Risk Management and Analysis: *Measuring and Modelling Financial Risk*
(C. Alexander, Ed) Wileys, (1998)

CHAPTER 4

Volatility and Correlation : Measurement, Models and Applications

Carol Alexander, University of Sussex UK

4.1 Introduction

The most widely accepted approach to 'risk' in financial markets focuses on the measurement of volatility in certain returns distributions.[1] The volatility of portfolio returns depends on the variances and covariances between the risk factors of the portfolio, and the sensitivities of individual assets to these risk factors. In linear portfolios, sensitivities are also measured by variances and covariances. So the primary objective of this chapter is to account for the standard methods of estimating and forecasting covariance matrices. The practical and theoretical advantages and limitations of each method is discussed in detail. Later it is shown how to employ them effectively in value at risk systems, capital allocation models, investment management and the pricing and hedging of derivative products. The chapter concludes with a short review of other methods for measuring market risk: downside risk and regret, expected maximum loss, and cointegration.

What is volatility? A standard non-parametric definition is that x is 'more volatile' than y if

$$P(|x|>c) > P(|y|>c)$$

for all c. So if y is a time series it becomes more volatile when $P(|y_{t+1}|>c) > P(|y_t|>c)$ for all c. But when y has a symmetric distribution, such as a normal distribution, this occurs if and only if σ

[1] Returns are given by the relative change in prices (or rates) over a fixed holding period h. The return at time t is given by $(P_t - P_{t-h})/P_{t-h}$ which, for small h, is well approximated by $\ln P_t - \ln P_{t-h}$. Value at Risk is, however, based on the volatility of a portfolio's change in market price (the profit and loss, P&L). Both returns and P&L are generally stationary (mean reverting) variables. But some confusion may be caused by the term 'price volatility', since prices (or rates) are random walks in efficient markets, and hence have infinite (unconditional) variance.

$^{2}{}_{t+1} > \sigma^2_t$ where σ^2_t denotes the conditional variance of y.[2] However, variance is not standardized. If we were to plot a term structure of variances, across different maturities, variance would simply increase with the holding period of returns. So volatility is usually quoted as an annualised percentage standard deviation:

$$\text{volatility at time } t = (100\, \sigma_t \sqrt{A})\ \% \qquad (1)$$

where A denotes the number of observations per year. We do this so that volatilities of different maturities may be compared on the same scale - as the variance increases with the holding period, so the annualising factor decreases. In this way, volatilities are standardized.

If one takes two associated returns series x and y - returns on two Gilts for example - we can calculate their correlation as

$$\text{CORR}(x,y) = \text{COV}(x,y) / \sqrt{V(x)}\, \sqrt{V(y)} \qquad (2)$$

that is,

$$\rho_{xy} = \sigma_{xy} / \sigma_x\ \sigma_y \qquad (3)$$

Note that joint stationarity[3] is necessary for the existence of correlation, and that it is the exception rather than the rule. Two arbitrary returns series, such as a Latin American Brady bond and a stock in the FTSE 100, should be unrelated and are not likely to be jointly stationary, so correlations between these two time series do not exist. Of course, one can calculate a number based on the correlation formulae given in this chapter, but it does not measure unconditional correlation unless the two series are jointly stationary. In such cases 'correlation' is found to be very unstable. Correlation estimates will jump about a lot over time - a sign of non-joint stationarity.

---

[2] We use the notations V(.) and σ^2 interchangeably, similarly COV(x,y) and σ_{xy} . The standard deviation is the square root of the variance. Conditional and unconditional variance are explained in section 4.3.1.

[3] Loosely speaking, this means that not only are the two individual returns series stationary (mean reverting) , but their joint distribution has stationarity properties such as constant autocorrelations.

Correlation does not need to be annualised, as does volatility, because it is already in a standardised form: correlation always lies between -1 and +1. A negative value means that returns tend to move in opposite directions and a positive value indicates synchronous moves in the same direction. The greater the absolute value of correlation, the greater the association between the series. [4]

Volatilities and correlations are just standardized forms of the variances and covariances between returns, so the information necessary to measure portfolio risk is usually summarized in a covariance matrix. Based on a set of n returns series y_1, ….. y_n , this is a square, symmetric matrix of the form

$$\begin{pmatrix} V(y_1) & COV(y_1,y_2) & \ldots & \ldots & COV(y_1,y_n) \\ COV(y_1,y_2) & V(y_2) & \ldots & \ldots & COV(y_2,y_n) \\ COV(y_1,y_3) & COV(y_2,y_3) & V(y_3) & \ldots & COV(y_3,y_n) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ COV(y_1,y_n) & \ldots & \ldots & \ldots & V(y_n) \end{pmatrix}$$

The next two sections of this chapter describe the two most commonly used methods of estimating and forecasting covariance matrices in financial markets. Section 2 assesses the moving average volatility and correlation estimation methods that are most financial institutions use today: the equally weighted 'historic' method and the exponentially weighted moving average method. Section 3 gives an overview of the huge technical literature on GARCH modelling in finance and section 4 covers implied volatility and correlation forecasts, their use in trading being covered elsewhere in this book (chapter 18). Section 5 surveys the use of volatility and correlation forecasts in risk management, with particular emphasis on value at risk (VAR) models. The last section of this chapter covers certain special issues, such as estimating volatilities from ‘fat-tailed’ distributions using normal mixtures, evaluation of the accuracy of different models, and new directions, such as ‘downside risk’ measures and cointegration.

---

[4] It is important to bear in mind that returns series can be perfectly correlated even when the prices are in fact moving in opposite directions. Correlation only measures short-term co-movements in returns, and has little to do with any long-term co-movements in prices. For the common trend analysis in prices, rates or yields the technique of *cointegration* offers many advantages, and this is reviewed in the last section of the chapter.

4.2 Moving Averages

A moving average is an arithmetic average over a rolling window of consecutive data points taken from a time series. Moving averages have been a useful tool in financial forecasting for many years. For example, in technical analysis, where they exist under the name of 'stochastics', the relationship between moving averages of different lengths can be used as a signal to trade. Traditionally they have also been used in volatility estimation. Usually volatility and correlation estimates are based on daily or intra-day returns, since even weekly data can miss some of the turbulence encountered in financial markets. Moving averages of squared (or cross products) of returns are estimates of variance (or covariance). These are converted to volatility and correlation as described above, or employed in a covariance matrix for measuring portfolio variance.

4.2.1 'Historic' Methods

This section describes the uses and misuses of the traditional 'historic' volatility and correlation forecasting methods. Recent advances in time series analysis allow a more critical view of the efficiency of these methods, and they are being replaced by exponentially weighted moving average or GARCH methods in most major institutions today.

The *n-period historic volatility* at time T is the quantity $\left(100\,\hat{\sigma}_T\,\sqrt{A}\right)\%$ where A is the number of returns r_t per year and

$$\hat{\sigma}_T^2 = \sum_{t=T-n}^{t=T-1} r_t^2 \,/\, n \qquad (4)$$

Thus $\hat{\sigma}_T$ is the unbiased estimate of standard deviation over a sample size n, assuming the mean is zero.[5]

[5] It is usual to apply moving averages to squared returns r_t^2 (t = 1,2,3,....n) rather than squared mean deviations of returns $(r_t - \bar{r})^2$ where $\bar{r}$ is the average return over the data window. Although standard statistical estimates of variance are based on mean deviations, empirical research on the accuracy of variance forecasts in financial

'Historic' correlations of maturity n are calculated in an analogous fashion: if x and y are two returns series, then n-period historic correlations may be calculated as[6]

$$r_T = \frac{\sum_{t=T-n}^{t=T-1} x_t y_t}{\sqrt{\sum_{t=T-n}^{t=T-1} x_t^2 \sum_{t=T-n}^{t=T-1} y_t^2}} \qquad (5)$$

Traditionally, the estimate of volatility or correlation over the last n periods has been used as a forecast over the next n periods. The rationale for this is that long-term volatility predictions should be unaffected by 'volatility clustering' behaviour, and so we need to take an average squared return over a long historic period - but short-term volatility predictions should reflect current market conditions, whether volatile or tranquil, which means that only the immediate past returns should be used.

However, when we examine the time-series properties of 'historic' volatilities and correlations we see that they have some undesirable qualities: A major problem with equally weighted averages is that extreme events are just as important to current estimates whether they occurred yesterday or a long time ago. Even just *one* unusual return will continue to keep volatility estimates high for exactly n days following that day, although the underlying volatility will have long ago returned to normal levels. Thus volatility estimates will be kept artificially high in periods of tranquillity, and they will be lower than they should be during the short bursts of volatility which characterise financial markets.[7]

---

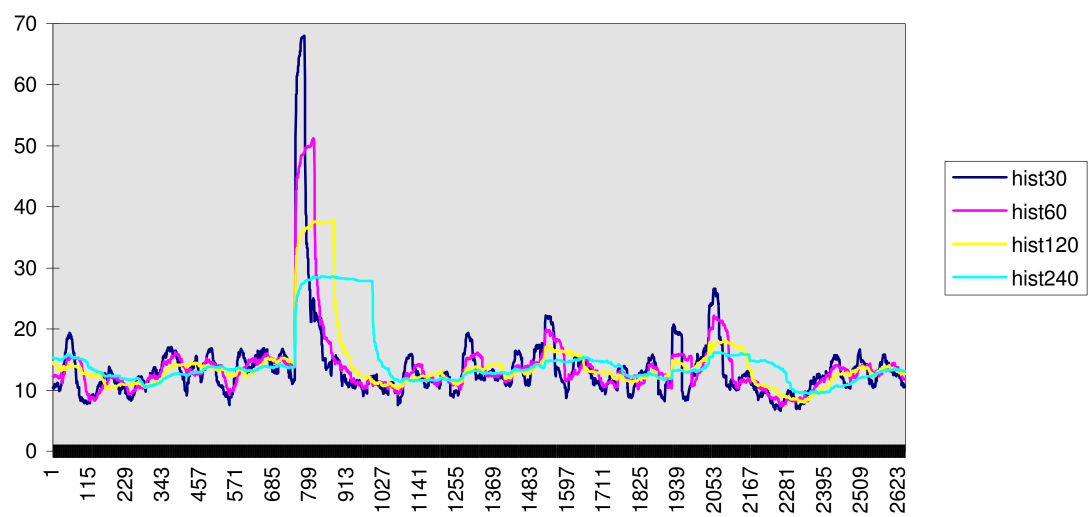
markets has shown that it is often better not to use mean deviations of returns, but to base variances on squared returns and covariances on cross products of returns (Figlewski, 1994, Alexander and Leigh, 1997).

[6] Again, assuming zero means is simpler, and there is no convincing empirical evidence that this degrades the quality of correlation estimates and forecasts in financial time series

[7] For this reason I suggest removing extreme events from the returns data before the moving average is calculated. This will give a better 'everyday' volatility estimate, for example, in VAR models. For stress testing portfolios, these extreme events can be put back into the series - indeed they could even be bootstrapped back into the series for more creative stress testing.

Figure 1 illustrates equally weighted averages of different lengths on squared returns to the FTSE. Daily squared returns are averaged over the last n observations for n= 30, 60, 120, 240, and this variance is transformed to an annualized volatility in figure 1. Note that the one-year volatility of the FTSE jumped up to 26% the day after Black Monday and it stayed at that level for a whole year because that one, huge squared return had exactly the same weight in the average. Exactly one year after the event the large return falls outs of the moving average, and so the volatility forecast returned to its normal level of around 13%. In shorter term equally weighted averages this 'ghost feature' will be much bigger because it will be averaged over fewer observations, but it will last for a shorter period of time.

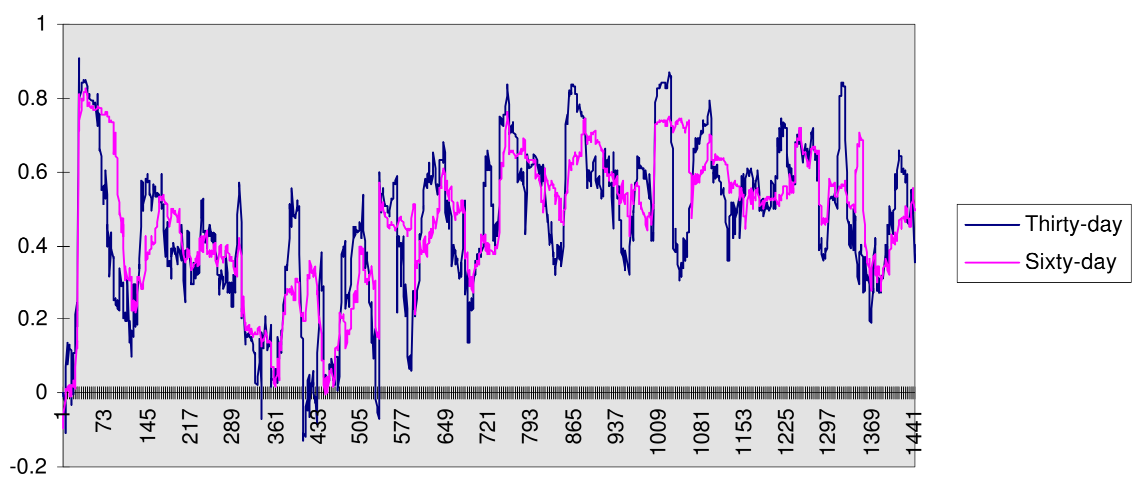
Figure 1: Historic Volatilities of the FTSE from 1984 to 1995, showing 'ghost features' of Black Monday and other extreme events



Ghost features are even more of an issue with equally weighted moving average correlation estimates, where they can induce an apparent stability in correlations. It may be that 'instantaneous' correlations are very unstable, because the two returns are not jointly stationary. But whatever the true properties of correlations between the two returns series, the longer the averaging period, the more stable will moving average correlations appear to be. It may also be that, by some fluke, the series both have large returns on the same day. This will cause a ghost feature in correlation, making it artificially high for the n periods following that day.

Even with two closely related series such as the FTSE 100 and the S&P 500, we will obtain correlation estimates which appear more stable as the averaging period increases. This point is illustrated in figure 2, where correlations between the two equity indices are shown for a 60-day averaging period, compared with the 30-day period. Stability of correlation estimates increases with the averaging period, rather than being linked to the degree of joint stationarity between the variables, because the equally weighted moving average method is masking the underlying nature of the relationship between variables. Thus we could erroneous conclude that correlations are 'stable' if this method of estimation is employed.

Figure 2: Historical correlations between UK and US equities from just after Black Monday to Spring 1993



Equally weighted moving averages should be used with caution, particularly for correlation, on the grounds of these ghost features alone. But there is another, finer point of consideration. Perceived changes in volatility and correlation can have important consequences, so it is essential to understand what is the source of variability in any particular model. In the 'historic' model, all variation is due only to differences in samples: a smaller sample size yields a less precise estimate, the larger the sample size the more accurate the estimate. So a short period moving average will be more variable than a longer moving average. But whatever the length of the averaging period we are still estimating the same thing: the unconditional volatility of the time series. This is one number, a constant, underlying the whole series. So variation in the n-period historic volatility model, which we perceive as variation over time, is actually due to sampling error alone. There is nothing else in the model that allows for variation. There is no estimated stochastic volatility model in any moving average method - they are simply estimates of unconditional moments, which are constants.[8] The estimated series does change over time, but as the underlying parameter of interest is a *constant* variance, all the observed variation in the estimate is simply due to sampling variation. The 'historic' model is also taking no account of the dynamic properties of returns, such as autocorrelation. It is essentially a 'static' model which has been forced into a time-varying framework. So, if you 'shuffle' the data within any given n-period window, you will get the same answer, provided of course for correlation the two returns series are shuffled 'in pairs'.

4.2.2 Exponentially Weighted Moving Averages

The 'historic' models explained above weight each observation equally, whether it is yesterdays return or the returns from several weeks or months ago. It is this equal weighting that induces the 'ghost features', which are clearly a problem. An exponentially weighted moving average (EWMA) places more weight on more recent observations, and this has the effect of eliminating the problematic 'ghost features'.

The exponential weighting is done by using a ‘smoothing constant’ λ: the larger the value of λ the more weight is placed on past observations and so the smoother the series becomes. An n-period EWMA of a time series x is defined as

---

[8] This is a limitation of moving average methods. When the estimation method also gives a model for stochastic volatility (as it does, for example, in GARCH models) the stochastic volatility model has very useful applications for pricing and hedging (see sections 4.5.2 and 4.5.3)

$$\frac{x_{t-1} + \lambda x_{t-2} + \lambda^2 x_{t-3} + \ldots.. + \lambda^{n-1} x_{t-n}}{1 + \lambda + \lambda^2 + \ldots\ldots + \lambda^{n-1}} \qquad (3)$$

Since the denominator converges to 1/(1-λ) as n → ∞, an infinite EWMA may be written

$$(1-\lambda) \sum_{i=1}^{\infty} \lambda^{i-1} x_{t-i-1} \qquad (4)$$

It is an EWMA that is used for volatility and correlation forecasts in JP Morgan and Reuter's RiskMetrics™ . The forecasts of volatility and correlation over the next day are calculated by taking λ=0.94 and using squared returns r^2 as the series x in (7) for variance forecasts and cross products of two returns r_1r_2 as the series x in (7) for covariance forecasts. Note that the same value of λ should be used for all variances and covariances in the matrix, otherwise it may not be positive semi definite (see JP Morgan, 1996) .

