1 Volatility Definition and Estimation

1.1 WHAT IS VOLATILITY?

It is useful to start with an explanation of what volatility is, at least for the purpose of clarifying the scope of this book. Volatility refers to the spread of all likely outcomes of an uncertain variable. Typically, in financial markets, we are often concerned with the spread of asset returns. Statistically, volatility is often measured as the sample standard deviation

$$\widehat{\sigma} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(r_t - \mu)^2}, \tag{1.1}$$

where r_t is the return on day t, and μ is the average return over the T-day period.

Sometimes, variance, σ^2, is used also as a volatility measure. Since variance is simply the square of standard deviation, it makes no difference whichever measure we use when we compare the volatility of two assets. However, variance is much less stable and less desirable than standard deviation as an object for computer estimation and volatility forecast evaluation. Moreover standard deviation has the same unit of measure as the mean, i.e. if the mean is in dollar, then standard deviation is also expressed in dollar whereas variance will be expressed in dollar square. For this reason, standard deviation is more convenient and intuitive when we think about volatility.

Volatility is related to, but not exactly the same as, risk. Risk is associated with undesirable outcome, whereas volatility as a measure strictly for uncertainty could be due to a positive outcome. This important difference is often overlooked. Take the Sharpe ratio for example. The Sharpe ratio is used for measuring the performance of an investment by comparing the mean return in relation to its 'risk' proxy by its volatility.

The Sharpe ratio is defined as

$$\text{Sharpe ratio} = \frac{\begin{pmatrix}\text{Average}\\ \text{return}, \mu\end{pmatrix} - \begin{pmatrix}\text{Risk-free interest}\\ \text{rate, e.g. T-bill rate}\end{pmatrix}}{\text{Standard deviation of returns}, \sigma}.$$

The notion is that a larger Sharpe ratio is preferred to a smaller one. An unusually large positive return, which is a desirable outcome, could lead to a reduction in the Sharpe ratio because it will have a greater impact on the standard deviation, σ, in the denominator than the average return, μ, in the numerator.

More importantly, the reason that volatility is not a good or perfect measure for risk is because volatility (or standard deviation) is only a measure for the spread of a distribution and has no information on its shape. The only exception is the case of a normal distribution or a lognormal distribution where the mean, μ, and the standard deviation, σ, are sufficient statistics for the entire distribution, i.e. with μ and σ alone, one is able to reproduce the empirical distribution.

This book is about volatility only. Although volatility is not the sole determinant of asset return distribution, it is a key input to many important finance applications such as investment, portfolio construction, option pricing, hedging, and risk management. When Clive Granger and I completed our survey paper on volatility forecasting research, there were 93 studies on our list plus several hundred non-forecasting papers written on volatility modelling. At the time of writing this book, the number of volatility studies is still rising and there are now about 120 volatility forecasting papers on the list. Financial market volatility is a 'live' subject and has many facets driven by political events, macroeconomy and investors' behaviour. This book will elaborate some of these complexities that kept the whole industry of volatility modelling and forecasting going in the last three decades. A new trend now emerging is on the trading and hedging of volatility. The Chicago Board of Exchange (CBOE) for example has started futures trading on a volatility index. Options on such futures contracts are likely to follow. Volatility swap contracts have been traded on the over-the-counter market well before the CBOE's developments. Previously volatility was an input to a model for pricing an asset or option written on the asset. It is now the principal subject of the model and valuation. One can only predict that volatility research will intensify for at least the next decade.

1.2 FINANCIAL MARKET STYLIZED FACTS

To give a brief appreciation of the amount of variation across different financial assets, Figure 1.1 plots the returns distributions of a normally

