Key Market Risk Factors: Identification and Applications

The Q-Group Seminar on Risk,

Florida, 2-5 April 2000

Carol Alexander[1]

During the last few years there have been many changes in the way that financial institutions model risk. New risk capital regulations have motivated a need for vertically integrated risk systems based on a unified framework throughout the whole office. If the risk exposures in all locations of a large institution are to be aggregated, the risk system must also be horizontally integrated. Internationally, regulators are pushing towards an environment where traders, quants and risk managers from all offices are referring to risk measures generated by the same models. This is a huge task which remains a challenge for many financial institutions, but the result should be useful to manage risks for allocation of capital between different areas of the firm, and to set traders limits as well as levels of capital reserves.

Following the Basle Accord Amendment in 1996 for the calculation of market risk capital using internal models, the Basle Committee on Banking Supervision (1995) have recommended two methods for generating a unified set of risk measures on a daily basis. These methods have become industry standards for measuring risk not only for external regulatory purposes, but also for internal risk management. The first approach is to calculate a Value-at-Risk (VaR) measure, which is a lower percentile of an unrealized profit and loss distribution. This distribution is based on movements of the market risk factors over a fixed risk horizon. The second approach is to quantify the maximum loss over a large set of scenarios for movements in the risk factors.

[1] Professor of Finance, Chair of the Risk Management Group, ISMA Centre, Reading University, UK

Given the huge number of market risk factors affecting the positions of a large financial institution, the VaR models and scenario-based loss models may become very complex indeed. In fact their implementation becomes extraordinarily cumbersome, if not impossible, without making assumptions that restrict the possibilities for movements in the risk factors. For example, at the heart of most risk models there is a covariance matrix that captures the volatilities and correlations between the risk factors. Typically hundreds of risk factors, such as all yield curves, interest rates, equity indices, foreign exchange rates and commodity prices, need to be encompassed by a very large dimensional covariance matrix. It is not easy to generate this matrix and so simplifying assumptions may be necessary. For example the RiskMetrics methodologies designed by JP Morgan use either simple equally weighted moving averages, or exponentially weighted moving averages with the same smoothing constant for all volatilities and correlations of returns. There are substantial limitations with both of these methods, described in Alexander (1996).

Another example of how the standard methods necessitate simplifying assumptions is in maximum loss calculations. The applicability of maximum loss measures depends on portfolio revaluation over all possible scenarios, including movements in both prices and implied volatilities of all risk factors. In complex portfolios the computational burden of full revaluation over thousands of scenarios would be absolutely enormous, and certainly not possible to achieve within an acceptable time frame unless analytic price approximations and advanced sampling techniques are employed in conjunction with a restriction of the possibility set for scenarios.

The problems outlined in both of the above examples have a common root: the computations, be they volatility and correlation calculations for a covariance matrix, or portfolio revaluation for the calculation of maximum loss, are being applied to the full set of risk factors. So the dimensions of the problem become too large to manage and the problem is intractable. But there is an alternative: to apply computations to only a few key market risk factors that capture the most important independent sources of information in the data. Such an approach is computationally efficient because it allows an enormous reduction in the dimension of the problem whilst retaining a very high degree of accuracy. Because the risk factors are independent it does not significantly increase the computational complexity even if a large number of key risk factors are employed. Normally a sufficient number of key risk factors will be generated so that any movements that are not captured by these factors are deemed to be insignificant 'noise' in the system, and by cutting out this noise the risk measures will become more stable and robust over time. Also, being able to quantify how much risk is associated with each key factor is an enormous advantage for risk managers, because their attention is more easily directed towards the most important sources of risk.

The method used here to identify key independent sources of risk within a large system is principal component analysis. Jamshidian and Zhu (1996) have shown how principal components may be used to improve computational efficiency for scenario based risk measures in large multi-currency portfolios. This paper extends these ideas to two other important areas: firstly the efficient computation of large positive semi-definite covariance matrices, and secondly the modelling of multivariate scenarios for the whole implied volatility smile surface as the underlying prices move. Both problems have immediate applications to internal models for measuring market risk.

Identification of the Key Risk Factors

Suppose a set of data with T observations on k asset or risk factor returns is summarized in a Txk matrix **Y**. Principal component analysis will give up to k uncorrelated stationary variables, called the principal components of **Y**, each component being a simple linear combination of the original returns as in (1) below. At the same time it is stated exactly how much of the total variation in the original system of risk factors is explained by each principal component, and the components are ordered according to the amount of variation they explain.

The first step in principal component analysis is to normalize the data in a Txk matrix **X** that represents the same variables as **Y**, but in **X** each column is standardized to have mean zero and variance 1. So if the ith risk factor or asset return in the system is $\mathbf{y}_i$, then the normalized variables are $\mathbf{x}_i = (\mathbf{y}_i - \mu_i)/\sigma_i$ where μ_i and σ_i are the mean and standard deviation of $\mathbf{y}_i$ for i = 1, … k. Now let **W** be the matrix of eigenvectors of **X'X**, and $\mathbf{\Lambda}$ be the associated diagonal matrix of eigenvalues, ordered according to decreasing magnitude of eigenvalue.[2] The principal components of **Y** are given by the Txk matrix

$$\mathbf{P} = \mathbf{XW} \tag{1}$$

Thus a linear transformation of the original risk factor returns has been made in such a way that the transformed risk factors are orthogonal, that is, they have zero correlation.[3]

The new risk factors are ordered by the amount of the variation they explain.[4] Hence only the first few, the most important factors may be chosen to represent the system as follows: Since **W** is orthogonal (1) is equivalent to **X = PW'**, that is

$$\mathbf{x}_i = w_{i1}\,\mathbf{p}_1 + w_{i2}\,\mathbf{p}_2 + \ldots\ldots + w_{ik}\,\mathbf{p}_k \tag{2}$$

so the matrix **W** is called the matrix of 'factor weights'. In terms of the original variables **Y** the representation (2) is equivalent to

$$\mathbf{y}_i = \mu_i + \omega^*_{i1}\,\mathbf{p}_1 + \omega^*_{i2}\,\mathbf{p}_2 + \ldots\ldots + \omega^*_{im}\,\mathbf{p}_m + \boldsymbol{\varepsilon}_i \tag{3}$$

where $\omega^*_{ij} = w_{ij}\sigma_i$ and the error term in (3) picks up the approximation from using only the first m of the k principal components. These m principal components are the 'key' risk factors of the system, and the rest of the variation is ascribed to 'noise' in the

---

[2] Thus $\mathbf{X'X\,W} = \mathbf{W\,\Lambda}$.

[3] Note that $\mathbf{P'P} = \mathbf{W'X'XW} = \mathbf{W'W\Lambda}$, but **W** is an orthogonal matrix so $\mathbf{P'P} = \mathbf{\Lambda}$, a diagonal matrix.

[4] The proportion of the total variation in **X** that is explained by the mth principal component is λ_m/k, where the eigenvalue λ_m of **X'X** corresponds to the mth principal component and the column labeling in **W** has been chosen so that $\lambda_1 > \lambda_2 > \ldots\ldots > \lambda_k$.

error term. The representation (3) indicates how, when covariance or scenario calculations are based only on the most important principal components, the effect may be easily translated back to the original system through a simple linear transformation.

Efficient Computation of Positive Semi-Definite Covariance Matrices

This section outlines the theory and methodology for using a few key market risk factors that represent only the most important independent sources of information to generate large covariance matrices. These matrices will be positive semi-definite, relatively stable over time, and may be computed easily using sophisticated models that have many advantages, but that are too complex for a direct application to large systems.

Since principal components are orthogonal their covariance matrix is simply the diagonal matrix of their variances. These variances can be quickly transformed into a covariance matrix of the original system using the factor weights as follows: Taking variances of (3) gives

$$\mathbf{V} = \mathbf{ADA'} + \mathbf{V}_\varepsilon \tag{4}$$

where $\mathbf{A} = (\omega^*_{ij})$ is the kxm matrix of normalized factor weights, $\mathbf{D} = \mathrm{diag}(V(P_1), \ldots V(P_m))$ is the diagonal matrix of variances of principal components and $\mathbf{V}_\varepsilon$ is the covariance matrix of the errors. Ignoring $\mathbf{V}_\varepsilon$ gives the approximation

$$\mathbf{V} \approx \mathbf{ADA'} \tag{5}$$

with an accuracy that is controlled by choosing more or less components to represent the system. This shows how the full kxk covariance matrix of asset or risk factor returns $\mathbf{V}$ is obtained from a just a few estimates of the variances of the principal components.

Note that **V** will be positive semi-definite, but it may not be strictly positive definite unless m = k.[5] Although **D** is positive definite because it is a diagonal matrix with positive elements, there is nothing to guarantee that **ADA'** will be positive definite when m < k. To see this write

$$\mathbf{x'ADA'x = y'Dy}$$

where **A'x = y**. Since **y** can be zero for some non-zero **x** , **x'ADA'x** will not be strictly positive for all non-zero **x**. It may be zero, and so **ADA'** is only positive semi-definite. When covariance matrices are based on (5) with m < k, they should be run through an eigenvalue check to ensure strict positive definiteness. However it is reasonable to expect that the approximation (5) will give a strictly positive definite covariance matrix if the representation (3) is made with a high degree of accuracy.

Advantages of the Orthogonal Method, Limitations of Direct Methods

The first advantage of using this type of orthogonal transformation to generate risk factor covariance matrices is clear. There is a very high degree of computational efficiency in calculating only m variances instead of the k(k+1)/2 variances and covariances of the original system. For example in a single yield curve with, say, 15 maturities, only the variances of the first 2 or 3 principal components need to be computed, instead of the 120 variances and covariances of the yields of 15 different maturities.[6]

---

[5] A symmetric matrix **A** is positive definite if **x'Ax** > 0 for all non-zero **x**. If **w** is a vector of portfolio weights and **V** is the covariance matrix of asset returns, then the portfolio variance is **w'Vw**. So covariance matrices must always be positive definite, otherwise some portfolios may have non-positive variance.

[6] In highly correlated systems the first principal component, which represents a common trend in the variables, will explain a large part of the variation. In term structures and other ordered systems the second principal component represents a 'tilt' from shorter to longer maturities. Often the majority of the variation in a term structure may be explained when the system is represented by these two components alone. It is common for over 90% of the variation to be explained when a third component, the 'curvature' is added, so the considerable dimension reduction achieved by using 2 or 3 principal components results in little loss of accuracy. More details and examples may be found in Alexander (2000).

Exponentially weighted moving averages of the squares and cross products of returns are a standard method for generating covariance matrices. But a limitation of this type of direct application of exponentially weighted moving averages is that the covariance matrix is only guaranteed to be positive semi-definite if the same smoothing constant is used for all the data.[7] That is, the reaction of volatility to market events and the persistence in volatility must be assumed to be the same in all the assets or risk factors that are represented in the covariance matrix. A major advantage of the orthogonal factor method described here is that it allows exponentially weighted moving average methods to be used without this unrealistic constraint. Each principal component exponentially weighted moving average variance would normally be applied with a different smoothing constant. So the degree of smoothing in the variance of any particular asset or risk factor that is calculated by the orthogonal method will depend on the factor weights in the principal component representation. Since the factor weights of an asset are determined by its correlation with other variables in the system, so also is the degree of smoothing. That is, the market reaction and volatility persistence of a given asset will not be the same at the other assets in the system, but instead it will be related to its correlation with the other assets.