In general this type of EWMA behaves in a reasonable way - see Alexander and Leigh, 1997 for an evaluation of their accuracy. In fact, an EWMA on squared returns is equivalent to an IGARCH model (see section 4.3.2). To see this, consider equation (7) :

$$\hat{\sigma}_t^2 \quad = \quad (1-\lambda) \sum_{i=1}^{\infty} \lambda^{i-1} r_{t-i-1}^2$$

This may be re-written in the form

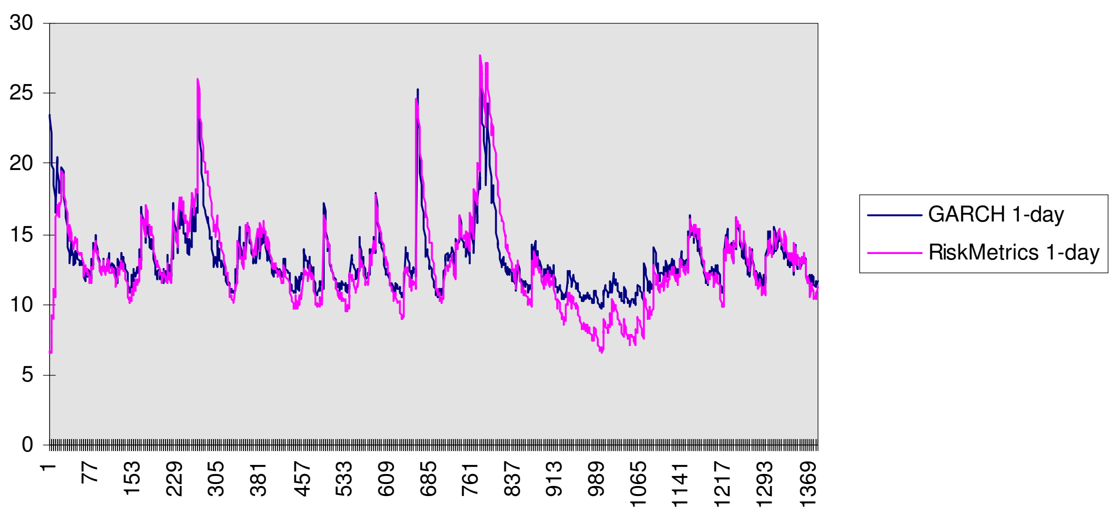
$$\hat{\sigma}_t^2 \quad = \quad (1-\lambda)\, r_{t-1}^2 \quad + \quad \lambda\, \hat{\sigma}_{t-1}^2 \qquad (8)$$

which shows the recursion normally used to calculate EWMAs. Comparison with equation (13) shows that an EWMA is equivalent to an IGARCH model without a constant term.

In section 4.3.3 we describe how the GARCH coefficients can be interpreted: the coefficient on the lagged squared return determines the speed of reaction of volatility to market events, and the coefficient on the lagged variance determines the persistence in volatility. In an EWMA these two coefficients are not independent - they sum to one. For the RiskMetrics data set, the persistence coefficient is 0.94 for all markets, and the reaction coefficient is 0.06 for all markets. But if one estimates these coefficients rather than imposes them, it appears that λ=0.94 is too high for most markets.

For example, in the FTSE series shown in figure 3, the GARCH persistence coefficient is 0.88, and so the GARCH volatility series dies out more quickly than the RiskMetrics. However their reaction coefficient is the same (0.06) and so both volatilities exhibit a similar size of market reaction in figure 3.

Figure 3: GARCH and Exponentially Weighted Moving Average 1-day Volatility Forecasts of the FTSE: 1988-1995



The RiskMetrics™ daily data do have some other problems, which are explained in Alexander, 1996, but these are not insurmountable.[9] However, the RiskMetrics™ forecasts of volatility over the next month behave in a rather strange fashion. Since the EWMA methodology is only really applicable to one-step-ahead forecasting the correct thing would be to smooth 25-day returns, but

---

[9] In fact all the RiskMetrics matrices have low rank - either because EWMA use insufficient data for the size of matrix, or because of the linear interpolation of yield curve data.

there is not enough data. Instead, JP Morgan have applied exponential smoothing with a value of λ=0.97 to the 25-day equally weighted variance. But this series will be full of 25-day 'ghost features' (see figure 1) so the exponential smoothing has the effect of augmenting the very 'ghost features' which they seek to diminish. After a major market movement the equally weighted 25-day series jumps up immediately - as would any sensible volatility forecast. But the RiskMetrics™ monthly data hardly reacts at all, at first, then it gradually increases over the next 25 days to reach a maximum exactly 25 days *after* the event. The proof of this is simple: denote by s_t^2 the 25-day 'historic' variance series, so the monthly variance forecast is $\hat{\sigma}_t^2 = (1-\lambda)\, s_{t-1}^2 + \lambda\, \hat{\sigma}_{t-1}^2$. Clearly $\hat{\sigma}_t^2 > \hat{\sigma}_{t-1}^2 \iff s_{t-1}^2 > \hat{\sigma}_{t-1}^2$. At the 'ghost feature' s_t^2 drops dramatically, and so the maximum value of $\hat{\sigma}_t^2$ will occur at this point.

Figure 4: 25-day 'Historic' and RiskMetrics monthly volatility forecasts of the FTSE: 1988-1995

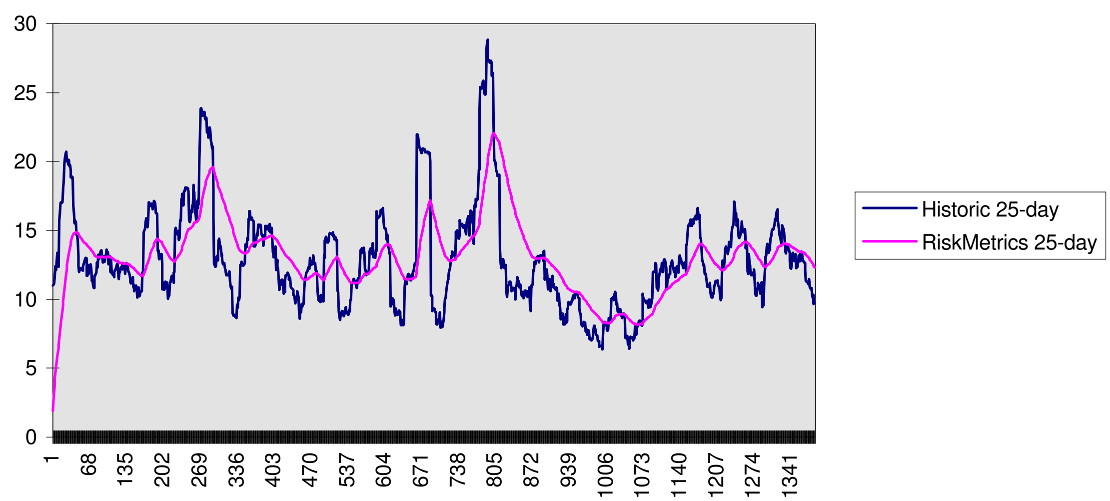


Figure 4 compares the RiskMetrics monthly forecasts for the FTSE with the equally weighted 25-day 'historic' forecasts, during the same period as in figure 3. Neither forecast is useful: the 'historic' forecast peaks at the right time, the day after a significant maret movement, but it stays high for 25-days, when it should probably be declining. The RiskMetrics 25-day forecast hardly reacts at all to a market movement, but it slowly increases during the next 25-days so that the largest forecast of the volatility over the next month occurs 25-days too late.

4.2.3 Volatility Term Structures in Moving Average Models: The Square Root of Time Rule

Term structure volatility forecasts which are consistent with moving average models are constant, because that is the underlying assumption on volatility. Moving average models are, after all, estimation methods, not forecasting methods. We can of course say that our current estimate of volatility is to be taken as all future forecasts, and this is what is generally done. The current one-day volatility estimate is taken to be the one-day forward volatility forecast at every point in the future.

Term structure forecasts - forecasts of the volatility of h-day returns for every maturity h - are then based on the 'square root of time' rule. This rule simply calculates h-day standard deviations as $\sqrt{h}$ times the daily standard deviation. It is based on the assumption that daily log returns are independently and identically distributed, so the variance of h-day returns is just h times the variance of daily returns. But since volatility is just an annualised form of the standard deviation, and since the annualising factor is - assuming 250 days per year - $\sqrt{250}$ for daily returns but $\sqrt{(250/h)}$ for h-day returns, the square root of time rule is equivalent to the Black-Scholes assumption that current levels of volatility remain the same.

The square root of time rule implies the assumption of constant volatility. This follows from the assumption that log returns are independent and identically distributed. To see this, introduce the notation $r_{t,h}$ for an h-day return at time t, and so approximately

$$r_{t,h} = \ln(P_{t+h}) - \ln(P_t)$$

where P_t denotes the price at time t. Clearly

$$r_{t,h} = r_{t,1} + r_{t+1,1} + \ldots\ldots + r_{t+h-1,1}$$

and if we assume that 1-day returns are independent and identically distributed with constant variance σ^2 we have $V(r_{t,h}) = h\sigma^2$. Annualising this into a volatility requires the use of an annualising factor 250/h, being the number of h-periods per year:

$$\text{h-period vol} = 100\ \sqrt{(250/h)}\ \sqrt{V(r_{h,t})} = 100\ \sqrt{(250/h)}\ \sqrt{(h\sigma^2)} = 100\ \sqrt{(250\sigma^2)} = \text{1-period vol}$$

That is, volatility term structures are constant. Constant term structures are a limitation of exponentially weighted moving average methods. The reason for this is that financial volatility tends to come in 'clusters', where tranquil periods of small returns are interspersed with volatile periods of large returns[10]. Thus volatility term structures should mean-revert, with short term volatility lying either above or below the long term mean depending on whether current conditions are high or low volatility. Clearly moving averages are quite limited in this respect. Substantial mispricing of volatility can result from these methods, as a number of large banks have recently discovered.

4.3 GARCH Models in Finance

4.3.1 Introduction

The unfortunate acronym 'GARCH' is nevertheless essential, since it stands for *generalised autoregressive conditional heteroscedasticity*! Heteroscedasticity means 'changing variance', so conditional heteroscedasticity means changing conditional variance. A time series displays conditional heteroscedasticity if it has highly volatile periods interspersed with tranquil periods: i.e. there are 'bursts' or 'clusters' of volatility. Autoregressive means 'regression on itself', and this refers to the method used to model conditional heteroscedasticity in GARCH models.

Most financial time series display autoregressive conditional heteroscedasticity. A typical example of conditionally heteroscedastic returns in high frequency data - minute by minute data on cotton

---

[10] As long ago as 1963 Benoit Mandlebrot observed that financial returns time series exhibit periods of volatility interspersed with tranquillity, where 'Large returns follow large returns, of either sign.....'

futures - is shown in figure 5. Note that two types of news events are apparent. The first volatility cluster shows an anticipated announcement, which turned out to be good news: the market was increasingly turbulent before the announcement, but the large positive return at that time shows that punters were pleased, and the volatility soon decreased. The later cluster of volatility shows increased turbulence following an unanticipated piece of bad news - the large negative return - what is often referred to as the 'leverage' effect (see section 4.3.2 (iv))

NOTE TO COPY EDITOR <Figure 5 here - same as figure 8.4 from 1st edition, and formatted with caption and frame as in the previous figures please>

At the root of understanding GARCH is the distinction between conditional (stochastic) and unconditional (constant) volatility. These ideas are based on different stochastic processes which are assumed to govern the returns data. Figure 6 illustrates the difference between conditional and unconditional distributions of returns. In figure 6(a) the stochastic process which generates the time series data on returns is assumed to be independent and identically distributed. This same distribution governs each of the data points, and since they are independent we may as well redraw the data without taking account of the dynamic ordering, along a line. The data are then considered to be random draws from a single distribution, called the unconditional distribution (figure 6(b)). However, in figure 6(c) the same data are assumed to be generated by a stochastic process with time varying volatility. In this case it is not realistic to collapse the data into a single distribution, ignoring the dynamic ordering. The conditional distribution changes at each point in time, and in particular the volatility process is stochastic.

NOTE TO COPY EDITOR <Figure 6 here please, please have the hand drawn figure attached properly set by printer>.

The first ARCH model, introduced by Rob Engle (1992) was later generalised by Tim Bollerslev (1986), and many variations on the basic 'vanilla' GARCH model have been introduced in the last ten years.[11] The idea of GARCH is to add a second equation to the standard regression model - an equation which models the conditional variance. The first equation in the GARCH model is the conditional mean equation.[12] This can be anything, but because the focus of GARCH is on the conditional variance equation[13] it is usual to have a very simple conditional mean equation, such as r_t = constant + ε_t.

Note that the dependent variable (the input to the GARCH model) is always the returns series, and in the simple case that r_t = constant + ε_t the unexpected return ε_t is just the mean deviation return, because the constant will be the average of returns over the data period. Of course we can put whatever explanatory variables we want in the conditional mean equation of a GARCH model, but should err on the side of parsimony if we want the model estimation procedure to converge properly (see section 4.3.4). The conditional mean equation r_t = constant + ε_t is fairly standard. A GARCH model conditional variance equation provides an easy analytic form for the stochastic volatility process in financial returns. GARCH models differ only because the conditional variance equations are specified in different forms, or because of different assumptions about the conditional distribution of unexpected returns.

In normal GARCH models we assume that ε_t is conditionally normally distributed with conditional variance σ_t^2 . The unconditional returns distributions will then be *leptokurtic* - that is, have fatter tails than the normal - because the changing conditional variance allows for more outliers or unusually large observations. However in high frequency data there may still be insufficient leptokurtosis in normal GARCH to capture the full extent of kurtosis in the data, and

---

[11] For excellent reviews of the enormous literature on GARCH models in finance see Bollerslev et.al. (1992) and (1993).

[12] The unconditional mean of a stationary time series y is a single number, a constant. It is denoted E(y) or μ, and is usually estimated very simply by the sample mean. On the other hand, the conditional mean varies over time, and is commonly measured by a linear regression model. The conditional mean is denoted $E_t(y_t)$ or μ_t or $E(y_t | \Omega_t)$, where Ω_t is the *information set* available at time t (so Ω_t includes y_{t-1} and any other values which are known at time t).

[13] The unconditional variance of a stationary time series y is a constant, denoted V(y) or σ^2. It is often estimated by a moving average, as explained in the previous section. On the other hand, the conditional variance σ_t^2 is often denoted $V_t(y_t)$ or $V_t(y_t | \Omega_t)$. Like the conditional mean, its' estimates will form a time-varying series.

in this case a t-distribution could be assumed (Baillie and Bollerslev, 1989, 1990), or a GARCH model defined on a mixture of normals (see section 4.6).

Square rooting the GARCH conditional variance series, and expressing it as an annualised percentage in the usual way yields a time-varying volatility estimate. But unlike the moving average methods just described, the current estimate is not taken to be the forecast of volatility over all future time horizons. Instead, by first estimating the GARCH model parameters, we can then construct mean-reverting forecasts of volatility as explained in section 4.3.3. GARCH is sufficiently flexible that these forecasts can be adapted to any time period. For example, when valuing Asian options, volatility options, or measuring risk capital requirements it is often necessary to forecast forward volatility, such as a 1-month volatility but in 6 months time. This flexibility is one of the many advantages of GARCH modelling over the moving average methods just described.

Very many different types of GARCH models have been proposed in the academic literature, but only a few of these have found good practical applications. The bibliography contains only a fraction of the most useful empirical research papers on GARCH. In the next section we review some of the univariate models which have received the most attention: ARCH, GARCH, IGARCH, AGARCH, EGARCH, Components GARCH and Factor ARCH. There is little doubt that these GARCH volatility models are easy and attractive to use. A summary of their useful applications in financial risk management is given in section 4.5. However, in a climate where firm-wide risk management is the key, there is a pressing need to model volatilities and correlations in the context of large covariance matrices which cover all the risk factors relevant to the operations of a firm. Unfortunately it is not easy to use GARCH for large systems. In the last part of section 3 we look at the problems with direct estimation of high dimensional multivariate GARCH models and propose a new method for generating large GARCH covariance matrices.

4.3.2 A Survey of GARCH Volatility Models

1. ARCH

The original model of autoregressive conditional heteroscedasticity introduced in Engle (1982) has the conditional variance equation

$$\begin{gathered}\sigma_t^2 \;=\; \alpha_0 + \alpha_1\,\varepsilon_{t-1}^2 \;+\; \ldots \;+ \alpha_p\,\varepsilon_{t-p}^2 \\ \alpha_0 > 0, \quad \alpha_1, \ldots, \alpha_p \geq 0\end{gathered} \tag{9}$$

where the constraints on the coefficients are necessary to ensure that the conditional variance is always positive. This is the ARCH(p) conditional variance specification, with a memory of p time periods. This model captures the conditional heteroscedasticity of financial returns by using a moving average of past squared unexpected returns: If a major market movement in either direction occurred m periods ago ($m \leq p$) the effect will be to increase today's conditional variance. This means that we are more likely to have a large market move today, so 'large movements tend to follow large movements ... of either sign'.