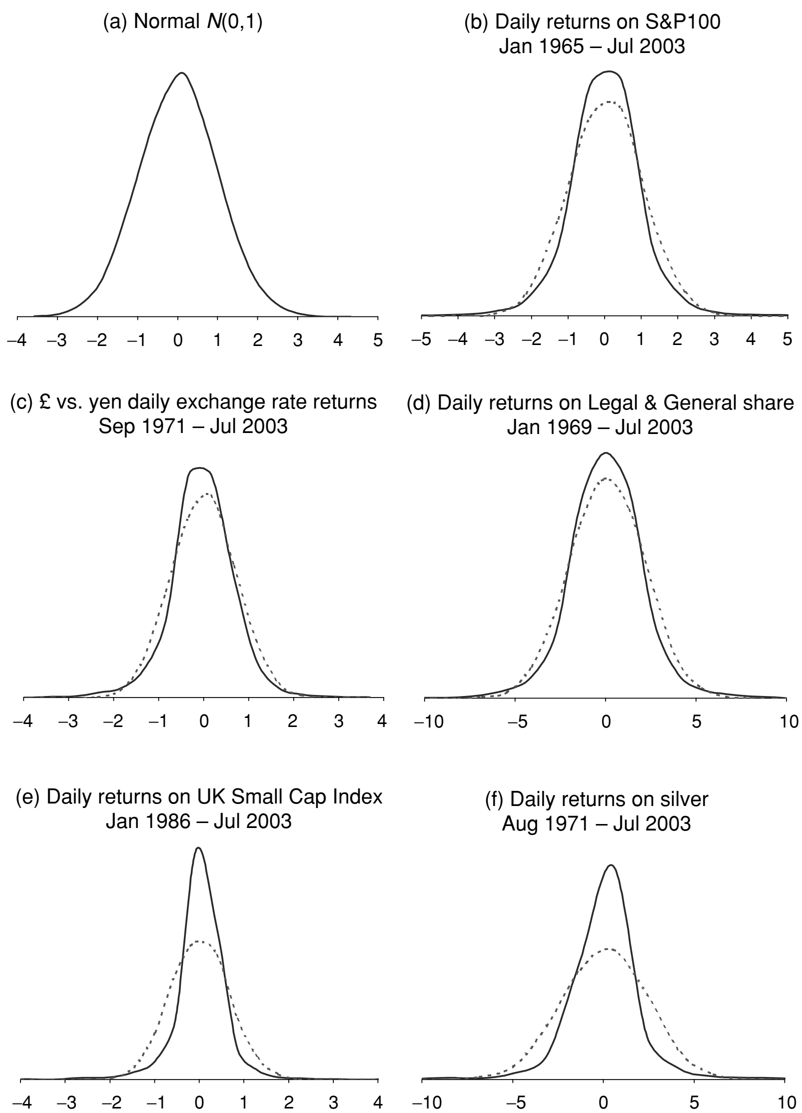

Figure 1.1 Distribution of daily financial market returns. (*Note*: the dotted line is the distribution of a normal random variable simulated using the mean and standard deviation of the financial asset returns)

distributed random variable, and the respective daily returns on the US Standard and Poor market index (S&P100),[1] the yen–sterling exchange rate, the share of Legal & General (a major insurance company in the UK), the UK Index for Small Capitalisation Stocks (i.e. small companies), and silver traded at the commodity exchange. The normal distribution simulated using the mean and standard deviation of the financial asset returns is drawn on the same graph to facilitate comparison.

From the small selection of financial asset returns presented in Figure 1.1, we notice several well-known features. Although the asset returns have different degrees of variation, most of them have long 'tails' as compared with the normally distributed random variable. Typically, the asset distribution and the normal distribution cross at least three times, leaving the financial asset returns with a longer left tail and a higher peak in the middle. The implications are that, for a large part of the time, financial asset returns fluctuate in a range smaller than a normal distribution. But there are some occasions where financial asset returns swing in a much wider scale than that permitted by a normal distribution. This phenomenon is most acute in the case of UK Small Cap and silver. Table 1.1 provides some summary statistics for these financial time series.

The normally distributed variable has a skewness equal to zero and a kurtosis of 3. The annualized standard deviation is simply $\sqrt{252}\sigma$, assuming that there are 252 trading days in a year. The financial asset returns are not adjusted for dividend. This omission is not likely to have any impact on the summary statistics because the amount of dividends distributed over the year is very small compared to the daily fluctuations of asset prices. From Table 1.1, the Small Cap Index is the most negatively skewed, meaning that it has a longer left tail (extreme losses) than right tail (extreme gains). Kurtosis is a measure for tail thickness and it is astronomical for S&P100, Small Cap Index and silver. However, these skewness and kurtosis statistics are very sensitive to outliers. The skewness statistic is much closer to zero, and the amount of kurtosis dropped by 60% to 80%, when the October 1987 crash and a small number of outliers are excluded.

Another characteristic of financial market volatility is the time-varying nature of returns fluctuations, the discovery of which led to Rob Engle's Nobel Prize for his achievement in modelling it. Figure 1.2 plots the time series history of returns of the same set of assets presented

[1] The data for S&P100 prior to 1986 comes from S&P500. Adjustments were made when the two series were grafted together.

