log_e 2]s^4 = 0.2274s^4 \\
E[udc^2] &= \left\{2 - 2\log_e 2 - \frac{7}{8}z(3)\right\}s^4 = -0.4381s^4 \\
E[ud^3] = E[u^3d] &= \left\{3 - 3\log_e 2 - \frac{9}{8}z(3)\right\}s^4 = -0.4318s^4
\end{aligned}$$

where $z(3) = \sum_{k=1}^{\infty} 1/k^3 = 1.2021$ is Riemann's zeta function. Substituting the above moments into (17a) and (17b) via (18), we find that $a_1^* = 0.511$ and $a_2^* = -0.019$. Employing these values in (15) yields the **best analytic scale-invariant estimator:**

$$\hat{s}_4^2 \equiv 0.511(u-d)^2 - 0.019\{c(u+d) - 2ud\} - 0.383c^2\,. \tag{19}$$

We find that $Eff(\hat{s}_3^2) \cong 7.4$. The recommended and "more practical" estimator

$$\hat{s}_5^2 \equiv 0.5(u-d)^2 - (2\log_e 2 - 1)c^2 \tag{19a}$$

possesses nearly the same efficiency but eliminates the small cross-product terms.

Now suppose that 0 < f < 1, that is, trading is both open and closed in [0, 1]. Then the opening price O_1, may differ from the previous closing price C_0, and so we may form the composite estimator

$$\hat{s}_6^2 \equiv a\frac{(O_1 - C_0)^2}{f} + (1-a)\frac{\hat{s}_4^2}{(1-f)} \qquad (20)$$

The variance of this estimator is minimized when a = 0.12, and in this case the efficiency is approximately 8.4. Thus, there exists an estimator possessing an efficiency factor which is more than eight times better than the classical estimator, given only high, low, open, and close prices.

VI. Volume Effects

The derivation of all of the high-low estimators of the previous sections depends critically upon the assumption of continuously monitored price paths. When the path can only be monitored at discrete transactions, all our statistics will be biased. Technically speaking, the knowledge that only a finite number of observations are available should lead us to commence a new search for the best estimator; however, we shall defer this task to a later paper.

TABLE 1: **Expected Values of Volatility Estimators (true variance = 1.0)**

No. transactions	$\hat{s}_0^2$	$\hat{c}^2$	$\hat{s}_2^2$	$\hat{s}_4^2$
5	1.03	0.48	0.55	0.38
10	1.01	0.67	0.65	0.51
20	1.00	0.82	0.74	0.64
50	1.00	0.92	0.82	0.73
100	1.00	0.97	0.86	0.80
200	1.00	0.99	0.89	0.85
500	1.00	1.00	0.92	0.89

We instead confine our considerations here to determining the extent of the bias in using the estimators already described, when only a finite set of observations is available. Simulation studies were employed to arrive at table 1.[10] Note that the close-to-close estimator $\hat{s}_0^2$ has only a slight positive bias.[11] On the other hand, the expected values of volatility estimators $\hat{c}^2 \equiv (C_1 - O_1)^2$, $\hat{s}_2^2$ and $\hat{s}_4^2$ are significantly less than 1.0 whenever a finite number of transactions take place. Moreover, there are two distinguishable reasons for the observed biases. The estimator $\hat{c}^2$ has downward bias because the "effective" time period over which it is estimated, since this is diminished to the span between the first and the last transaction, being an open-to-close estimator. The latter two estimators are also subject to this effect when trading is closed during some portion of the day. (In the absence of other considerations, finite transaction volume will make the opening appear to be later and the closing appear to be earlier. However, many exchanges will collect orders during the night for execution at the opening; additionally, some exchanges have a "closing rotation" in which firm-offer prices are quoted at the closing. Each of these procedures might tend to obscure the effective-time bias.) The second reason affects only the latter high-low estimators, which take on a downward bias because the observed highs and lows will be less in absolute magnitude than the actual highs and lows. In addition, the high-low estimators are indirectly affected by the effective-time bias since highs and lows tend to occur at the first and last transactions.

As a practical procedure, one should divide the corresponding empirical estimators by the numbers in table 1. Since the composite estimators are linear combinations of the given estimators, their bias and that of several others may be quickly computed therefrom.

Random transaction volume is one source of predictable bias somewhat within the scope of the current model. But there are other important sources of bias which can be made predictable only by significant extension of the current model. Some of these unpredictable bias sources are the following. (1) To the extent that transactions themselves may convey new information, daytime volatilities may be different from nighttime volatilities. (2) Bid-ask spreads usually exist, within which the transactions process may be quite complicated (Garman 1976). (3) In addition, volatilities are though be nonstationary, possibly in a variety of fashions.

VII. Conclusions

We have examined a number of estimators of price volatility. Efficiency factors more than eight times better than the classical estimators have been demonstrated. These same estimators are also subject to more sources of predictable bias, one of the most evident of which is finite transaction volume.

Appendix A

Estimator Invariance Properties

***Lemma**:* Let Θ be a parameter space. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a vector of (not necessarily independent) observations whose joint density $f_q(\mathbf{X})$ depends on an unknown parameter $q \in \Theta$ to be estimated. Let $T: \mathbf{R^n} \to \mathbf{R^n}$ be a fixed measure-preserving transformation. Suppose that, for all $q \in \Theta$ and all $\mathbf{X}$ in the support of $f_q(\mathbf{X})$,

$$f_q(T\mathbf{X}) = f_q(\mathbf{X}) \,. \tag{A1}$$

Let $D(\mathbf{X})$ be any decision rule which estimates q. Let $L(q, D(\mathbf{X}))$ be any loss function such that $L(q, D(\mathbf{X}))$ is a convex function for each fixed $q \in \Theta$. Defining $T^j \equiv TT^{j-1}$, where T^0 is the identity operator, let A_k be an averaging operator which maps decision rules into decision rules according to the prescription

$$A_k(D(\mathbf{X})) = \frac{1}{k}\sum_{j=1}^{k} D(T^{j-1}\mathbf{X}) \tag{A2}$$

Then, for all $q \in \Theta$,

$$E_q L(q, A_k(D(\mathbf{X}))) \leq E_q L(q, D(\mathbf{X})) \,. \tag{A3}$$

***Proof**:* [Deleted for brevity. Use the convexity, take expectations.]
***Note**:* The relevant transformations used in the text, namely price symmetry and time symmetry in formulas (7), (8), both have the property that $T^2 \equiv I$, and clearly these will be measure-preserving.

Appendix B

Analytic Estimators are Quadratic

***Lemma**:* Estimators $D(u,d,c)$ of s^2 which are analytic in the neighborhood of the origin are quadratic in form
***Proof**:* If $D(u,d,c)$ is analytic in the neighborhood of the origin, we may write its Taylor series expansion as

$$D(u,d,c) = \sum_{i,j,k \geq 0} a_{ijk} u^i d^j c^k \tag{B1}$$

Next define

$$F_{udc}(l) \equiv l^2 \sum_{i,j,k \geq 0} a_{ijk} u^i d^j c^k - \sum_{i,j,k \geq 0} a_{ijk} l^{i+j+k} u^i d^j c^k \tag{B2}$$

From the scale invariance property (12), this latter function must be identically zero. It may also be extended to an analytic function in some open neighborhood of the origin. Thus, by uniqueness, all coefficients of powers of l in (B2) must be identically zero. It follows that $a_{ijk}u^i d^j c^k = 0$ for $i + j + k \neq 2$, that is, D is quadratic.

Appendix C

Generating Function of High, Low, and Close[12]

The expectation of the moments of u, d, and c is given by

$$E(u^p d^q c^r) = (-1)^{p+r} \frac{s^n}{2^{n/2}\Gamma((n/2)+1)} \left[\frac{\P^n \mathbf{H}(t,s,z)}{\P t^p \P s^q \P z^r} \right]_{t=s=z=0} \tag{C1}$$

where $n = p + q + r$ and our generating function $\mathbf{H}$ is

$$\begin{aligned} \mathbf{H}(t,s,z) = \frac{st}{1-z^2} \sum_{k=1}^{\infty} \{ & -\frac{1}{(2k+t)(2k+s+1-z)} - \frac{1}{(2k+s)(2k+t+1+z)} \\ & + \frac{1}{(2k+t+2)(2k+s+1-z)} + \frac{1}{(2k+s+2)(2k+t+1+z)} \} \\ & + \frac{1}{1-z^2} \left[1 - \frac{s}{1+s-z} - \frac{t}{1+t+z} + \frac{ts}{(2+t)(1+s-z)} + \frac{ts}{(2+s)(1+t+z)} \right] \end{aligned} \tag{C2}$$

---

[1] Mark B. Garman is Emeritus Professor of Finance at the Haas School of Business, University of California at Berkeley (garman@haas.berkeley.edu); Michael J. Klass is Professor of Statistics at the University of California at Berkeley (garman@haas.berkeley.edu);.

[2] This paper is an improved and updated version of work originally published in the Journal of Business, vol. 53, no. 1, 1980, pp. 67-78. The original research work was supported in part by the Dean Witter Foundation, administered by the Institute of Business and Economic Research, and by the National Science Foundation, grants MCS-75-1036-A01 and SOC-77-18O87, the latter administered by the Center for Research in Management Science.

[3] Monotonicity and time-independence are employed to assure that the same set of sample paths generates the sample maximum and minimum values of *B* and *P*

[4] See Thorpe (1976) for the arguments and empirical evidence on this point.

[5] This has become known as the "market microstructure" topic; see Garman (1976) for a foundational treatment of such a model.

[6] As Boyle and Anantbanarayan (1977) have observed, any estimation procedure for a random variable will produce bias in the estimation of any nonlinear function of that random variable. Their example was the use of the Black-Scholes (1973) option pricing formula, which is a nonlinear function of volatility. Thus unbiased estimators of volatility will produce biased estimates of option prices, and vice versa.

[7] Note: since we are dealing here with the variance of variance estimators, fourth moments are naturally involved.

[8] Parkinson actually provides two estimators of volatility, the one described in formula (5) and another which employs the square of the sum of the differences of high and low. However, this latter estimator is biased, and so we ignore it here.

[9] See Appendix C for the calculation of moments.

[10] Our finite-volume simulations assumed that all transactions are scattered "randomly" (i.e. uniformly) throughout the interval [0, 1].

[11] Scholes and Williams (1977) also consider this source of bias.

[12] This generating function was kindly supplied to the authors by Oldrich Vasicek.