2. VANILLA GARCH

The generalisation of Engle's ARCH(p) model by Bollerslev (1986, 1987) adds q autoregressive terms to the moving averages of squared unexpected returns: it takes the form

$$\begin{gathered}\sigma_t^2 = \omega \;+ \alpha_1\varepsilon_{t-1}^2 + \ldots + \alpha_p\varepsilon_{t-p}^2 + \beta_1\sigma_{t-1}^2 + \ldots + \beta_q\sigma_{t-q}^2 \\ \omega > 0, \quad \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \geq 0\end{gathered} \tag{10}$$

The parsimonious GARCH(1,1) model, which has just one lagged error square and one autoregressive term, is most commonly used:

$$\begin{gathered}\sigma_t^2 = \omega \;+ \alpha\,\varepsilon_{t-1}^2 + \beta\,\sigma_{t-1}^2 \\ \omega > 0, \quad \alpha\;\beta \geq 0\end{gathered} \tag{11}$$

It is equivalent to an infinite ARCH model, with exponentially declining weights on the past squared errors:

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha\ \varepsilon_{t-1}^2 + \beta\ \sigma_{t-1}^2 \\ &= \omega + \alpha\ \varepsilon_{t-1}^2 + \beta\ (\omega + \alpha\,\varepsilon_{t-2}^2 + \beta\ (\omega + \alpha\ \varepsilon_{t-3}^2 + \beta\ (\ldots.. \\ &= \omega\ /(1-\beta\ ) + \alpha\ (\varepsilon_{t-1}^2 + \beta\ \varepsilon_{t-2}^2 + \beta^2\ \varepsilon_{t-3}^2 + \ldots.)\end{aligned}$$

The above assumes that the GARCH(1,1) lag coefficient β is less than 1. In fact a few calculations show that the unconditional variance corresponding to a GARCH(1,1) conditional variance is

$$\sigma^2 = \omega\ /(1-\alpha - \beta\ ) \qquad (12)$$

and so not only must the GARCH return coefficient α also be less than 1, the sum $\alpha + \beta \leq 1$.

In financial markets it is common to find GARCH lag coefficients in excess of 0.7 but GARCH returns coefficients tend to be smaller, usually less than 0.25. The size of these parameters determine the shape of the resulting volatility time series: large GARCH lag coefficients indicate that shocks to conditional variance take a long time to die out, so volatility is ‘persistent’; large GARCH reurns coefficients mean that volatility is quick to re-act to market movements, and volatilities tend to be more 'spiky'. Figure 7 shows US dollar rate GARCH(1,1) volatilities for Sterling and the Japanese Yen. Cable volatility is more persistent (its lag coefficient is 0.931 compared with 0.839 for the Yen/$) - the Yen/$ is more spiky (its return coefficient is 0.094 compared with 0.052 for Cable).[14] See Alexander (1995a) for more details.

NOTE TO COPY EDITOR <Figure 7 here - same as figure 8.6 of 1st edition but please reduce size and line thickness in charts>

The constant ω determines the long-term average level of volatility to which GARCH forecasts converge (see section 4.3.3). Unlike the lag and return coefficients, its value is quite sensitive to

---

[14] These are slightly different from the values given in table 1 below, because firstly they use daily data (table 1 exchange rates are based on weekly data) and secondly a different data estimation period is used.

the length of data period used to estimate the model.[15] If a period of many years is used, during which there were extreme market movements, the estimate of ω will be high. So current volatility term structures will converge to a higher level. Consider for example, the generation of a GARCH volatility term structure on the FTSE today. Long term average volatility levels in the FTSE are around 13%, but if we include the Black Monday period in the data used to estimate today's GARCH model, we would find current long term volatility forecasts at around 15%.

3. INTEGRATED GARCH

When $\alpha + \beta = 1$ we can put $\beta = \lambda$ and re-write the GARCH(1,1) model as

$$\sigma_t^2 = \omega + (1-\lambda)\,\varepsilon_{t-1}^2 + \lambda\,\sigma_{t-1}^2 \qquad 0 \le \lambda \le 1 \tag{13}$$

Note that the unconditional variance (12) is now undefined - indeed we have a non-stationary GARCH model called the Integrated GARCH (I-GARCH) model, for which term structure forecasts do not converge. Our main interest in the I-GARCH model is that when $\omega = 0$ it is equivalent to an infinite EWMA, such as those used by RiskMetrics™. This may be seen by repeated substitution in (13):

$$\begin{aligned} \sigma_t^2 &= \omega + (1-\lambda)\,\varepsilon_{t-1}^2 + \lambda\,(\omega + (1-\lambda)\,\varepsilon_{t-2}^2 + \lambda\,(\omega + (1-\lambda)\,\varepsilon_{t-3}^2 + \lambda\,(\ldots. \\ &= \omega/(1-\lambda) + (1-\lambda)(\varepsilon_{t-1}^2 + \lambda\,\varepsilon_{t-2}^2 + \lambda^2\,\varepsilon_{t-3}^2 + \ldots..) \end{aligned} \tag{14}$$

Currency markets commonly have close to integrated GARCH models, and this has prompted major players such as Salomon Bros. to formulate new models for currency GARCH, such as the components GARCH model described below (see Chew, 1993).

4. ASYMMETRIC GARCH

The asymmetric GARCH (A-GARCH) model (Engle and Ng, 1993) has conditional variance equation

---

[15] For this reason it is common to impose a value of $\omega = (1 - \alpha - \beta)\,\sigma^2$ from equation (12), using an unconditional volatility estimate for σ^2

$$\sigma_t^2 \quad = \quad \omega + \alpha\ (\varepsilon_{t-1} - \xi)^2 + \beta\ \sigma_{t-1}^2 \qquad \omega > 0, \quad \alpha, \beta, \xi \geq 0 \tag{15}$$

In this model, negative shocks to returns ($\varepsilon_{t-1} < 0$) induce larger conditional variances than positive shocks. Thus the A-GARCH model is appropriate when we expect more volatility following a market fall than following a market rise. This ‘leverage effect’ is a common feature of financial markets, particularly equities.

5. EXPONENTIAL GARCH

The non-negativity constraints of the GARCH models considered so far can unduly restrain the dynamics of conditional variances so obtained. Nelson (1991) eliminated the need for such constraints in his exponential GARCH model by formulating the conditional variance equation in logarithmic terms

$$\log \sigma_t^2 = \alpha \quad + g\left(z_{t-1}\right) + \beta\ \log \sigma_{t-1}^2 \tag{16}$$

where $z_t = \varepsilon_t / \sigma_t$, so z_t is standard normal, and

$$g\left(z_t\right) = \omega\ \ z_t\ + \lambda \left(\left| z_t \right| - \sqrt{\frac{2}{\pi}} \right). \tag{17}$$

The asymmetric response function g(.) , which is illustrated in figure 8, provides the leverage effect just as in the AGARCH model. Many studies have found that the E-GARCH model fits financial data very well (for example, see Taylor, 1994, Heynen et. al., 1994, and Lumsdaine, 1995) but it is much more difficult to obtain volatility forecasts using this model.

NOTE TO COPY EDITOR <Figure 8 here - same as figure 8.7 from 1st edition>

6. COMPONENTS GARCH

In practice, estimation of a GARCH model over a rolling data window will generate a term structure of volatility forecasts for each day. In each of these term structures, as volatility maturity increases, GARCH forecasts should converge to a long-term level of volatility (see section 4.3.3). As the data window is rolled different long-term levels will be estimated, corresponding to different estimates of the GARCH parameters. The components model extends this idea of time-varying 'baseline' volatility to allow variation within the estimation period (Engle and Lee, 1993 and Engle and Mezrich, 1995).

To understand the components model, note that when $\alpha + \beta < 1$ the GARCH(1,1) conditional variance is often estimated by imposing ω. The model may be written in the form

$$\begin{aligned} \sigma_t^2 &= (1-\alpha-\beta)\sigma^2 + \alpha\,\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \\ &= \sigma^2 + \alpha\,(\varepsilon_{t-1}^2 - \sigma^2) + \beta\,(\sigma_{t-1}^2 - \sigma^2) \end{aligned} \tag{18}$$

where σ^2 is defined by (12). We now replace σ^2 by a time-varying permanent component in conditional variance:

$$q_t = \overline{\omega} + \rho\,(q_{t-1} - \overline{\omega}) + \phi\,(\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \tag{19}$$

and then the formulation (18) becomes

$$\sigma_t^2 = q_t + \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}) \tag{20}$$

The equations (19) and (20) together define the components model. If $\rho = 1$, the permanent component - to which long-term forecasts mean revert - is just a random walk.

7. FACTOR GARCH

Factor GARCH allows individual volatilities and correlations to be estimated and forecast from a single GARCH volatility - the volatility of the market. Consider the simple capital asset pricing model, where individual asset or portfolio returns are related to market returns M_t by the regression equation

$$r_{it} = \alpha_i + \beta_i M_t + \varepsilon_{it} \qquad i = 1,2,\ldots\ldots,n \tag{21}$$

Denoting by σ_{it} the standard deviation of asset i at time t and by σ_{ijt} the covariance between assets i and j at time t, equation (21) yields

$$\begin{aligned} \sigma_{it}^2 &= \beta_i^2 \sigma_{Mt}^2 + \sigma_{\varepsilon_{it}}^2 \\ \sigma_{ijt}^2 &= \beta_i \beta_j \sigma_{Mt}^2 + \sigma_{\varepsilon_{it}\varepsilon_{jt}} \end{aligned} \tag{22}$$

From a simultaneous estimation of the n linear regression equations in (22) we can obtain estimates of the factor sensitivities β_i and the error variances and covariances. These, and the univariate GARCH estimates of the market volatility σ_M are then used to generate individual asset volatilities and correlations using (22). The idea is easily extended to a model with more than one risk factor, see Engle, Ng and Rothschild (1990).

4.3.3 GARCH Volatility Term Structure Forecasts

One of the beauties of GARCH is that volatility and correlation forecasts for any horizon can be constructed from the one estimated model. First we use the estimated GARCH model to give us forecasts of instantaneous forward volatilities, that is the volatility of r_{t+j}, made at time t and for every step ahead j. The instantaneous GARCH forecasts are calculated analytically: for example, in the GARCH(1,1) model

$$\hat{\sigma}_{t+1}^2 = \hat{\omega} + \hat{\alpha}\, \varepsilon_t^2 + \hat{\beta}\, \hat{\sigma}_t^2 \tag{23}$$

and the j-step ahead forecasts are computed iteratively[16] as

---

[16] We only know the unexpected return at time t, not ε_{t+j} for j>0. But $E(\varepsilon_{t+j}^2) = \sigma_{t+j}^2$.

$$\hat{\sigma}^2_{t+j} = \hat{\omega} + (\hat{\alpha} + \hat{\beta}\ )\,\hat{\sigma}^2_{t+j-1} \qquad (24)$$

To get a term structure of volatility forecasts from these forward volatilities note that the (logarithmic) return at time t over the next n periods is

$$r_{t,n} \quad = \quad \sum_{j=1}^{n} r_{t+j}$$

The volatility term structure is a plot of the volatility of these returns for n=1,2,3,……. Since

$$V_t\,(r_{t,n}) \quad = \quad \sum_{i=1}^{n} V_t\,(r_{t+i}) \quad + \quad \sum_i \sum_j COV_t\,(r_{t+i}\ , r_{t+j}) \qquad (25)$$

the GARCH forecast of n-period variance is the sum of the instantaneous GARCH forecast variances, plus the double sum of the forecast autocovariances between returns. This double sum will be very small compared to the first sum on the right hand side of (25), indeed in the majority of cases the conditional mean equation in a GARCH model is simply a constant, so returns are independent and the double sum is zero.[17] Hence we ignore the autocovariance term in (25) and construct n-period volatility forecasts simply by adding the j-step-ahead GARCH variance forecasts (and then square-rooting and annualising in the usual way).[18]

---

[17] Even in an AR(1)-GARCH(1,1) model (with autocorrelation coefficient ρ in the conditional mean equation) (25) becomes

$$\hat{\sigma}^2_{t,n} = \sum_{i=1}^{n} \hat{\sigma}^2_{t+i} \quad + \ \hat{\sigma}^2_{\mathrm{t}}[\rho\ (1-\rho^n)\,/\,(1-\rho)]^2$$

and the first term clearly dominates the second.

[18] In this way we can also construct any sort of forward volatility forecasts, such as 3-month volatility but for a period starting six months from now.

Figure 9: GARCH volatility term structure forecasts of the FTSE: 1988-1995

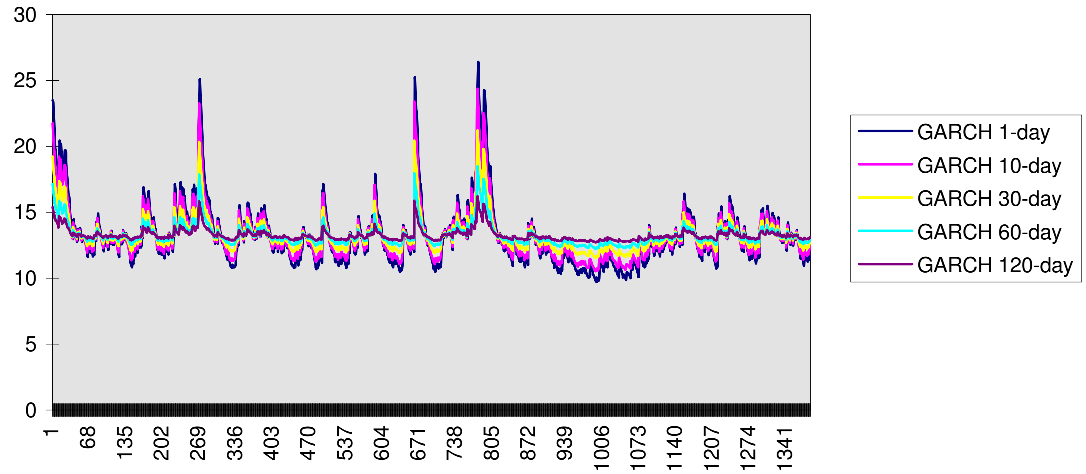


Figure 9 shows 1-day, 10-day 30-day 60-day and 120-day volatility forecasts for the FTSE from 1988-1995. Note how the forecasts of different maturities converge to the long-term volatility level of around 13%. During a volatile period GARCH term structures converge to this level from above (figure 10(a)) and during a tranquil period they converge from below (figure 10(b)) .

Figure 10(a): GARCH volatility term structures of US Dollar rates on April 13 1995

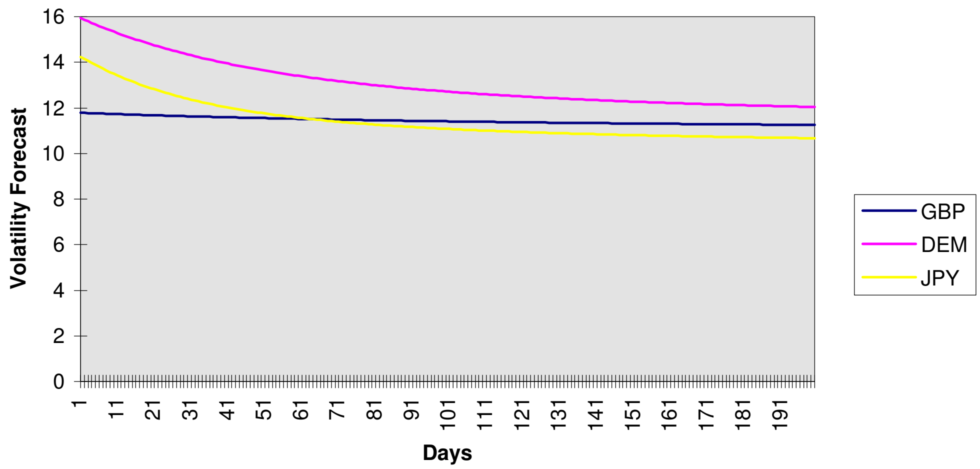
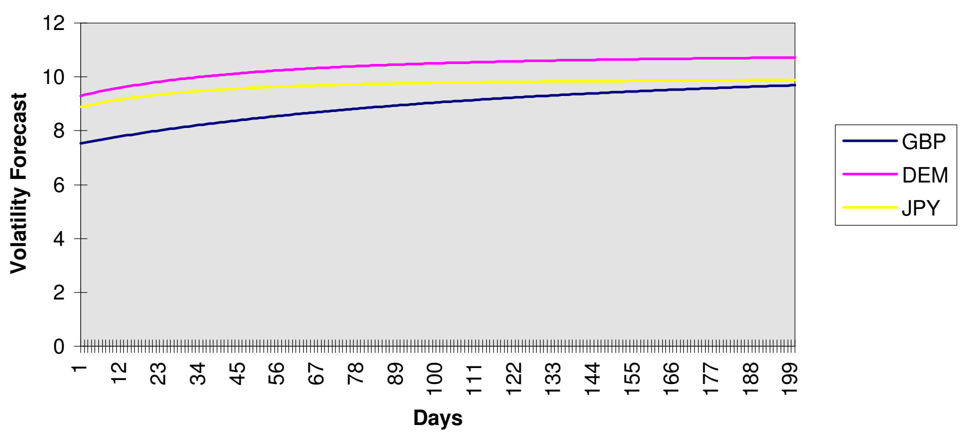


Figure 10(b): GARCH volatility term structures of US Dollar rates on March 2 1995



The speed of convergence in GARCH(1,1) depends on $\alpha + \beta$. Currency markets generally have the highest values of $\alpha + \beta$, and hence the slowest convergence. The speed of convergence in equity and commodity GARCH models tends to be faster, and bond markets often have the lowest $\alpha + \beta$ and the fastest converge of volatility term structures to the long-term level (see section 4.3.4).

