Generalised Variance

Research Proposal for Quantitative Business Analysis Honours

Maxwell Sutton

May 28, 2014

1 Introduction

2 Literature Review

This section comprises of a review of literature for two historical volatility estimators, realized volatility and realized range. The following common notation is used throughout these sections. All prices discussed have been log-transformed. C_t is the closing price at the end of period t and O_t is the opening price. H_t is the highest price observed during the period from $t-1$ to t and L_t is the lowest price. There are n days, indexed by i; C_i is the closing price for the ith day. Within each day, there are M intervals, indexed by j; $C_{i,j}$ is the closing price of the jth period on the ith day.

Introducing the mathematical concept of market microstructures, we can write the observed price of an asset at time t as

$$\tilde{p}_t = p_t \vartheta_t$$

where p_t is the true or equilibrium price that is unobserved and ϑ_t denotes market microstructure noise. The equilibrium prices are assumed to come from the following process

$$\log(p_t) - \log(p_0) = \int_0^t \phi_s \, ds + \int_0^t \sigma_s \, dW_s$$

where ϕ_t is a continuous predictable drift process, σ_t is a cadlag spot volatility process and W_t is a standard brownian motion. The object of interest is the integrated variance, IV

$$IV = \int_0^h \sigma_s^2 \, ds$$
$$IQ = \int_0^h \sigma_s^4 \, ds$$

where IQ is the integrated quarticity that is necessary in deriving the asymptotic distributions.

2.1 Realized Volatility

The traditional method to estimate volatility is the close-close estimator, which is the sample variance of returns based on the closing prices. Using squared returns to estimate volatility was first seen in **taylor1986modelling** and then the traditional form was settled by **French1987** and **Schwert1989**

$$\hat{\sigma}_{cc}^2 = \sum_{i=1}^{n} \left(C_i - C_{i-1}\right)^2$$

Garman1980 suggested that its disadvantage is that it does not take into account all information available, thus losing efficiency. The intraday equivalent was formalised by **Andersen2001** as the summation of intraday squared returns

$$\hat{\sigma}_{rv,t}^2 = \sum_{j=1}^{M} \left(C_{j,t} - C_{j-1,t}\right)^2$$

It was suggested that by examining the returns over smaller and smaller intervals, realized volatility would accurately measure integrated volatility. The authors investigated the distribution of the estimator on only two exchange rates; the deutschemark-US dollar (USDDEM) and and the yen-US dollar (USDJPY). A more thorough examination of the properties of realized volatility was conducted on 30 stocks from the Dow Jones Industrial Average (DJIA) by **Andersen2001a**

Due to the lack of contemporaneous competing estimators, the authors in both cases can only examine the distribution of the realized volatility estimator. In addition, as volatility of empirical data is not known a priori, the accuracy of the estimator was not considered in the empirical section. Neither paper provides a simulation study of the properties of realized volatility, which would allow a comparison of a known true value with the estimated value.

In the absence of noise and market microstructures, **Andersen2003** show that as the frequency of observations collected increases and the time interval between observations decreases to 0, the realized volatility estimator converges to the integrated variance. The asymptotic distribution of realized volatility, under a stochastic Geometric Brownian Motion process was shown by **Bandi2008** to be Gaussian in the multivariate case. It was reported for the univariate case by **McAleer2008**

$$\sqrt{n}\left(\hat{\sigma}_{rv,t}^2 - IV\right) \xrightarrow[n \to \infty]{d} \mathcal{N}\left(0,\, 2IQ\right)$$

A good review of the effects of microstructure noise on the realized volatility estimator is given by **McAleer2008** **Bandi2008** showed that when there was noise, the estimate provided by the realized volatility estimator would be infinite as the frequency of observations increased, resulting in vast overestimation of the true volatility.

Attempts to alleviate the effects of market microstructures include: considering multiple time scales (**Zhang2005**), using kernels (**Zhou1996**; **Barndorff-Nielsen2008a**) and scaling realized volatility (**Martens2007**). **McAleer2008** provide a good review of the techniques used.