Table 1.1 Summary statistics for a selection of financial series

	$N(0, 1)$	S&P100	Yen/£ rate	Legal & General	UK Small Cap	Silver
Start date		Jan 65	Sep 71	Jan 69	Jan 86	Aug 71
Number of observations	8000	9675	7338	7684	4432	7771
Daily average[a]	0	0.024	−0.021	0.043	0.022	0.014
Daily Standard Deviation	1	0.985	0.715	2.061	0.648	2.347
Annualized average	0	6.067	−5.188	10.727	5.461	3.543
Annualized Standard Deviation	15.875	15.632	11.356	32.715	10.286	37.255
Skewness	0	−1.337	−0.523	0.026	−3.099	0.387
Kurtosis	3	37.140	7.664	6.386	42.561	45.503
Number of outliers removed		1			5	9
Skewness[b]		−0.055			−0.917	−0.088
Kurtosis[b]		7.989			13.972	15.369

[a] Returns not adjusted for dividends.
[b] These two statistical measures are computed after the removal of outliers.
All series have an end date of 22 July, 2003.

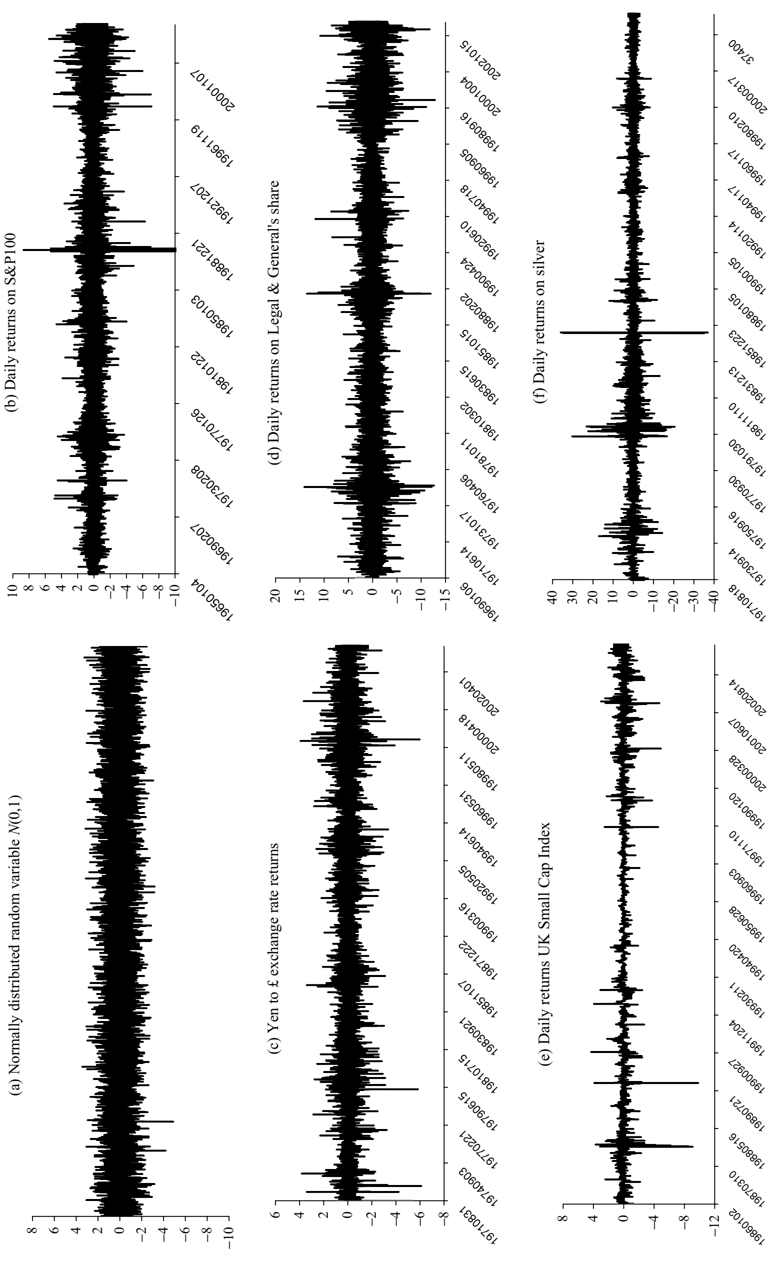

Figure 1.2 Time series of daily returns on a simulated random variable and a collection of financial assets