4.3.4 Estimating GARCH Models: Methods and Results

Most of the models described above are available as pre-programmed procedures in econometric packages such as S-PLUS, TSP, EVIEWS and MICROFIT, and in GAUSS and RATS, GARCH procedures can be written, as explained in the manuals. The method used to estimate GARCH model parameters is maximum likelihood estimation, which is a powerful and general statistical procedure, widely used because it always produces consistent estimates.

The idea is to choose parameter estimates to maximise the likelihood of the data under an assumption about the shape of the distribution of the data generation process. For example, if we assume a normal data generation process with mean μ and variance σ^2 then the likelihood of getting returns r_1 , r_2 , ..., r_T is

$$L(\mu, \sigma^2 \mid r_1, r_2, \ldots, r_T) = \prod_{t=1}^{T} f(r_t) \qquad (26)$$

where

$$f(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2}\left(\frac{r-\mu}{\sigma}\right)^2\} \qquad (27)$$

Choosing μ and σ^2 to maximise L (or equivalently to minimise -2logL) yields the maximum likelihood estimates of these parameters.

In GARCH models there are more than just two parameters to estimate, and the likelihood functions are more complex (for example, see Engle, 1982 and Bollerslev, 1986) but the principle is the same. Problems arise though, because the more parameters in the likelihood function the 'flatter' it becomes, and therefore more difficult to estimate. For this reason the GARCH(1,1) model is preferred to an ARCH model with a long lag, and parameterizations of conditional mean equations are as parsimonious as possible - often we use just a single constant in the conditional mean.

Convergence problems with GARCH models can arise because gradient search algorithms used to maximize the likelihood functions fall off a boundary. Sometimes this problem can be mitigated by changing the starting values of the parameters, or knocking a few data points off the beginning of the data set so the likelihood function has a different gradient at the beginning of the search. The time taken for GARCH models to converge will be greatly increased unless analytic derivatives are used to calculate the gradient in the search. For leptokurtic data, t-distributed GARCH models have lost ground in favour of normal mixture GARCH models since they require numerical derivatives to be calculated at each iteration.

However, most univariate GARCH models should encounter few convergence or robustness problems if they are well-specified. Care should be taken over the re-specification of initial conditions if the coefficients estimates hit a boundary value, and sometimes minor changes in the returns data can induce the occasional odd value of coefficients. These will be evident in rolling GARCH models, where convergence conditions may need re-setting. In some cases the coefficient values take some time to settle down but once the model is properly tuned to the data it can be updated daily or weekly with few problems.

Table 1: Approximate GARCH(1,1) parameters for equity markets and USD exchange rates

	alpha	beta
Equities		
UK	0.105	0.810
GE	0.188	0.712
US	0.271	0.641
JP	0.049	0.935
NL	0.146	0.829
USD rates		
DEM	0.113	0.747
JPY	0.102	0.763
GBP	0.028	0.935
NLG	0.125	0.735
ESP	0.160	0.597
AUD	0.241	0.674

Table 1 gives an idea of what to expect when estimating GARCH(1,1) parameters in some of the major currency and equity markets. Note that very few of these markets have persistence parameters which are as large as the value 0.94 used for the RiskMetrics data. Of course the GARCH parameters depend on the data frequency and estimation period, but they should be fairly robust to these differences. When rolling the estimation of a GARCH model day by day, significant changes in the parameters should occur consequent to major market movements only. Bond market GARCH models are more difficult to estimate, particularly since some maturities can be quite illiquid. To estimate and forecast volatilities and correlations for an entire yield curve, the orthogonal GARCH model is highly recommended (see section 4.3.7).

4.3.5 Choosing the Data Period and the Appropriate GARCH Model

The plain 'vanilla' GARCH(1,1) model - even without asymmetric or leptokurtic effects - already offers many advantages over the moving average methods described in section 4.2, and many financial institutions are currently basing their systems on this model. One of the questions that senior management will want to address is whether such a simple GARCH model does the trick - is it capturing the right type of volatility clustering in the market - or should we be using some sort of complex fractionally integrated GARCH model like the house next door? In this section we show how to diagnose whether you have a good GARCH model or not, and how to employ data to the best advantage.

A test for the presence of ARCH effects in returns is obtained by looking at the autocorrelation in the time series of squared returns . Standard autocorrelation test statistics may be used, such as the Box-Pierce Q $\sim \chi^2$ (p) *:*

$$Q = T \sum_{n=1}^{p} \varphi(n)^2 \qquad (28)$$

where φ(n) is the nth order autocorrelation coefficient in squared returns

$$\varphi(n) = \frac{\sum_{t=n+1}^{T} r_t^2 \; r_{t-n}^2}{\sum_{t=1}^{T} r_t^4} \qquad (29)$$

One of the main specification diagnostics in GARCH models is to first standardise the returns by dividing by the estimated GARCH standard deviation, and then test for autocorrelation in squared standardised returns. If it has been removed, the GARCH model is doing its job. But what if several GARCH models account equally well for GARCH effects? In that case choose the GARCH model which gives the highest likelihood, either in-sample or in post-sample predictive tests.

The two important considerations in choosing data for GARCH modelling are the data frequency and the data period. It is usual to employ daily or even intra-day data rather than weekly data since convergence problems could be encountered on low frequency data due to insufficient ARCH effects. If tic data are used the time bucket should be sufficiently large to ensure there are no long periods of no trades, and by the same token bank holidays may cause problems so it is often better to deviate from the standard time-series practise of using equally spaced daily data, and not fill in the zero returns caused by back holidays.

When it comes to choosing the amount of historical data for estimating GARCH models, the real issue is whether you want major market events from several years ago to influence your forecasts today. As we have already seen, including Black Monday in equity GARCH models has the effect of raising long-term volatility forecasts by several percent. In the orthogonal GARCH model of section 4.3.7 it is also important not to take too long a data period, since the principal components are only unconditionally orthogonal so the model will become ill-conditioned if too long a data period is chosen.[19] On the other hand, a certain amount of data is necessary for the likelihood fuction to be sufficiently well defined. Usually at least one or two years of daily data are necessary to ensure proper convergence of the model.

4.3.6 Multivariate GARCH

Univariate GARCH models commonly converge in an instant, the only real problems being lack of proper specification by the user, or inappropriate data. But there are very serious computational problems when attempting to build large positive definite GARCH covariance matrices which are necessary if one is to net the risks from all positions in a large trading book.

This section reviews the basic multivariate GARCH models, discussing the inevitable computational problems if one attempts direct estimation of full GARCH models in large dimensional systems. The unconditional variance of a multivariate process is a positive definite

---

[19] For risk management purposes, currently one does not need to go further back than 1st January 1993. The major disruptions to financial markets during 1992 are best left out of ordinary everyday volatilities, but can easily be re-created when the model is being stressed.

matrix, the covariance matrix, already defined in the introduction. The conditional variance of a multivariate process is a time-series of matrices, one matrix for each point in time. It is not surprising therefore that estimation of these models can pose problems! The convergence problems outlined in section 4.3.4 can become insurmountable even in relatively low dimensional systems, so parameterizations of multivariate GARCH models should be as parsimonious as possible. There are many ways that multivariate GARCH can be constrained in order to facilitate their estimation but these methods often fall down on at least one of two counts: either they are only applicable to systems of limited dimension (something like 5-10 factors being the maximum) or they need to make unrealistic assumptions on parameters that are not confirmed by the data (see Engle and Kroner, 1995). However, section 4.3.7 presents a new method which falls down on neither count. It uses orthogonal approximations to generating arbitrary large GARCH covariance matrices using only univariate GARCH models.

Consider first the bivariate GARCH model, appropriate only if we are just interested in the correlation between two returns series, r_1 and r_2 . There will now be two conditional mean equations, which can be anything we like but for the sake of parsimony we shall assume that each equation gives the return as a constant plus error:

$$
\begin{aligned}
r_{1,t} &= \varphi_{11} + \varepsilon_{1,t} \\
r_{2,t} &= \varphi_{21} + \varepsilon_{2,t}
\end{aligned}
$$

In a bivariate GARCH there are three conditional variance equations, one for each conditional variance and one for the conditional covariance. Since all parameters will be estimated simultaneously the likelihood can get *very* flat indeed, so we need to use as few parameters as possible. In this section we introduce the two standard parameterizations of multivariate GARCH models: the vech and the BEKK.

In the vech parameterization each equation is a GARCH(1,1):

$$
\begin{aligned}
\sigma_{1,t}^2 &= \omega_1 + \alpha_1 \varepsilon_{1,t-1}^2 + \beta_1 \sigma_{1,t-1}^2 \\
\sigma_{2,t}^2 &= \omega_2 + \alpha_2 \varepsilon_{2,t-1}^2 + \beta_2 \sigma_{2,t-1}^2 \qquad (30) \\
\sigma_{12,t} &= \omega \ + \alpha_3 \varepsilon_{1,t-1} \varepsilon_{2,t-1} + \beta_3 \sigma_{12,t-1}
\end{aligned}
$$

As usual, constraints on the coefficients in (30) are necessary to ensure positive definiteness of the covariance matrices. To obtain time-series of GARCH correlations we simply divide the estimated covariance by the product of the estimated standard deviations, at every point in time (see figure 11).

<Figure 11 here - same as fig 8.8 from 1st edition>

When considering systems with more than two returns series we need matrix notation. The matrix form of equations (30) is

$$\text{vech}(H_t) = A + B\,\text{vech}(\xi_{t-1}\xi_{t-1}') + C\,\text{vech}(H_{t-1}) \qquad (31)$$

where H_t is the conditional variance matrix at time t. So $\text{vech}(H_t) = (\sigma_{1t}^2, \sigma_{2t}^2, \sigma_{12t})'$, and $\xi_t = (\varepsilon_{1t}, \varepsilon_{2t})'$, $A = (\omega_1, \omega_2, \omega_3)'$, $B = \text{diag}(\alpha_1, \alpha_2, \alpha_3)$ and $C = \text{diag}(\beta_1, \beta_2, \beta_3)$, with the obvious generalisation to higher dimensions.

There are severe cross equation restrictions in the vech model, for example the conditional variances are not allowed to affect the covariances and vice-versa. In some markets these restrictions can lead to substantial differences between the vech and BEKK estimates so the vech model should be employed with caution.

A more general formulation, which involves the minimum number of parameters whilst imposing no cross equation restrictions and ensuring positive definiteness for any parameter values is the BEKK model (after Baba, Engle, Kraft and Kroner who wrote the preliminary version of Engle and Kroner, 1995). This parameterization is given by

$$H_t = A'A + B'\xi_{t-1}\xi_{t-1}'B + C'H_{t-1}C \qquad (32)$$

where A, B and C are general nxn matrices (A is triangular). The BEKK parameterization for a bivariate model involves eleven parameters, only two more than the vech. But for higher dimensional systems the extra number of parameters increases, and completely free estimation becomes very difficult indeed. Often it is necessary to let B and C be scalar matrices, to reduce the number of parameters needing estimation and so improve the likelihood surface and, hopefully, achieve convergence. More details are given in Bollerslev et. al. (1992) and Bollerslev et. al. (1994) .

A useful technique for calibrating multivariate GARCH is to compare the multivariate volatility estimates with those obtained from direct univariate GARCH estimation. The multivariate GARCH volatility term structure forecasts are computed as outlined in section 4.3.4. and correlation forecasts are calculated in a similar fashion: simply by iterating conditional covariance forecasts and summing these to get n-period covariance forecasts.[20] These are then divided by the product of n-period volatility forecasts for correlation term structures:

$$\hat{\rho}_{t,n} = \hat{\sigma}_{12,t,n} / \hat{\sigma}_{1,t,n}\hat{\sigma}_{2,t,n}$$

4.3.7 Generating Large GARCH Covariance Matrices: Orthogonal GARCH

The computations required to estimate very large GARCH covariance matrices by direct methods are simply not feasible at the present time. However indirect 'orthogonalization' methods can be used to produce the large covariance matrices necessary to measure risk in large portfolios. This method is introduced, explained and verified for all major equity, currency and fixed income markets by Alexander and Chibumba (1997).

In orthogonal GARCH the risk factors from all positions across the entire firm are first divided into reasonably highly correlated categories, according to geographic locations and instrument types. Principal components analysis is then used to orthogonalize each sub-system of risk factors, univariate GARCH is applied to the principal components to obtain the (diagonal) covariance matrix, and then the factor weights from the principal components analysis are used to 'splice' together the large covariance matrix for the whole enterprise.

An example explaining the method for just two risk factor categories can easily be extrapolated to any number of risk factor categories: Suppose **P** = (P_1 , ... P_n) are the principal components of the first system (n risk factors) and let **Q** = (Q_1 ,...Q_m) be the principal components of the second system (m risk factors).[21] Denote by **A** (nxn) and **B** (mxm) the factor weights matrices of the first and second systems. Within factor covariances are given by **AV(P)A'** and **BV(Q)B'** respectively and cross factor covariances are **ACB'** where **C** denotes the mxn matrix of covariances of principal components.

An illustration of this method applies to a system of equity indices and USD exchange rates (Alexander and Chibumba, 1997). The system is just small enough to compare the 'splicing' method with full multivariate GARCH estimation and figure 12 shows one of the resulting

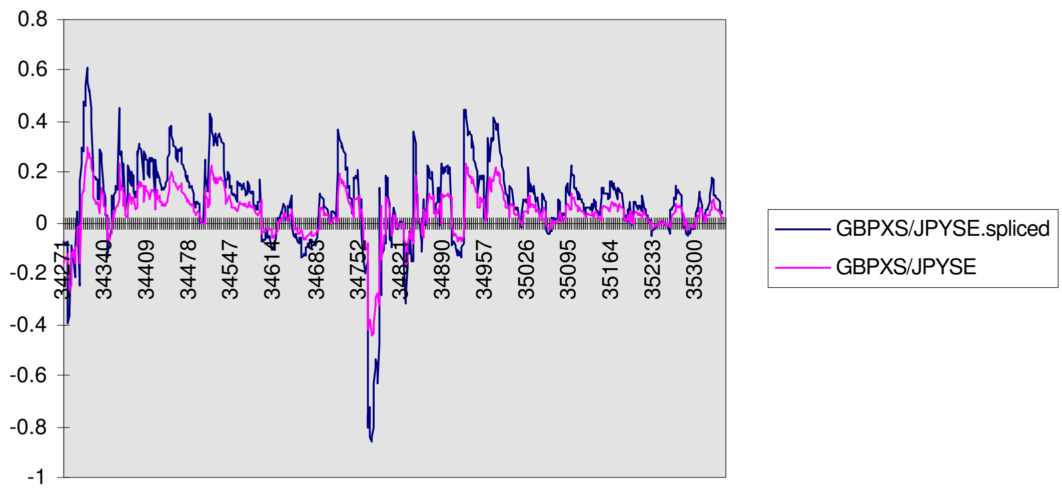
---

[20] To see this, note that $COV_t\,(r_{1,t,n}, r_{2,t,n}) \;=\; \sum_{i,j=1}^{n} COV_t\,(r_{1,t+i}, r_{2,t+j})$. Ignoring non-contemporaneous covariances gives $\hat{\sigma}_{12,t,n} = \sum_{i=1}^{n} \hat{\sigma}_{12,t+i}$.

[21] One of the advantages of principal components analysis is to reduce dimension, and this is certainly an attractive proposition for the yield curve (see chapter ?), or for systems when the reduction in 'noise' obtained by using only a few principal components makes correlations more stable. However it is the orthogonality property of principal components which is the primary attraction here, and if one does not retain the full number of principal components, we cannot ensure positive definiteness of the final covariance matrix.

correlations obtained by each method - between the GBP/USD exchange rate and the Nikkei during the period 1 Jan 1993 and 17 Dec 1996.

Figure 12: Comparison of orthogonal GARCH and BEKK correlations between the Cable rate and the Nikkei index: 1993-1996



Some care must be taken with the initial calibration of orthogonal GARCH, but once calibrated it is a useful technique for generating large GARCH covariance matrices, with all the advantages offered by these models. It is particularly useful for yield curves, where the more illiquid maturities can preclude the direction estimation of GARCH volatilities. The orthogonal method not only provides estimates for maturities with inadequate data - a substantial reduction in dimensionality is also possible.