2.2 Realized Range

Parkinson1980 introduced the first range-based estimator of volatility. The author proposes that there is a clear relationship between the range and the spot-volatility process and finds that the estimator

$$\hat{\sigma}_p^2 = \frac{1}{4\log(2)} \sum_{i=1}^{n} (H_i - L_i)^2$$

is more efficient than an estimator based on squared returns. However this estimator assumes zero drift and Geometric Brownian Motion (**Kunitomo1992**). **Garman1980** suggested improvements in efficiency could be made when also taking into account the open and the close price for each day, resulting in a family of estimators. The practical estimator as stated by **Haug2007** was

$$\hat{\sigma}_{gk}^2 = \frac{1}{2} \sum_{i=1}^{n} (H_i - L_i)^2 - (2\log(2) - 1) \sum_{i=1}^{n} (C_i - C_{i-1})^2$$

as a linear combination of Parkinson's estimator and the traditional close-close estimator. This estimator was found to be even more efficient than Parkinson's, however it was still dependent on the drift. Finally, **Rogers1991** introduced an estimator that included the open price and was independent of the drift term

$$\hat{\sigma}_{rs}^2 = \sum_{i=1}^{n} (H_i - C_i)(H_i - O_i) + (L_i - C_i)(L_i - O_i)$$

DennisYang2000 and **Kunitomo1992** have both introduced other estimators of volatility that use the range. The Yang and Zhang estimator uses estimates over multiple periods while Kunitomo's estimator considers a Brownian Bridge to form an adjusted estimate of the range, removing the drift. These two estimators are not considered. The reader is directed to **Molnar2012** for further reading.

Range-based estimators share some common properties. Prices are not observed continuously; rather, they are observed at discrete time intervals. In addition, a transaction is a discrete event and therefore the change in price due to this transaction is also discrete. This means that the continuous price process is observed only at discrete times. The effect of this is that during an interval, the observed highest price may be lower than the highest true price. This effect can bias range-based estimators downwards (**Jacob2008**).

Realized range was introduced in two concurrent concurrent papers by **Martens2007** and **Christensen2007** In both papers, Parkinson's estimator is employed, however, instead of daily high and low data, the highest and lowest price observed within a subinterval are considered.

$$\hat{\sigma}^2_{rrv,t} = \frac{1}{4\log(2)} \sum_{j=1}^{M} (H_{j,t} - L_{j,t})^2 \tag{2.1}$$

The realized range-based volatility can reduce the effect of drift on Parkinson's range estimator, as for smaller intervals, the drift will become negligible. Thus $\hat{\sigma}^2_{gk}$ and $\hat{\sigma}^2_{rs}$ are not explored. **Martens2007** report that they do improve results, however the main results are only shown for the form seen in Eq.(2.1).

Martens2007 consider various bias correction factors to improve both realized volatility and realized range. To compare the performance of the estimators, the authors conduct a simulation; 100 price observations are simulated for each second during 24 hour trading over 5000 days (**Martens2007**) in an attempt to produce as close to a continuous process as possible. The prices followed a Geometric Brownian Motion process with volatility chosen to be 0.21. This choice of volatility is not justified or explained with reference to average volatility found empirically or previous research assumptions. In addition, empirical data does not exhibit constant volatility; **Alizadeh2002** found that Geometric Brownian Motion is not sufficient to explain the autocorrelation or variance of volatility as found in foreign exchange data.

The effect of two market microstructures; bid-ask bounce and infrequent trading, on the accuracy of realized volatility and realized range are examined in the simulation. The authors find that in the absence of noise, the realized volatility estimator is superior, however for both market microstructures, the realized range estimator is more accurate at higher frequencies (**Martens2007**). This accuracy is measured using root mean squared error.

Martens2007 examine the distributional properties of realized volatility and realized range on an empirical dataset of the S&P 500 futures index, replicating **Andersen2001a** The authors, however, expand on this methodology, incorporating ideas from **Beckers1983** and **Fleming2003** Comparing the explanatory power of each estimator, realized range is superior to realized variance (**Martens2007**). The authors then consider individual stocks of the S&P 100 and find similar results.

