# Business School
# Discipline of Finance

# Discussion Paper
# 2015-002

# "A Simple Overnight/Intraday
# Volatility Estimator"

## Robert Krause
Chief Executive Officer /
The Volatility Exchange

## Stephen Satchell
Trinity College Cambridge /
University of Sydney Business School

**Introduction**

Our motivation for this paper is not to create the "best" volatility estimator. There are already several very good intraday measures in existence. Our method sets out to provide the 24-hour volatility that would best correspond to the experience of a delta hedger who follows up his or her position throughout the day. Of course, we should be able to get very close to this value if we use tick data. However, in order to provide a method that could furnish a very long data record and could be used on virtually any underlying market, we need only readily available data such as open, high, low, and close. Tick data are simply not available for a host of markets and certainly not going back to the start of trading.

The open value is key to the first possible follow-up (overnight); then the high and low give us the best possible idea of the market range for trading hours. Different formulas add other terms, but, while improving accuracy on the "true" value of realized volatility, they do not improve the accuracy of the estimate of volatility that an intraday hedger may experience. Our strategy, therefore, is to sacrifice some accuracy for a value that is more representative of an intraday follow-up hedge.

In addition, we wanted the same built-in assumptions as interday formulas. These can vary from circumstance to circumstance but the assumptions that interest us include a zero mean for the returns process, and a 252-day annualisation period, which is the standard method for volatility and variance swaps in the over-the-counter (OTC) marketplace. It is the same approach that will soon launch for exchange-listed tradable instruments on realized volatility. Therefore, we want the intraday formula to correspond to all those assumptions/constraints so that if one looks at interday to intraday values, they will be comparable.

We list six assumptions below linking both measurements.

1.    They should assume no drift.

2.    They have a 252-day annualisation factor.

3.    They use a simple set of daily price information (Open, High, Low, and Close).

4.    They are measured over a full 24-hour day.

5.    There is no need for tick data.

6.   The estimators can be easily calculated and independently verified (whereas intraday volatility based on tick data would be difficult for others to verify).

In addition, we treat two distinct portions of the day separately: the period from the close of the market until the open ("overnight") and the period when the market is open ("intraday").

Our approach is to estimate standard deviation, rather than variance, directly; this approach is not entirely new. Although virtually all the published literature is concerned with direct estimation of variance, one unpublished exception to this is Buescu et al. (2011), which partially draws upon Kone (1996). Our resulting estimator has some similarities to the Garman–Klass estimator but differs in a number of key aspects. It is first-moment based and it avoids cross-terms by using non-overlapping information. This present approach is motivated by the work of Garmann and Klass who find that the cross-terms add very little to the efficiency of the estimator. We readily acknowledge that there may be more efficient versions of our estimator but they will come with added complexity.

Another interesting feature of our approach is that it is implicitly weighted for a 24-hour day without the need to weight the two terms with any coefficient, making it independent of a weighting method requiring time-of-day values (this adds to its simplicity and reduces the chance of error).

There are a large number of intraday variance estimators already in existence in the literature, and we shall not attempt to review them in any detail, as good reviews are readily available. We do note the contributions of Parkinson (1980), Garman and Klass (1980), Rogers and Satchell (1991), and Rogers et al. (1994). More recent contributions include Sutrick et al. (1997), Yang and Zhang (2000), and Magdon-Ismail et al. (2000) and (2004). Examples of data applications are numerous; see, inter alia, Rogers et al. (1994) and Chan and Lien (2003). In Section 2 we present our estimator and discuss some of its properties relative to alternative procedures. In Section 3 we provide further analysis of our estimator, considering it as a method of moments estimator. Section 4 concludes.

