Volatility Estimation

Ser-Huang Poon

August 11, 2008

1 Introduction

Consider a time series of returns r_{t+i}, $i = 1, \cdots, \tau$ and $T = t + \tau$, the sample variance, σ^2,

$$\widehat{\sigma}^2 = \frac{1}{\tau - 1} \sum_{i=1}^{\tau} (r_{t+i} - \mu)^2 \,, \tag{1}$$

where r_t is the return at time t, and μ is the average return over the τ-period, and $\sigma = \sqrt{\sigma^2}$ is the *unconditional volatility* for the period t to T. If $T - t$ is e.g. a ten-year period and t is measured in daily interval, then $\widehat{\sigma}^2$ in (1) is the daily variance, $\widehat{\sigma}^2_d$, over the ten-year period. If t is measured in weekly interval, then $\widehat{\sigma}^2$ in (1) is the weekly variance, $\widehat{\sigma}^2_w$, over the ten-year period. Since variance is linear in time and can be aggregated but not standard deviation,

$$\widehat{\sigma}^2_w = 5 \times \widehat{\sigma}^2_d$$

with a multiplier of 5 since there are 5 trading days in a week.[1] To derive volatility, which is often linked to the standard deviation, we have the weekly

[1] Note that such scaling property is not very accurate if the return, r_t, are not *iid* (indenpendent and identically distributed).

volatility

$$\begin{aligned}\widehat{\sigma}_w &= \sqrt{5 \times \widehat{\sigma}_d^2} \\ &= \sqrt{5} \times \widehat{\sigma}_d\end{aligned}$$

and daily volatility is simply $\widehat{\sigma}_d$.

It is a well known fact that volatility does not remain constant through time, the *conditional volatility*, σ_t, is a more relevant information for asset pricing and risk management at time t. So it is a common practice to break $T-t$ up into smaller superiods such that

$$\begin{aligned}T-t &= (T_n - T_{n-1}) + (T_{n-1} - T_{n-2}) + (T_{n-2} - T_{n-3}) + \cdots + (T_1 - t) \\ &= \tau_n + \tau_{n-1} + \tau_{n-2} + \cdots + \tau_1\end{aligned}$$

and (1) becomes

$$\widehat{\sigma}_t^2 = \frac{1}{\tau_j - 1} \sum_{i=1}^{\tau_j} (r_{t+i} - \mu)^2, \qquad j = 1, \cdots, n$$

Figlewski (1997) noted that the sample mean μ is a very inaccurate estimate of the true mean especially for small samples; taking deviations around zero instead of the sample mean as in equation (1) typically increases volatility forecast accuracy. Hence,

$$\widehat{\sigma}_t^2 = \frac{1}{\tau_j - 1} \sum_{i=1}^{\tau_j} r_{t+i}^2, \qquad j = 1, \cdots, n \tag{2}$$

Volatility estimation procedure varies a great deal depending on how much information we have at each sub-interval t,[2] and the length of the volatility reference period, τ_j, i.e. the period to which the volatility estimate

[2] Recently, intraday transaction data has become more widely available providing a channel for more accurate volatility estimate and forecast. This is the area where much research effort is concentrated in the last two years.

is applied. Many financial time series are available at the daily interval, while the volatility reference period, τ_j, could vary from 1 to 10 days (for risk management), months (for option pricing) and years (for investment analysing).

When monthly volatility is required and daily data is available, volatility can simply be calculated based on equation (2), where τ_j is one month, and r_{t+i} for $i = 1, \cdots, \tau_j$ are the daily observations in that month. Many macroeconomic series are available only at the monthly interval, so the current practice is to use absolute value to proxy for volatility since $\tau_j = 1$ (Note that we will not divide the estimator by $\tau_j - 1 = 0$ as it will make the volatility infinite). The same applies to financial time series when daily volatility estimate is required and only daily data is available.

The use of daily absolute return to proxy daily volatility will produce a very noisy volatility estimator. This is explained in Section 1.1 later. Engle (1982) was the first to propose the use of an ARCH (Autoregressive Conditional Heteroskedasticity) model below to produce conditional voaltility for inflation rate r_t;

$$\begin{aligned} r_t &= \mu + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \sqrt{h_t}\right) \\ \varepsilon_t &= z_t \sqrt{h_t}, \\ h_t &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots . \end{aligned} \tag{3}$$

where h_t is the conditional variance and $\sigma_t = \sqrt{h_t}$ is the conditional volatility. The ARCH model is estimated by maximising likelihood of observing $\{\varepsilon_t\}$. This approach of estimating conditional volatility is less noisy than the absolute return approach but it relies on the assumption that the ARCH model in (3) is the true return generating process, ε_t is gaussian and the time series is long enough for such a maximum likelihood estimation.

Moreover, while $\widehat{\sigma}^2$ in equation (1) is an unbiased estimator for σ^2, the square root of $\widehat{\sigma}^2$ is a biased estimator for σ due to Jensen inequality.[3] Ding, Granger and Engle (1993) suggest measuring conditional volatility directly from absolute returns. Davidian and Carroll (1987) show absolute returns volatility specification is more robust against asymmetry and non-normality. There is some empirical evidence that deviations or absolute returns based models produce better volatility forecasts than models based on squared returns (Taylor (1986), Ederinton and Guan (2000a) and McKenzie (1999)) but the majority of time series volatility models especially the ARCH class models are squared-returns models. There are methods for estimating volatility that are designed to exploit or reduce the influence of extremes.[4] Again these methods would require the assumption of gaussian variable or a particular distribution function for returns.

1.1 Using squared return as a proxy for daily volatility

Volatility is a latent variable. Before high frequency data became widely available, many researchers resorted to using daily squared return, calculated from market closing prices, to proxy daily volatility. Lopez (2001) shows that ε_t^2 is an unbiased but extremely imprecise estimator of σ_t^2 due to its asymmetric distribution. Let

$$r_t = \mu + \varepsilon_t \ , \qquad \varepsilon_t = \sigma_t z_t \ , \tag{4}$$

[3] If $r_t \sim N\left(0, \sigma_t^2\right)$, then $E\left(|r_t|\right) = \sigma_t \sqrt{2/\pi}$. Hence, $\widehat{\sigma}_t = \frac{|r_t|}{\sqrt{2/\pi}}$ if r_t has a conditional normal distribution.

[4] For example, the Maximum likelihood method proposed by Ball and Torous (1984), the high-low method proposed by Parkinson (1980) and Garman and Klass (1980).

and $z_t \sim N(0,1)$. Then

$$E\left[\left.\varepsilon_t^2\right|\Phi_{t-1}\right] = \sigma_t^2 E\left[\left.z_t^2\right|\Phi_{t-1}\right] = \sigma_t^2$$

since $z_t^2 \sim \chi^2_{(1)}$. However, since the median of a $\chi^2_{(1)}$ distribution is 0.455, ε_t^2 is less than $\frac{1}{2}\sigma_t^2$ more than 50% of the time. In fact

$$\Pr\left(\varepsilon_t^2 \in \left[\frac{1}{2}\sigma_t^2, \frac{3}{2}\sigma_t^2\right]\right) = \Pr\left(z_t^2 \in \left[\frac{1}{2}, \frac{3}{2}\right]\right) = 0.2588\ ,$$

which means that ε_t^2 is 50% greater or smaller than σ_t^2 nearly 75% of the time!

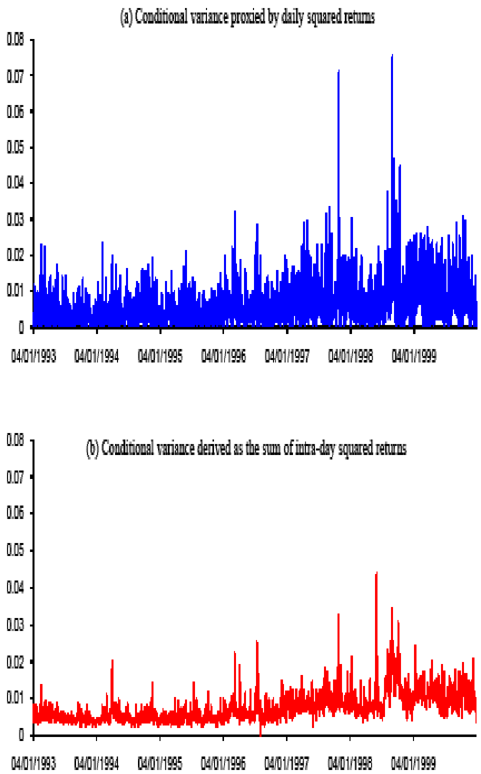
Under the null hypothesis that r_t in (4) is generated by a GARCH(1,1) process, Andersen and Bollerslev (1998) show that the population R^2 for the regression

$$\varepsilon_t^2 = \alpha + \beta\widehat{\sigma}_t^2 + \upsilon_t$$

is equal to κ^{-1} where κ is the kurtosis of the standardized residuals, z_t, and κ is finite. For conditional Gaussian error, the R^2 from a correctly specified GARCH(1,1) model is bounded from above by $\frac{1}{3}$. Christodoulakis and Satchell (1998) extend the results to include compound normals and the Gram-Charlier class of distributions and show that the mis-estimation of forecast performance is likely to be worsened by non-normality.

It is clear that the use of ε_t^2 as a volatility proxy will lead to low R^2 and undermine the inference regarding forecast accuracy. Blair, Poon and Taylor (2001) report an increase of R^2 by 3 to 4 times for the one-day ahead forecast when intra-day 5-minutes squared returns instead of daily squared returns are used to proxy the actual volatility. Figure 1 shows the time series of these two volatility estimates over the seven-year period from January 1993 to December 1999. Although the overall trends look similar, the R^2 of the regression of $|\varepsilon_t|$ on σ_t^{intra} is only 28.5%. Hence,

Figure 1: Aurocorrelation of daily returns and proxies of daily volatility of S&P100



unless there is no other choice, one should always refrain from using daily absolute return to proxy daily volatilty, or using daily squared return to proxy daily conditional variance.

1.2 Using high-low measure to proxy volatility

The high-low, also known as range based or extreme-value, method of estimating volatility is very convenient because daily high, low, opening and closing prices are reported by major news papers, and the calculation is easy to programme even with a hand held calculator. The high-low (H-L) volatility estimator was studied in Parkinson (1980), Garman and Klass (1980), Beckers (1983), Rogers and Satchell (1991), Wiggins (1992), Rogers, Satchell and Yoon (1994) and Alizadeh, Brandt and Diebold (2002). It is based on the assumption that return is conditionally normally distributed with conditional volatility σ_t. Let H_t and L_t denote, respectively, the highest and the lowest prices on day t. Applying the Parkinson (1980) H-L measure to a price process that follows a geometric Brownian motion results in the following volatility estimator (see Bollen and Inder, 2002);

$$\widehat{\sigma}_t^2 = \frac{(\ln H_t - \ln L_t)^2}{4 \ln 2}$$

The German and Klass's (1980) estimator is an extension of Parkinson (1980) where information about opening, p_{t-1}, and closing, p_t, prices are incorporated as follow:

$$\widehat{\sigma}_t^2 = 0.5 \left(\ln \frac{H_t}{L_t} \right)^2 - 0.39 \left(\ln \frac{p_t}{p_{t-1}} \right)^2 .$$

We have already shown that financial market returns are not likely to be normally distributed and have long tail distribution. The H-L volatility estimator is very sensitive to outliers. It will be useful to apply the trimming procedures, specifies in equation (10), to the entire data set first if the H-L volatility estimator is used. Provided there are no destabilising large values, the H-L volatility estimator is very efficient and, unlike the realized volatility estimator introduced in the next section, it is less affected by market microstructure noise.

1.3 Realised volatility, quadratic variation and jumps

More recently and with the increased availability of tick data (i.e. prices recorded at transaction level), the term ***realised volatility*** is now used to refer to volatility estimates calculated as the sum of intraday squared returns at short intervals such as 5 or 15 minutes.[5] For a series that has zero mean and no jumps, the realised volatility converges to the continuous time volatility, known as the ***quadratic variation***. To understand this, we assume for the ease of exposition that the instantaneous returns are generated by the continuous time martingale,

$$dp_t = \sigma_t dW_t \tag{5}$$

where dW_t denotes a standard Wiener process. While asset price p_t can be observed at time t, the volatility σ_t is an unobservable latent variable that scales the stochastic process dW_t continuously through time. From (5) the conditional variance for the one-period returns, $r_{t+1} \equiv p_{t+1} - p_t$, is

$$\sigma_t^2 = \int_t^{t+1} \sigma_s^2 ds$$

which is known as the ***integrated volatility*** over the period t to $t+1$, and p_t is the logarithmic of stock price at time t.

Let m be the sampling frequency within each period t, i.e. there are m continuously compounded returns between t-1 and t

$$\begin{aligned} r_{m,t+1/m} &\equiv p_{t+1/m} - p_t \qquad (6) \\ r_{m,t+2/m} &\equiv p_{t+2/m} - p_{t+1/m} \end{aligned}$$

[5] See Fung and Hsieh (1991) and Andersen and Bollerslev (1998). In the foreign exchange markets, quotes for major exchange rates are available round the clock. In the case of stock markets, close-to-open squared return is used in the volatility aggregation process during market close.

and ***realised volatility***

$$RV_{t+1} = \sum_{j=1,\cdots,m} r^2_{m,t+j/m}.$$

If the discretely sampled returns are serially uncorrelated and the sample path for σ_t is continuous, it follows from the theory of quadratic variation (Karatzas and Shreve (1988)) that

$$\lim_{m\to\infty}\left(\int_t^{t+1}\sigma_s^2 ds - \sum_{j=1,\cdots,m} r^2_{m,t+j/m}\right) = 0.$$

Hence, time t volatility is theoretically observable from the sample path of the return process so long as the sampling process is frequent enough. Characteristics of financial market data suggest that returns measured at an interval shorter than 5 minutes are plagued by spurious serial correlation caused by various market microstructure effects including nonsynchronous trading, discrete price observations, intraday periodic volatility patterns and bid-ask bounce.[6] Bollen and Inder (2002), Ait-Sahalia, Mykland and Zhang (2003) and Bandi and Russell (2004) gave suggestions on how to isolate microstructure noise from realised volatility estimator.

1.3.1 Problem with jumps

When there are jumps, the price process in (5) becomes

$$dp_t = \sigma_t dW_t + J_t dq_t, \tag{7}$$

[6] The bid-ask bounce for example induces negative autocorrelation in tick data and causes the realised volatility estimator to be upwardly biased. Theoretical modelling of this issue so far assumes the price process and the microstructure effect are not correlated which is open to debate since market microstructure theory suggests that trading has an impact on the efficient price. I am grateful to Frank de Jong for explaining this to me at a conference.

where dq_t is a poison process with $dq_t = 1$ corresponding to a jump at time t, and zero otherwise, and κ_t is the jump size at time t when there is a jump. In this case, the quadratic variation for the return process in (7) is then given by

$$\int_t^{t+1} \sigma_s^2 ds + \sum_{t<s\leq t+1}^{2} J_s^2, \tag{8}$$

which is the sum of the integrated volatility and jumps.

In the absence of jumps, the second term on the RHS of (8) disappears, and the ***quadratic variation*** is simply equal to the integrated volatility. In the presence of jumps, the realised volatility continues to converge to the quadratic variation in (8)

$$\lim_{m\to\infty} \left(\int_t^{t+1} \sigma_s^2 ds + \sum_{t<s\leq t+1} J_s^2 - \sum_{j=1}^{m} r_{m,t+j/m}^2 \right) = 0. \tag{9}$$

1.3.2 Bipower variations

If there is a jump between t and $t + 1/m$, then $r_{m,t+1/m}^2$ will be very big. Barndorff-Nielsen and Shephard (2003) show that the standardised *realised bipower variation* measure

$$BV_{m,t+1}^{[a,b]} = \left(\frac{1}{m}\right)^{1-(a+b)/2} \sum_{j=1}^{m-1} \left|r_{m,t+j/m}\right|^a \left|r_{m,t+(j+1)/m}\right|^b, \quad a,b \geq 0.$$

with $a = b = 1$ converges to integrated volatility

$$\mu_1^{-2} BV_{m,t+1}^{[1,1]} = \mu_1^{-2} \sum_{j=1}^{m-1} \left|r_{m,t+j/m}\right| \left|r_{m,t+(j+1)/m}\right| \to \int_t^{t+1} \sigma_s^2 ds$$

where $\mu_1 = \sqrt{\frac{2}{\pi}}$ provided that jumps are rare and there is no 'leverage' effect.[7] Hence, the realised volatility and the realised bipower variation

[7] This refers to the empirically observed negative relationship between stock price and stock volatility.

can be substituted into (9) to estimate the jump component, J_t. Barndorff-Nielsen and Shephard (2003) suggest imposing a non-negative constraint on κ_t. This is perhaps too rstrictive. For non-negative volatility, $J_t + \mu_1^{-2} BV_t > 0$ will be sufficient.

1.4 Scaling and actual volatility

The forecast of multi-period volatility $\sigma_{t,t+\tau}$ (i.e. for τ period) is taken to be the sum of individual multi-step point forecasts $\sum_{s=1}^{\tau} h_{t+\tau|t}$. These multi-step point forecasts are produced by recursive substitution and using the fact that $\varepsilon_{t+s|t}^2 = h_{t+s|t}$ for $s > 0$ and $\varepsilon_{t+s|t}^2 = \varepsilon_{t+s}^2$ for $s \leq 0$. Since volatility of financial time series has complex structure, Diebold, Hickman, Inoue and Schuermann (1998) warn that forecast estimates will differ depending on the current level of volatility, volatility structure (e.g. the degree of persistence and mean reversion etc.) and the forecast horizon, τ. These will be made clearer in the discussions below.

If returns are *iid* (independent and identically distributed, or strict white noise), then variance of returns over a long horizon can be derived as a simple multiple of single period variance. But, *iid* is clearly not the case for many financial time series because of the volatility stylized facts discussed previously. On the other hand, while a point forecast of $\widehat{\sigma}_{T-1,T\,|\,t-1}$ becomes very noisy as $T = t + \tau \to \infty$, a cumulative forecast, $\widehat{\sigma}_{t,T\,|\,t-1} = \sqrt{\widehat{\sigma}_{t,T\,|\,t-1}^2}$, becomes more accurate because of errors cancellation and volatility mean reversion unless there is a fundamental change in the volatility level or structure.[8]

[8] $\widehat{\sigma}_{t,T\,|\,t-1}$ denotes a volatility forecast formulated at time $t-1$ for volatility over the period from t to T. In pricing options, the required volatility parameter is the expected volatility over the life of the option. The pricing model relies on a riskless hedge to be

Complication in relation to the choice of forecast horizon is partly due to volatility mean reversion. As far as sampling frequency is concerned, Drost and Nijman (1993) prove, theoretically and for a special GARCH(1,1) case, that volatility structure should be preserved through intertemporal aggregation. This means that whether one models volatility at the hourly, daily or monthly intervals, the volatility structure should be the same. But, this is not the case in practice; volatility persistence, which is highly significant in daily data, weakens as the frequency of data decreases.[9]

In general, volatility forecast accuracy improves as data sampling frequency increases relative to forecast horizon (Andersen, Bollerslev and Lange (1999)). However, for forecasting volatility over a long horizon, Figlewski (1997) finds forecast error doubled in size when daily data, instead of monthly data, is used to forecast volatility over 24 months. In some cases, where application is of very long horizon e.g. over 10 years, volatility estimate calculated using low frequency data such as weekly or monthly is better because volatility mean reversion is difficult to adjust using high frequency data. In general, model based forecasts lose supremacy when the forecast horizon increases with respect to the data frequency. For forecast horizons that are longer than 6 months, a simple historical model using low frequency

followed through until the option reaches maturity. Therefore the required volatility input, or the implied volatility derived, is a cumulative volatility forecast over the option maturity and not a point forecast of volatility at option maturity. The interest in forecasting $\sigma_{t,T\,|\,t-1}$ goes beyond the riskless hedge argument however.

[9] See Diebold (1988), Baillie and Bollerslev (1989) and Poon and Taylor (1992) for examples. Note that Nelson (1992) points out separately that as the sampling frequency becomes shorter, volatility modelled using discrete time model approaches its diffusion limit and persistence is to be expected provided that the underlying returns is a diffusion or a near diffusion process with no jumps.

data over a period at least as long as the forecast horizon works best (Alford and Boatsman (1995) and Figlewski (1997)).

2 The treatment of large numbers

To a statistician, there are always two 'extremes' in each sample, namely the minimum and the maximum. Here, a large number refers generally to extreme values, outliers and jumps, a group of observations that do not belong to the same distribution as the majority of the observations in data sample. These large numbers have undue influence on modelling and estimation (Huber, 1986). Unless extreme value techniques are used, where scale and marginal distribution are removed, it is advisable that outliers are removed or trimmed before modelling volatility. One such outlier in stock market returns is the October 1987 crash that produced a one-day loss of over 20% in stock markets worldwide.

The ways that outliers were tackled in the literature very much depend on their sizes, the frequency of their occurence and if these outliers produce an additive or a multiplicative impact. For the rare and additive outliers, the most common treatment is simply by removing them from the sample or omit them in the likelihood calculation (Kearns and Pagan, 1993). Franses and Ghijsels (1999) find forecasting performance of the GARCH model is substantially improved in 4 out of 5 stock markets studied when the additive outliers are removed. For the rare multiplicative outliers that produced a residual impact on volatility, a dummy variable could be included in the conditional volatility equation after the outlier returns has been dummied out in the mean equation (Blair, Poon and Taylor, 2001).

$$
\begin{aligned}
r_t &= \mu + \psi_1 D_t + \varepsilon_t, \qquad \varepsilon_t = \sqrt{h_t} z_t \\
h_t &= \omega + \beta h_{t-1} + \alpha \varepsilon_{t-1}^2 + \psi_2 D_{t-1}
\end{aligned}
$$

where D_t is 1 when t refers to 19 October 1987 and 0 otherwise. Personally, I find simple method like trimming rule below

$$
Tr = Min\left[r, r_{0.99}\right], \text{ or } Max\left[r, r_{0.01}\right]. \tag{10}
$$

(10) very easy to implement and effective. Such a jump removal procedure will underestimate volatility if jump is expected to occur again. But jump, by nature, is hard to predict. Until we have a better understand of and model for stock market jumps, the recommendation is to remove them in the estimation and modelling of volatility.

The removal of outliers does not remove volatility persistence. In fact, the evidence in the previous section shows trimming the data using (10) actually increases the 'long memory' in volatility making it appears to be extremely persistence. Since autocorrelation is defined as

$$
\rho\left(r_t, r_{t-\tau}\right) = \frac{Cov\left(r_t, r_{t-\tau}\right)}{Var\left(r_t\right)},
$$

the removal of outliers has a great impact on the denominator, reduces $Var\left(r_t\right)$ and increases the individual and the cumulative autocorrelation coefficients.