In contrast, **Christensen2007** examine the theoretical properties of the realized range estimator in more depth. They first show consistency of the estimator to integrated variance and derive the asymptotic distribution as

$$\sqrt{n}\left(\hat{\sigma}^2_{rrv,t} - IV\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0,\, \lambda IQ\right)$$

where $\lambda \approx 0.4$. Thus the asymptotic variance of the realized range estimator is five times smaller than the realized volatility estimator. **Christensen2007** evaluate the realized range estimator through both simulation and empirical analysis. The authors use a stochastic volatility model which is an improvement over **Martens2007** however their methods are inferior.

The authors simulate 1,000,000 days, each with 100, 500 or 1000 increments per day (**Christensen2007**). The result is a poorer approximation of the continuous price process. The authors only examine the asymptotic distribution of realized range, instead of how accurately it measures the true value that was simulated. There is also no comparison to other available estimators. Finally, the authors do not examine the effects of market microstructures within the simulation.

The empirical study undertaken by **Christensen2007** is lacking as it only considers one stock; General Motors, and the authors only examine the distributions of realized volatility and range and their correlation. The methodology found in **Martens2007** is more substantial.

3 Hypothesis Development

This proposal suggests an extension to realized volatility in an attempt to remove the effect of market microstructures. While realized volatility considers the squared return from period 1 to period 2 and then period 2 to period 3, the suggestion is to also consider the return from period 1 to period 3. Such an estimator would then take into account all possible returns and perhaps give a better indication of volatility. By expanding the interval the return is calculated over, the effect of market microstructures may also be reduced. A general form for this estimator could be written as

$$\hat{\sigma}^2_{g,t} = Var\left(\frac{C_{t_2} - C_{t_1}}{f(t_1, t_2)}\right)$$

where $t-1 \leq t_1 < t_2 \leq t$, $f(i,j)$ is a function to correct for taking different sized intervals, $Var(x)$ is the variance of x. The problem with this estimator is that at a frequency of 1 second, there might be up to 86,400 intervals in per day. Considering all possible sized intervals, there are over 3.7 billion combinations for each day. Thus the problem is not tractable.

In order to reduce the number of combinations, the first proposal is to fix t_1 as the opening price for each day. The first problem with fixing the first price as the opening price, is that this introduces a source of error. Changes in the opening price will affect all other terms in the estimator. Preliminary simulations have been run with the following estimator

$$\hat{\sigma}^2_{gva,t} = \frac{1}{2}Var\left(\frac{C_j - C_0}{j}\right) + \frac{1}{2}Var\left(\frac{C_M - C_j}{M - j}\right) \tag{3.1}$$

The second term is an attempt to remove some of the error. By fixing the closing price of the day as the final point and taking the average of the two estimates, this is thought to remove some of the error. The function chosen, $f(t_1, t_2) = t_1$ and $f(t_1, t_2) = t_2 - t_1$ were chosen to scale the multiple period returns back to one period. However, the preliminary results suggest this is not the correct functional form. As the frequency of observations increases, it is found that $\hat{\sigma}^2_{gva,t}$ approaches 0.

A cursory examination of the theoretical properties of $Var(C_j - C_0)$ suggests that swapping the denominators of Eq.(3.1) may yield better results, however this is not included in the simulation results at the time of writing.

The second proposal is to consider a Newey-West type estimator, where the maximum number of periods in each interval is truncated. **Newey1987** introduced an estimator of the covariance matrix for regression coefficients. Its main use is in time-series regression as it takes into account autocorrelation of residuals. This option is has not been examined at all at the time of writing, but is an interesting avenue of research.

4 Proposed Methodologies

The properties of generalised variance will be examined through a simulation study and an empirical study. This section outlines the methods that will be used and the reasons for each choice of method.