## Section 2

Let $(O_t, H_t, L_t, C_t)$ be the opening, high, low, and closing prices of some financial entity observed at day t. Of particular interest to us is the daily log range.

$$R_t = \ln(H_t) - \ln(L_t)$$

2

We assume that the log-price process is a scaled Brownian motion without drift. This is known as a Samuelson Process, and it involves a diffusion process for prices P(s), whose equation of motion is given below.

$$dP(s) = \mu P(s)ds + \sigma P(s)dW(s) \tag{1}$$

and W(s) is standard Brownian motion. Solving (1) to arrive at

$$P(s) = P(0)\exp((\mu - \tfrac{1}{2}\sigma^2)s + \sigma W(s)),$$

we see that $\ln(P(s))$ is a mean-zero process if $\mu = \tfrac{1}{2}\sigma^2$.

We now consider the properties of the range for a mean-zero diffusion with constant variance. Parkinson (1980, pg. 62) provides formulae for the population moments of $R_t$ based on results of Feller (1951). These are, for a daily unit-time period,

$$E(R_t^p) = \frac{4}{\sqrt{\pi}}\Gamma(\tfrac{p+1}{2})(1 - \tfrac{4}{2^p})\zeta(p-1)(2\sigma^2)^{p/2}$$

where $\sigma$ is the daily standard deviation whilst $\zeta(x)$ is the Riemann Zeta function; see Abramovitz and Stegum (1964). The index p can be any real number greater than or equal to one.

Of particular interest to us are the moments when p = 1 or 2. Here:

$$E(R_t) = 2\sigma\sqrt{\frac{2}{\pi}} \text{ and} \tag{2}$$

$$E(R_t^2) = (4\ln(2))\sigma^2$$

We see that the use of the range allows us to estimate and analyse both the daily standard deviation and variance separately; we shall discuss here the estimation of the standard deviation.

## Direct Estimation of the Standard Deviation

Given N data-points for estimation, we note that the properties of lognormal Brownian motion ensure that our daily log-returns are independent and identically distributed. Thus a natural unbiased estimator of $\sigma$ will be

$$\tilde{\sigma} = \frac{\sum_{t=1}^{N} R_t}{2N\sqrt{\frac{2}{\pi}}}.$$

This can be construed as a method of moments estimator as we are equating the first sample and population moments. This estimator will have a variance that equals

$$Var(\tilde{\sigma}) = \frac{\pi\sigma^2}{2N}(\ln(2) - \frac{2}{\pi}) = 0.0089\frac{\sigma^2}{N}$$

The natural comparison will be with the classical estimator based on daily close-to-close data; we use $X_t = \ln(C_t) - \ln(C_{t-1})$. Given the assumption of zero-drift,

$$\hat{\sigma} = \sqrt{\frac{\sum_{t=1}^{N} X_t^2}{N}}$$

the estimator $\hat{\sigma}$ is the appropriate estimator if we assume that the process has a mean of zero. Its distribution can be described as $\frac{\sigma}{\sqrt{N}}$ times the square root of a Chi-squared with N degrees of freedom, where a Chi-squared random variable is denoted as ($\chi^2(N)$). The square root of a Chi-squared is referred to as a Chi distribution. Its moments can be deduced from the moments of a Chi-squared since it is straightforward to compute fractional moments. Indeed,

$$E(\hat{\sigma}) = \frac{\sigma\Gamma(\frac{N+1}{2})}{\sqrt{N}}\frac{\sqrt{2}}{\Gamma(\frac{N}{2})} \text{ and } E(\widehat{\sigma^2}) = \sigma^2$$

We can calculate an unbiased estimator as

$$\bar{\bar{\sigma}} = \frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N+1}{2})}\sqrt{\frac{\sum_{t=1}^{N} X_t^2}{2}}$$

From this, we can see that the variance is that of $\frac{\Gamma(\frac{N}{2})\sigma}{\sqrt{2}\Gamma(\frac{N+1}{2})}$ times the variance of a Chi-squared variable with N degrees of freedom. The final result is given by the following:

$$Var(\bar{\bar{\sigma}}) = \frac{\sigma^2}{2}(\frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N+1}{2})})^2(N - 2(\frac{\Gamma(\frac{N+1}{2})}{\Gamma(\frac{N}{2})})^2)$$