The univariate generalised autoregressive conditional heteroscedasticity (GARCH) models that were introduced by Engle (1982) and Bollerslev (1986) have been very successful for short term volatility estimation and forecasting in financial markets. The mathematical foundation of GARCH models compares favourably with some of the alternatives used by financial practitioners, and this mathematical coherency makes GARCH models easy to adapt to new financial applications. There is also evidence that GARCH models generate more realistic long-term forecasts than exponentially weighted moving averages. This is because the GARCH volatility and correlation term structure forecasts will converge to the long-term average level, which may be imposed on the model, whereas the exponentially weighted moving average model forecasts average volatility to be same for all risk horizons (see

[7] See the RiskMetrics Technical Document, 3rd Edition, 1996 (www.riskmetrics.com)

Alexander, 1998). As for short-term volatility forecasts, statistical results are mixed (see for example Brailsford and Faff, 1996, Dimson and Marsh, 1990, Figlewski, 1994, Alexander and Leigh (1997)). This is not surprising since the whole area of statistical evaluation of volatility forecasts is fraught with difficulty. Another test of volatility forecasting models is in their hedging performance. There is much to be said for using the GARCH volatility framework for pricing and hedging options (see Duan 1995, 1996). Engle and Rosenberg (1995) provide an operational evaluation of GARCH models in option pricing and hedging, demonstrating a clear superiority to the Black-Scholes methods with an extensive empirical study. The beauty of the GARCH approach stems from the fact that a stochastic volatility is built into the model, which is closer to the real world, yet it does not introduce an additional source of uncertainty and therefore delta hedging is still sufficient.

Large covariance matrices that are based on GARCH models would, therefore, have clear advantages over those generated by exponentially or equally weighted moving averages. But previous research in this area has met with rather limited success. It is straightforward to generalize the univariate GARCH models to multivariate parameterizations, as in Engle and Kroner (1993). But the actual implementation of these models is extremely difficult. With so many parameters, the likelihood function becomes very flat, and so convergence problems are very common in the optimization routine. If the modeler also needs to 'nurse' the model for systems with only a few variables, there is little hope of a fully functional implementation of a direct multivariate GARCH model to work on large risk systems.

The idea of using factor models with GARCH is not new. Engle, Ng and Rothschild (1990) use the capital asset pricing model to show how the volatilities and correlations between individual equities can be generated from the univariate GARCH variance of the market risk factor. Their results have a straightforward extension to multi-factor models, but unless the factors are orthogonal a multi-variate GARCH model will be required, with all the associated problems.

A principal component representation is a multi-factor model. In fact the orthogonal GARCH model introduced in Alexander (2000) is a generalization of the factor GARCH model introduced by Engle, Ng and Rothschild (1990) to a multi-factor model with orthogonal factors. The orthogonal GARCH model allows kxk GARCH covariance matrices to be generated from just m univariate GARCH models. It may be that m, the number of principal components can be much less than k, the number of variables in the system - and quite often one would wish m to be less than k so that extraneous 'noise' is excluded from the data. But since only univariate GARCH models are used it does not really matter: there no dimensional restrictions as there are with the direct parameterizations of multivariate GARCH.

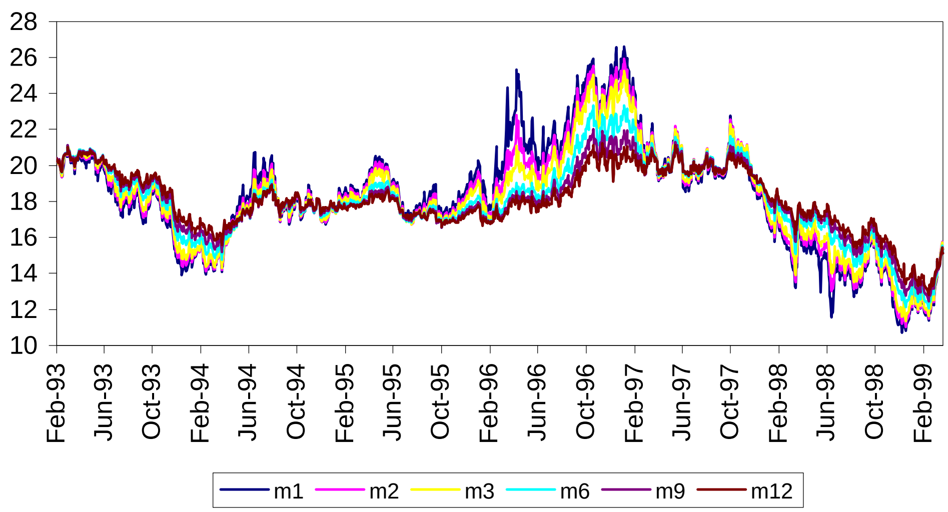
Of course, the principal components are only unconditionally uncorrelated, so a GARCH covariance matrix of principal components is not necessarily diagonal. However the assumption of zero conditional correlations has to be made, otherwise it misses the whole point of the model, which is to generate large GARCH covariance matrices from GARCH volatilities alone. The degree of accuracy that is lost by making this assumption is investigated by a thorough calibration of the model, comparing the variances and covariances produced with those from other models such as exponentially weighted moving averages or, for small systems, with multivariate GARCH. Care needs to be taken with the initial calibration, in terms of the number of components used and the time period used to estimate them, but once calibrated the orthogonal GARCH model may be run very quickly and efficiently on a daily basis.

Another advantage is that the orthogonal method, applied with either GARCH or exponentially weighted moving average variances, allows one to generate estimates for volatilities and correlations of variables in the system even when data are sparse and unreliable, for example in illiquid markets. For example, the direct estimation of a time-varying variance of a 12-year bond may be difficult, but the orthogonal method allows its variance to be calculated from the variances of the key risk factors used in its representation.

Some Examples [8]

The orthogonal method is ideally suited to highly correlated ordered systems such as a term structure. The first example uses (a) exponentially weighted moving average variances and (b) GARCH(1,1) variances of just two principal components for the WTI crude oil futures from 1 month to 12 months, sampled daily between 4th February 1993 and 24th March 1999. The 1, 2, 3, 6, 9 and 12-month maturity futures prices are shown in figure 1[9] and the results of a principal component analysis on daily returns are given in table 1.

Figure 1: NYMEX Sweet Crude Prices



---

[8] The results in this section are reported in more detail, along with several other examples, in Alexander (2000).
[9] See Alexander (1999) for a full discussion of these data and of correlations in energy markets in general. Many thanks to Enron for providing these data.

Of course the factor weights show that, as with any term structure, the interpretations of the first three principal components are the trend, tilt and curvature components respectively. In fact this particular system is so highly correlated that over 99% of its variation may be explained by just two principal components and the first principal component alone explains almost 96% of the variation over the period.

Table 1a: Eigenvalue Analysis

Component	Eigenvalue	Cumulative R^2
P1	11.51	0.9592
P2	0.397	0.9923
P3	0.069	0.9981

Table 1b: Factor Weights

	P1	P2	P3
1mth	0.89609	0.40495	0.18027
2mth	0.96522	0.24255	-0.063052
3mth	0.98275	0.15984	-0.085002
4mth	0.99252	0.087091	-0.080116
5mth	0.99676	0.026339	-0.065143
6mth	0.99783	-0.020895	-0.046369
7mth	0.99702	-0.062206	-0.023588
8mth	0.99451	-0.098582	0.000183
9mth	0.99061	-0.13183	0.020876
10mth	0.98567	-0.16123	0.040270
11mth	0.97699	-0.19269	0.064930
12mth	0.97241	-0.21399	0.075176

The GARCH(1,1) model defines the conditional variance at time t as

$$s_t^2 = w + a e_{t-1}^2 + b s_{t-1}^2 \qquad (6)$$

where $\omega > 0$, α, $\beta \geq 0$. This simple GARCH model effectively captures volatility clustering and provides convergent term structure forecasts to the long-term average level of volatility $100\sqrt{250\omega/(1-\alpha-\beta)}$. The coefficient α measures the intensity of reaction of volatility to yesterday's unexpected market return ε^2_{t-1}, and the coefficient β measures the persistence in volatility.[10]

Applying (6) to the first two principal components of these data gives the parameter estimates reported in table 2. Note that the first component has low market reaction but high persistence, and the opposite is true for the second component. This reflects much of what is already known about the data from the principal component analysis: the system is very highly correlated indeed, in fact price decoupling occurs for only very short periods of time. Now in the orthogonal model all the variation in correlations will come from the second or higher principal components because with only one component all variables are assumed to be perfectly correlated. The second component here has a 'spiky' volatility, and this gives rise to orthogonal GARCH correlations that also have only temporary deviations from normal levels. Thus the orthogonal GARCH model is capturing the true nature of crude oil futures markets. Unfortunately the exponentially or equally weighted moving average correlations that are in standard use have a substantial bias that arises from only very temporary price decoupling.

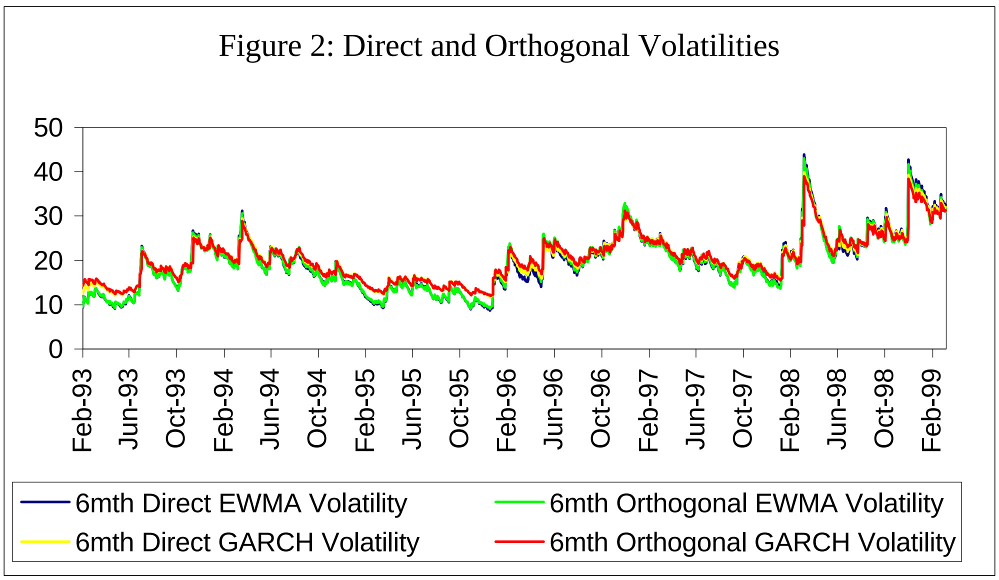
Table 2: GARCH(1,1) models of the first two principal components

	1st Principal Component		*2nd Principal Component*	
	Coefficient	*t-stat*	*Coefficient*	*t-stat*
constant	.650847E-02	.304468	.122938E-02	.066431
ω	.644458E-02	3.16614	.110818	7.34255
α	.037769	8.46392	.224810	9.64432
β	.957769	169.198	.665654	21.5793

Figure 2 shows how closely the volatilities that are obtained using the orthogonal method compare with those obtained by the direct application of (a) exponentially

[10] Note that these are determined independently in the GARCH(1,1) model, subject only to the constraint that $\alpha+\beta<1$. In the exponentially weighted moving average model these parameters are not independent because they always sum to 1, and the constant is zero, so there is not long-term average level in the model and volatility terms structures are constant.

weighted moving averages and (b) GARCH(1,1) models.[11] Of course there is no space here to graph all 78 volatilities and correlations from the 12x12 covariance matrix. But interested readers may use the programs provided with Alexander (2000) to verify that all volatilities, not just those shown in figure 2, are very similar. But there is a difference in correlations. Not depending on whether a direct or an orthogonal approach is used, but depending on whether exponentially weighted moving averages or GARCH(1,1) models are used. As mentioned above, the GARCH correlations more accurately reflect the true nature of the data.