in Figure 1.1. The amplitude of the returns fluctuations represents the amount of variation with respect to a short instance in time. It is clear from Figures 1.2(b) to (f) that fluctuations of financial asset returns are 'lumpier' in contrast to the even variations of the normally distributed variable in Figure 1.2(a). In the finance literature, this 'lumpiness' is called volatility clustering. With volatility clustering, a turbulent trading day tends to be followed by another turbulent day, while a tranquil period tends to be followed by another tranquil period. Rob Engle (1982) is the first to use the ARCH (autoregressive conditional heteroscedasticity) model to capture this type of volatility persistence; 'autoregressive' because high/low volatility tends to persist, 'conditional' means time-varying or with respect to a point in time, and 'heteroscedasticity' is a technical jargon for non-constant volatility.[2]

There are several salient features about financial market returns and volatility that are now well documented. These include fat tails and volatility clustering that we mentioned above. Other characteristics documented in the literature include:

(i) Asset returns, r_t, are not autocorrelated except possibly at lag one due to nonsynchronous or thin trading. The lack of autocorrelation pattern in returns corresponds to the notion of *weak form market efficiency* in the sense that returns are not predictable.
(ii) The autocorrelation function of $|r_t|$ and r_t^2 decays slowly and $corr\left(|r_t|, |r_{t-1}|\right) > corr\left(r_t^2, r_{t-1}^2\right)$. The decay rate of the autocorrelation function is much slower than the exponential rate of a stationary AR or ARMA model. The autocorrelations remain positive for very long lags. This is known as the long memory effect of volatility which will be discussed in greater detail in Chapter 5. In the table below, we give a brief taste of the finding:

	$\sum \rho(\lvert r \rvert)$	$\sum \rho(r^2)$	$\sum \rho(ln\lvert r \rvert)$	$\sum \rho(\lvert Tr \rvert)$
S&P100	35.687	3.912	27.466	41.930
Yen/£	4.111	1.108	0.966	5.718
L&G	25.898	14.767	29.907	28.711
Small Cap	25.381	3.712	35.152	38.631
Silver	45.504	8.275	88.706	60.545

[2] It is worth noting that the ARCH effect appears in many time series other than financial time series. In fact Engle's (1982) seminal work is illustrated with the UK inflation rate.

(iii) The numbers reported above are the sum of autocorrelations for the first 1000 lags. The last column, $\rho(|Tr|)$, is the autocorrelation of absolute returns after the most extreme 1% tail observations were truncated. Let $r_{0.01}$ and $r_{0.99}$ be the 98% confidence interval of the empirical distribution,

$$Tr = Min\left[r, r_{0.99}\right], \text{ or } Max\left[r, r_{0.01}\right]. \tag{1.2}$$

The effect of such an outlier truncation is discussed in Huber (1981). The results reported in the table show that suppressing the large numbers markedly increases the long memory effect.

(iv) Autocorrelation of powers of an absolute return are highest at power one: $corr\left(|r_t|, |r_{t-1}|\right) > corr\left(r_t^d, r_{t-1}^d\right), d \neq 1$. Granger and Ding (1995) call this property the Taylor effect, following Taylor (1986). We showed above that other means of suppressing large numbers could make the memory last longer. The absolute returns $|r_t|$ and squared returns r_t^2 are proxies of daily volatility. By analysing the more accurate volatility estimator, we note that the strongest autocorrelation pattern is observed among realized volatility. Figure 1.3 demonstrates this convincingly.

(v) Volatility asymmetry: it has been observed that volatility increases if the previous day returns are negative. This is known as the leverage effect (Black, 1976; Christie, 1982) because the fall in stock price causes leverage and financial risk of the firm to increase. The phenomenon of volatility asymmetry is most marked during large falls. The leverage effect has not been tested between contemporaneous returns and volatility possibly due to the fact that it is the previous day residuals returns (and its sign dummy) that are included in the conditional volatility specification in many models. With the availability of realized volatility, we find a similar, albeit slightly weaker, relationship in volatility and the sign of contemporaneous returns.

(vi) The returns and volatility of different assets (e.g. different company shares) and different markets (e.g. stock vs. bond markets in one or more regions) tend to move together. More recent research finds correlation among volatility is stronger than that among returns and both tend to increase during bear markets and financial crises.

The art of volatility modelling is to exploit the time series properties and stylized facts of financial market volatility. Some financial time series have their unique characteristics. The Korean stock market, for

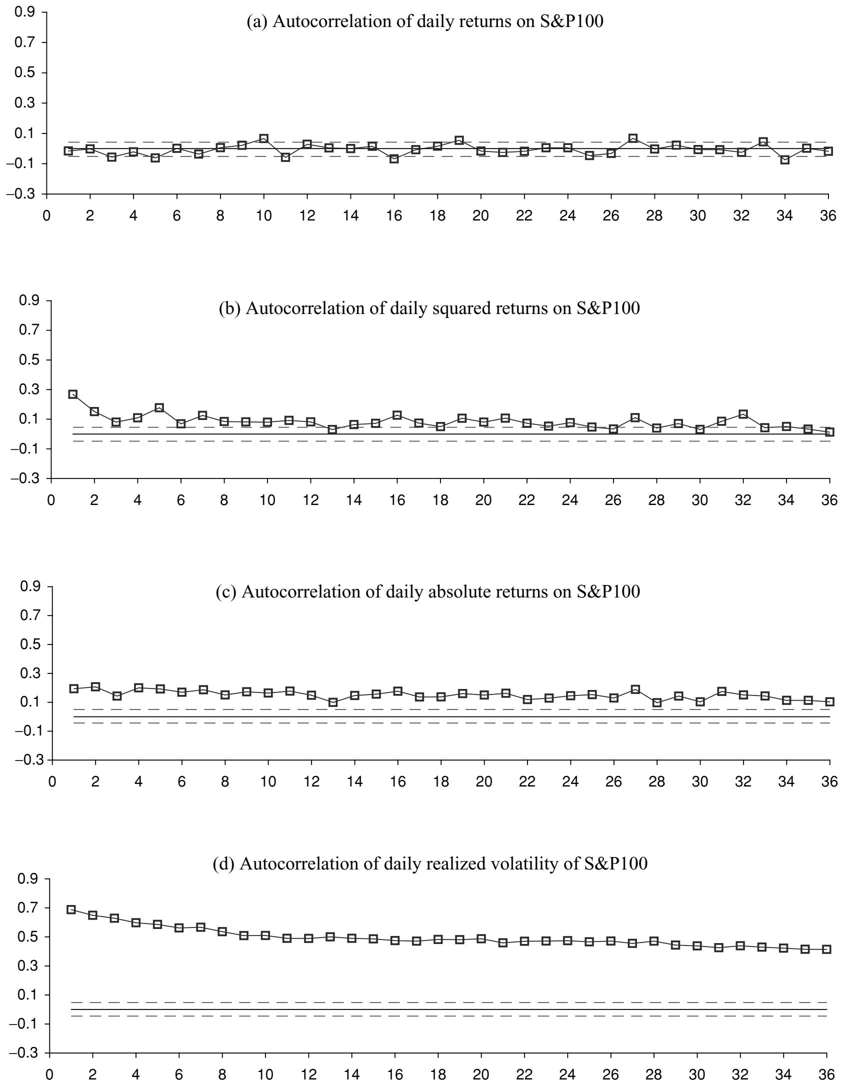

Figure 1.3 Aurocorrelation of daily returns and proxies of daily volatility of S&P100. (*Note*: dotted lines represent two standard errors)

example, clearly went through a regime shift with a much higher volatility level after 1998. Many of the Asian markets have behaved differently since the Asian crisis in 1997. The difficulty and sophistication of volatility modelling lie in the controlling of these special and unique features of each individual financial time series.

1.3 VOLATILITY ESTIMATION

Consider a time series of returns r_t, $t = 1, \cdots, T$, the standard deviation, σ, in (1.1) is the *unconditional volatility* over the T period. Since volatility does not remain constant through time, the *conditional volatility*, $\sigma_{t,\tau}$ is a more relevant information for asset pricing and risk management at time t. Volatility estimation procedure varies a great deal depending on how much information we have at each sub-interval t, and the length of τ, the volatility reference period. Many financial time series are available at the daily interval, while τ could vary from 1 to 10 days (for risk management), months (for option pricing) and years (for investment analysis). Recently, intraday transaction data has become more widely available providing a channel for more accurate volatility estimation and forecast. This is the area where much research effort has been concentrated in the last two years.

When monthly volatility is required and daily data is available, volatility can simply be calculated using Equation (1.1). Many macroeconomic series are available only at the monthly interval, so the current practice is to use absolute monthly value to proxy for macro volatility. The same applies to financial time series when a daily volatility estimate is required and only daily data is available. The use of absolute value to proxy for volatility is the equivalent of forcing $T = 1$ and $\mu = 0$ in Equation (1.1). Figlewski (1997) noted that the statistical properties of the sample mean make it a very inaccurate estimate of the true mean especially for small samples. Taking deviations around zero instead of the sample mean as in Equation (1.1) typically increases volatility forecast accuracy.

The use of daily return to proxy daily volatility will produce a very noisy volatility estimator. Section 1.3.1 explains this in a greater detail. Engle (1982) was the first to propose the use of an ARCH (autoregressive conditional heteroscedasticity) model below to produce conditional volatility for inflation rate r_t;

$$\begin{aligned} r_t &= \mu + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \sqrt{h_t}\right). \\ \varepsilon_t &= z_t \sqrt{h_t}, \\ h_t &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots. \end{aligned} \tag{1.3}$$

The ARCH model is estimated by maximizing the likelihood of $\{\varepsilon_t\}$. This approach of estimating conditional volatility is less noisy than the absolute return approach but it relies on the assumption that (1.3) is the true return-generating process, ε_t is Gaussian and the time series is long enough for such an estimation.

While Equation (1.1) 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 volatility directly from absolute returns. Davidian and Carroll (1987) show absolute returns volatility specification is more robust against asymmetry and nonnormality. There is some empirical evidence that deviations or absolute returns based models produce better volatility forecasts than models that are based on squared returns (Taylor, 1986; Ederington and Guan, 2000a; McKenzie, 1999). However, 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 a Gaussian variable or a particular distribution function for returns.

1.3.1 Using squared return as a proxy for daily volatility

Volatility is a latent variable. Before high-frequency data became widely available, many researchers have resorted to using daily squared returns, calculated from market daily 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

$$Y_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \tag{1.4}$$

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

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

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!

---

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

Under the null hypothesis that returns in (1.4) are 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 and κ is finite. For conditional Gaussian error, the R^2 from a correctly specified GARCH(1,1) model cannot be greater than 1/3. For thick tail distribution, the upper bound for R^2 is lower than 1/3. Christodoulakis and Satchell (1998) extend the results to include compound normals and the Gram–Charlier class of distributions confirming that the mis-estimation of forecast performance is likely to be worsened by nonnormality known to be widespread in financial data.

Hence, the use of ε_t^2 as a volatility proxy will lead to low R^2 and undermine the inference on forecast accuracy. Blair, Poon and Taylor (2001) report an increase of R^2 by three to four folds for the 1-day-ahead forecast when intraday 5-minutes squared returns instead of daily squared returns are used to proxy the actual volatility. The R^2 of the regression of $|\varepsilon_t|$ on σ_t^{intra} is 28.5%. Extra caution is needed when interpreting empirical findings in studies that adopt such a noisy volatility estimator. Figure 1.4 shows the time series of these two volatility estimates over the 7-year period from January 1993 to December 1999. Although the overall trends look similar, the two volatility estimates differ in many details.

1.3.2 Using the high–low measure to proxy volatility

The high–low, also known as the range-based or extreme-value, method of estimating volatility is very convenient because daily high, low, opening and closing prices are reported by major newspapers, and the calculation is easy to program using a hand-held calculator. The high–low volatility estimator was studied by Parkinson (1980), Garman and Klass (1980), Beckers (1993), 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 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

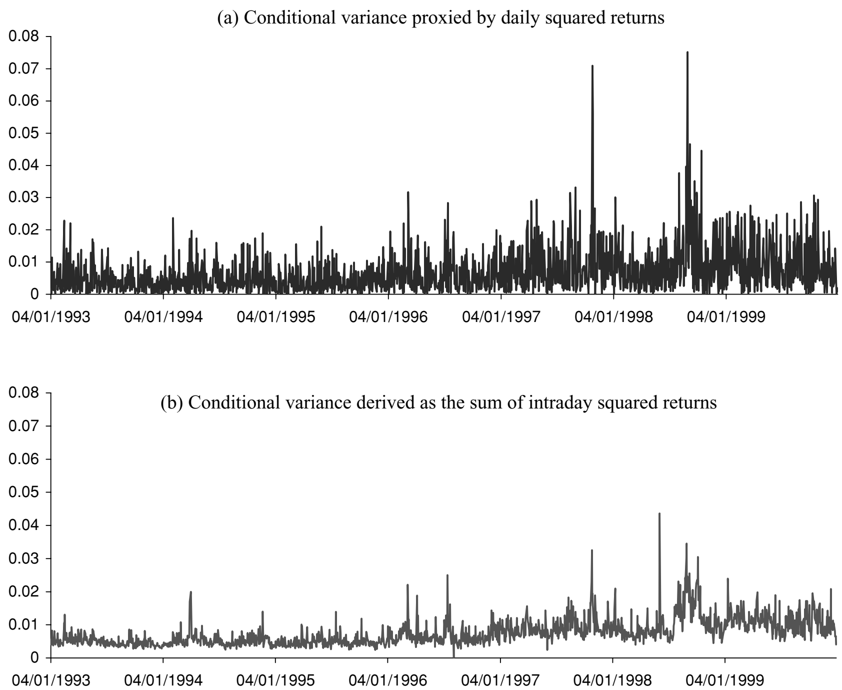

Figure 1.4 S&P100 daily volatility for the period from January 1993 to December 1999

motion results in the following volatility estimator (Bollen and Inder, 2002):

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

The Garman and Klass (1980) estimator is an extension of Parkinson (1980) where information about opening, p_{t-1}, and closing, p_t, prices are incorporated as follows:

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

We have already shown that financial market returns are not likely to be normally distributed and have a long tail distribution. As the H-L volatility estimator is very sensitive to outliers, it will be useful to apply the trimming procedures in Section 1.4. Provided that there are no destabilizing large values, the H-L volatility estimator is very efficient and, unlike the realized volatility estimator introduced in the next section, it is least affected by market microstructure effect.

1.3.3 Realized volatility, quadratic variation and jumps

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

$$dp_t = \sigma_t dW_t, \tag{1.5}$$

where dW_t denotes a standard Wiener process. From (1.5) the conditional variance for the one-period returns, $r_{t+1} \equiv p_{t+1} - p_t$, is $\int_t^{t+1} \sigma_s^2 ds$ which is known as the *integrated volatility* over the period t to $t+1$. Note that 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.

Let m be the sampling frequency such that there are m continuously compounded returns in one unit of time and

$$r_{m,t} \equiv p_t - p_{t-1/m} \tag{1.6}$$

and realized 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

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

Hence time t volatility is theoretically observable from the sample path of the return process so long as the sampling process is frequent enough.

---

[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.

When there are jumps in price process, (1.5) becomes

$$dp_t = \sigma_t dW_t + \kappa_t dq_t,$$

where dq_t is a Poisson 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 cumulative return process is then given by

$$\int_t^{t+1} \sigma_s^2 ds + \sum_{t<s\leq t+1} \kappa^2(s), \tag{1.7}$$

which is the sum of the integrated volatility and jumps.

In the absence of jumps, the second term on the right-hand side of (1.7) disappears, and the quadratic variation is simply equal to the integrated volatility. In the presence of jumps, the realized volatility continues to converge to the quadratic variation in (1.7)

$$p\lim_{m\to\infty} \left(\int_t^{t+1} \sigma_s^2 ds + \sum_{t<s\leq t+1} \kappa^2(s) - \sum_{j=1}^{m} r_{m,t+j/m}^2 \right) = 0. \tag{1.8}$$

Barndorff-Nielsen and Shephard (2003) studied the property of the standardized *realized bipower variation* measure

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

They showed that when jumps are large but rare, the simplest case where $a = b = 1$,

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

where $\mu_1 = \sqrt{2/\pi}$. Hence, the realized volatility and the realized bipower variation can be substituted into (1.8) to estimate the jump component, κ_t. Barndorff-Nielsen and Shephard (2003) suggested imposing a nonnegative constraint on κ_t. This is perhaps too restrictive. For nonnegative volatility, $\kappa_t + \mu_1^{-2} BV_t > 0$ will be sufficient.

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) have given suggestions on how to isolate microstructure noise from realized volatility estimator.

1.3.4 Scaling and actual volatility

The forecast of multi-period volatility $\sigma_{T,T+j}$ (i.e. for j period) is taken to be the sum of individual multi-step point forecasts $\Sigma_{j=1}^{s} h_{T+j|T}$. These multi-step point forecasts are produced by recursive substitution and using the fact that $\varepsilon^2_{T+i|T} = h_{T+i|T}$ for $i > 0$ and $\varepsilon^2_{T+i|T} = \varepsilon^2_{T+i}$ for $T + i \leq 0$. 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.

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, this is clearly not the case for many financial time series because of the stylized facts listed in Section 1.2. While a point forecast of $\widehat{\sigma}_{T-1,T\,|\,t-1}$ becomes very noisy as $T \to \infty$, a cumulative forecast, $\widehat{\sigma}_{t,T\,|\,t-1}$, becomes more accurate because of errors cancellation and volatility mean reversion except when there is a fundamental change in the volatility level or structure.[7]

Complication in relation to the choice of forecast horizon is partly due to volatility mean reversion. 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

[6] The bid–ask bounce for example induces negative autocorrelation in tick data and causes the realized 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.

[7] $\widehat{\sigma}_{t,T\,|\,t-1}$ 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 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.

weekly or monthly data 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 method using low-frequency data over a period at least as long as the forecast horizon works best (Alford and Boatsman, 1995; and Figlewski, 1997).

As far as sampling frequency is concerned, Drost and Nijman (1993) prove, theoretically and for a special case (i.e. the GARCH(1,1) process, which will be introduced in Chapter 4), that volatility structure should be preserved through intertemporal aggregation. This means that whether one models volatility at hourly, daily or monthly intervals, the volatility structure should be the same. But, it is well known that this is not the case in practice; volatility persistence, which is highly significant in daily data, weakens as the frequency of data decreases. [8] This further complicates any attempt to generalize volatility patterns and forecasting results.

1.4 THE TREATMENT OF LARGE NUMBERS

In this section, I use large numbers to refer generally to extreme values, outliers and rare jumps, a group of data that have similar characteristics but do not necessarily belong to the same set. To a statistician, there are always two 'extremes' in each sample, namely the minimum and the maximum. The H-L method for estimating volatility described in the previous section, for example, is also called the extreme value method. We have also noted that these H-L estimators assume conditional distribution is normal. In extreme value statistics, normal distribution is but one of the distributions for the tail. There are many other extreme value distributions that have tails thinner or thicker than the normal distribution's. We have known for a long time now that financial asset returns are not normally distributed. We also know the standardized residuals from ARCH models still display large kurtosis (see McCurdy and Morgan, 1987; Milhoj, 1987; Hsieh, 1989; Baillie and Bollerslev, 1989). Conditional heteroscedasticity alone could not account for all the tail thickness. This is true even when the Student-t distribution is used to construct

[8] 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.

the likelihood function (see Bollerslev, 1987; Hsieh, 1989). Hence, in the literature, the extreme values and the tail observations often refer to those data that lie outside the (conditional) Gaussian region. Given that jumps are large and are modelled as a separate component to the Brownian motion, jumps could potentially be seen as a set similar to those tail observations provided that they are truly rare.

Outliers are by definition unusually large in scale. They are so large that some have argued that they are generated from a completely different process or distribution. The frequency of occurrence should be much smaller for outliers than for jumps or extreme values. Outliers are so huge and rare that it is very unlikely that any modelling effort will be able to capture and predict them. They have, however, undue influence on modelling and estimation (Huber, 1981). Unless extreme value techniques are used where scale and marginal distribution are often 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 1-day loss of over 20% in stock markets worldwide.

The ways that outliers have been tackled in the literature largely depend on their sizes, the frequency of their occurrence and whether these outliers have an additive or a multiplicative impact. For the rare and additive outliers, the most common treatment is simply to remove 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 four out of five 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 a simple method such as the trimming rule in (1.2) very quick to implement and effective.

The removal of outliers does not remove volatility persistence. In fact, the evidence in the previous section shows that trimming the data using (1.2) actually increases the 'long memory' in volatility making it appear to be extremely persistent. 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.

Once the impact of outliers is removed, there are different views about how the extremes and jumps should be handled vis-à-vis the rest of the data. There are two schools of thought, each proposing a seemingly different model, and both can explain the long memory in volatility. The first believes structural breaks in volatility cause mean level of volatility to shift up and down. There is no restriction on the frequency or the size of the breaks. The second advocates the regime-switching model where volatility switches between high and low volatility states. The means of the two states are fixed, but there is no restriction on the timing of the switch, the duration of each regime and the probability of switching. Sometimes a three-regime switching is adopted but, as the number of regimes increases, the estimation and modelling become more complex. Technically speaking, if there are infinite numbers of regimes then there is no difference between the two models. The regime-switching model and the structural break model will be described in Chapter 5.