The relative efficiency, $R(\hat{\sigma}, \bar{\bar{\sigma}})$, of the two estimators is given by

$$R(\hat{\sigma}, \bar{\bar{\sigma}}) = \frac{\frac{\pi}{N}(\ln(2) - \frac{2}{\pi}).}{(\frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N+1}{2})})^2(N - 2(\frac{\Gamma(\frac{N+1}{2})}{\Gamma(\frac{N}{2})})^2)}$$

This rather messy expression can be evaluated directly for small N or by using Stirling's formula for large N. This seems to work very badly in small samples.

Another more natural competitor is to construct an estimator based on the absolute value $|X_t|$. This has mean

$$E(|X_t|) = \sigma\sqrt{\frac{2}{\pi}}$$

As before we construct an unbiased estimator,

$$\tilde{\sigma} = \sqrt{\frac{\pi}{2}}\sum_{t=1}^{N}|X_t|/N$$

This will have a variance,

$$Var(\tilde{\sigma}) = \frac{(\pi-2)\sigma^2}{2N}$$

and the relative efficiency, $R(\hat{\sigma}, \tilde{\sigma})$, is given by

$$R(\hat{\sigma}, \tilde{\sigma}) = \frac{\pi(\ln(2) - \frac{2}{\pi})}{\pi - 2} = 0.1555.$$

This gives us a six- to seven-fold increase in efficiency of our estimator relative to $\tilde{\sigma}$.

## Estimating the Standard Deviation (SD) When the Market Is Closed for Part of the Day

In the above we are implicitly assuming that the market is open for 24 hours a day. In reality, virtually all markets will be closed some of the time. For such a market, consider the volatility contribution from the time the market closed last night ($price = C_{t-1}$) to the time it opens the next day ($price = O_t$). This can sensibly be measured by the quantity $(\ln(\frac{O_t}{C_{t-1}}))^2$; here we do not even need to assume that the volatility per unit time when the market is closed is the same as the volatility per unit time when the market is open, which is really an assumption about the arrival of information.

With the addition of this second, overnight, contribution to the previously discussed intraday portion of our formula, our final estimator of annualised SD, DVOL, is given by

$$\text{DVOL} = \sqrt{\frac{252\sum_{t=1}^{N}\left(\ln\left(\frac{O_t}{C_{t-1}}\right)\right)^2 + 252\pi\left(\frac{\sum_{t=1}^{N}\ln\left(\frac{H_t}{L_t}\right)}{2\sqrt{2N}}\right)^2}{N}} \quad (3)$$

This slightly simplifies to

$$\text{DVOL} = \sqrt{\frac{252}{N}\sum_{t=1}^{N}\left(ln\left(\frac{O_t}{C_{t-1}}\right)\right)^2 + \frac{252\pi}{8}\left(\frac{\sum_{t=1}^{N}\left(\frac{H_t}{L_t}\right)}{N}\right)^2} \quad (4)$$

Of course, when the market is closed, no data are available. The next best thing is to use the close of the night before and the open the next day ($C_{t-1}$ and $O_t$) to get an estimate of the overnight volatility and the first potential follow-up opportunity for a delta hedger. Also our measure of annualised daily volatility will still be correct even if volatility is different in the two intraday periods. Whilst it might be argued that we should use a second moment estimator inside (4), or that we should rescale the second term by a different function of N, we prefer the above representation because in the important special case where the market never closes and the first term is zero, we recover our range-based estimator. Furthermore for the values of N envisaged, the difference in scaling is likely to be negligible.

Finally, we stress the advantages of the above formula in being independent of the actual times the market is open and closed over the day. This is because the overnight range is directly affected by the time the market is closed and the intraday high/low range is directly affected by the time the market is open. This eliminates the need to go back in time and collect exchange opening and closing times as these times may have changed. In other words, irrespective of the actual times concerned the two components naturally capture that component of volatility attributable to that sub-period and the two added together give us the daily volatility scaled up to an annualised measure by multiplying by 252. Finally, there is no need for any covariance terms as the two components do not overlap and, because of the properties of (1), the fact that they do not overlap implies that they are independent.

## Section 3: Interpretation as a Method of Moments Estimator

In this section, we elaborate further on the properties of our estimator as a method of moments estimator in the face of misspecification. The situation we shall consider is when the drift of the process is non-zero, since it may be argued that there are periods when the drift is non-zero even at a daily frequency.

Buescu et al. (2011) provide a thorough analysis of method of moments estimation for the case of log Brownian motion with drift in which they provide an implicit estimation of daily standard deviation. They also correct a number of published and unpublished results in the intraday volatility literature.

They also derive the expected value of the range with a drift parameter $\mu$, and a time (elapsed value) of t, and they show (see Theorem 2.1, page 6, op. cit.), for $UR_t$, the range of the non-zero mean process.

$$E(UR_t) = (\mu + \frac{\sigma^2}{\mu})(1 - 2\Phi(\frac{-\mu\sqrt{t}}{\sigma})) + 2\frac{\sigma\sqrt{t}}{\sqrt{2\pi}}\exp(-\frac{t\mu^2}{2\sigma^2}),$$

where $\Phi(x)$ is the distribution function of the standardised normal. In the context of our problem, t = 1, corresponding to daily data, whilst $\mu = 0$. It may be thought that this estimator might be preferred but in fact it has no closed form expression based on data, as the authors discuss on page 14. Furthermore, it involves knowing explicitly what proportion of the day the market is open, a parameter we do not need to specify explicitly in our equation.

For t = 1, we can expand the above expression in Taylor's series about 0 for the $\Phi$ and exp functions. Upon simplification we arrive back at the formula in (2).

The expansion is given below.

$$E(UR_t) = (\mu + \frac{\sigma^2}{\mu})(\frac{2\mu}{\sqrt{2\pi\sigma^2}} - \frac{2\mu^3}{\sigma^3\sqrt{2\pi}}) + 2\frac{\sigma}{\sqrt{2\pi}}\left(1 - \frac{\mu^2}{2\sigma^2}\right) + O(\mu^4) \qquad (5)$$

If we now subtract (2) from (5), we see that

$$E(UR_t) - (E(R_t) = O(\mu^4).$$

This represents a very small amount for typical daily drifts that we might see in practise. For example, if $252\,\mu = 10\%$ per annum, a fairly high annual drift in the current environment, then $\mu = 0.004$ and $\mu^4 = 256\ x\ 10^{-12}$ approximately; it is clear that this should be very small in most realistic cases.

We now provide some empirical results. In Table 1 below, we present the maximum, average, and minimum values for 1-month, 3-month and 12-month volatilities based on interday (close to close) data for the SPDR® S&P 500® ETF on NYSE Arca® from 29 Jan 1993 (at start of trading) to 31 March 2015. These are denoted by the acronym VOL. We also present in Table 2, for the same data, equivalent intraday measures based on our estimator; these are denoted DVOL. The corresponding volatility of volatility measures are denoted by VOV and DVOV respectively. The symbol key can be found immediately below the two tables. We find clear evidence of a substantial reduction in the volatility of volatility as predicted by theory. Whilst the reductions are not exactly equal to efficiency gains derived earlier in the paper, they nevertheless provide convincing evidence that there are tangible benefits to adopting our estimator. We also observe that the range of our range-based estimator is smaller over 1, 3, and 12 months.

Table 1

|  | 1VOL | 3VOL | 12VOL | 1VOV | 3VOV | 12VOV |
|---|---|---|---|---|---|---|
| Max | 91.25% | 73.86% | 45.56% | 317.93% | 131.41% | 54.61% |
| Avg | 16.29% | 16.74% | 17.57% | 99.97% | 35.83% | 10.71% |
| Min | 5.14% | 6.42% | 8.52% | 32.66% | 5.33% | 0.99% |

Table 2

|  | 1DVOL | 3DVOL | 12DVOL | 1DVOV | 3DVOV | 12DVOV |
|---|---|---|---|---|---|---|
| Max | 89.87% | 69.90% | 41.28% | 272.68% | 80.81% | 11.94% |
| Avg | 16.22% | 16.38% | 16.79% | 53.61% | 21.13% | 6.48% |
| Min | 5.27% | 6.40% | 7.45% | 19.21% | 10.20% | 3.24% |

**Key**

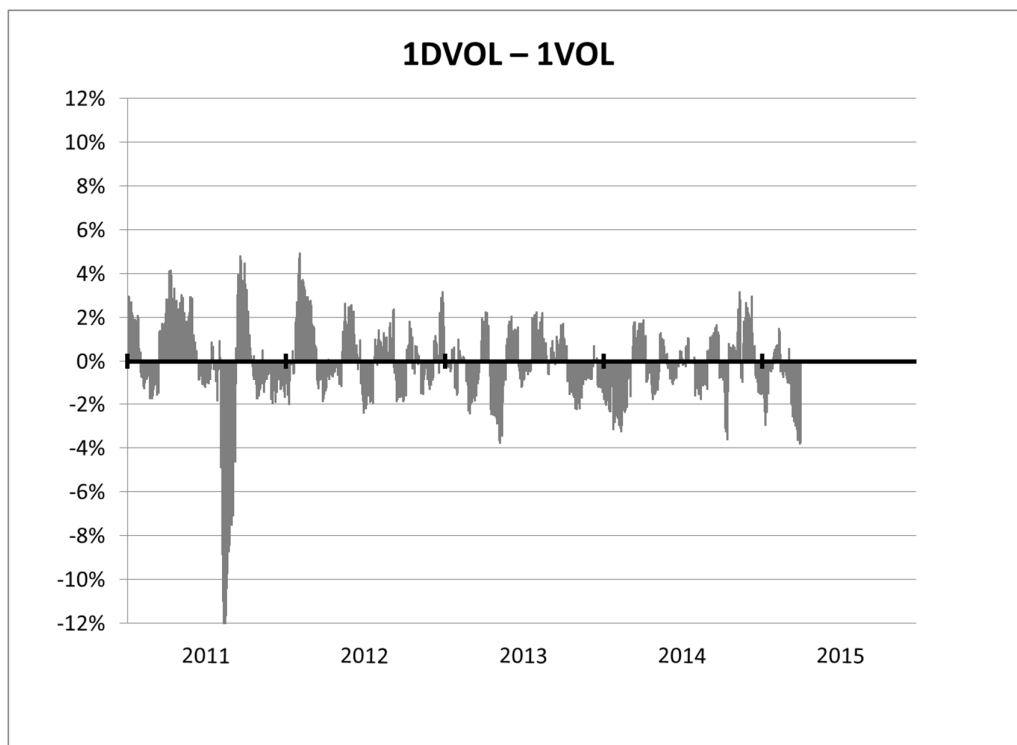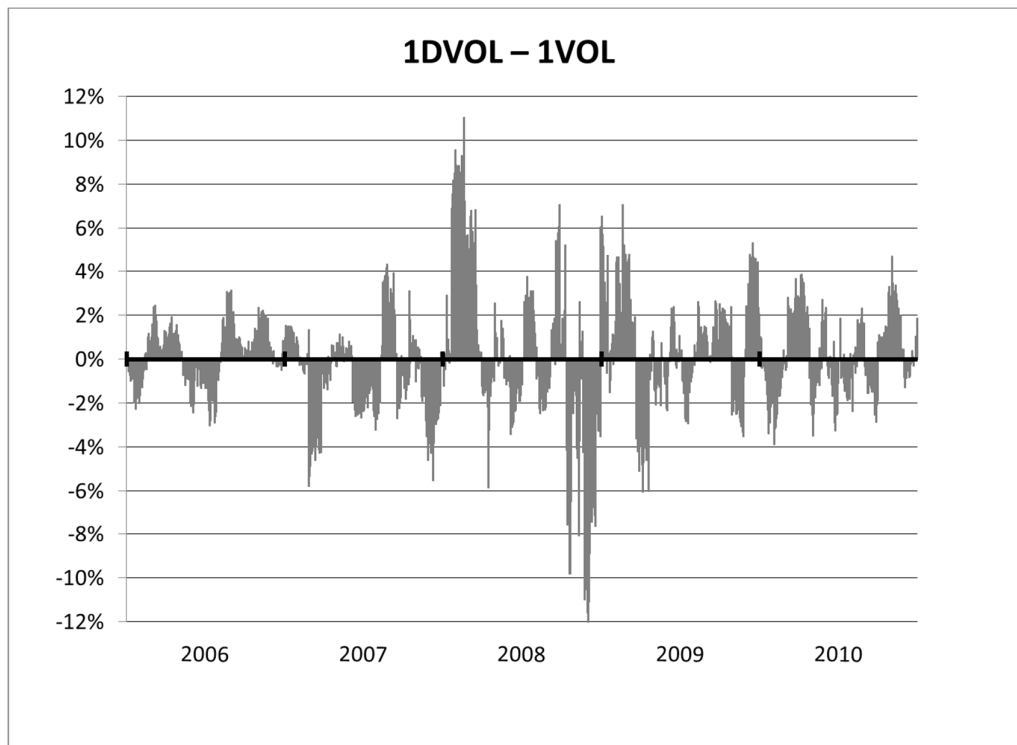| 1VOL | 1-month/21-trading-day interday (close-to-close) realised volatility |
|---|---|
| 3VOL | 3-month/63-trading-day interday realised volatility |
| 12VOL | 12-month/252-trading-day interday realised volatility |
| 1VOV, 3VOV, and 12VOV | 1-month interday realised volatility of 1-, 3-, and 12-month interday realised volatility, respectively |
| 1DVOL, 3DVOL, and 12DVOL | 1-, 3-, and 12-month overnight/intraday realised volatility, respectively |
| 1DVOV, 3DVOV, and 12DVOV | 1-month interday realised volatility of 1-, 3-, and 12-month overnight/intraday realised volatility, respectively |