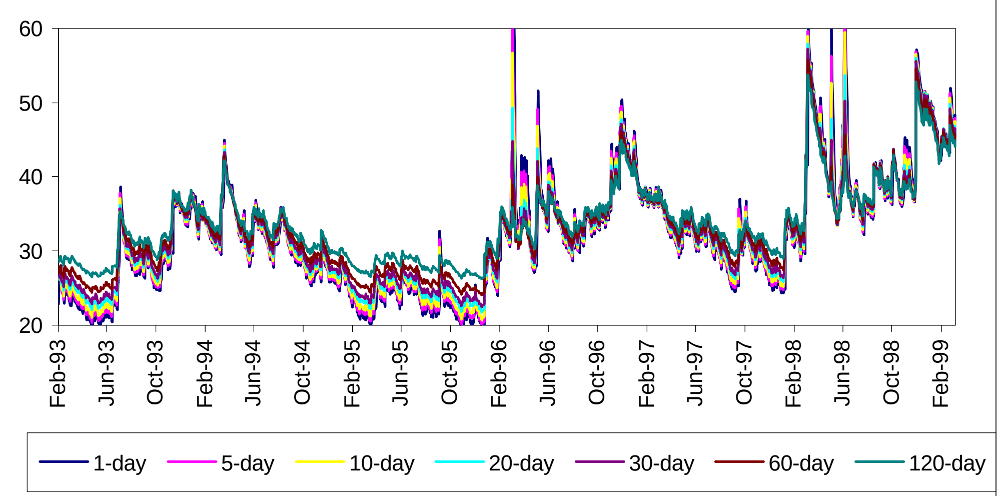
The main disadvantage of the direct method is that it requires estimating 78 volatilities and correlations, using (a) the same value of the smoothing constant for the exponentially weighted moving average model, or (b) a 12-dimensional multivariate GARCH model. Both of these approaches have substantial limitations as described above. However using the orthogonal method only two moving average variances, or two univariate GARCH(1,1) variances, of the trend and tilt principal components need to be generated. The entire 12x12 covariance matrix of the original system is

[11] There is no optimal method for choosing a value for the smoothing in these exponentially weighted moving averages. A value of 0.95 has been used throughout, but the reader may experiment with different values by adjusting the programs that are provided with Alexander (2000).

simply a transformation of these two variances, as defined in (5) above, and it may be recovered in this way with negligible loss of precision.

Several good reasons to prefer GARCH models to exponentially weighted moving averages have already been mentioned, and one of the most attractive reasons is that only the GARCH approach will give convergent term structure forecasts. In the orthogonal GARCH model these forecasts, for volatilities and correlations of all maturities, are obtained from the simple transformations (5) where the diagonal matrix **D** contains the n-period GARCH(1,1) variance forecasts of the principal components.[12] Some of these are illustrated for volatilities of the 1-mth oil future in figure 3.

Figure 3: Orthogonal GARCH Term Structure Volatility
Forecasts for 1mth Crude Oil Futures

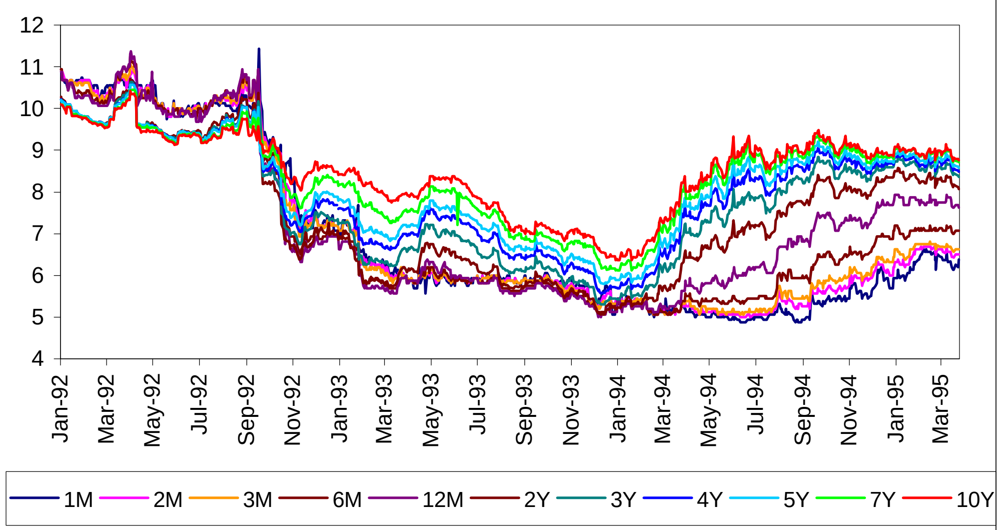


The next example applies the orthogonal GARCH(1,1) model to another term structure, but a rather difficult one. Daily zero coupon yield data in the UK with 11 different maturities between 1mth and 10 years from 1st Jan 1992 to 24th Mar 1995 are

[12] The n-period GARCH(1,1) variance forecast is the sum of n forward variances for j = 1, ….n:
$\hat{s}^2_{t+j} = \hat{w} + (\hat{a} + \hat{b})\hat{s}^2_{t+j-1}$

shown in figure 4. It is not an easy task to estimate univariate GARCH models on these data directly because the yields may remain relatively fixed for a number of days. Particularly on the more illiquid maturities, there is insufficient conditional heteroscedasticity for univariate GARCH models to converge well. So an 11-dimensional multivariate GARCH model is completely out of the question.

Figure 4: UK Zero-Coupon Yields



Again two principal components were used in the orthogonal GARCH, but the principal component analysis reported in table 3 shows that these two components only account for 72% of the total variation. Also the 10yr yield has a very low correlation with the rest of the system, as reflected by its factor weight on the 1st principal component, which is quite out of line with the rest of the factor weights on this component. So the fit of the orthogonal model could be improved if the 10yr bond were excluded from the system. Despite these difficulties the volatilities obtained using the orthogonal GARCH model are very similar to those obtained by direct estimation of exponentially weighted moving averages.[13]

[13] The smoothing constant for all exponentially weighted moving averages was again set at 0.95.

Table 3a: Eigenvalue Analysis

Component	Eigenvalue	Cumulative R^2
P1	5.9284117	0.53894652
P2	1.9899323	0.71984946
P3	0.97903180	0.80885235

Table 3b: Factor Weights

	P1	P2	P3
1mth	0.50916	0.60370	0.12757
2mth	0.63635	0.62136	-0.048183
3mth	0.68721	0.57266	-0.10112
6mth	0.67638	0.47617	-0.10112
12mth	0.83575	0.088099	-0.019350
2yr	0.88733	-0.21379	0.033486
3yr	0.87788	-0.30805	-0.033217
4yr	0.89648	-0.36430	0.054061
5yr	0.79420	-0.37981	0.14267
7yr	0.78346	-0.47448	0.069182
10yr	0.17250	-0.18508	-0.95497

The GARCH (1,1) parameter estimates of the principal components are given in table 4. This time both components have fairly persistent volatilities, and both are less reactive than the volatility models reported in table 2. Combine this with the fact that almost 28% of the variation has been ascribed to 'noise' by using only these first two principal components, and it is not unsurprising that the orthogonal GARCH model produces quite stable correlation estimates: more stable than those obtained by direct application of exponentially weighted moving averages.

Table 4: GARCH(1,1) models of the first two principal components

	1st Principal Component		*2nd Principal Component*	
	Coefficient	*t-stat*	*Coefficient*	*t-stat*
constant	.769758E-02	.249734	.033682	1.09064
ω	.024124	4.50366	.046368	6.46634
α	.124735	6.46634	.061022	9.64432
β	.866025	135.440	.895787	50.8779

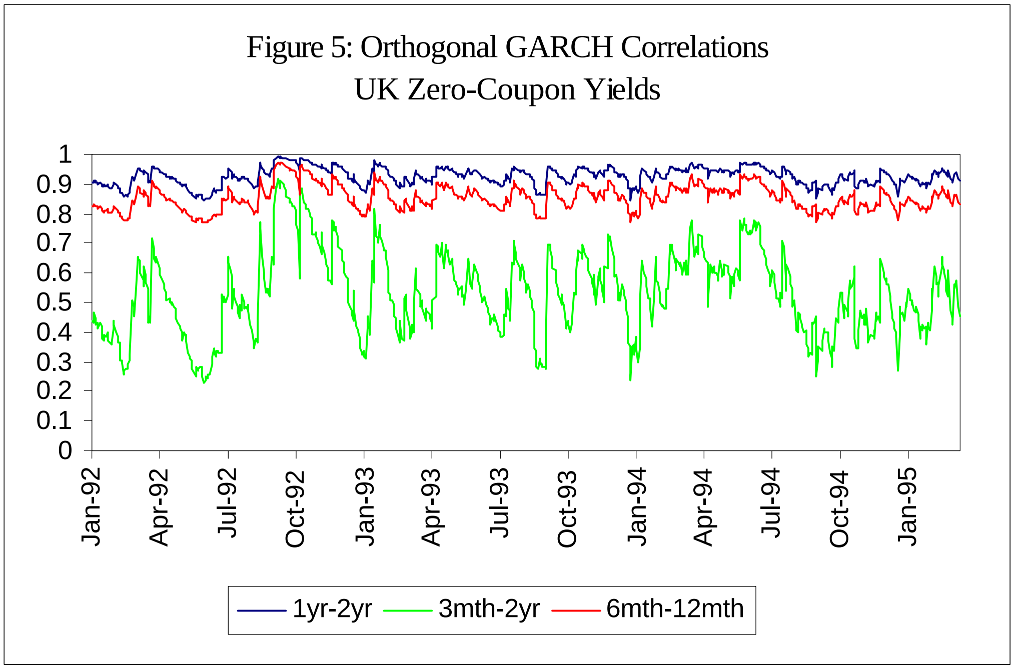


Figure 5 shows some of the orthogonal GARCH correlations for the UK zero coupon yields. So not only does the orthogonal method provide a way of estimating GARCH volatilities and volatility term structures that may be difficult to obtain by direct univariate GARCH estimation. They also give very sensible GARCH correlations, which would be very difficult indeed to estimate using direct multivariate GARCH. And all these are obtained from just two principal components, the key market risk factors that are representing the most important sources of information - all the rest of the variation is ascribed to 'noise' and is not included in the model.

Generating a Large Covariance Matrix across All Risk Factor Categories

The risk factors - equity market indices, exchange rates, commodities, government bond and money market rates and so on - are first divided into reasonably highly correlated categories, according to geographic locations and instrument types. Principal component analysis is then used to extract the key risk factors from each sub-system and their diagonal covariance matrix is obtained using one of the methods outlined above. Then the factor weights from the principal component analysis are used to 'splice' together a large covariance matrix for the original system.

The method is explained for just two categories, then the generalization to any number of categories is straightforward. Suppose there are n variables in the first system, say it is European equity indices, and m variables in the second system, European exchange rates say. It is not the dimensions that matter. What does matter is that each system of risk factors is suitably co-dependent, so that it justifies the categorization as a separate and coherent sub-system. The first step is to find the principal components of each system, $\mathbf{P} = (P_1 , \ldots P_r)$, and separately $\mathbf{Q} = (Q_1 ,\ldots Q_s)$ where r and s are number of principal components that are used in the representation of each system. Denote by $\mathbf{A}$ (nxr) and $\mathbf{B}$ (mxs) the normalized factor weights matrices obtained in the principal component analysis of the European equity and exchange rate systems respectively. Then the 'within factor' covariances, i.e. the covariance matrix for the equity system, and for the exchange rate system separately, are given by $\mathbf{AD}_1\mathbf{A}'$ and $\mathbf{B}\,\mathbf{D}_2\mathbf{B}'$ respectively. Here $\mathbf{D}_1$ and $\mathbf{D}_2$ are the diagonal matrices of the variances of the principal components of each system. The cross factor covariances are $\mathbf{ACB}'$ where $\mathbf{C}$ denotes the rxs matrix of covariances of principal components across the two systems, that is

$$\mathbf{C} = \{\mathrm{COV}(P_i, Q_j)\}$$

Then the full covariance matrix of the system of European equity and exchange rate risk factors is:

$$\begin{pmatrix} \mathbf{AD}_1\mathbf{A}' & \mathbf{ACB}' \\ (\mathbf{ACB}')' & \mathbf{BD}_2\mathbf{B}' \end{pmatrix}$$

The within factor covariance matrices $\mathbf{AD}_1\mathbf{A}'$ and $\mathbf{BD}_2\mathbf{B}'$ will always be positive semi-definite. But it is not always possible to guarantee positive semi-definiteness of the full covariance matrix of the original system, unless the off diagonal blocks $\mathbf{ACB}'$ are set to zero. This is not necessarily a silly thing to do; in fact it may be quite sensible in the light of the huge instabilities often observed in cross-factor covariances.[14]

The method is illustrated using four European equity indices and their associated sterling foreign exchange rates. The graphs in figure 6 are based on daily return data from 1st April 1993 to 31st December 1996 on France (CAC40), Germany (DAX30), Holland (AEX), and the UK (FTSE100). In this 7-dimensional system of equity indices and foreign exchange rates there are 28 volatilities and correlations in total. Figure 6 shows just two of the correlations from an orthogonal GARCH(1,1) model of the system compared with those obtained from two different direct parameterizations of a multivariate GARCH(1,1) model: (a) the Vech model, and (b) the BEKK model. These multivariate GARCH models were only possible to estimate on each sub-system separately. In fact convergence problems with the BEKK model for the foreign exchange system were encountered, so only the Vech model correlations, which have severe cross equation restrictions[15] are shown in figure 6b. These two graphs, which indicate a close similarity between the correlations, were chosen at random from the correlations for which multivariate GARCH models also produce results. Principal component analysis and orthogonal GARCH, Vech and BEKK model parameter estimates are not reported here due to lack of space, but full details of these models and the results are given in Alexander (2000).

[14] For non-zero cross-factor covariances it is possible to estimate the covariance between principal components of different risk factor sub-systems using exponentially weighted moving averages or bivariate GARCH, giving the required estimate for **C**.

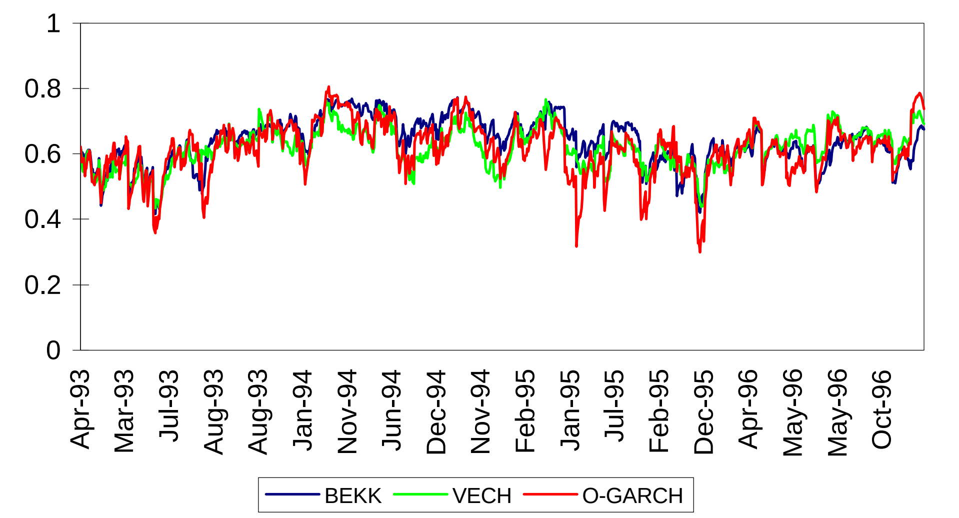


Figure 6a: Equity Correlation Comparison (CAC-FTSE)

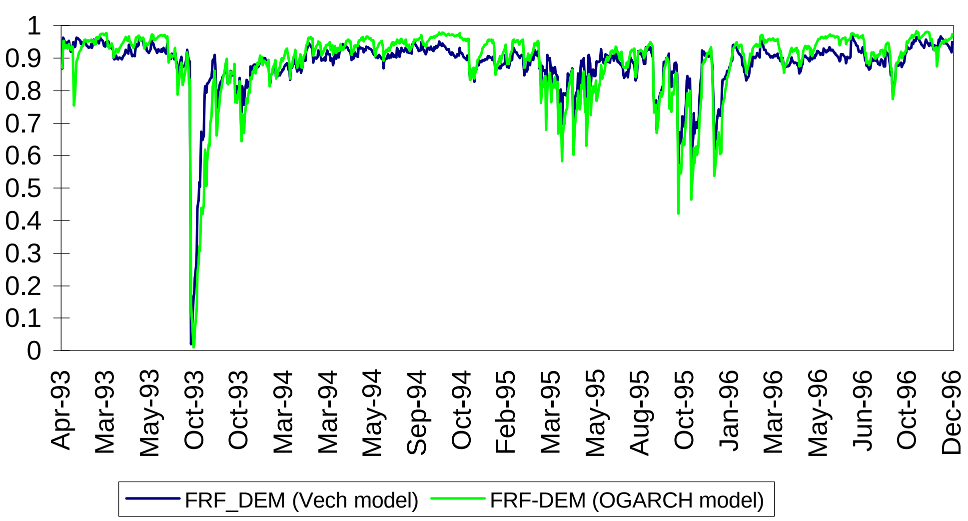


Figure 6b: FX Correlation Comparison

The example has been mentioned here to illustrate the scope and flexibility of the approach to all types of asset class. It shows that it is possible to estimate these covariance matrices when direct methods are not possible, or require unrealistic

[15] In the Vech model all variances and covariances depend only on their own lag, and not the lags of other variances and covariances in the system.

restrictions. Provided the assets are first divided into reasonably highly correlated categories, principal component analysis provides a way to extract the important independent sources of information in each category. The covariance matrices for each category are generated from the variances of these key risk factors, and then a large covariance matrix that encompasses all categories is spliced together.

Using Key Risk Factors of Volatility Skews to Identify Different Market Regimes and the Price-Volatility Scenarios that Apply

Scenario based maximum loss calculations require at least the definition, if not the joint distribution, of scenarios for implied volatilities and underlying asset prices. In the absence of an effective model of how implied volatilities change with market price, these scenarios may be rather simplistic. The base scenario that the smile surface remains unchanged over all risk horizons is often augmented by a only a few simple scenarios, such as parallel shifts in all volatilities that are assumed to be independent of movements in underlying prices.

But for equity options there is often a negative correlation between at-the-money volatility and the underlying price. This is clear from figure 7 which shows, for three different two month periods during 1998, a scatter plot of the daily changes in 1mth at-the-money volatility vs daily changes in index price for the FTSE100 European option. The periods chosen were (a) May and June 1998; (b) February and March 1998; and (c) August and September 1998.[16]

[16] The fixed maturity implied volatility data used in this section have been obtained by linear interpolation between the two adjacent maturity option implied volatilities. However this presents a problem for the 1mth volatility series because often during the last few working days before expiry data on the near maturity option volatilities are totally unreliable. So the 1mth series rolls over to the next maturity, until the expiry date of the near-term option, and thereafter continues to be interpolated linearly between the two option volatilities of less than and greater than 1 month.

Figure 7a: At-the-Money Volatility vs FTSE 100 (Daily Changes) May and June 1998

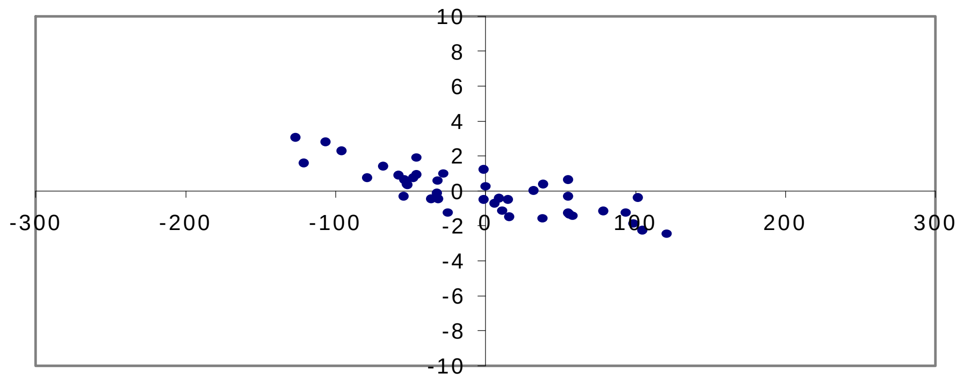


Figure 7b: At-the-Money Volatility vs FTSE 100 (Daily Changes) February and March 1998

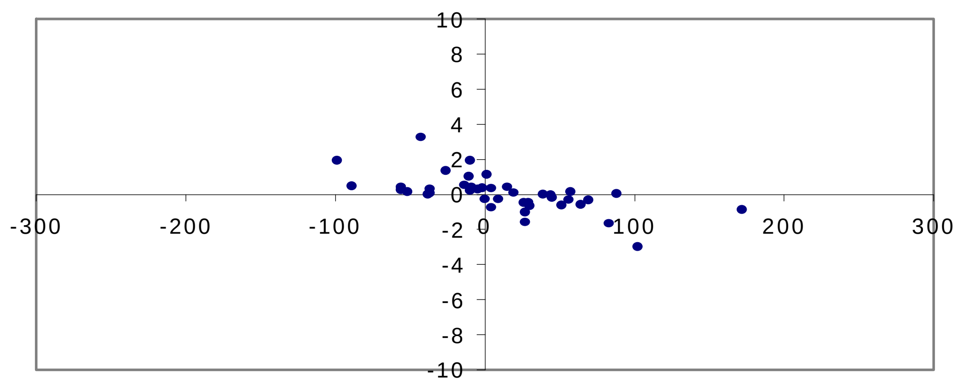
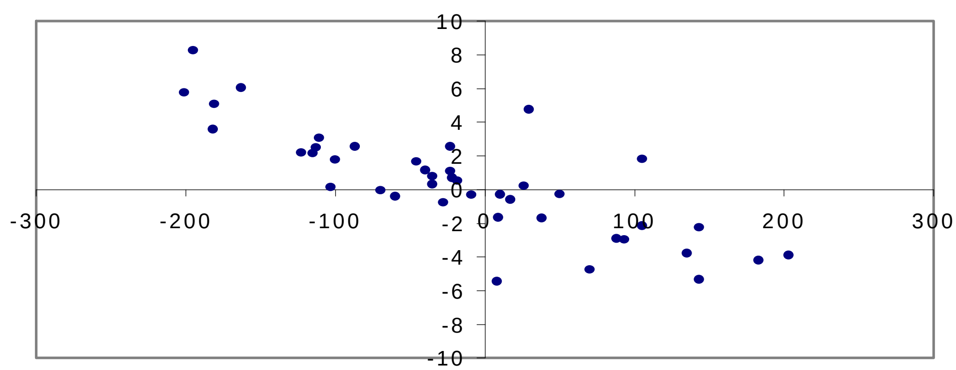


Figure 7c: At-the-Money Volatility vs FTSE 100 (Daily Changes) August and September 1998



Casual observation of these scatter plots indicates a significant negative correlation between the 1mth implied volatility and the index price, but the strength of this correlation depends on the data period. Period (b) when the UK equity market was very stable and trending, shows less correlation than period (a), when daily movements in the FTSE100 index were limited to a ‘normal’ range; but the negative correlation is most obvious during the mini-crash period (c). These observations are not peculiar to the 1mth at-the-money FTSE100 volatilities, and not just during the periods shown: negative correlations, of more or less strength depending on the data period, are also evident in other fixed term at-the-money volatilities and in other equity markets.

So realistic scenarios for at-the money volatility and index prices would be for movements in at-the-money volatility to occur in the opposite direction to the index price movements. But how large should these movements be in relation to each other? Does the answer depend on current market conditions? If so, how can we model the current market conditions to quantify the correlation effect? And what about the fixed-strike volatilities? Since positions are likely to move in- or out-of-the-money during the risk horizon, we need to know what scenarios are most probable for the whole volatility skew.

Derman's Volatility Regimes

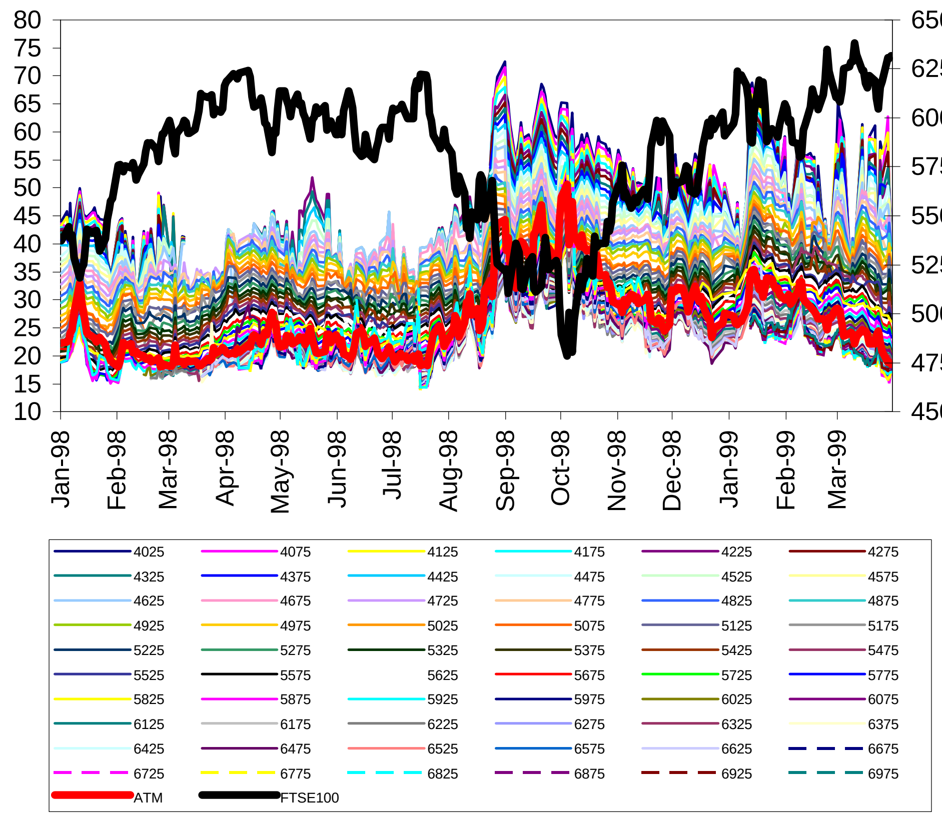
Figure 8a shows the 1mth implied volatilities for European options of all strikes on the FTSE100 index for the period 4th January 1998 to 31st March 1999. The bold red line indicates the at-the-money volatility and the bold black line the FTSE100 index price (on the right-hand scale). Look at the movements in the index and the way that at-the-money volatility is behaving in relation to the index during the three different periods chosen in figure 7. Observation of data similar to these, but on the S&P500 index option 3mth volatilities, has motivated Derman (1999) to formulate three different market regimes:

(a) *Range-bounded*, where future price moves are likely to be constrained within a certain range and there no significant change in realized volatility;

(b) *Trending*, where the level of the market is changing but in a stable manner so there is again little change in realized volatility in the long run; and

(c) *Jumpy*, where the probability of jumps in the price level is particularly high so realized volatility increases.

Figure 8a: Fixed-Strike Volatilities, At-the-Money Volatility and the Index Level



Different linear parameterizations of the volatility skew for pricing and hedging options apply in each regime. These are known as Derman's 'sticky' models, because each parameterization implies a different type of 'stickiness' for the local volatility in a binomial tree.[17] Denote by $\sigma_K(t)$ the implied volatility of an option with maturity t and strike K, $\sigma_{ATM}(t)$ the volatility of the t-maturity at-the-money option, S the current value of the index and σ_0 and S_0 the initial implied volatility and price used to calibrate the tree:

(a) In a range bounded market Derman proposes that skews are parameterized by the 'sticky strike' model:

$$\sigma_K(t) = \sigma_0 - b(t)\,(K-S_0) \tag{7a}$$

So fixed strike volatility $\sigma_K(t)$ is independent of the index level S.

Since $\sigma_{ATM}(t) = \sigma_0 - b(t)\,(S-S_0)$ this model implies that σ_{ATM} decreases as index increases.

(b) For a stable trending market skews are parameterized by the 'sticky delta' model:

$$\sigma_K(t) = \sigma_0 - b(t)\,(K-S) \tag{7b}$$

So fixed strike volatility $\sigma_K(t)$ increases with the index level S.

Since $\sigma_{ATM}(t) = \sigma_0$ this model implies that $\sigma_{ATM}(t)$ is independent of the index.

(c) In jumpy markets skews are parameterized by the 'sticky tree' model:

$$\sigma_K(t) = \sigma_0 - b(t)\,(K+S) \tag{7c}$$

So fixed strike volatility $\sigma_K(t)$ decreases as the index increases.

Since $\sigma_{ATM}(t) = \sigma_0 - 2b(t)S$, the at-the-money volatility $\sigma_{ATM}(t)$ also decreases as index increases, and twice as fast as the fixed strike volatilities.

Fixed-Strike Volatility Deviations from At-the-Money Volatility

Time series data such as that shown in figure 8a should contain all the information necessary to estimate the skew parameterization that is appropriate for the current market regime. But there are around 60 different strikes represented there, and their volatilities form a correlated, ordered system that is similar to a term structure. It is therefore natural to consider using principal component analysis to identify the main

[17] The 'sticky strike' is so called because local volatilities are constant with respect to strike, changing only with moneyness; the 'sticky delta' model has local volatilities that are not constant with strike, but are constant with respect to moneyness or delta; and only in the 'sticky tree' model is there one, unique tree for all strikes and moneyness.

independent sources of information. Both analytic simplicity and computational efficiency would result from a model that is based only on these key risk factors.

Principal component analysis of the volatility skew has been used before, by Derman and Kamal (1997). However their work is based on quite different data to that shown in figure 8a.[18] Time series data on fixed strike or fixed delta volatilities often display very much negative autocorrelation, possibly because markets over-react, so the 'noise' in daily changes of fixed strike volatilities is a problem. Therefore a principal components analysis of daily changes in fixed-strike volatilities may not give very good results.

But look at the *deviations* of fixed strike volatilities from at-the-money volatility, shown in figure 8b. These display less negative autocorrelation, they are even more highly correlated and ordered than the fixed strike volatilities themselves, and their positive correlation with the index is very evident indeed during the whole period. The reason for this becomes evident when (7a) – (7c) are rewritten in terms of fixed-strike volatility deviations from at-the-money volatility $\sigma_K(t) - \sigma_{ATM}(t)$. Each of Derman's models yields the same relationship between fixed-strike volatility deviations from at-the-money volatility and the current index price, viz.:

$$\sigma_K(t) - \sigma_{ATM}(t) = -b(t)\,(K-S) \tag{8}$$

So all three models imply the same, positive correlation between the index and the skew deviations $\sigma_K(t) - \sigma_{ATM}(t)$. In fact an alternative formulation of Derman's sticky models is (8) with a different specification for the behaviour of at-the-money volatility in relation to the index in each regime, viz.

(a) Range-bounded: $\sigma_{ATM}(t) = \sigma_0 - b(t)\,(S - S_0)$

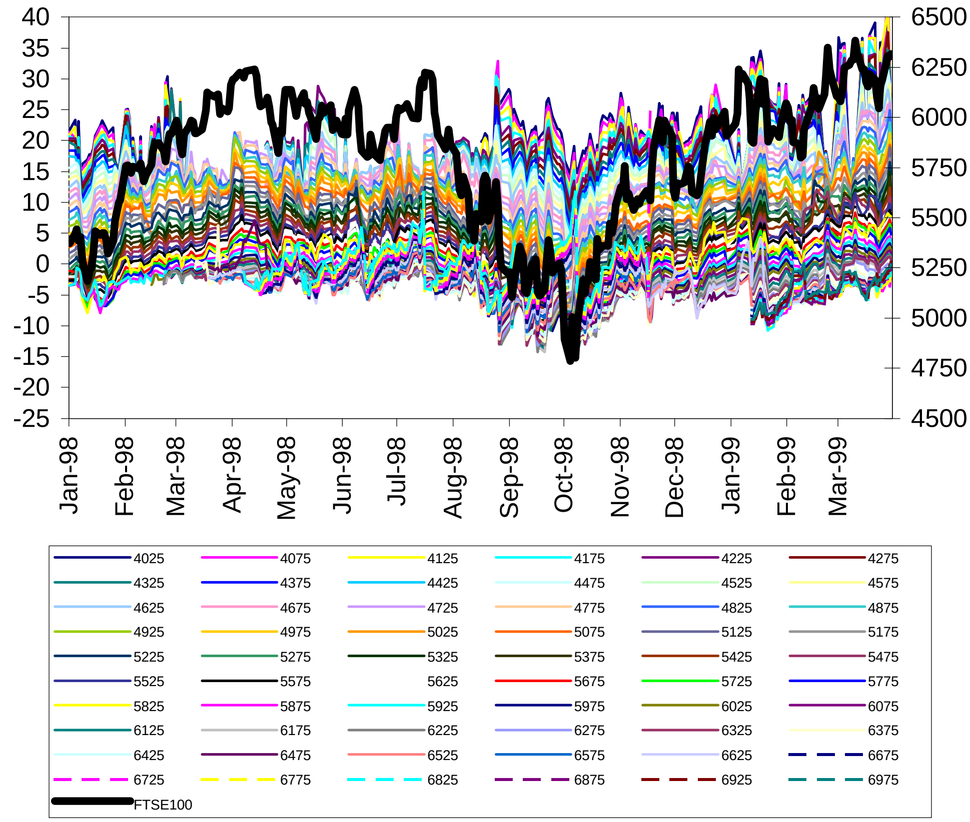
(b) Stable trending: $\sigma_{ATM}(t) = \sigma_0$

(c) Jumpy: $\sigma_{ATM}(t) = \sigma_0 - 2b(t)S$.

---

[18] Dermand and Kamal use weekly mid-market volatility of S&P500 index options from May 1994 to September 1997 where the surface is specified by 12 numbers corresponding to three different deltas for 1mth, 3mth, 6mth and 12mth maturities; and daily Nikkei 225 index volatility from September

Figure 8b: Deviations of Fixed-Strike Volatility from At-the-Money Volatility (1mth)



Effective Methods for Identification of the Current Market Regime

The above formulation of Derman's regime models suggests that one might perform an empirical investigation into which regime currently prevails by estimating linear regressions of the form:

$$\Delta\sigma_{ATM}(t) = \alpha(t) + \beta(t)\Delta S + \varepsilon(t) \qquad (9)$$

where $\Delta\sigma_{ATM}(t)$ denotes the daily change in at-the-money volatility of maturity t and ΔS is the daily change in the index. In general, due to the negative correlation, each

1994 to May 1997 for 9 deltas and 5 different maturities. For each of these markets they analyze the principal components of the changes in the whole implied volatility surface.

β(t) will be negative. But if all the coefficients β(t) are insignificantly different from zero the market is stable and trending, so the sticky delta model should be used. A signal that the market has entered a different regime occurs when β(t) undergoes a significant change in value. In a jumpy market that is characterized by the sticky tree model, the value of β(t) will be approximately twice the value that it takes in a range-bounded market where the sticky strike model is valid.

Figure 9a: At-the-Money Volatility Sensitivity to Change in the Index

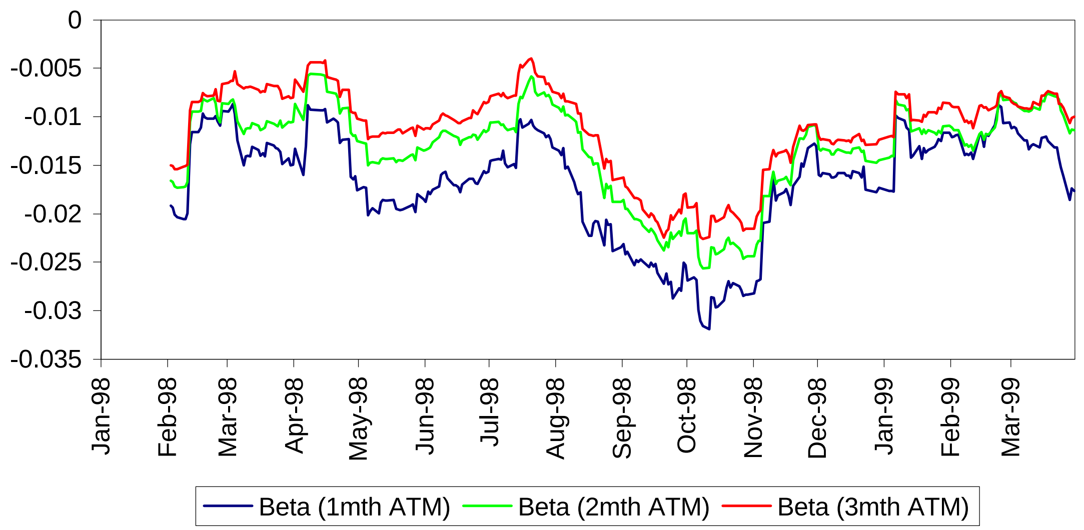
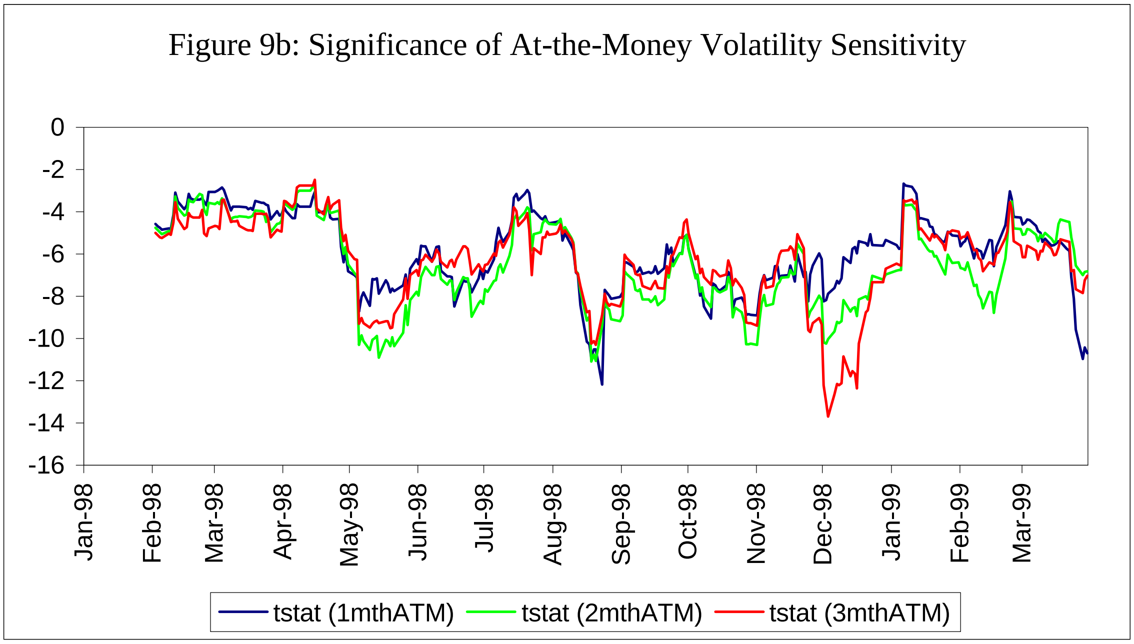


Figure 9a shows the values obtained for β(t) for t = 1mth, 2mths and 3mths. In order to capture the current market conditions one month of daily data is used in each regression. These regressions were rolled over the whole period from 4th January 1998 to 31st March 1999 and each time the coefficient and its t-statistic are recorded.

The response of at-the-money volatility to changes in the underlying index level increases as options approach expiry, and this fact is reflected in figure 9a since at all times

$$|\beta(1mth)| > |\beta(2mth)| > |\beta(3mth)|.$$

However no such order is apparent in the accompanying t-statistics, shown in figure 9b, so the negative correlation between at-the-money volatility and index price is not a simple function of the maturity of volatility.



Casual observation of figure 8a has indicated that February and March 1998 might be characterized as a stable and trending market. Figure 9b provides quantifiable evidence of this, because during February and March 1998 the t-statistics on the β coefficients are less significant than at other times.[19] Two other periods were picked out in the earlier discussion: May and June 1998, when the market seemed to be operating in a range-bounded regime, and the mini-crash period that began in August 1998 and initiated a very jumpy market until the November of that year. From figure 9a it is apparent that the values of the β coefficients during the mini-crash period, although not exactly double their values during May and June 1998, were far greater than at any other time. A rapid decline in β, for all maturity volatilities, occurred at the end of July 1998. Thus the model is providing a signal of a change in market regime *before the mini-crash occurred.* It was not until November 1998 that the level

[19] The 99% significance level is approximately 2.5.

of β returned to more normal levels, when the market appears to pass back into a range-bounded regime.

Using Key Risk Factors to Formulate Appropriate Skew Scenarios

The simple regressions just described may be used to identify the current volatility regime, and to forewarn risk managers of any change in market conditions. It is now shown how such information may be put to practical advantage. The applicability of maximum loss calculations will depend upon the construction of appropriate price-volatility scenarios. So we now ask, which type of skew scenarios should accompany the scenarios on movements in the underlying? Are simple static or parallel shift scenarios for the volatility skew appropriate at the moment? If so, is it the volatility by strike that should remain static, so the volatility by moneyness or delta has a parallel shift? Or is it volatility by delta that is static, which is equivalent to a parallel shift in volatility by strike? But perhaps one should be placing more importance on scenarios that encompass changes in the tilt or curvature of the volatility skew? If so, at which end: should in-the-money volatilities be changed as much as out-of-the-money volatilities? The following discussion illustrates how all these questions can be answered by an empirical model of the relationship between the index and the key risk factors of the skew.

Derman's models are based a linear parameterization of the skew given by (8). For any given maturity, the deviations of all fixed strike volatilities from at-the-money volatility will change by the same amount b(t) as the index level changes, as shown in figure 10a. Four strikes are marked on this figure: a low strike K_L, the initial at-the-money strike K_1, the new at-the-money strike after the index level moves up K_2, and a high strike K_H. The volatilities at each of these strikes are shown in figure 10b, before and after an assumed unit rise in index level ($\Delta S = 1$). In each of the three market regimes the range of the skew between K_L and K_H, that is $\sigma_L - \sigma_H$, will be the same after the rise in index level. Thus all of Derman's model's imply a parallel shift scenario for the skew by strike.

Figure 10a: Parallel Shift in Skew Deviations as Price Moves Up

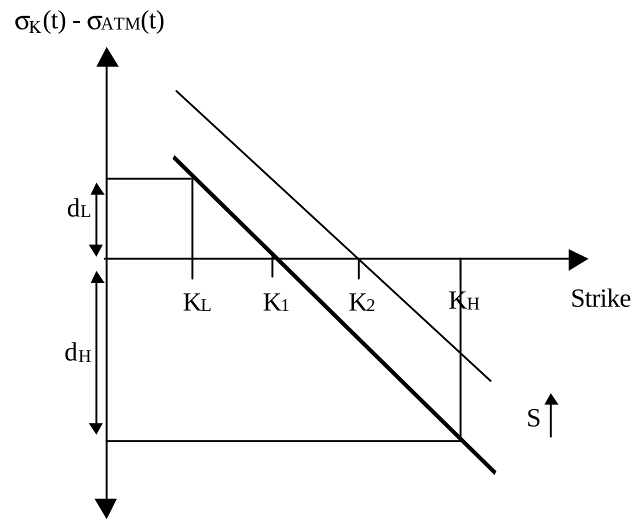
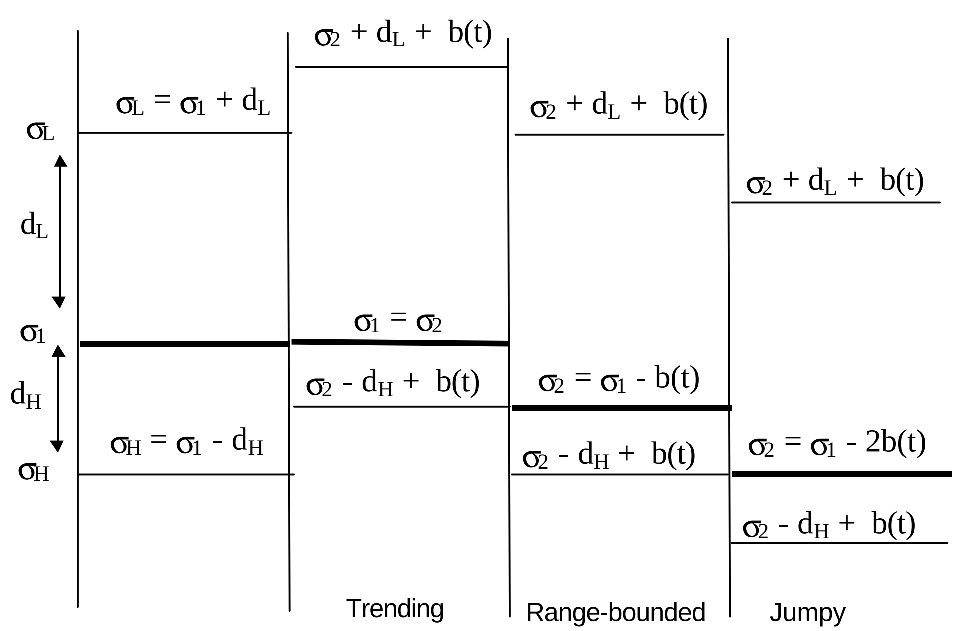
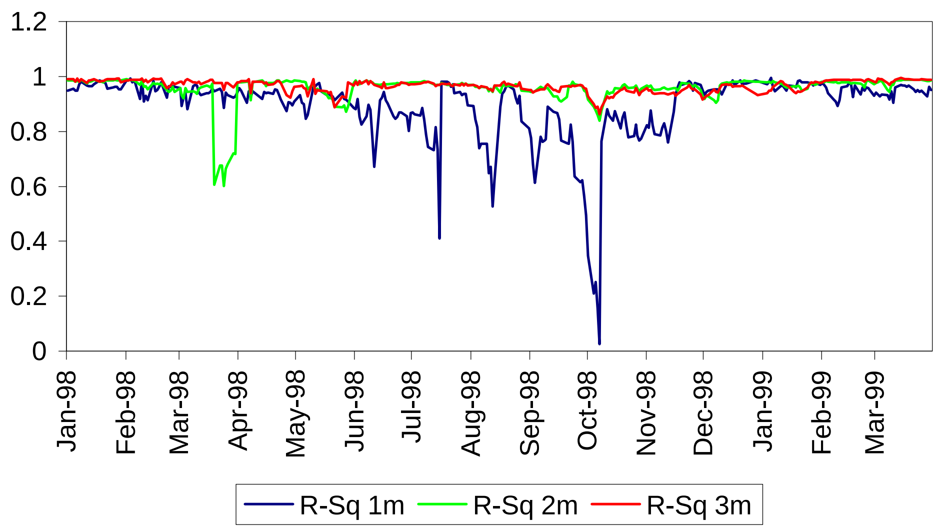


Figure 10b: Parallel Shifts in Fixed-Strike Volatilities as Price Moves Up



The extent of the parallel shift depends on the relationship between the original at-the-money volatility σ_1 and the new at-the-money volatility σ_2, and this will be defined by the current market regime. In a range bounded market $\sigma_2 = \sigma_1 - b(t)$, but fixed-strike volatilities have all increased by the same amount b(t), so a static scenario for the skew by strike should be applied, as depicted in figure 10b. When the market is stable and trending, $\sigma_2 = \sigma_1$ and there is an *upwards* shift of b(t) in all fixed-strike volatilities. Finally, in a jumpy market $\sigma_2 = \sigma_1 - 2b(t)$, so a parallel shift *downwards* of the skew by strike should be applied.

Figure 11: R-Squared from Linear Skew Parameterization



Whilst a linear parameterization of the skew may be good approximation for the 3mth or longer maturities, empirical observations show that it may not be very realistic at the shorter end. Figure 11 shows the correlations from simple regressions based on (8). Daily changes of fixed-strike deviations from at-the-money volatility, $\Delta(\sigma_K - \sigma_{ATM})$ are regressed on daily changes in the index price, using one month of data. These regressions are then rolled over the entire data period. It is clear that whilst the skew is fairly linear at the 3mth maturity, it becomes quite non-linear at the 1mth maturity, particularly during the summer of 1998.

So the parallel shift scenarios for volatility skews that are a consequence of Derman's models may be reasonable for 3mth volatilities, but for shorter-term volatilities a simple, effective non-linear model of the skew would be advantageous. Such a model can be based on a principal component analysis of $\Delta(\sigma_K(t) - \sigma_{ATM}(t))$, the daily changes in t-maturity fixed-strike volatility deviations from t-maturity at-the-money volatility. In this way the key risk factors for the volatility skew will be identified and consequently they will be used in an empirical justification for skew scenarios that encompass more change at either or both of the wings. Whether one should change volatilities at the out-of-the-money wing or at the in-the-money wing of the skew, or both, will be shown to depend on the current market conditions.

Principal component analysis of $\Delta(\sigma_K(t) - \sigma_{ATM}(t))$ has given some excellent results. For fixed maturity volatility skews in the FTSE100 index option market during most of 1998, the parallel shift component accounted for around 65-80% of the variation, the tilt component explained a further 5 to 15% of the variation, and the curvature component another 5% or so of the variation. The precise figures depend on the maturity of the volatility (1mth, 2mth or 3mth) and the exact period in time that the principal components were measured. But generally speaking 80-90% of the total variation in skew deviations can be explained by just three key risk factors: parallel shifts, tilts and curvature changes.[20]

[20] For example, the principal component analysis for 3mth implied volatility skew deviations over the whole data period gives the following output. Note that sparse trading in very out-of-the money options implies that the extreme low strike volatilities show less correlation with the rest of the system, and this is reflected by their lower factor weights on the first component.

Component	Eigenvalue	Cumulative R^2
P1	13.3574	0.742078
P2	2.257596	0.8675
P3	0.691317	0.905906

This identification of the important risk factors allows one to quantify the expected movements in the volatility skew as the index moves under different market circumstances. The first stage is to represent fixed-strike skew deviations by three principal components:

$$\Delta(\sigma_K(t) - \sigma_{ATM}(t)) = \omega_{K,1}(t)\ P_1(t) + \omega_{K,2}(t)\ P_2(t) + \omega_{K,3}(t)\ P_3(t) \qquad (10a)$$

The second part of the model employs simple linear regressions of each component P_i (i = 1, 2, or 3) on the daily changes ΔS in the index, viz.:

$$P_i\,(t) = \gamma_{0,i}\,(t) + \gamma_i\,(t)\,\Delta S + \eta_i\,(t) \qquad (10b)$$

where t is the volatility maturity (1mth, 2mth or 3mth). Thus the movements a t-maturity volatility at strike K consequent to a change in index level will be determined by the factor weights $\omega_{K,i}$ and the sensitivities of the key risk factors to

---

Factor Weights

	P1	P2	P3
4225	0.53906	0.74624	0.26712
4325	0.6436	0.7037	0.1862
4425	0.67858	0.58105	0.035155
4525	0.8194	0.48822	-0.03331
4625	0.84751	0.34675	-0.19671
4725	0.86724	0.1287	-0.41161
4825	0.86634	0.017412	-0.43254
4925	0.80957	-0.01649	-0.28777
5025	0.9408	-0.18548	0.068028
5125	0.92639	-0.22766	0.13049
5225	0.92764	-0.21065	0.12154
5325	0.93927	-0.22396	0.14343
5425	0.93046	-0.25167	0.16246
5525	0.90232	-0.20613	0.017523
5625	0.94478	-0.2214	0.073863
5725	0.94202	-0.22928	0.073997
5825	0.93583	-0.22818	0.074602
5925	0.90699	-0.22788	0.068758

index movements, $\gamma_i(t)$ for i = 1, 2, 3. Note that Derman's models are a special case of this model, where there is just one principal component in the representation (10a) and so perfect correlation is assumed between all fixed-strike volatility deviations from at-the-money volatility.

In order to capture the current market conditions, the regressions (10b) have been performed using just one month of the FTSE 100 index data. These regressions were rolled over the whole period from 4th January 1998 to 31st March 1999, and each time the coefficients $\gamma_i(t)$ are recorded, for i = 1, 2, and 3 and t = 1mth, 2mths and 3mths. The statistical significance of these coefficients is as interesting as their actual value. In fact it is the significance levels that provide the important information for risk managers when coming to a decision about which types of risk should be the current focus.

Figure 12: Significance of the Key Risk Factors

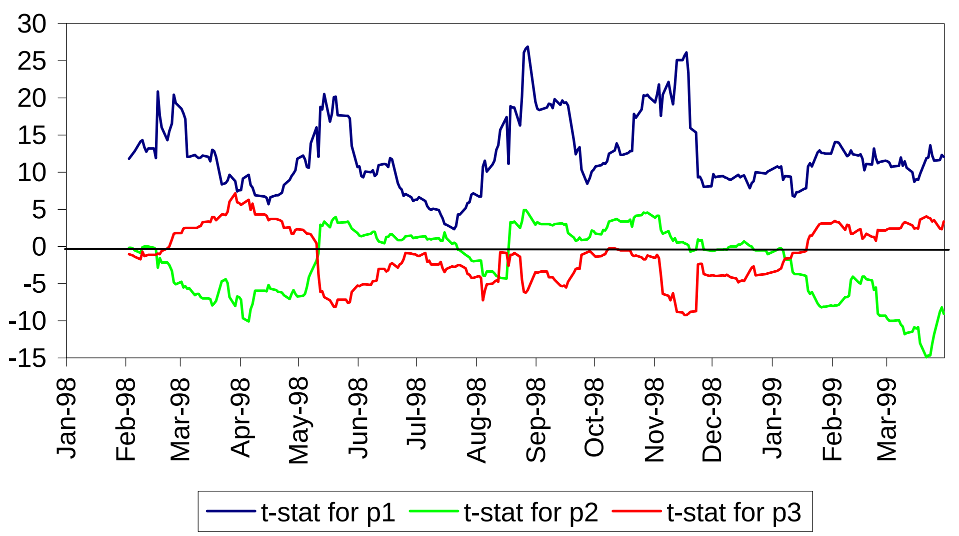


Figure 12 shows the t-statistics on $\gamma_i(t)$ for i = 1, 2 and 3 and t = 1mth from one month rolling regressions (10b). Clearly γ_1, which captures a parallel shift in all fixed-strike volatility deviations, is significant throughout the period, always positive and particularly important during the mini-crash period and the consequent market recovery. But the tilt component γ_2 is much less significant. It is only playing a really important role during the spring of 1998 and again in the spring of 1999. At both these times the tilt has a negative relationship with index moves, indicating that as the index moves up the low strike deviations will decrease and the high strike deviations will increase. It is interesting to see that γ_3, which captures the curvature component of the skew deviations, almost always has the opposite sign to the tilt coefficient. The implication of these observations, for constructing scenarios to model the likely behaviour of the volatility skew as the index moves will now be explained.

Figure 13a: Non-Parallel Shift in Skew Deviations as Price Moves Up

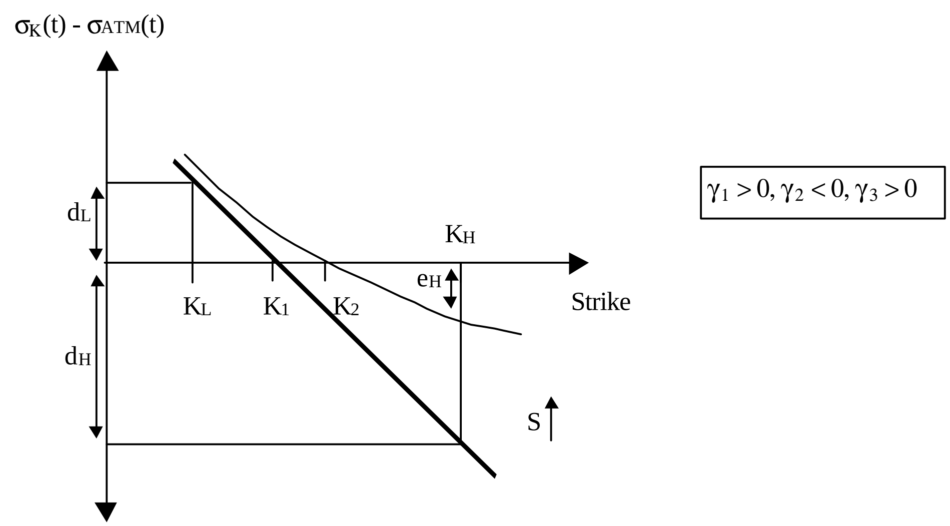


Figure 13a illustrates how the skew deviations move in response to an *upward* movement in the index when $\gamma_1 > 0$, $\gamma_2 < 0$ and $\gamma_3 > 0$. Note that the upward movements in volatility deviations from at-the-money volatility are far greater at high strikes than at low strikes. In fact a result of the upward movement in the index is that one of the high strike deviations, at strike K_2 say, will change from a negative value to a value of zero because the at-the-money strike has moved from K_1 to K_2. Strikes above K_2 will still have volatilities that are lower than the at-the-money volatility, strikes between K_1 and K_2 now have volatilities that are above at-the-money volatility, and strikes below K_1 always have and remain to have volatilities above the at-the-money volatility. For the lowest strikes there will be little change: their volatility deviation from the new at-the-money volatility is about the same as it was before the index move.

Figure 13b: Range Narrowing of Fixed-Strike Volatilities as Price Moves Up

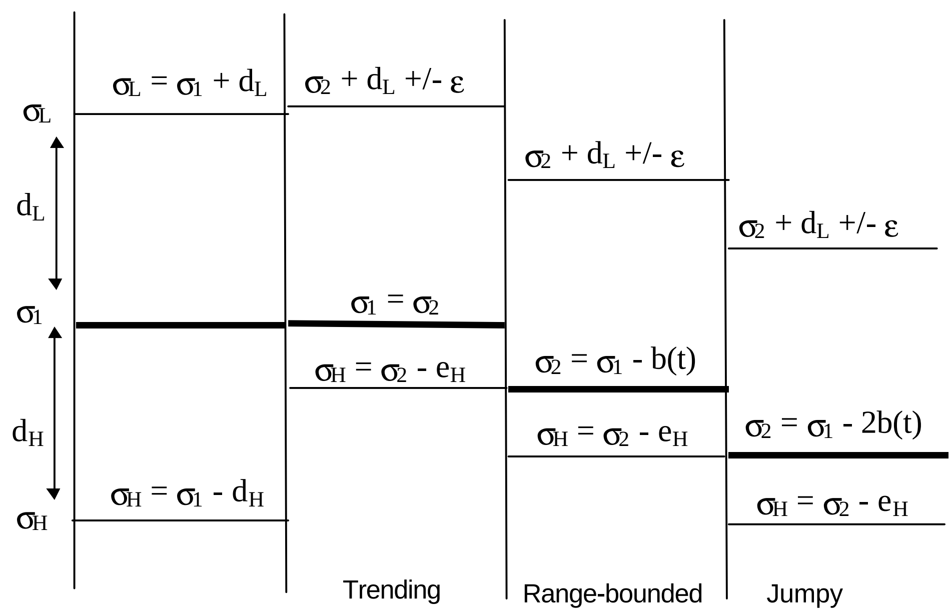


Figure 13b translates the effect of index moves on fixed-strike volatility deviations from at-the-money volatility, into movements in the actual fixed strike volatilities. It is a generalization of figure 10b, using the non-linear model (10) of the skew, to accommodate scenarios that are more general than simple parallel shifts. As before the three volatility regimes are shown according as, after a unit rise in the index level, the new at-the-money volatility σ_2 equals the original at-the-money volatility σ_1 (in a stable trending market), or $\sigma_2 = \sigma_1 - b(t)$ (for a range-bounded market), or $\sigma_2 = \sigma_1 - 2b(t)$ (a jumpy market).

The difference between this figure and figure 10b is that there is no longer a uniform response b(t) for all fixed-strike volatility deviations when the index level changes. In fact figure 13a shows that there will in fact be little change in low-strike volatility deviations from at-the-money volatility, whereas high strike volatility deviations from at-the-money volatility will change considerably. Therefore the range of the skew between K_L and K_H, that is σ_L - σ_H, will become narrower after the rise in index level. Figure 13b shows that it is the current volatility regime that determines whether the movement should occur at the high in-the-money strikes, the low out-of-the-money strikes, or both.

Similar remarks apply to the effect of a *downward* move in the index. It is left to the reader to depict the effect of a unit decrease in the index level on (a) fixed-strike deviations from at-the-money volatility, and (b) fixed-strike volatilities themselves, again when $\gamma_1 > 0$, $\gamma_2 < 0$ and $\gamma_3 > 0$. The net effect is the same as for an upward move in the index: most of the movement in fixed-strike volatilities comes from the low strikes and the high strike volatilities move very little. The range of the skew will widen as the index moves down and the movement will occur at high strikes, low strikes or both depending on the current market regime, just as it does in figure 13b.

From the above discussion it is clear that in a stable trending market regime the low strike out-of-the-money volatilities tend to adjust very little to changes in the index level and most of the movement will come from the higher strikes. But if one refers back to figure 8a it is clear that much of the time the low strike volatilities are moving down and up considerably as the index moves up and down. There is much less movement in the high strike volatilities, except possibly during the mini-crash period in the late summer of 1998. The model has shown that when $\gamma_1 > 0$, $\gamma_2 < 0$ and $\gamma_3 > 0$ there will be less movement in high strike volatilities and more in the low strikes when markets are either range-bounded or jumpy. Therefore these are the regimes that have prevailed for most of the period.

The widening and narrowing effects in the skew are also quite obvious in figure 8a, particularly during the last few months of the data period. At times like this the simple parallel shift scenarios for the skew, as implied by Derman's model, would not be sufficiently general. Instead, the non-linear model (10) can be used to build non-parallel shift skew scenarios as described above, that are more appropriate for these market conditions.

Now consider what happens when the major risk factor is still the trend component, but when the tilt and curvature components of the skew deviations have the opposite influence to that just discussed. That is when $\gamma_1 > 0$, $\gamma_2 > 0$ and $\gamma_3 < 0$. For example, during the mini-crash period of the summer of 1998 it is evident from figure 10 that parallel shifts in skew deviations were the dominant risk factor and the two other types of movement, though much less significant, had both changed signs.

Figure 14a: Non-Parallel Shift in Skew Deviations as Price Moves Up

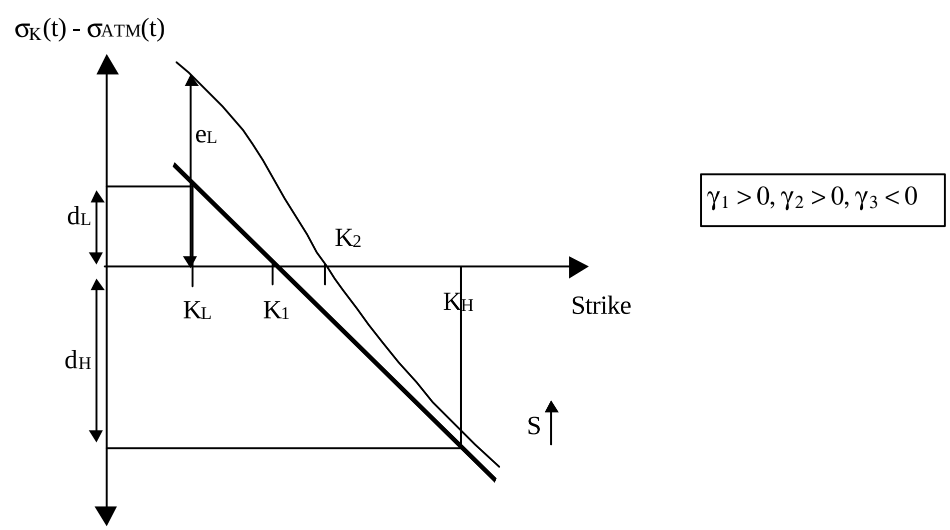


Figure 14b: Range Widening of Fixed-Strike Volatilities as Price Moves Up

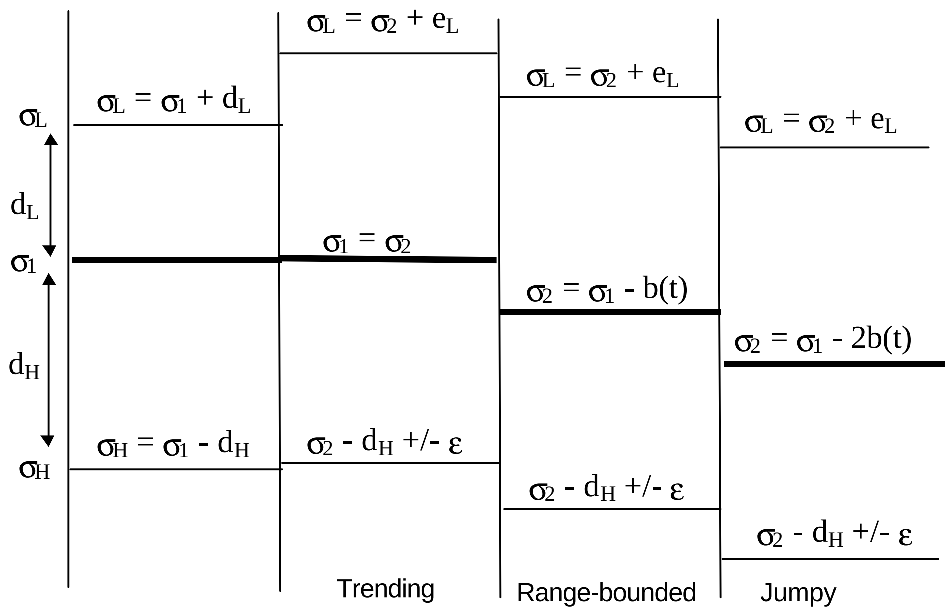


Figure 14 shows the effect of a unit increase in the index level on (a) fixed-strike volatility deviations from at-the-money volatility, and (b) fixed-strike volatilities themselves, when $\gamma_1 > 0$, $\gamma_2 > 0$ and $\gamma_3 < 0$. The net effect from all three principal components is for high strike volatility deviations from at-the-money volatility to change very little, whereas the low strike deviations will increase further. Thus the range of the skew will widen as the index moves up and narrow as the index moves down. The at-the-money volatility response, or equivalently the market regime, will determine whether the movement occurs at low strikes, high strikes or both. In fact, since we have already seen that the parameter values $\gamma_1 > 0$, $\gamma_2 > 0$ and $\gamma_3 < 0$ only occurred during the mini-crash period, it is safe to assume that the jumpy market regime model holds. So when the index level increases high strike volatilities should be adjusted down, about the same amount as the at-the-money volatility. But low strike volatilities should be adjusted less far down, or even upwards.[21] Similarly, if the index level falls, high strike volatilities should be adjusted up, about the same amount

[21] From figure 14b, it is clear that low strike volatilities will move down as the index increases (and up as the index decreases) if and only if $d_L < e_L - 2b(t)$.

as the at-the-money volatility, but low strike volatilities should be adjusted less, and they may even move downwards.

To summarize the modelling procedure: First regression models that are based only on recent market data are used to indicate which volatility regime is likely to prevail in the near future, and the relevant sensitivity of at-the-money volatility to changes in the index level. Then the key risk factors of a volatility skew are quantified by the trend, tilt and curvature components of the deviations of fixed-strike volatilities from at-the-money volatility. The response of fixed-strike volatilities to changes in the index level depends on which of these key risk factors are important in the current market conditions. Typically the trend component will always be the most significant risk factor. If it is the only significant risk factor then the parallel shift scenarios that are implied by Derman's models will apply. But when the tilt or curvature are also significant risk factors, adjustments should be made for greater changes at out-of-the-money volatilities and perhaps also at in-the-money volatilities in the skew. The magnitude and direction of such changes are determined by the sensitivities of the three key risk factors to changes in the index level. When measured, these sensitivities are found to depend very much on the current market regime.

Application of this model to daily data on the FTSE 100 European index option has produced some likely scenarios for FTSE 100 volatility skews and indicated the circumstances in which they should be applied. Typically the range of volatility in the skew with respect to strike will widen as the index level decreases and narrow as the index level increases. When the market is in a range-bounded regime most of the change should be coming from the low strike out-of-the money volatilities. But if a market crash is feared, the high strike in-the-money volatilities will also move in the opposite direction to the index, although not as much as the low strike volatilities.

The general method used here may be applied to other equity index markets and to other types of options, and this is the subject of ongoing research. It is possible that these methods could be used to determine the swaption volatility skew as a function of the key risk factors of cap volatility skews. And the use of three key risk factors in a non-linear model of price-volatility scenarios should be particularly useful in currency option markets, where smile models are unlikely to be based on a linear parameterization.

Summary and Conclusions

It is a common problem in risk management today that risk measures and pricing models are being applied to a very large set of scenarios based on movements in all possible risk factors. The dimensions are so large that the computations become extremely slow and cumbersome, so it is quite common that over-simplistic assumptions will be made. This paper presents an alternative. Large covariance matrices and price-volatility scenarios are generated from only a few key market risk factors that capture the most important independent sources of information in the data.

The first part of this paper shows how orthogonal methods for generating covariance matrices are applied to several different types of asset class: commodity futures prices, yield curves, equity indices and foreign exchange rates. Large covariance matrices that are based on the volatility of a few, independent key market risk factors alone are calculated, and are shown to have many advantages over the other methods in standard use:

- Positive semi-definiteness is assured, without severe constraints such as using the same model parameters for all assets and all markets;
- Stochastic volatility and correlation models such as multivariate GARCH, that have many advantages but that are usually difficult to apply in higher dimensions, may be employed;

- Correlations are more stable because the 'noise' in the system may be measured and, if required, be ignored;
- The method should conform to the standard regulatory requirements on historic data if at least one year of data is used in the principal component analysis;
- Periods of sparse trading on some (but not all) assets do not present a problem because their current volatilities and correlations will be inferred from their historic relationship with the other variables in the system.

The second part of this paper presents an empirical model of price-volatility scenarios that is based on three key risk factors of the implied volatility skew. The analysis is greatly simplified, because it is based on only a few risk factors. But these risk factors are still capturing most of the risk, so there is little loss of accuracy. Non-linear skew parameterizations and non-parallel shift scenarios for the volatility skew are accommodated very easily in this framework. And the empirical nature of the model allows the actual quantification of appropriate moves in the volatility skew as the underlying price changes:

- The price-volatility scenarios that are most likely at any given time will depend on the market regime that is expected to prevail during the risk horizon;
- The model first provides a leading indicator of the expected market regime;
- Then, given the expected regime, and for a given change in underlying price, the model provides a numerical forecast of the most likely change in the at-the-money volatility, and in all fixed-strike volatilities, of any maturity.

To conclude, the method outlined in this paper is computationally efficient because it allows an enormous reduction in the dimension of the scenario set, whilst retaining a very high degree of accuracy in the risk measures and prices obtained. Since the key risk factors are independent, the method is computational efficient even when many factors are used to represent the system. In most cases only a few key factors are necessary, and any movements that are not captured by these factors are ascribed to 'noise' in the system. In fact, by cutting out this noise the model produces risk measures and prices that are more robust. Finally, it is quite straightforward to quantify how much risk is associated with each key factor. So risk managers will be able to focus their attention on the most important sources of risk.

Acknowledgement

I am greatly indebted to Emmanuel Derman, Goldman Sachs, Head of Risk Management in New York for many very useful discussions; and to Jacques Pezier, General Manager of CAL-FP Bank, London for his insightful comments.

References

Alexander C.O. (1996) "Evaluating the use of RiskMetrics™ as a risk measurement tool for your operation : What are its advantages and limitations" *Derivatives Use, Trading and Regulation* 2:3 pp277-285

Alexander, C.O. and C. