AMaMeF 2013 & Banach Center, Warszawa

An application of the method of moments to range-based volatility estimation using daily high, low, opening, and closing prices

Cristin Buescu
King's College, London

Joint work with M. Taksar and F.J.Koné

Disclaimer

The following expresses the views of its author(s) only, and should not be taken to represent views of institutions with which the authors are or have been based. In particular, the information provided is personal opinion and should not be relied upon for financial advice. If you require financial advice or guidance please contact an appropriate professional.

Dedication

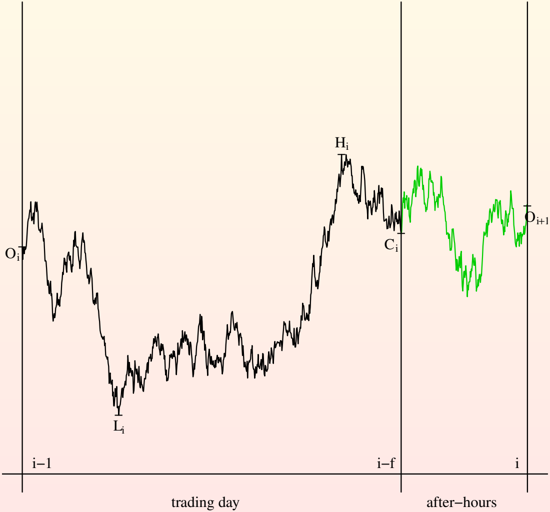
In memoriam Michael Taksar (1949 - 2012)

Outline

1. Motivation (THE STORY)
   The Volatility Estimation Problem
2. Our approach (THE MATHEMATICS)
   1. Expectation of range of arithmetic Brownian motion
   2. Method of moments using HLOC data
   3. Solving implicit equation and including opening jumps
3. Comparison results (SIMULATED DATA)
4. Application (REAL DATA)
5. Summary

Outline

Notation (HLOC: high, low, close, open stock price)



Volatility Estimation Problem in Black-Scholes I

- Black-Scholes setting: stock price $dS_t = \mu S_t dt + \sigma S_t W_t$ (GBM)
- Volatility estimation σ (equivalently, variance estimation σ^2)
- Literature on this:

$$\begin{cases} \text{assumes security price has no drift} \Rightarrow \text{OVERESTIMATION} \\ \text{assumes no price jump at opening} \Rightarrow \text{UNDERESTIMATION} \end{cases}$$

- Garman and Klass (1980): variance estimator V_{GK} assumes no drift
- Parkinson (1980): variance estimator V_P uses only high-low data

Volatility Estimation Problem in Black-Scholes II

- Rogers and Satchell (1991) and Rogers et al (1994): V_{RS} assumes no opening jumps

$$V_{RS} = \frac{1}{n}\sum_{i=1}^{n}\Big(\log\frac{H_i}{O_i}\Big(\log\frac{H_i}{O_i}-\log\frac{C_i}{O_i}\Big)+\log\frac{L_i}{O_i}\Big(\log\frac{L_i}{O_i}-\log\frac{C_i}{O_i}\Big)\Big). \tag{1}$$

- JUMP ($C_i \to O_{i+1}$): the sample variance V_0 of $\log(O_{i+1}/C_i)$

$$V_0 := \frac{1}{n-1}\sum_{i=1}^{n}\Big(\log\frac{O_{i+1}}{C_i}-\frac{1}{n}\sum_{j=1}^{n}\log\frac{O_{j+1}}{C_j}\Big)^2. \tag{2}$$

- Incorporate opening jumps in V_{RS} by using $V_0 + V_{RS}$?

Volatility Estimation Problem in Black-Scholes III

Yang and Zhang (2000)

- caution against easy fix (adding V_0 to V_{RS}), not min variance
- propose estimator assuming stock has drift and opening jumps:

$$V_{YZ} = V_0 + kV_C + (1-k)V_{RS}, \tag{3}$$

V_0 sample variance of $\log(O_{i+1}/C_i)$
V_C is the sample variance of $\log(C_i/O_i)$
k is constant chosen to minimize estimator variance for fixed n

Volatility Estimation Problem in Black-Scholes IV

In contrast, our approach is based in part on Koné (1996):

1. derive expectation of range of arithmetic Brownian motion
2. estimate range using method of moments and HLOC daily data
3. solve implicit equation for intra-day volatility and include opening jumps

We propose the variance estimator:

$$V_Z := V_0 + V_i, \tag{4}$$

where V_i is solution of an implicit equation.

Compare to $$V_{YZ} = V_0 + kV_C + (1-k)V_{RS}.$$

Outline

1. Motivation (THE STORY)
   The Volatility Estimation Problem

2. Our approach (THE MATHEMATICS)
   1. Expectation of range of arithmetic Brownian motion
   2. Method of moments using HLOC data
   3. Solving implicit equation and including opening jumps

3. Comparison results (SIMULATED DATA)

4. Application (REAL DATA)

5. Summary

Range of arithmetic Brownian motion

In a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_t, P)$ consider:

- Arithmetic Brownian motion: $dX_t = \mu dt + \sigma dW_t$, $X_0 = 0$
- running max: $M_t := \sup_{0 \le s \le t} X_s$
- running min: $m_t := \inf_{0 \le s \le t} X_s$
- Range: $R_t := M_t - m_t$

Joint density of aBM and max: (X_t, M_t)

$$P(X_t \in da, M_t \in db) = \frac{2(2b-a)}{\sqrt{2\pi t^3}\sigma^3} \exp\left\{-\frac{(2b-a)^2}{2t\sigma^2} + \frac{\mu}{\sigma^2}a - \frac{1}{2}\frac{\mu^2}{\sigma^2}t\right\} da\, db.$$

- Derived directly.
- Can be obtained from equation (1.8.8) of Harrison (1985)
- For $\sigma = 1$ this was used in Example E5 of Karatzas and Shreve (1998) in relationship to Clark's formula to obtain explicitly the hedging portfolio.

Density and expectation of half-range: $M_t - X_t$

Density of half-range via Jacobian (2-dim transformation):

$$f_{M_t - X_t}(c) = 2\frac{\mu}{\sigma^2}\,\Phi\Big(\frac{\mu t - c}{\sigma\sqrt{t}}\Big)\exp\Big(-2\frac{\mu}{\sigma^2}c\Big) + \frac{2}{\sigma\sqrt{2t\pi}}\exp\Big\{-\frac{(\mu t + c)^2}{2t\sigma^2}\Big\}.$$

Expectation of half-range:

$$\begin{aligned} E(M_t - X_t) &= \frac{\sigma^2}{2\mu}\,\Phi\Big(\frac{\mu}{\sigma}\sqrt{t}\Big) - \Big(\mu t + \frac{\sigma^2}{2\mu}\Big)\Big(1 - \Phi\Big(\frac{\mu}{\sigma}\sqrt(t)\Big)\Big) \\ &+ \frac{\sigma\sqrt{t}}{\sqrt{2\pi}}\exp\Big(-\frac{t\mu^2}{2\sigma^2}\Big). \end{aligned}$$

Expectation of range of arithmetic Brownian motion

$$R_t = M_t - m_t = (M_t - X_t) + (X_t - m_t)$$

Symmetry between half-ranges gives:

$$E[R_t] = \Big(\mu t + \frac{\sigma^2}{\mu}\Big)\left(1 - 2\Phi\Big(-\sqrt{t}\frac{\mu}{\sigma}\Big)\right) + 2\frac{\sigma\sqrt{t}}{\sqrt{2\pi}}\exp\Big(-\frac{t\mu^2}{2\sigma^2}\Big).$$

Equivalently:

$$E[R_t] = h\Big(\frac{\mu t}{\sigma\sqrt{t}}, \frac{\sigma^2}{\mu}\Big) =: ER(\mu, \sigma, t), \qquad (6)$$

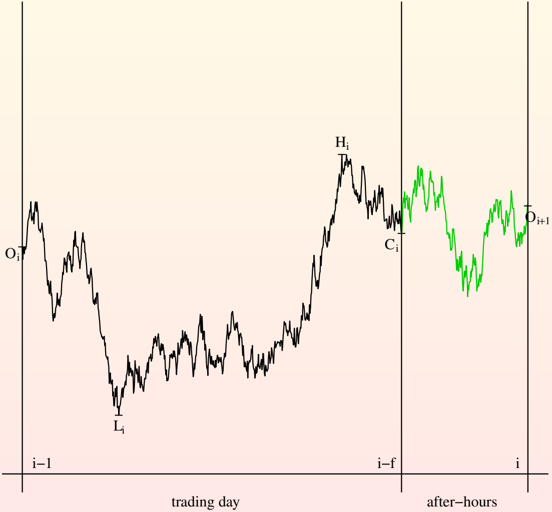
where the function h is defined by:

$$h(x, y) := \left\{(x^2 + 1)(2\Phi(x) - 1) + \frac{2x}{\sqrt{2\pi}}\exp\Big(-\frac{x^2}{2}\Big)\right\} y. \qquad (7)$$

Corollaries

- Derive also joint density of (M_t, m_t) (see also Borodin and Salminen (1996))
- and the density of the range R_t of an arithmetic Brownian motion

1 day = intra-day + after-hours



$$1 = (1 - f) + f$$

Price S_t (GBM):

$$\frac{dS_t}{S_t} = \mu_s \, dt + \sigma \, dW_t, \quad t \geq 0. \tag{8}$$

Log-price $X_t = \log S_t$ (aBM):

$$dX_t = \mu dt + \sigma dW_t, \quad \mu = \mu_s - \frac{\sigma^2}{2}$$

$1 = (1 - f) + f$

Price S_t (GBM):

$$\frac{dS_t}{S_t} = \mu_s \, dt + \sigma \, dW_t, \quad t \geq 0. \tag{8}$$

Log-price $X_t = \log S_t$ (aBM):

$$dX_t = \mu dt + \sigma dW_t, \quad \mu = \mu_s - \frac{\sigma^2}{2}$$

Assume after-hours= fraction f of 1 day.

Intra-day data (fraction $1 - f$ of 1 day):

$$O_i = S_{i-1}, \; C_i = S_{i-f}, \; H_i = \sup_{t \in [i-1, i-f]} S_t, \quad L_i = \inf_{t \in [i-1, i-f]} S_t.$$

Application of the method of moments

Use intra-day range data to derive the implicit equation in $x = \widehat{\sigma(1-f)}$:

$$R_{1-f} = M_{1-f} - m_{1-f} = \log H_1 - \log L_1 = \log \frac{H_1}{L_1} \Rightarrow k_1 := \frac{1}{n}\sum_{i=1}^{n} \log \frac{H_i}{L_i}.$$

$$\text{Recall} \quad E(R_{1-f}) = ER(\mu, \sigma, 1-f) = ER(\mu(1-f), \sigma\sqrt{1-f}, 1).$$

$$\text{Compute} \quad E\left[\log \frac{C_i}{O_i}\right] = E\left[\log \frac{S_{i-f}}{S_{i-1}}\right] = \mu\,(1-f) \Rightarrow k_2 := \frac{1}{n}\sum_{i=1}^{n} \log \frac{C_i}{O_i}.$$

$$\text{Implicit equation} \quad k_1 = h\left(\frac{k_2}{x}, \frac{x^2}{k_2}\right). \;\; \text{Solution:} \quad V_i = x^2. \tag{9}$$

Full one-day variance

Compute:

$$\text{VAR}\Big[\log\frac{C_i}{O_i}\Big] = \text{VAR}\left[\log\frac{S_{i-f}}{S_{i-1}}\right] = \sigma^2\,(1-f),$$
$$\text{VAR}\Big[\log\frac{O_{i+1}}{C_i}\Big] = \text{VAR}\left[\log\frac{S_i}{S_{i-f}}\right] = \sigma^2\,f.$$

Thus, we can write heuristically:

$$\sigma^2 = \text{VAR}\left[\log\frac{C_i}{O_i}\right] + \text{VAR}\left[\log\frac{O_{i+1}}{C_i}\right] = \text{VAR(intra-day)} + \text{VAR(after hours)}$$

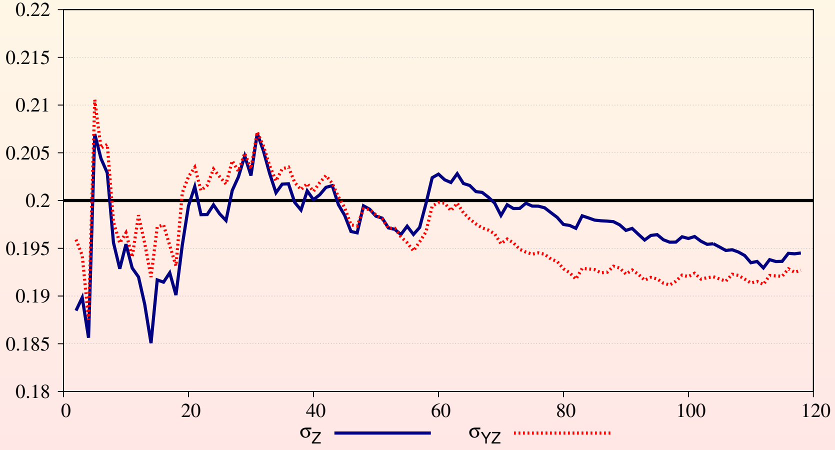
leading to

$$V_Z = V_i + V_0 \quad \text{(annualised } \sigma_a^2 := 252 V_Z\text{)}.$$

Outline

Monte Carlo simulation for known vol

Eg 1: Strikingly similar pattern for the two estimated vols when true $\sigma_a = 0.2$ for varying number of data points ($\mu_s = 0.015$, $f = 0.25$, 250 trading days, 50k time points)



Compare the two vol estimators: σ_Z vs σ_{YZ}

Goal: investigate MVUE

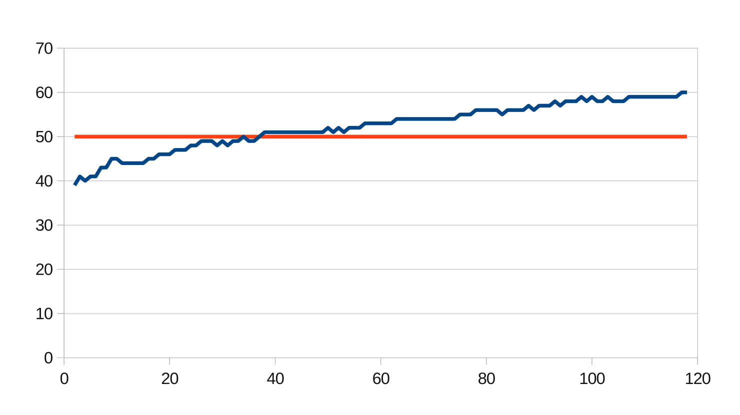
Due to the implicit nature of the equation we achieve this by simulation.

Repeat previous simulation 5k times and look at:

- unbiased
  - percentage of times σ_Z is closer to true σ than σ_{YZ}
  - Mean Absolute Error (in L^1-norm) comparison
  - average σ_Z vs average σ_{YZ} (averaged over the number of scenarios)
- min variance
  - efficiency

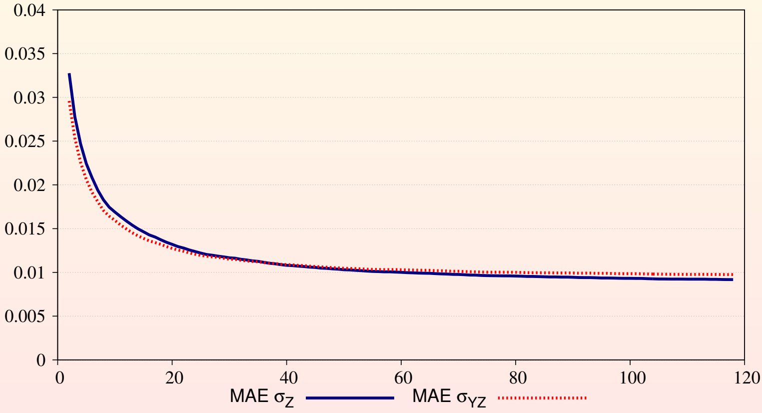
Unbiased I

Percentage of scenarios where our estimator σ_Z is closer than σ_{YZ} to the true value. For more than 37 data points in the estimation σ_Z is closer to the true value more than half the time.



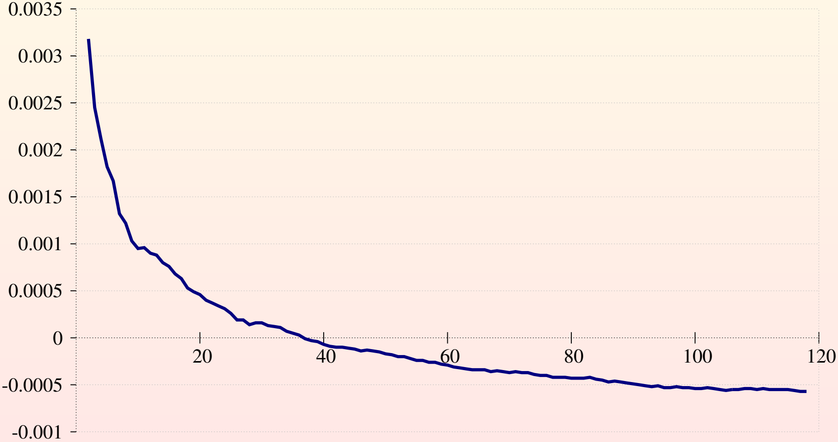
Unbiased II

The Mean Absolute Error (MAE) for our estimator σ_Z (continuous line) follows closely that of σ_{YZ} (dotted line) and both stabilize for $n \geq 55$.



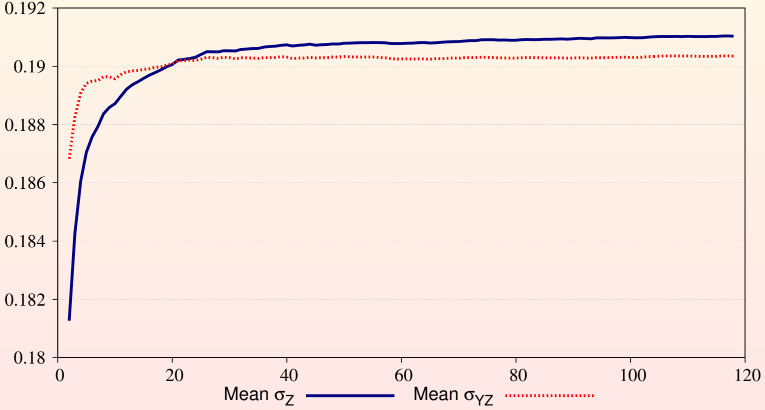
Unbiased III

The difference of the Mean Absolute Errors of σ_Z and σ_{YZ} goes below 0 for $n \geq 37$, suggesting that our σ_Z is better than σ_{YZ} in this range.



Unbiased IV

The volatility σ_Z averaged over the number of scenarios (continuous line) is closer to the true value $\sigma = 0.2$ than the averaged σ_{YZ} (dotted line) when the number of data points is $n \geq 21$.



Min variance I

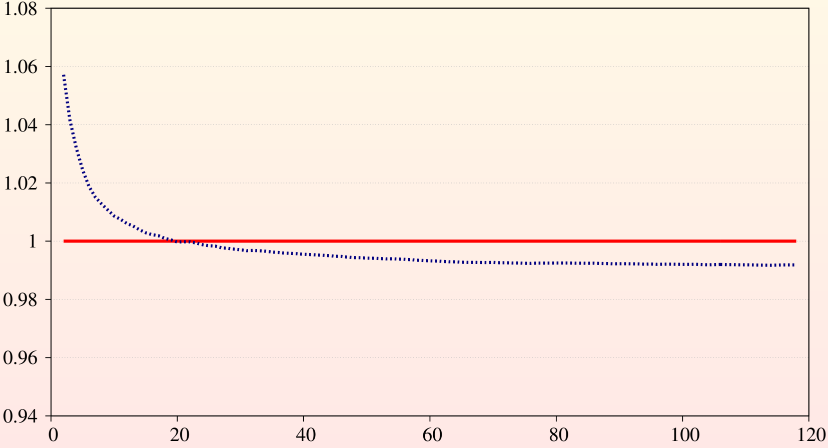
- Following Garman and Klass (1980) we define the efficiency of our estimator with respect to V_{YZ} as:

$$Eff := \frac{\mathrm{VAR}(V_{YZ})}{\mathrm{VAR}(V_Z)}. \tag{10}$$

- Numerical approximation shows that the efficiency is higher than 1 with fewer data points, while overall it doesn't drop by more than 1%.

Min variance II

Efficiency vs number of data points:



Impact of choice of f: none

- Yang and Zhang (2000): range for f is $[0.18, 0.3]$ (typically 0.25).
- For this reason we have considered not only the case $f = 0.25$, but also $f = 0, f = 0.18$ and $f = 0.3$, but the results were similar.
- Even in the driftless case ($\mu = 0$) the values $f = 0$, $f = 0.18$, $f = 0.25$ and $f = 0.3$ resulted in findings that were qualitatively similar.

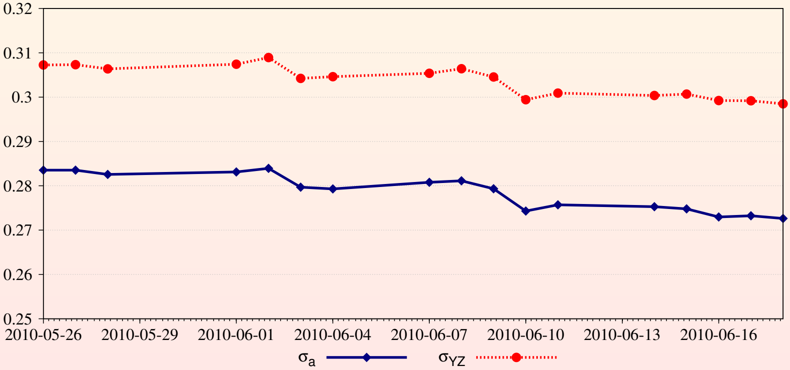
Conclude: satisfactory performance on simulated data.

Outline

1. Motivation (THE STORY)
   The Volatility Estimation Problem
2. Our approach (THE MATHEMATICS)
   1. Expectation of range of arithmetic Brownian motion
   2. Method of moments using HLOC data
   3. Solving implicit equation and including opening jumps
3. Comparison results (SIMULATED DATA)
4. Application (REAL DATA)
5. Summary

Application to algorithmic trading

Eg 2: IBM estimated volatility σ_a (continuous line) is similar to, but below, the corresponding volatility σ_{YZ} (dotted line) for each trading day between May 26 and June 18, 2010. Can be used in algo trading.



Outline

Summary

- Derived expectation of range of Brownian motion (and density, see also Borodin and Salminen (1996))
- Proposed a volatility estimator σ_Z based on daily range data, which includes opening jumps and does not assume zero drift for stock price

Summary

- Derived expectation of range of Brownian motion (and density, see also Borodin and Salminen (1996))
- Proposed a volatility estimator σ_Z based on daily range data, which includes opening jumps and does not assume zero drift for stock price
- Advantages over estimator of Yang and Zhang (2000):
  - no need to estimate the constant k that achieves min variance
  - captures volatility using range data (fewer measurements than normalized highs and lows)

Summary

- Derived expectation of range of Brownian motion (and density, see also Borodin and Salminen (1996))
- Proposed a volatility estimator σ_Z based on daily range data, which includes opening jumps and does not assume zero drift for stock price
- Advantages over estimator of Yang and Zhang (2000):
  - no need to estimate the constant k that achieves min variance
  - captures volatility using range data (fewer measurements than normalized highs and lows)
- Disadvantages:
  - implicit equation gives only numerical approximation of variance
  - may lose up to 1% of efficiency (could have slightly larger variance)
- Works well and can be used in algorithmic trading.

Bibliography I

A.N. Borodin and P. Salminen.
Handbook of Brownian motion - facts and formulae.
Birkhäuser, Basel, 1996.

C. Buescu, M. Taksar and F.J.Koné
An application of the method of moments to range-based volatility estimation using daily high, low, opening, and closing (HLOC) prices.
(forthcoming in Int. J. Theor. Appl. Finance)

M. Garman and M. Klass
On the estimation of security price volatilities from historical data.
Journal of Business 53(1), 67–78, 1980.

Bibliography II

F.J. Koné
Estimation of the volatility of stocks using the high, low and closing prices.
Ph.D. thesis, State University of New York at Stony Brook, 1996.

M. Parkinson
The extreme value method for estimating the variance of the rate of return.
Journal of Business 53(1), 61–65, 1980.

L.C.G. Rogers and S. Satchell
Estimating variance from high, low and closing prices.
Annals of Applied Probability 1(4), 504–512, 1991.

Bibliography III

L.C.G. Rogers, S. Satchell and Y. Yoon
Estimating the volatility of stock prices: a comparison of methods that use high and low prices.
Applied Financial Economics 4, 241–247, 1994.

K. Sutrick, J. Teall, A. Tucker and J. Wei
The range of Brownian motion processes: density functions and derivative pricing applications.
The Journal of Financial Engineering 6(1), 31–46, 1997.

D. Yang D. and Q. Zhang
Drift-independent volatility estimation based on high, low, open, and close prices.
Journal of Business 73, 477–491, 2000.