The vol of vol calculation is performed on the rolling 1VOL, 3VOL, 12VOL, 1DVOL, 3DVOL, and 12DVOL time series using the standard interday formula outlined earlier.

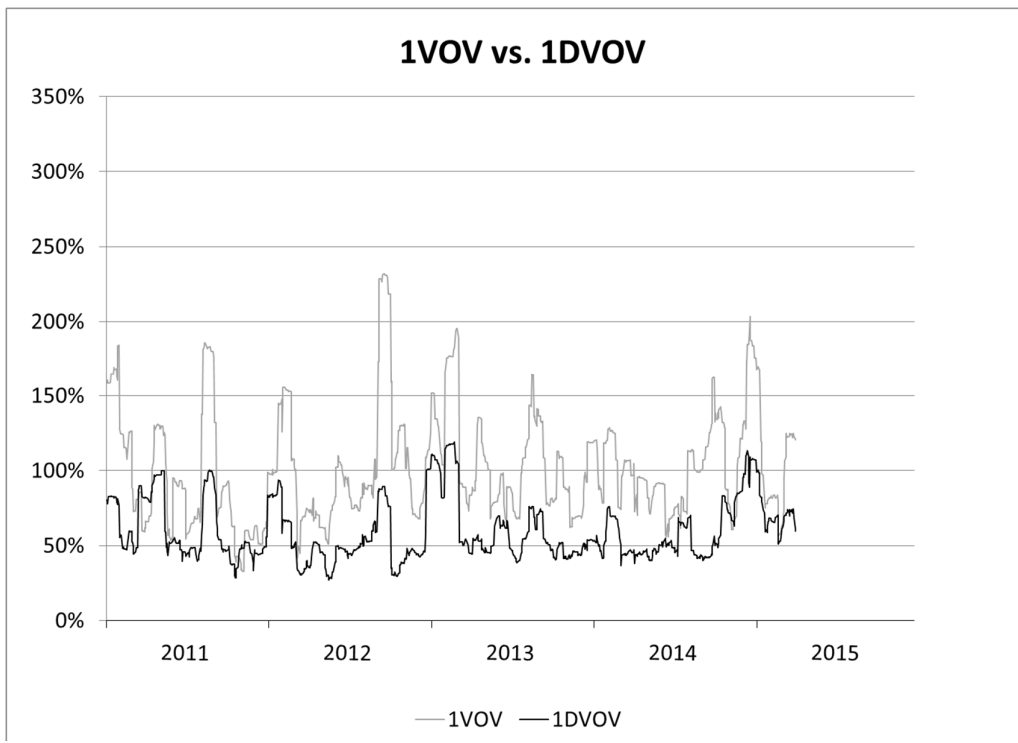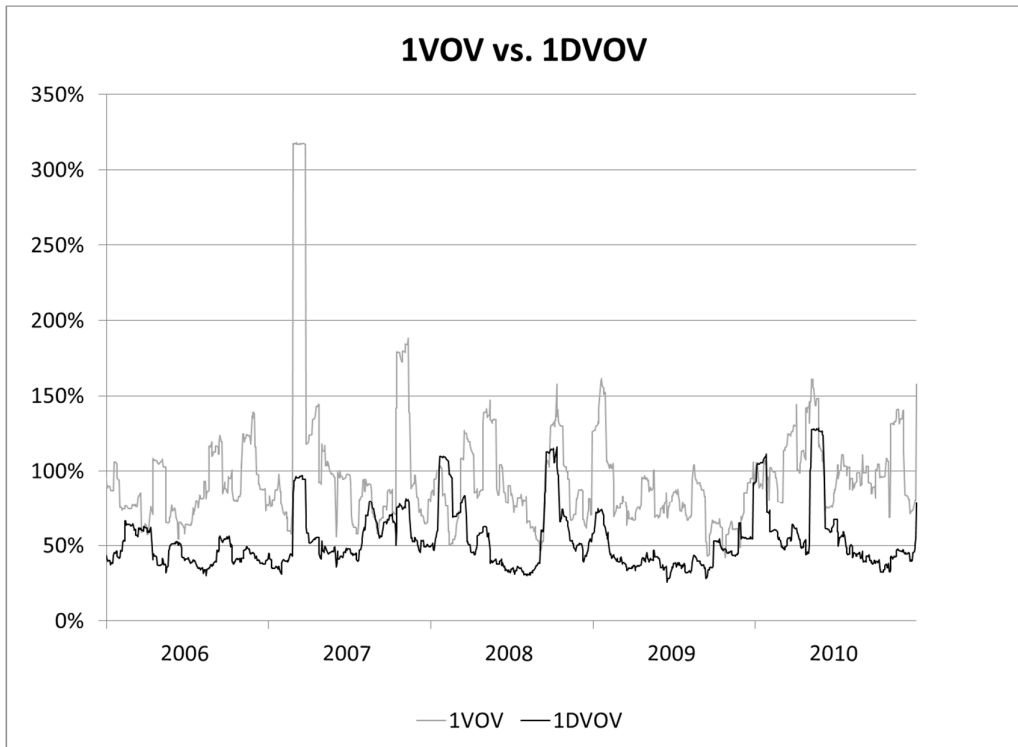$$\sigma = \sqrt{\frac{252 \sum_{t=1}^{N} X_t^2}{N}}$$

Since the underlying time series are realized volatility, the result of performing the interday realized volatility calculation is the vol of vol. It should be noted that even though the underlying time series uses 1, 3, and 12 months, the vol of vol calculation uses a 1-month (21-day) time frame for all six series. It should also be noted that we cannot use our new DVOL calculation on the VOL or DVOL series to get vol of vol because there is no open, high, and low data available for a VOL series or DVOL series. We therefore resort to using the standard interday formula on daily (i.e., closing) values.

Graph 1 plots the difference in the two estimators whilst Graph 2 plots the two vols of vol. Both graphs were split into two five-year time periods in order to show more detail. It is immediately evident that the two estimators do not coincide and that, in virtually every time period, 1DVOV is noticeably smaller than 1VOV. This reduction in the vol of vol is the feature we hoped to find in our new estimator. A full empirical characterisation in terms of market conditions of when the intraday estimator "works better" is a topic for future study.

Graph 1

Graph 2



1VOV vs. 1DVOV



1VOV vs. 1DVOV

## Conclusion

In this paper, we have examined the properties of a simple estimator based on the overnight and high/low range and shown it to be easy to calculate and highly efficient compared to conventional estimators of the daily standard deviation of returns. We felt that it more closely resembles the results an intraday hedger would experience given the assumption that tick data is not readily available. Unlike other intraday estimators, DVOL utilizes the same assumptions as the interday volatility formula that is prevalent in the over-the-counter volatility and variance swaps marketplace, and the soon-to-be-launched instruments in the exchange-traded arena.

We have also presented an adaptation of the estimator that allows the use of open and close price data when the market is closed for part of the day without the need to weight each term. Furthermore, our estimator seems robust to situations where the drift, which we have assumed to be zero, is in fact, non-zero. We demonstrate this by considering an approximation based on a comparison of the zero mean drift estimator versus the more sophisticated but also much more complex estimator based on the expected value of the range when the drift is non-zero.

Finally we provide evidence of risk reduction by looking at the volatility of volatility of our estimator and the conventional close to close estimator using data based on the SPDR® S&P 500® ETF on NYSE Arca® from 29 Jan 1993 (at start of trading) to 31 March 2015. The empirical results clearly support the theoretical calculations provided in the paper.

## References

ABRAMOVITZ and STEGUM (1964): Handbook of Mathematical Functions National Bureau of Standards, Applied Mathematics Series, Number 55, US GPO, Washington.

BUESCU, C, M. TAKSAR1, and F. J. KON'E (2011): An application of the method of moments to volatility estimation using daily high, low, opening and closing prices, arXiv:1112.4534v1 [q-fin.ST] 20 Dec 2011. CHAN, L. and D.

LIEN (2003): Using high, low, open, and closing prices to estimate the effects of cash settlements on futures prices. *International Review of Financial Analysis* 12, 35–47.

FELLER, W. (1951): The asymptotic distribution of the range of sums of independent random variables. *Annals of Mathematical Statistics* 22, 427–432.

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

KON′E, F. J. (1996): *Estimation of the volatility of stocks using the high, low and closing prices.* Ph.D. thesis, State University of New York at Stony Brook.
.
MAGDON-ISMAIL, M. and A. ATIYA (2000): Volatility estimation using high, low and close data — a maximum likelihood approach. *Computational Finance*, June CF2000 Proceedings.

MAGDON-ISMAIL, M., A. ATIYA, A. PRATAP, and Y. ABU-MUSTAFA (2004): On the maximum drawdown of a Brownian motion. *Journal of Applied Probability*, 41(1), 147–161.

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

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

ROGERS, L.C.G., S. SATCHELL, and Y. YOON (1994): Estimating the volatility of stock prices: a comparison of methods that use high and low prices. *Applied Financial Economics* 4, 241–247.

ROGERS, L.C.G. and F. ZHOU (2008): Estimating correlation from high, low, opening, and closing prices. *The Annals of Applied Probability* 18(2), 813–823.

SUTRICK K., J. TEALL, A. TUCKER, and J. WEI (1997): The range of Brownian motion processes: density functions and derivative pricing applications. *The Journal of Financial Engineering* 6(1), 31–46.

YANG D. and Q. ZHANG (2000): Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business* 73, 477–491.