The factors that must be taken into account when calibrating an orthogonal GARCH are the time period used for estimation, and the sub-division into risk factor categories. For example, if a market that has very idiosyncratic properties is taken into a sub-system of risk factor categories, the volatilities and correlations of all other markets in the system will be erroneously affected. Therefore, one has to compare the volatilities and correlations obtained by direct GARCH - or

EWMA - with those obtained from the orthogonal GARCH, to ensure successful calibration of the model.

4.4 'Implied' Volatility and Correlation

Implied volatility and correlation are those volatilities and correlations which are implicit in the prices of options. When an explicit analytic pricing formula is available (such as the Black-Scholes formula - see Volume 1 Chapter 2) the quoted prices of these products, along with known variables such as interest rates, time to maturity, exercise prices and so one, can be used in an implicit formula for volatility. The result is called the implied volatility. It is a volatility forecast - not an estimate of current volatility - the volatility forecast which is implicit in the quoted price of the option, with an horizon given by the maturity of the option.

Although they forecast the same thing (the volatility of the underlying assets over the life of the option) implied volatilities must be viewed differently from statistical volatilities because they use different data and different models[22]. If the option pricing model were an accurate representation of reality, and if investors expectations are correct so that there is no over- or under- pricing in the options market, then any observed differences between implied and statistical volatility would reflect inaccuracies in the statistical forecast. Alternatively, if statistical volatilities are correct, then differences between the implied and statistical measure of volatility would reflect a mispricing of the option. In fact, rather than viewing implied volatility and statistical methods as complementary forecasting procedures, when implied volatilities are available they should be taken along side the statistical forecasts. For example in the ‘volatility cones’ described below, or in Salomon Brothers 'Gift' (GARCH index forecasting tool) the relationship between implied volatility and GARCH volatility is used to predict future movements in prices (see Chew, 1993).

4.4.1 Black-Scholes Implied Volatility

---

[22] Implied methods use current data on market prices of options - hence implied volatility contains all the forward expectations of investors about the likely evolution of the underlying. Statistical methods use only historic data on the underlying asset price. Apart from differences in data, the mathematical models used to generate these

The Black-Scholes formula for the price of a call option with strike price K and time to maturity t on an underlying asset with current price S and t-period volatility σ is

$$C = SN(x) \; - \; Kr^{-t} N\left(x - \sigma\sqrt{t}\right) \qquad (33)$$

where r denotes the 'risk-free' rate of interest, N(.) is the normal distribution function and

$$x \; = \; \ln\left(S \,/\, Kr^{-t}\right) \big/ \sigma\sqrt{t} + \sigma\sqrt{t} \big/ 2$$

.

Option writers would estimate a statistical volatility and use the Black-Scholes or some other option model to price the option. But if C is observed from the market, and so also are S, K, r and t, then (33) may be used instead to ‘back-out’ a volatility implied by the model. This is called the (Black-Scholes) implied volatility, σ. Options of different strikes and maturities on the same underlying attract different prices, and therefore different implied volatilities. A plot of these volatilities against strike price and maturity is called the ‘volatility smile’ (see figure 13).

Long term options have little variation in prices, but as maturity approaches it is typical that out-of-the money options would imply higher volatility in the underlying than at-the-money options, otherwise they would not be in the market. Thus implied volatility is often higher for out-of-the money puts and calls than for at-the money options, an effect which is termed the ‘smile’. Equity markets exhibit a ‘leverage effect’, where volatility is often higher following bad news than good news (see section 4.3.2 (4)). Thus out-of-the money puts require higher volatility to end up in-the-money than do out-of-the money calls, and this gives rise to the typical ‘skew’ smiles of equity markets, such as in figure 13.

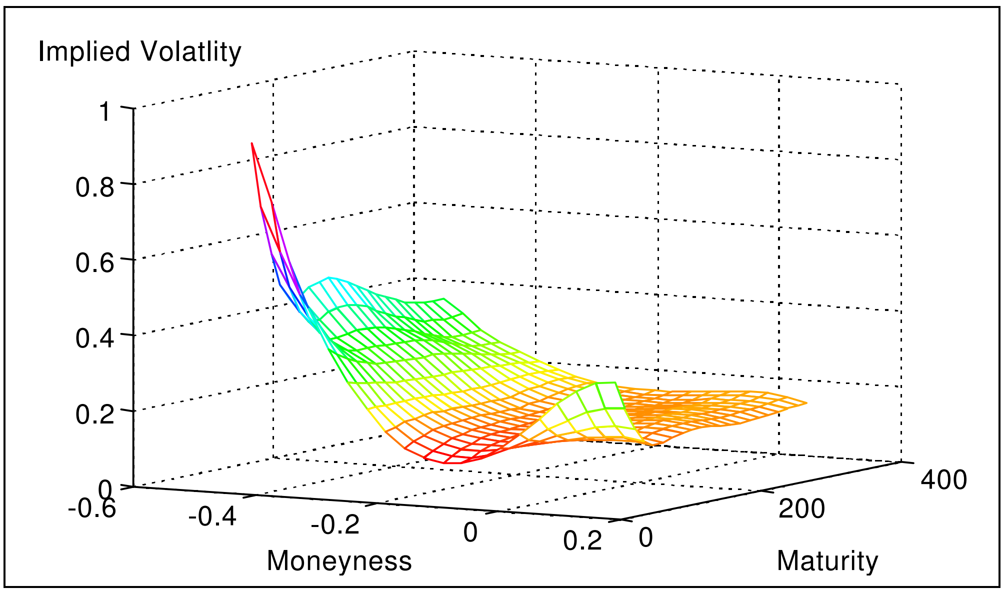
The volatility smile is a result of pricing model bias, and would not be found if options were priced using appropriate stochastic volatility models (see Dupire, 1997). For example, when

---

quantities are totally different: Implied methods assume risk neutrality and use a diffusion process for the underlying asset price.

options are priced using the Black-Scholes formula, which assumes log-normal prices, we observe volatility smiles because empirical returns distributions are not normal. Most noticeable in currency markets, returns distributions have much fatter tails than normal distributions, so out-of-the money options have more chance of being in-the-money than is assumed under Black-Scholes. This underpricing of the Black-Scholes model compared to observed market behaviour yields higher Black-Scholes implied volatilities for out-of-the money options.

Figure 13: Black-Scholes smile surface for FTSE options, December 1997



'Volatility cones' are graphic representations of the term structure of volatility, used for visual comparison of current implied volatility with an empirical historical distribution of implied volatilities. To construct a cone, fix a time to maturity t and estimate a frequency distribution of implied volatility from all t-maturity implied volatilities during the last two years (say), recording the upper and lower 95% confidence limits. Repeat for all t. Plotting these confidence limits yields a cone-like structure because the implied volatility distribution becomes more peaked as the option approaches maturity (see volume 1 chapter 8). Cones are used to track implied volatility over the life of a particular option, and under or overshooting the cone can signal an opportunity to trade. Sometimes cones are constructed on 'historic' volatility because historic data on implied volatility may be difficult to obtain. In this case cones should be used with caution, particularly if overshooting is apparent at the long end. Firstly, differences between long-term 'historic' and implied volatility are to be expected, since transactions costs are included in the implied volatilities but not the statistical. But also the ‘Black-Scholes bias, which tends to under price ATM options and over price OTM options, can increase with maturity.

4.4.2 Implied Correlation

The increase in derivatives trading in OTC markets enables implied correlations to be calculated from three implied volatilities by rearrangement of the formula for the variance of a difference: if we denote by ρ the correlation between x and y

$$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y\rho$$

or

$$\rho = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y} \qquad (34)$$

Putting implied volatilities in formula (34) gives the associated implied correlation - we just need traded options on three associated assets or rates X, Y and X-Y. For example X and Y could be two US dollar FX rates (in logarithms) so X-Y is the cross rate. In the above formulae, the implied correlation between the two FX rates is calculated from the implied volatilies of the two USD rates σ_x and σ_y and the implied volatility of the cross rate σ_{x-y} .

Similar calculations can be used for equity implied correlations*:* Assuming the correlation between all pairs of equities in an index is constant allows this correlation to be approximated from implied volatilities of stocks in the index - see Kelly (1994). Of course, the assumption is very restrictive and not all equities will be optionable, so the approximation is very crude and can lead to implied correlations which are greater than 1. Another way in which implied correlations can be obtained is by inverting the quanto pricing formula. When we know the quanto price, the interest rate, the equity and FX implied volatilities and so on we can invert the Black-Scholes quanto pricing formula to obtain an implied correlation between the equity and the FX rate.

There are a limited number of ways in which implied correlations can be backed out from option prices, and even when it is possible to obtain such implied correlations, they will be very unstable. Not only are the underlying correlations often unstable, but with implied correlations there are a number of model assumptions which may be questionable. In short, these correlations should be used with even more caution than implied volatilities.

4.5 A Short Survey of Applications to Measuring Market Risk

4.5.1 Factor Models

Simple factor models describe returns to an asset or portfolio in terms of returns to risk factors and an idiosyncratic or 'specific' return. For example, in the simple Capital Asset Pricing Model (CAPM) the return to a portfolio, denoted r_t, is approximated by the linear regression model

$$r_t = \alpha + \beta R_t + \varepsilon_t$$

where R_t is the return to the market risk factor, the error process ε_t denotes the idiosyncratic return and β denotes the sensitivity (or 'beta') of the portfolio to the risk factor. Standard regression models would then estimate this beta by the ratio of the covariance (between the portfolio and the risk factor) to the variance of the risk factor:

$$\beta = \mathrm{COV}(r,R)/\mathrm{V}(R)$$

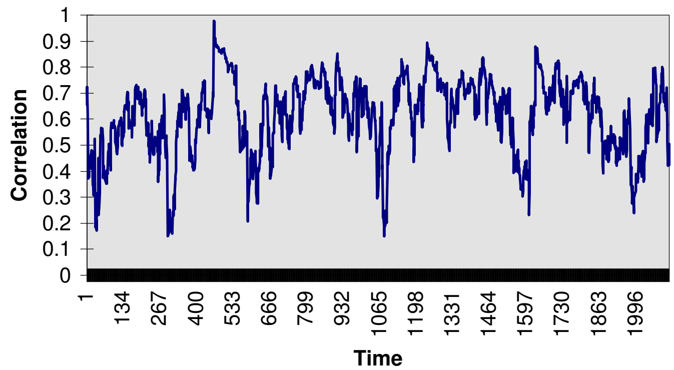
Thus beta is the relative volatility times the correlation between the risk factor and the portfolio. Although relative volatilities may be quite stable, the same cannot be said of the correlation. This means that portfolio betas tend to jump around quite a bit, although this will not be apparent if equally weighted averaging methods are used to estimate them.[23] But this is the case in standard ordinary least squares (OLS) regression. And many analytics firms will be using OLS for their

---

[23] It is exactly the same problem of 'ghost features', that have already been described with respect to 'historic' volatilities and correlations, that makes market betas appear more stable than they really are.

betas, so their estimates just reflect an average beta over whatever data period is used for their estimation. If, on the other hand, we use a time-varying estimate such as EWMA or GARCH for the covariance and variance (as in Kroner and Claessens, 1991) the beta estimate will more accurately reflect current market conditions.

Figure 14: EWMA Correlation and Beta for NatWest bank vs the FTSE: 1985-1991



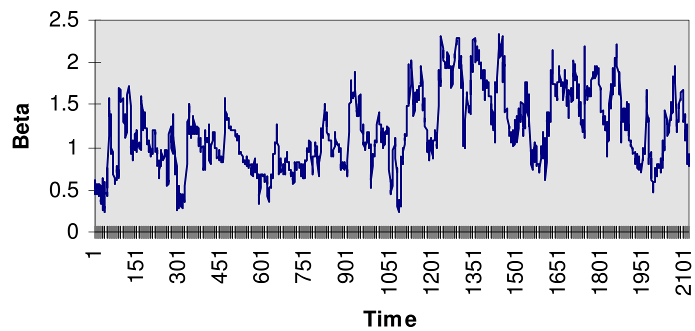


Figure 14 shows the time-varying beta for National Westminster Bank in the FTSE, calculated using an exponentially weighted moving average for correlation and the volatility of the FTSE. It shows significant variation over time, ranging from less than 0.2 to nearly 1 (on Black Monday).

Unfortunately many analytics firms still employ equally weighted averages in the OLS procedures commonly used to calculate market betas, so they provide only an average of the beta during the estimation period, which is 0.613 for figure 14. Clearly this can produce very inaccurate measures of the current beta when compared with time-varying methods.

More generally, in multivariate factor models, portfolio returns are attributed to several risk factors: the portfolio return is given by

$$r_t = \alpha + \beta_1 R_{1t} + \beta_2 R_{2t} + \ldots\ldots + \beta_n R_{nt} + \varepsilon_t$$

where R_1 , ..., R_n are the returns to different risk factors, β_1 , ..., β_n are the net portfolio sensitivities (i.e. the weighted sum of the individual asset sensitivities) and ε is the idiosyncratic or specific return of the portfolio (i.e. that part of the return not attributed to the different risk factors). The net betas with respect to each risk factor can be calculated from the covariances (between risk factors and between the portfolio and risk factors) and the variances of the risk factors: Denote by **X** the Txn matrix of data on the risk factors and by **y** the Tx1 vector of data on the portfolio returns. Then we estimate β as the inverse of the covariance matrix between risk factors, times the vector of covariances between the portfolio and the risk factors:

$$\beta = (\mathbf{X'X})^{-1}\,\mathbf{X'y}$$

Again, time-varying methods for calculating these variances and covariances should be employed in order to obtain a true reflection of current market conditions. The use of OLS in factor models is a major source of error, which has implications for risk management whenever these models are used - in mean-variance analysis, in VAR models, and in other methods for capital allocation.

4.5.2 Capital Allocation

Suppose first that capital allocation decisions are made on the basis of risk characteristics alone. The standard problem is of choosing a vector of weights $\mathbf{w} = (w_1 , \ldots, w_n)'$ in a portfolio to minimise its variance (although variance is not necessarily the best measure of risk - see Dembo (1997) and chapter 7 of this volume). In mathematical notation:

$$\text{Min}_{\mathbf{w}} \quad \mathbf{w'Vw} \quad \text{such that} \quad \Sigma w_i = 1$$

where **V** is the covariance matrix of asset returns. Assuming that weights can be negative or zero this is a straightforward linear programming problem, having the solution

$$w_i^* = \psi_i \;/\; \Sigma\psi_i \quad \text{where} \quad \psi_i = \;\Sigma \text{ ith column of } \mathbf{V}^{-1}.$$

Remarks about the appropriate use of covariance matrices **V** are made below. For example we may take **V** to be a time-varying covariance matrix such as the GARCH matrix. This will give time-varying weights $w_{i,t}^*$ that can be used to re-balance the portfolio within certain limits as volatilities and correlations between the assets change.

But capital allocation decisions made on the basis of risk alone may be inappropriate since the return characteristics are ignored. More risk may be acceptable if it is accompanied by higher returns, and managers are in danger of under utilising capital if insufficient capital is allocated to high risk, high return activities. Hence capital allocation is commonly viewed within the framework of risk-return analysis. The problem is to allocate capital between assets $\mathbf{R} = (R_1 , ..., R_n)$'with optimal weights $\mathbf{w} = (w_1 , …., w_n)$. Efficient allocations lie on the 'efficient frontier', where it is not possible to adjust allocations $w_1 , …., w_n$ to gain higher return for the same level of risk, or less risk for the same level of return (see figure 15). Assuming markets are efficient, so that it is not possible to gain return for no risk, the optimal allocation for a risk neutral investor will be at the point X in figure 16, and the risk adjusted return on capital is given by the reciprocal of the slope of the tangent at that point.

Figure 15 here - efficient frontier

In simple portfolios which can be described by a weighted sum of the constituent assets, the current return to the portfolio is **w'R** where **w** is the vector of portfolio weights and **R** is the vector of asset returns. The portfolio risk is commonly measured as variance, although this has its drawbacks, as described in chapter 7. The portfolio variance is $\mathbf{w'Vw}$ where $\mathbf{V}$ is the covariance matrix of $\mathbf{R}$.[24] In larger linear portfolios, which are best described not at the asset level but by the factor models described above, the portfolio variance is measured using the covariance matrix of risk factor returns and the vector of sensitivities to different risk factors. If we ignore the idiosyncratic risk, the portfolio return is $\boldsymbol{\beta}'\mathbf{R}$ and its variance is simply $\boldsymbol{\beta}'\mathbf{V}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the vector of net portfolio sensitivities and $\mathbf{V}$ is the covariance matrix of risk factor returns. Both $\mathbf{V}$ and $\boldsymbol{\beta}$ are obtained from variances and covariances, and again the remarks already made about the dangers of using equally weighted measures for these apply. The problem is that covariance matrices, and in particular correlations, are often unstable. This instability may be masked by inappropriate use of equally weighted averages (see section 4.2.1), but when more realistic measures such as EWMA or GARCH are used, the efficient frontier and hence also the optimal capital allocation, may vary significantly as correlations change. This implies that efficient portfolios chosen according to risk-return analysis may need constant rebalancing to remain optimal. Realistic transactions costs should therefore be factored in to this type of capital allocation framework.

4.5.3 Option Pricing and Hedging

There are many basic applications of volatility and correlation to pricing and hedging options. For example:

1. Statistical forecasts of average volatility and/or correlation over the life of an option can be ‘plugged’ into the closed form solution (e.g. into the Black-Scholes formula for vanilla options);

2. Volatilities are used in numerical methods (lattices, or partial differential equations);

3. The covariance matrix is used to simulate correlated risk factor returns for a terminal price distribution (see chapter 5).

---

[24] Provided $\mathbf{V}$ is a positive definite matrix, the portfolio risk will always be positive, whatever the weights in the portfolio.

4. Two factor models, where the underlying price process has stochastic volatility (see Hull, and White,1987)

This section outlines some uses of GARCH models in option pricing and hedging.

First consider an example of the fourth application above: a simulation method for an option based on a single risk factor where a GARCH model is employed in a two factor model with stochastic volatility. To fix ideas, first recall from chapter 5 how simulation can be can be used to price and hedge a call option on an underlying price S(t) which follows the Geometric Brownian Motion diffusion

$$dS(t)/S(t) = \mu dt + \sigma dZ$$

where Z is a Wiener process. Since volatility σ is constant this is a one factor model, and it is only necessary to use Monte Carlo on the independent increments dZ of the Wiener process to generate price paths S(t) over the life of the option. This is done on the discrete form of Geometric Brownian Motion,[25] viz.

$$S_t = S_{t-1} \exp (\mu - 0.5\sigma^2 + \sigma z_t)$$

where $z_t \sim NID(0,1)$. So, starting from the current price S_0 , Monte Carlo simulation of an independent series on z_t for t=1,2,....T will generate a terminal price S_T. Thousands of these terminal prices should be generated starting from the one current price S_0 , and the discounted expectation of the option pay-off function gives the price of the call. For example for a plain vanilla option

$$C(S_0) = \exp(-rT)\ E(\max\{S-K,0\})$$

where r is the risk-free rate, T the option maturity and K is the strike. So from the simulated distribution $S_{T,i}$ (i = 1,...N) of terminal prices we get the estimated call price

[25] To derive this from the continuous form use Ito's lemma on logS and then make time discrete.

$$\hat{C}(S_0) = \exp(-rT)\,(\Sigma_i \max\{S_{T,i} - K, 0\} / N)$$

Simulation deltas and gammas are calculated using finite difference approximations, such as the central differences

$$\delta = [C(S_0+\eta) - C(S_0-\eta)]/2\eta$$

$$\gamma = [C(S_0+\eta) - 2C(S_0) + C(S_0-\eta)]/\eta^2$$

(see Broadie and Glasserman, 1996). In this example, with constant volatility GBM for the simulations, the delta and gamma should be the same as those obtained using the Black-Scholes formula. But in practice, simulation errors can be very large unless the time is taken to run large numbers of simulations for each option price[26].

Now consider how to extend the standard GBM model to a two factor model, where the second factor is GARCH(1,1) stochastic volatility.[27] The two diffusion processes in discrete time are

$$S_t = S_{t-1} \exp(\mu - 0.5\sigma_t^2 + \varepsilon_t)$$

$$\sigma_t^2 = \omega + \alpha\, \varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

where $\varepsilon_t = \sigma_t z_t$. Starting with current price S_0 and unconditional standard deviation σ_0 , an independent set of Monte Carlo simulations on z_t (t = 0,1, ... T) will now generate σ_t at the same time as S_t for t = 1 ,..., T. Note that the simulated price paths are already based on expected volatility levels, so the GARCH delta and gamma hedge ratios do not require additional vega hedging (see Duan, 1993 and Engle and Rosenberg, 1995).[28]

---

[26] Simulation errors are reduced by using correlated random numbers - the variance of delta estimates is reduced when $C(S_0+\eta)$ and $C(S_0-\eta)$ are positively correlated.

[27] Note that moving average methods do not yield stochastic volatility models and so these methods cannot be used to generate option prices and hedge ratios which already account for stochastic volatility.

[28] Note that in a world with stochastic volatility there does not always exist a perfect hedge, so the risk neutral pricing assumption which underlies this method does not, in fact, hold.

GARCH models can also be used to forecast option prices from the volatility smile. The idea is to estimate the GARCH parameters using cross-section data on the market implied smile surface, and then the dynamics of the GARCH model can be used to predict the smile. Initial values for the GARCH model parameters are fixed, and then GARCH option prices obtained, as explained above, for options of different strikes and maturity. These prices are put into the Black-Scholes formula, and the GARCH 'implied' volatility is then backed out of the formula (just as one would do with ordinary market implied volatilities, only this time the GARCH price is used instead of the price observed in the market). Comparison of the GARCH smile surface with the observed market smile surface leads to a refinement of the GARCH model parameters (that is, iteration on the root mean square error between the two smiles) and so the GARCH smile is fitted. It turns out that the GARCH parameters estimated this way are very similar to those obtained from time series data, so using the GARCH smile to predict future smiles leads to sensible results (see Duan, 1996).

4.5.4 Value-at-Risk Models

Value At Risk (VAR) has become central to financial decision making - not just for risk managers and regulators, but for anyone concerned with the actual numbers produced: traders, fund managers, corporate treasuries, accountants etc.. Not only is capital set aside for regulatory compliance on the basis of these measures - such things as trading limits or capital allocation decisions may be set. Providing accurate VAR measures therefore becomes a concern for many.

VAR is the level of loss that is expected to be exceed, in normal market circumstances, if the portfolio is left unmanaged during a fixed holding period. Since we cannot state this level with certainty, we must provide a measure of confidence associated with this figure. For example, a 1% 10-day VAR of $10,000 means that if the portfolio is not managed for 10 days, in normal market circumstances we would be 99% sure that we would not lose more than 10,000$.

A VAR measure depends on two parameters, the holding period h and the significance level, α (0.01 in the above example). A good VAR report should provide a convergent and consistent sequence of VAR measures for every holding period from one day (or even less than a day) to several years. Uncertainties generally increase with time, so the VAR measure will increase as the holding period increases. However the significance level is just a matter of personal choice: Is it relevant to reflect potential losses one day in twenty (α=0.05) or one day in a hundred (α=0.01)? The VAR measure will of course increase with its associated significance level. Current recommendations of the Basle committee are that 1% VAR measures, calculated over a holding period of 10 working days are used to calculate capital reserves for the central banks.

Two of the VAR models in common use ('Covariance' methods for linear portfolios and Monte Carlo methods for options portfolios) require an accurate, positive definite covariance matrix. The 'covariance' method is commonly referred to as the 'RiskMetrics' method, since it was popularized by JP Morgan in their RiskMetrics database. This, and other common methods for calculating VAR are fully described in chapter 6. In this section we only give a brief overview of the two methods that employ covariance matrices.

Denote by ΔP_t the forecast P&L over the next h days and by μ_t and σ^2_t its mean and variance. A mathematical formulation of VAR is that number x such that

$$\mathrm{Prob}(\Delta P_t < -x) = \alpha \ .$$

That is, the VAR is the lower 100α% quantile of the P&L distribution. The covariance method assumes P&Ls are conditionally normally distributed, and in that case we have[29]

$$\mathrm{VAR} = Z_\alpha \, \sigma_t - \mu_t$$

where Z_α denotes the critical value from the standard normal distribution corresponding to the choice of significance level. It is often assumed that μ_t is zero,[30] and Z_α is just looked up in tables, so the sole focus of this method is on σ_t, the standard deviation of forecast P&L over the holding period. This is given by

[29] Apply the standard normal transformation to ΔP_t

$$\sigma^2_t = \mathbf{P'VP}$$

where **V** is the covariance matrix of asset returns (or risk factor returns) over the holding period and **P** is a vector of the current mark-to-market values of the assets (or the current mark-to-market values of the risk factors times their respective sensitivities or cash flows amounts)[31]. Since **V** is the main stochastic part of the model[32] it is important to obtain good estimates of the parameters in **V**, i.e. the variance and covariance forecasts of asset (or factor) returns over the holding period. For example, the model may give zero or negative VAR measures if **V** is not positive definite.[33]

The covariance method is subject to a number of errors, or inappropriate usage. For example, factor models may be mis-specified, sensitivites or covariances may be inaccurately measured, and the assumption of normality may be violated. Also it is only applicable to linear portfolios, and the obvious non-linearities of options pay-offs have prompted the BIS to require different methods for VAR in options portfolios. An example of the inappropriate use of this linear method is given in figure 16. It shows the P&L distribution of a barrier option, where the covariance method would gives the same 1% 1-day VAR whether long or short the position (0.26$ in both cases). A simulation method, such as 'historical' or Monte Carlo simulation would give a much more accurate measure of VAR (0.05$ for the short and 0.63$ for the long).[34]

---

[30] However, this may be an unrealistic assumption for insurance companies, corporates and pension funds, who commonly look at very long holding periods in VAR. In this case, an exogenously determined mean P&L may be used to offset the final VAR figure.

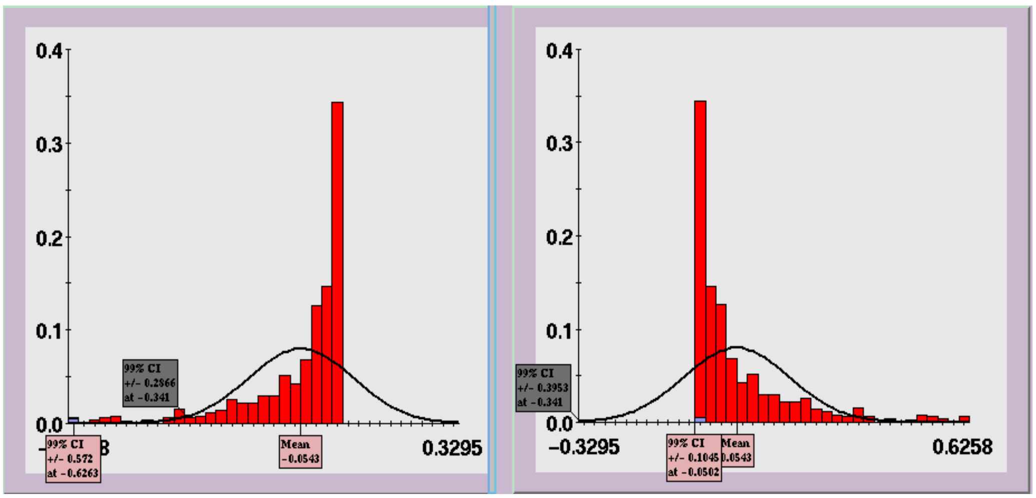
[31] Large portfolios of cash or futures will be represented either as a linear model of risk factor returns, or as a map in cash-flow space. The **P** vector is ($w_1\mathbf{P}_1$, $w_2\mathbf{P}_2$, ….., $w_n\mathbf{P}_n$) where $\mathbf{P}_i$ are the MTM values of the n risk factors (I=1, …n). In the linear factor model w_i is the sensitivity of the portfolio to the ith risk factor. In the cash-flow model the w_i are the amount invested in the ith risk factor.

[32] But the accurate measurement of sensitivites is also crucial, as explained earlier in this section.

[33] The RiskMetrics data are not always positive definite. See Alexander (1996).

[34] Many thanks to Algorithmics Inc for permitting the use of figures 16, 17 and 18.

Figure 16: VAR of a single barrier option using covariance and simulation methods

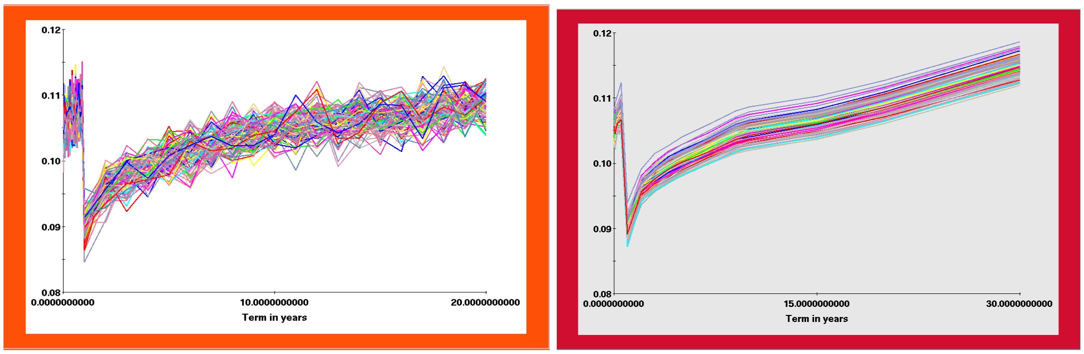


VAR measures for options portfolios may be calculated using Monte Carlo methods to simulate a more realistic distribution of P&Ls over the forthcoming holding period than that which would be obtained using normal approximations. To calculate this distribution, correlated vectors of underlying returns are simulated over the holding period. These are generated by first simulating independent returns to the risk factors and then applying the Cholesky decomposition of the covariance matrix of returns to the underlyings to obtain correlated paths for risk factors[35], as illustrated in figure 17.

---

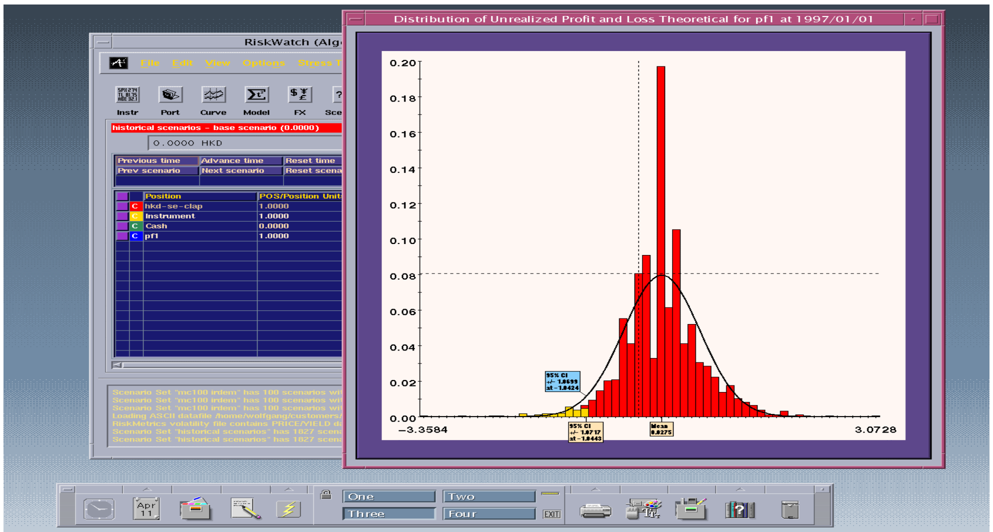
[35] If **r** is a vector of standard independent returns, and **C'C = V** where **C** is the (triangular) Cholesky matrix, then **C'r** is a vector of returns with variance covariance matrix **V**. If the covariance matrix is not positive semi-definite, the Cholesky decomposition does not exist .So what can be done with the non-positive semi-definite matrices, produced by RiskMetrics™? Ad hoc methods - such as shooting the negative or zero eigenvalues to something small and positive - must be resorted to. Unfortunately this changes the original covariance matrix in an arbitrary way, without control over which volatilities and correlations are being affected.

Figure 17: Uncorrelated and correlated yield curve scenarios



Each set of correlated simulations - e,g. each yield curve in figure 17 - gives one value of the portfolio at the end of the holding period. Lots of such curves are simulated and then used to calculate a simulated P&L distribution from which the VAR measure can be read off directly as the lower 100α% quantile as in figure 18.

Figure 18: Simulated P&L distribution showing VAR compared with normal curve VAR



Many financial institutions are using the 'historic simulation' method for computing VAR, which has many advantages - and limitations. Since this method does not employ covariance matrices, it is not discussed here, but is fully described in chapter 6.

The main problem with simulation VAR methods is the time it takes to re-value the portfolio for each simulation of the underlying factors. There are two ways in which computational time can be reduced. Firstly, advanced sampling techniques such as low-discrepancy sequences may be applied to reduce the number of simulation required (see chapter 5). Secondly, a form of portfolio compression may be used: for example delta-gamma approximations could be used instead of full portfolio revaluation at each stage, or the portfolio may be replicated using more simple instruments that are quick to re-value (as in the Algorithmics product RiskWatch).

The current emphasis on integrated risk systems, to measure VAR across all the risk positions of a large bank, requires very large covariance matrices indeed. Regulators currently require the use of at least one year of historic data, so exponentially weighted moving average methods are ruled out.[36] Currently there are only two realistic alternatives: equally weighted moving averages or GARCH. Direct multivariate GARCH models of enterprise-wide dimensions would be computationally impossible. One way that one can construct large dimension covariance matrices using GARCH is to employ the orthogonal GARCH model . However much care should be taken over the calibration of the orthogonal GARCH model, as outlined in section 4.3.7. Equally weighted moving average covariance matrices are easy to construct, and should be positive definite (assuming no linear interpolation of data along a yield curve) but these forecasts will be contaminated by any stress events which may have occurred during the last year, as already explained in section 4.2.1. Therefore the stress events should be filtered out of the data before the moving average is applied, and saved for later use when investigating the effect of real stress scenarios on portfolio P&Ls.

4.6 Special Issues and New Directions

4.6.1 Evaluating Volatility and Correlation Forecasts

It is not unusual to read statements such as '…we employ fractionally integrated GARCH volatilities because they are more accurate'. But the accuracy of volatility forecasts is extremely difficult to assess. Results will depend on the method of evaluation employed, so no single answer can be given to the question 'which method is more accurate?' Assessing the accuracy of correlation forecasts is an even more difficult problem, not only because correlations can be so unstable, but also because a proper statistical evaluation requires the full multivariate distribution to be known - which it is not.

If one decides to embark on the treacherous path of forecast evaluation, the first point to note is that different models should be ranked according to how they perform according to a certain

---

[36] Unless one tailors the tolerance level to the smoothing constant. For example, a smoothing constant of 0.9 would mean that 250 days of past data is taken into account for the EWMA if the tolerance level is 10^{-12}.

benchmark. But problems arise when applying this type of ranking: unlike prices, volatility is unobservable, so what should the volatility benchmark be? I would consider taking the unconditional standard deviation over a very long data period as the benchmark against which to test different volatility forecasting models - if a model cannot improve upon this simplest of models there is no point in using any greater degree of sophistication!

The next problem along the route of evaluating a volatility model is ‘what do we mean by 'better''? Is it preferable to use a statistical evaluation criterion or to assess models purely on a P&L basis, when possible? Much research has been published in this area, mostly on statistical evaluation, because this has a greater degree of subjectivity.[37] P&L evaluations are based on specific trading metrics so this type of operational evaluation is quite objective - and it is difficult to come to any broad conclusions.

If statistical evaluation is used it should be emphasised that however well a model fits *in-sample* (i.e. within the data period used to estimate the model parameters) the real test of its forecasting power is in *out-of-sample* predictive tests. A certain amount of historic data should be withheld from the period used to estimate the model, and then forecasts made from the model may then be evaluated by comparing them to the out-of-sample returns data. The appropriate statistical criteria to use is the maximization of out-of-sample likelihoods, but the effectiveness of this method relies on correct specification of the returns distributions. Some of the literature uses *root mean square errors* to evaluate volatility and correlation forecasts, since this is a simple distance metric, not based on a statistical distribution of returns which in many cases is erroneously assumed to be normal. [38]

---

[37] Alexander and Leigh, 1997, Brailsford and Faff (1996), Dimson and Marsh (1990), Figlewski (1994), Tse and Tung (1992), and West and Cho (1995).

[38] As a statistical criterion the root mean square error criterion comes from a normal likelihood function - minimising the sum of squared differences between the actual observations and the predicted means will maximise the normal likelihood when volatility is constant (see (27)). Hence root mean square errors are applicable to *mean* predictions rather than variance or covariance predictions. Of course, a variance is a mean, but the mean of the *squared* random variable, which is chi-squared distributed, not normally distributed, so the likelihood function is totally different and does not involve any sum of squared errors

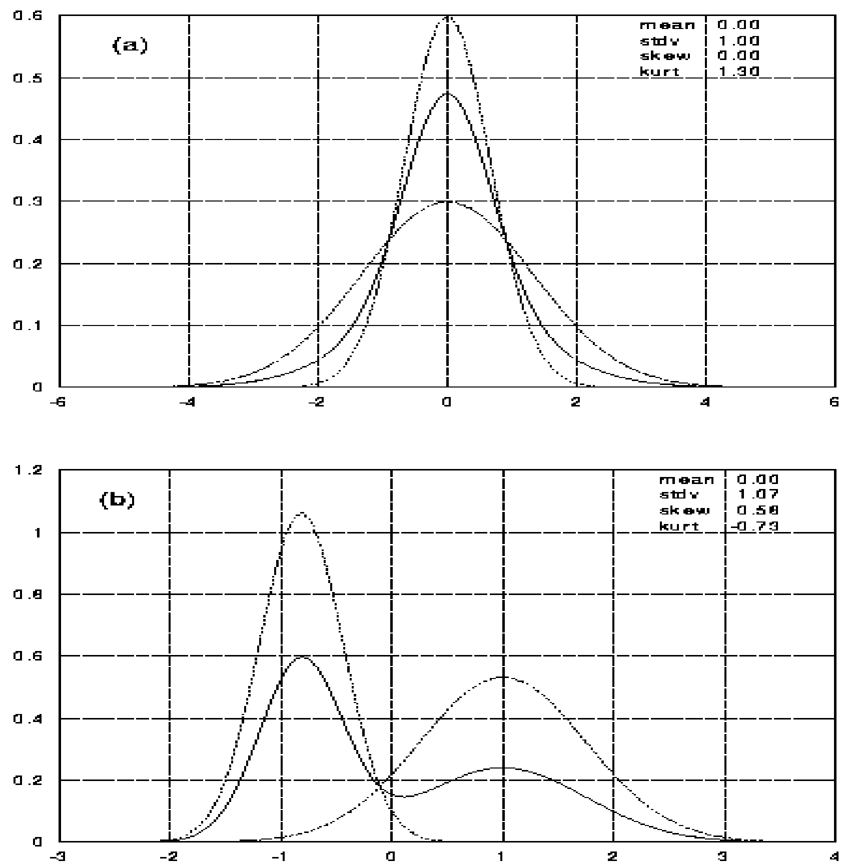
Alexander and Leigh (1997) look at both statistical and operational evaluation of the three types of volatility forecasts which are in standard use: equally weighted moving averages, exponentially weighted moving averages and GARCH. Given the remarks just made, it is impossible to draw any firm conclusions about the relative effectiveness of any volatility forecasting method for an arbitrary portfolio. But, using data from the major equity indices and foreign exchange rates during 1996 it appears that whilst EWMA methods perform well for predicting the 'centre' of a normal distribution the GARCH and equally weighted moving average methods are generally more accurate for the 'tails' prediction required by VAR models.

4.6.2 Modelling 'Fat-Tailed' Distributions

One way of retaining normality assumptions whilst allowing for 'fat-tails' is to model returns distributions using mixtures of normals. Mixtures of normals with different variances generate leptokurtic distributions and if means are also allowed to vary this will capture the skewness or multi-modal nature of some portfolio returns distributions (see figure 19).[39]

[39] Many thanks to Peter Williams of Sussex University for providing these figures

Figure 19: Mixtures of normal distributions



Hull and White (1997, 1998) have developed a model for estimating mixtures of two normal distributions with constant parameters. Their means are both assumed to be zero, and their standard deviations are $u\sigma_t$ and $v\sigma_t$ where $u < 1$ and $v > 1$ and σ_t is the standard deviation of market returns at time t. In the empirical work of Hull and White, on 12 major USD exchange rates between 1988 and 1994, σ_t is either assumed constant or an exponentially weighted moving average estimate of σ_t is obtained (for example, using RiskMetrics data). The proportions of the two normals in the mixture are assumed constant through time, denoted p and 1-p respectively, so that

$$pu^2 + (1-p)v^2 = 1 \qquad (35)$$

Hull and White propose a maximum likelihood method to estimate the model parameters which are consistent with (35). Direct maximization of the likelihood of the normal mixture is not robust to outliers, so the empirical data are divided into four categories: the first category consists of returns which are less than one standard deviation, the second category contains returns which are between one and two standard deviations, the third category of returns is between three and four standard deviations and the fourth is for those returns more than four standard deviations. A maximum likelihood method is employed where p, u and v are chosen to match the observed number of returns in each category with the predicted number of returns from the normal mixture in that category. Results are tested out-of-sample using standard diagnostics. The model fits better when EWMA rather than constant volatilities are assumed, and when parameters are estimated for individual exchange rates rather than using the same parameters for all. It has natural extensions to modelling correlations and the calculation of VAR using simulation techniques. Further details are given in Hull and White (1998).

Another robust statistical technique for estimating mixtures of normals is with a neural network which trains on the whole likelihood function rather than the simple error surface used in standard neural networks packages. Many neural networks output only a single vector, and back-propagation is then used to minimise the root mean square error of the output, compared with the actual data in the training set. But minimizing a root mean square error is equivalent to maximum likelihood estimation of the mean of a single normal distribution with constant variance, and is therefore very restrictive. Packaged neural networks also have the reputation of fitting well in sample, but predicting poorly out-of-sample.[40] This can be due to over-fitting the data: a neural network is a universal approximator, and so needs to be constrained in its complexity if sufficient flexibility is to be retained for generating future scenarios.

A method for constraining complexity which is more advanced than simply cutting off the network after a certain point, is to use a penalized likelihood function where the value of the in-sample likelihood is offset by an amount proportional to the number of weights and connections in

[40] See the Proceeding of the First International Non-linear Financial Forecasting Competition (Caldwell, 1997) , for example.

the network. In Williams, (1995, 1996) a network is trained by maximizing the likelihood function of a normal mixture. The training is automatically regularized to avoid over-fitting, by using a penalized likelihood function to estimate weights and biases. Simulation of the conditional returns distributions from such a neural network then produces term structure forecasts of all the moments of the normal mixture - including the variance, skewness and kurtosis. Alexander and Williams, (1997) show that for US Dollar exchange rates, kurtosis term structures generated by this neural network are much closer to empirically expected levels than the GARCH kurtosis term structures, although their volatility term structures are similar.

4.6.3 Downside Risk and Expected Maximum Loss

One of the problems with using variance to measure 'risk' is that it does not distinguish between upside and downside risk - that is, positive returns and negative returns are regarded as equally unfavourable. Short-term hedging based on mean-variance analysis can lead to very skewed distributions of longer term returns without frequent re-balancing (see section 4.6.4) and for these portfolios the standard variance operator is totally inadequate. A risk measure which focuses purely on downside risk, proposed by Markovitz (1959) is the semi-variance

$$SV = E(\min(0, r-E(r))^2)$$

When returns are greater than average min(0, r-E(r)) = 0, so only below average returns are counted in this risk measure. Many software and data providers are now incorporating more flexibility in their risk measures (see chapter 2), including some measure of downside risk such as semi-variance.

Algorithmics Inc. have patented an optimizer based on 'regret', a well known concept from the decision sciences which is similar to semi-variance. Returns are measured relative to a benchmark B, which can be measured empirically or simulated based on user requirements. The regret of a portfolio is given by

$$\text{Regret} = -E(\min(0, r-B))$$

Portfolios which appear to have similar risk attributes when standard variance or VAR is employed can be distinguished by their regret - if variance is concentrated on the 'upside' the regret is small, but 'bad' portfolios with distributions which are skewed on the downside will have large regret. The use of the benchmark also has attractive applications in many areas of allocation and risk measurement - see chapter 7 for more details.

Finally, a brief mention of a risk measure which focuses on the actual decisions faced by risk managers. VAR only gives the frequency with which a fixed loss (or worse) should occur, but risk managers are more likely to be concerned with the nature of losses - are they consecutive small losses or intermittent large ones? What is the maximum loss I can expect to incur from this portfolio during a given holding period? Both these questions are addressed with a risk measure called 'expected maximum loss' (Acar and Prieul, 1997). By utilizing more information about the dynamic nature of returns, this risk measure can be usefully employed in the simultaneous allocation of capital and setting of trading limits.

4.6.4 Cointegration

Cointegration has had an astounding impact on applied economic modelling during the last decade. Since the seminal work of Engle and Granger (1996), the many thousands of papers published on the theory and application of cointegration to economics bears testimony to the huge resources now devoted to this approach. Every modern econometrics text book will contain details of the statistical theory necessary to master the application of cointegration to time series data, Hamilton (1996) and Feller (1996) being two excellent sources.

However, the financial community has been less aware of the potential for cointegration modelling. Traditionally, a portfolio risk analysis takes as its starting point the distribution of financial returns, so the log price, rate or yield data are differenced before the analysis is even begun. This removes any long term trends in the price, rate or yield data and, although these trends are implicit in the returns data, the analysis of any common trends (cointegration) in the raw data is often precluded.

Cointegration and correlation are related, but different concepts. Correlation measures co-movements in returns, so it is intrinsically a short run measure, liable to instabilities over time since returns typically have low autocorrelation. Cointegration measures long run co-movements in prices (rates or yields) and high correlation of returns does not necessarily imply high cointegration in prices. An example is given in figure 20, with a 10 year daily series on US dollar spot exchange rates of the German Mark (DEM) and the Dutch Guilder (NLG) from, say 1975 to 1985. Their returns are very highly correlated: the correlation coefficient is approximately 0.98 (figure 20(a)). So also do the rates move together over long periods of time, and they appear to be cointegrated (figure 20(b)). Now suppose that we add a very small daily incremental return of, say, 0.0002 to NLG. The returns on this NLG 'plus' and the DEM still have correlation of 0.98, but the price series diverge more and more as time goes on - they are not cointegrated (figure 20(c)).[41]

[41] Many thanks for Wayne Weddington of Pennoyer Capital Management for providing these figures

Figure 20(a): NLG and DEM returns

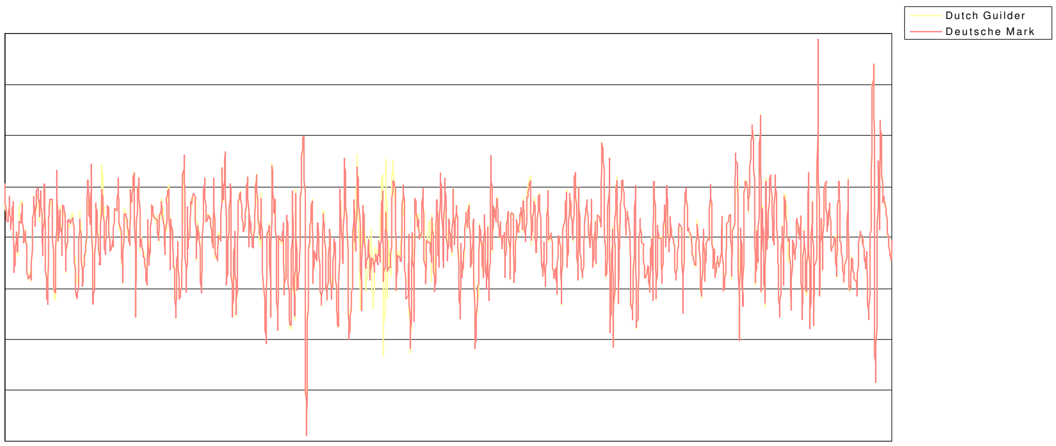


Figure 20(b): NLG and DEM rates

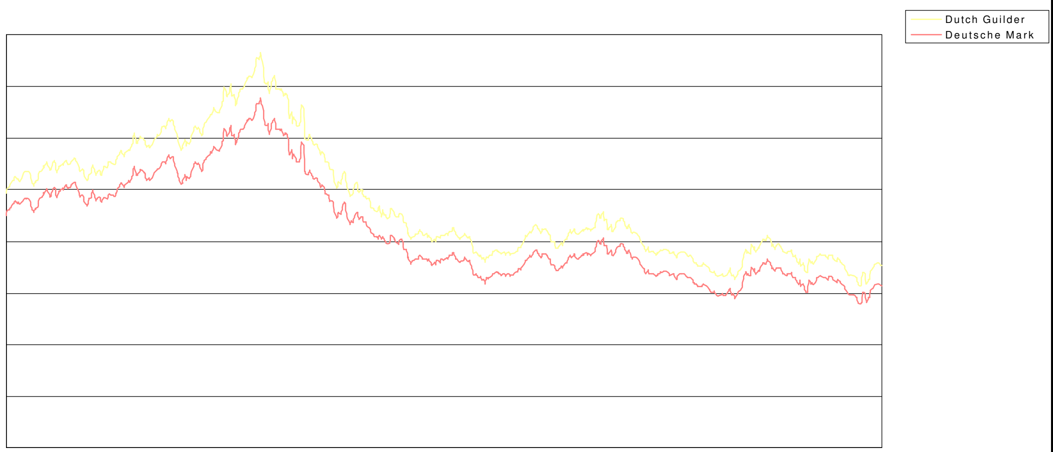
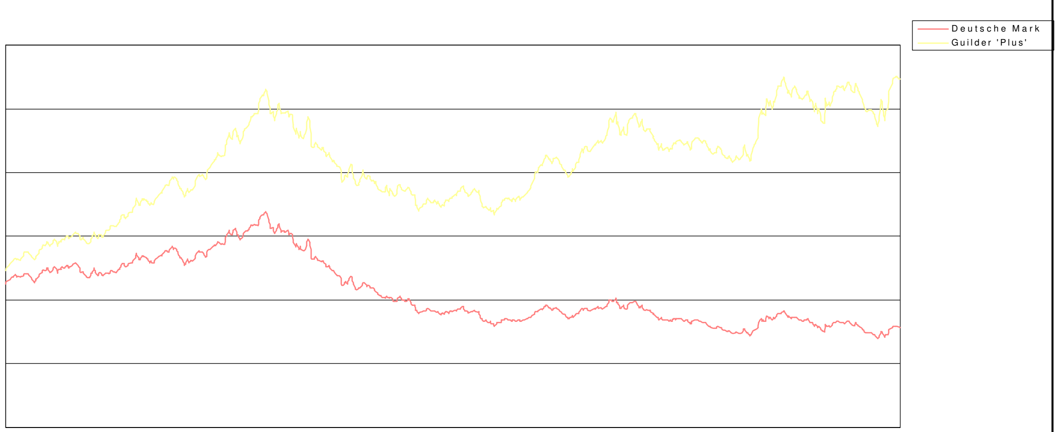


Figure 20(c): NLG 'Plus' vs. DEM



To introduce cointegration we first define the notion of an *integrated* series. An example of an integrated series is a random walk

$$y_t = \alpha + y_{t-1} + \varepsilon_t$$

where α denotes the drift and the error process ε_t is an independent and identically distributed (i.i.d) stationary process.[42] If financial markets are efficient, then prices (or log prices) rates or yields are all random walks. In the random walk model we could take y_t= log P_t, where P_t is a financial price series, and in this case ε_t denotes the return at time t.

Now if returns are stationary i.i.d processes, then log prices are random walks, but what if returns are not i.i.d, because they are not independent - there is some autocorrelation in returns? Provided returns are still stationary, log prices will be an integrated series of order 1. A time series is integrated of order 1, denoted I(1) if the stochastic part is non-stationary, but it is stationary after first differencing.[43]

Two or more integrated series are termed 'cointegrated' if there is a weighted sum of these series (called the *cointegrating vector*) which is stationary. Consider a simple case: x and y are random walk, but x-y is stationary. This means that x and y cannot drift too far apart, because x-y is stationary, mean-reverting. So cointegrated series are 'tied together' by the cointegrating vector in the long run, even though they can drift apart in the short run.

The mechanism which ties cointegrated series together is a 'causality' - not in the sense that if we make a structural change to one series the other will change too, but in the sense that turning points on one series precede turning point in the other. This is called *Granger causality* (see

---

[42] A (covariance) stationary series has constant, finite mean and variance, and an autocorrelation which depends only on the lag. The constant, finite variance implies that stationary series are 'mean-reverting'. Typically, financial returns data are stationary. We use the notation I(n) to denote a series which is non-stationary, and only stationary after differencing a minimum of n times. Stationary series are denoted I(0), random walks are I(1).

[43] Series with deterministic trends are not integrated, they are I(0)+trend. If a deterministic trend is removed from a random walk, the result is another random walk, without a drift term. Thus standard technical analysis methods of 'detrending' data by fitting a line are not removing the stochastic trend - the result will be a random walk and not a mean-reverting series.

Engle and Granger, 1987). The strength and directions of Granger causality can change over time, there can be bi-directional causality, or the direction of causality can change. For example in the 'price discovery' relationship between spot and futures there may be times when futures lead spot, whereas at other time spot prices can lead futures prices. The necessity for causality between cointegrated series is revealed by the *error correction model*, which is a dynamic model of returns where deviations from the long-run equilibrium are corrected. More details may be found in Alexander and Johnson (1992, 1994) and in any standard econometrics text.

Consider two cointegrated series. Each is individually a random walk - which means that given enough time they could be anywhere (because they have infinite unconditional variance) and the best forecast of any future value is the value today. But since they are cointegrated we know that, wherever one series is in x years time, the other series will be right there along with it.

When only two integrated series are considered for cointegration, there can be at most one cointegrating vector, because if there were two cointegrating vectors the original series would have to be stationary. More generally, taking n series in cointegration analysis, the maximum number of cointegrating vectors is n-1. Each cointegrating vector is a 'bond' in the system, and so the more cointegrating vectors found the more the coherence and co-movements in the prices. Yield curves have very high cointegration - if each of the n-1 independent spreads is mean-reverting, there are n-1 cointegrating vectors.

Examples of cointegrated series abound in financial markets:

- Spot and futures prices: If spot and futures prices are tied together, the basis can never get too large - it will be mean-reverting. The basis is the cointegrating vector (see Beck, 1994, Schwarz and Szakmary, 1994)

- Yields of different maturities: Yields are random walks - they could be anywhere given enough time. But wherever the 1mth yield is in x years time, the 3mth yield will be right along there with it, because the spread can never get too large. The spreads are mean-reverting, so they are the cointegrating vectors (see Bradley and Lumpkin, 1992, Hall et. Al., 1992)

- Any pair of series with a mean-reverting spread: For example two different equity indices (or bond indices) or pairs of bond and equity indices in the same country can be cointegrated, but they are not always so. It depends on the time series properties of the spread (see Alexander, 1995, Alexander and Thillainathan, 1996, Clare et. al., 1995)

- Related commodities: Carry costs represent the difference in prices between related commodities. If carry costs are mean-reverting, then commodities based on the same underlying market will be cointegrated (see Brenner and Kroner, 1995.)

- Equities within an index: since the index is by definition a weighted sum of the constituents, there should be some sufficiently large basket which is cointegrated with the index. The cointegrating vector is the tracking error, between that basket and the index. Baskets found by cointegration will have mean-reverting tracking errors, so the basket cannot drift away from the index, they are 'tied together' in the long run (see Alexander, Giblin and Weddington, 1998)

4.7 Conclusion

This chapter has covered a very wide subject area, which is of central importance for effective measurement of market risks. It is a very active area for both practical and academic research, and in places the chapter gives only a brief indication of the key issues, but with supporting references for further reading whenever possible.

The first part of the chapter provides a critical overview of some of the most commonly used methods for estimating volatility and correlation: moving averages and GARCH models. Common pitfalls in the use of these models are explained, and the chapter should provide the reader with enough information to employ these statistical models appropriately for market risk measurement.

After a short discussion of implied volatility and correlation, its calculation and use in trading and risk management, the chapter has focused on the main applications of volatility and correlation for market risk systems. Special issues, such as the use of GARCH models in option pricing and smile fitting, and the proper implementation of covariance matrices in Value at risk models have been explored. The chapter closes with an overview of some of the key areas of current research: in particular we have examined the use of alternative risk measures such as regret and cointegration for market risk analysis.

Acknowledgements

I would like to thank Dr Aubrey Chibumba of Citibank, UK, Dr Ian Giblin of Pennoyer Capital Management, US, Christopher Leigh of Robert Flemings, UK, Wayne Weddington of Pennoyer Capital Management, US and Dr Peter Williams of the University of Sussex, UK. Also a huge thank you to Dr Ron Dembo of Algorothmics Inc, from whose vision I have benefited enormously, and to Professor John Hull of University of Toronto, CA, for his encouragement and support.

Bibliography

Acar, E. and D. Prieul (1997) “Expected maximum loss” *NetExposure* 1 (www.netexposure.co.uk)

Alexander, C. (1994) "History Debunked" *RISK* 7 No. 12 pp59-63.

Alexander, C. (1995) "Common volatility in the foreign exchange market" *Applied Financial Economics* 5 No. 1 pp 1-10

Alexander, C. (1996) “Evaluating RiskMetrics as a risk measurement tool for your operation: What are its advantages and limitations?” *Derivatives: Use, Trading and Regulation* 2 No. 3 pp277-285.

Alexander, C. (1996) Ed. The Handbook of Risk Management and Analysis *John Wileys, Chichester, UK.*

Alexander, C. (1997a) “Splicing methods for generating large covariance matrices*”* Learning curve of Derivatives Week Volume 3 *Institutional Investor*

Alexander, C. (1997b) “Estimating and forecasting volatility and correlation: methods and applications” in Risk Management and Financial Derivatives: A Guide to the Mathematics’ (S. Das, Ed) pp 337 -354 LBC Publications

Alexander, C. and A. Johnson (1992) "Are foreign exchange markets really efficient?" Economics Letters 40 (1992) 449-453

Alexander, C. and N. Riyait (1992) "The world according to GARCH" RISK 5:8 pp120-125

Alexander, C. and A. Johnson (1994) "Dynamic Links" *RISK* 7 no. 2 pp 56-61

Alexander, C. and R. Thillainathan (1996) "The Asian Connections" Emerging Markets Investor, 2 No. 6 pp42-47

Alexander, C. and C. Leigh (1997) ‘On the Covariance Matrices used in VAR Models*’ Journal of Derivatives* 4:3 pp50-62

Alexander C. and I. Giblin (1997) "Multivariate embedding methods: Forecasting high-frequency data in the first INFFC" *Journal of Computational Intelligence in Finance* Vol5 No 6 pp17-24

Alexander, C. and A. Chibumba (1997) ‘Orthogonal GARCH: An Empirical Validation in Equities, Foreign Exchange and Interest Rates’ *Mimeo*

Alexander, C. and P. Williams (1997) 'Modelling the Term Structure of Kurtosis: A comparison of neural network and GARCH methods' *Forthcoming in RISK* August 1998

Alexander, C., I. Giblin and W. Weddington (1998) "Cointegration in international equity markets: Application to tracking EAFE" *In Prep*

Andersen, T.G. (1994) "Stochastic autoregressive volatility: A framework for volatility modelling" *Mathematical Finance* 4 vol2 pp75-102

Baillie, R.T. and T. Bollerslev (1989) "The message in daily exchange rates: a conditional-variance tale". *Journal of Business and Economic Statistics* 7 No. 3 pp 297-305.

Baillie, R.T. and T. Bollerslev (1990)'Intra-day and inter-market volatility in foreign exchange rates' *Review of Economic Studies* , 58, pp 567-585

Beck, S.E. (1994) "Cointegration and market efficiency in commodities futures markets" *Applied Economics* 26:3, pp. 249-57

Bollerselv, T (1988) 'On the correlation structure for the generalised autoregressive conditional heteroskedastic process' *Journal of Time Series Analysis* , 9, pp 121-131

Bollerslev, T (1986) 'Generalised autoregressive conditional heteroskedasticy' *Journal of Econometrics* , 31, pp307-327

Bollerslev, T (1987) 'A conditional heteroskedasticity time series model for security prices and rates of return data' *Review of Economics and Statistics* , 69, pp 542-547

Bollerslev, T, R.Y. Chou and K.F. Kroner (1992) "ARCH modeling in finance". *Journal of Econometrics* 52 pp 5-59.

Bollerslev, T, RF Engle and D Nelson (1994) "ARCH models" in *Handbook of Econometrics* volume 4 (Edited by R.F. Engle and D.L. McFadden , North Holland) pp2959-3038

Bradley, M. and S. Lumpkin (1992) "The Treasury yield curve as a cointegrated system" *Journal of Financial and Quantitative Analysis* 27 (1992) pp 449-63.

Brailsford T.J. and Faff R.W. (1996) " An evaluation of volatility forecasting techniques" *Journal of Banking and Finance* 20 pp419-438

Brenner, R.J. and K.F. Kroner (1995) "Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets" *Journal of Financial and Quantitative Analysis* 30:1, March 1995, pp. 23-42

Brenner, R.J., R.H. Harjes and K.F. Kroner (1996) “Another look at alternative models of the short term interest rate” *Journal of Financial and Quantitative Analysis* 31:1 pp85-108

Cerchi, M. and A. Havenner (1988) "Cointegration and stock prices". *Journal of Economic Dynamics and Control* 12 pp 333-46.

Chew, L. (1993) "Summer of content" *RISK* 6 No.8 pp28-35

Chowdhury, A.R. (1991) "Futures market efficiency: evidence from cointegration test" *The Journal of Futures Markets* 11:5 pp 577-89. John Wiley and Sons Inc., 1991.

Clare, A.D., M. Maras and S.H. Thomas (1995) "The integration and efficiency of international bond markets" *Journal of Business Finance and Accounting* 22:2, March 1995, pp. 313-22.

Dembo, R (1997) ‘Value at Risk and return’ *NetExposure* 1 (www.netexposure.co.uk)

Dickey, D.A. and W.A. Fuller (1979) "Distribution of the estimates for autoregressive time series with a unit root" *Journal of the American Statistical Association* 74 pp. 427-9

Dimson, E. and P. Marsh (1990) “Volatility forecasting without data-snooping” *Journal of Banking and Finance* 14 pp399-421

Duan, J.C. (1993) ‘The GARCH option pricing model’ *Mathematical Finance* , 4:2

Duan, JC (1996) ‘Cracking the smile’ *RISK* 9 pp 55- 59

Dupire, B. (1997) ‘A unified theory of volatility’ *Mimeo*

Dwyer, G.P. and M.S. Wallace (1992) "Cointegration and market efficiency" *Journal of International Money and Finance* 11 (1992), 318-327.

Engle, R.F. (1982) "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". *Econometrica* 50:4 pp 987-1007.

Engle, R.F. , Ng, V. and Rothschild, M. (1990) “Asset pricing with a factor ARCH covariance structure : Empirical estimates for treasury bills” *Journal of Econometrics* , 45 pp 213-238

Engle, R.F. and C.W.J. Granger (1987) "Co-integration and error correction: representation, estimation, and testing". *Econometrica* 55:2 pp 251-76.

Engle, R.F. and G. Lee (1993) ‘A permanent and transitory component model of stock return volatility’ *Journal of Econometrics*

Engle, R.F. and K.F. Kroner (1995) "Multivariate Simultaneous Generalized ARCH" *Econometric Theory* 11 pp122-150

Engle, R.F. and V. Ng (1993) ‘Measuring and testing the impact of news on volatility’ *Journal of Finance* , 48 pp 1749-1778

Engle, R.F. ed (1995) ‘ARCH : Selected readings’ Advanced texts in econometrics , *Oxford University Press*

Engle, RF and J Mezrich (1995) "Grappling with GARCH" *RISK* 8 No9 pp 112-117

Figlewski, S. (1994) "Forecasting volatility using historical data" *New York University Salomon Center (Leonard N. Stern School of Business) Working Paper Series* no. S-94-13

Hall, A.D., H.M. Anderson and C.W.J. Granger (1992) "A cointegration analysis of Treasury bill yields". *The Review of Economics and Statistics* pp 116-26.

Heynen, R., A. Kemna and T. Vorst (1994) "Analysis of the term structure of inplied volatilities" *Journal of Financial and Quantitative Analysis* 29 vol1 pp31-56

Ho, T., M. Chen and F. Eng (1996) ‘VAR Analytics: Portfolio Structure, Key Rate Convexities and VAR Betas’ *Journal of Portfolio Management* Vol 23 no3 pp89-98

Hull, J. and W White (1987) "The pricing of options on assets with stochastic volatilities". *Journal of Finance* 42, 281-300.

Hull, J and A. White (1997) ‘Evaluating the Impact of Skewness and Kurtosis on Derivatives Prices’ *NetExposure* 3 (www.netexposure.co.uk)

Hull, J and A. White (1998) ‘Taking account of the kurtosis in market variables when calculating VAR’ Forthcoming in *Journal of Derivatives*

Johansen, S. and K. Juselius (1990) "Maximum likelihood estimation and inference on cointegration - with applications to the demand for money". *Oxford Bulletin of Economics and Statistics* 52:2 pp 169-210.

Jorion, P. (1996) Value at Risk: The New Benchmark for Controlling Market Risk *Irwin*

Jorion. P. (1996) ‘Risk2 : Measuring the Risk in VAR’ *Financial Analysts Journal*

Kelly M. (1994) "Stock Answer" *RISK* 7 No. 8 pp40-43

Kroner, K.F. and J. Sultan (1991) "Exchange rate volatility and time varying hedge ratios". *Pacific-Basin Capital Markets Research* II pp 397-412.

Kroner, K.F. and S. Claessens (1991) "Optimal dynamic hedging portfolios and the currency composition of external debt". *Journal of International Money and Finance* 10 pp 131-48.

Kroner, K.F., D.P. Kneafsey and S. Claessens (1993) "Forecasting volatility in commoditiy markets" *World Bank Policy Research Paper 1226.*

Lumsdaine, R.L. (1995) "Finite sample properties of the maximum likelihood estimation in GARCH(1,1) and IGARCH(1,1) models: a Monte Carlo investigation" *Journal of Business and Economic Statistics* 13 vol1 pp1-10

Magdon-Ismail, M. and Y.S. Abu-Mostafa (1996) "Validation of volatility models" *Caltech discussion paper*

Nelson, D B (1991) "Conditional heteroskedasticity in asset returns: a new approach". *Econometrica* 59 No.2 pp 347-70.

Schwarz, T.V. and A.C. Szakmary (1994) "Price discovery in petroleum markets: arbitrage, cointegration, and the time interval of analysis". *Journal of Fututres Markets* 14:2, pp 147-167.

Taylor, S.J. (1994) "Modelling stochastic volatility" *Mathematical Finance* 4 No.2 pp183-204.

Tse, Y.K. and S.H. Tung (1992) "Forecasting volatility in the Singapore stock market" *Asia Pacific Journal of Management* 9, pp1-13

Wang, G.H.K. and J. Yau (1994) "A time series approach to testing for market linkage: Unit root and cointegration tests" *Journal of Futures Markets* 14:4, June 1994, pp. 457-74

West, K. D. and D. Cho (1995) "The predictive ability of several models of exchange rate volatility" *Journal of Econometrics* 69 pp367-391

Williams, P. (1995) "Bayesian regularization and pruning using a Laplace prior" *Neural Computation* 7 pp117-143

Williams, P. (1996) "Using neural networks to model conditional multivariate densities" *Neural Computation* 8 pp?