4.1 Simulation Study

Following **Martens2007** an attempt will be made to generate as close to a continuous price process as possible. Therefore, 100 price observations will be simulated per second, in a 24 hour trading period. The starting price will be set at $100; this is chosen arbitrarily, and the prices will be simulated for 1000 days following this first observation. By convention, there are generally 250 to 252 trading days in a year; this report assumes 252, providing 4 years of simulation.

The expected return will be set at 20% per annum; as a comparison, the ASX 200 posted a yearly return of 17.29% in the year up to June 2013 (22.2% when adjusted for dividends) (**Yeow2013**). This could be expanded to consider different expected returns to identify if the drift has a significant effect on the estimators, however the drift over the small time scales considered would be negligible, as discussed in Section 2.2.

In order to capture different periods of volatility, four different starting values for the volatility; 5%, 10%, 20% and 30%. The realized volatility of the All Ordinaries over the past five years as found by **Herber2014** seems to fall between these values.

To improve upon **Martens2007** the cases of constant volatility and dynamic volatility will be considered. Three different volatility processes will be simulated, using the properties above. The first will be Geometric Brownian Motion, with constant volatility. A GARCH(1,1) process will then be considered, which takes the form

$$
\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
a_t &= \sigma_t \varepsilon_t \\
\varepsilon_t &\sim \mathcal{N}(0,\, 1)
\end{aligned}
$$

The unconditional variance; the variance in the long-run, of σ_t^2 in the GARCH(1,1) model is given by

$$
v = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}
$$

The intial value, $\sigma_{t=1}^2$, is set equal to the geometric brownian motion σ_{gbm}^2. In addition, α_0 is chosen so that $v = \sigma_{gbm}^2$. That is, the long run variance of the GARCH(1,1) model is equal to the variance of the Geometric Brownian Motion model.

Finally, a popular stochastic volatility model introduced by **Heston1993** is used, where the variance is a Cox-Ingersoll-Ross (CIR) process given by **Wilmott2007** as

$$
\begin{aligned}
d\sigma_t^2 &= \kappa \left(\theta - \sigma_t^2\right) dt + \xi \sigma_t dW_t^v \\
dW_t dW_t^v &= \rho dt
\end{aligned}
$$

The second equation forces that the Wiener process for the CIR process is correlated with the Wiener process in the price equation. θ is the long-run variance, which again is equal to σ_{gbm}^2. κ gives the speed of reversion to this long-run variance level, while ξ gives the variance of the variance process. Again, the initial value of σ_t^2 is chosen as the Geometric Brownian Motion value, as it θ.

For each series, estimates of the daily variance were calculated at several frequencies: 1 sec, 5 sec, 10 sec, 20 sec, 1 min, 5 min, 10 min and 20 min.

4.2 Empirical Study

The performance of each estimator will be examined at the same frequencies of observations as used in the simulation study. The date range selected for this empirical report was

01/01/10 until 31/12/13. This date range was chosen as there are 999 working or trading days within this period, which matches up quite closely with the 1000 days used in the simulation study.

The stocks that were in the ASX 20 as of 13/05/14 were chosen as the pool of stocks to examine. It is the 20 stocks on the Australian Stock Exchange with the highest market capitilisations. This list includes the top 10 most traded shares of 2013, as reported by **MostTradesASX10** Three of the stocks did not have the required data, and so were excluded from the analysis. The data was sourced from Sirca.

The first step will be to examine the distributions of the estimators for different frequencies in the manner of **Christensen2007** A normal distribution is desirable for use in stochastic volatility models (**Christensen2007**). As such, the first four moments of the distributions will be examined; mean, variance, skewness and kurtosis, followed by tests for normality. The second step will follow **Martens2007** in attempting to find the explanatory of each estimator with regards to each other in an attempt to find the best estimator without introducing bias.

Finally, the performance of the estimators as inputs to forecasting Value-at-Risk; a prevalent measure of risk, and the related Expected Shortfall will give an indication as to the ability of the estimators to help analysts measure and forecast risk.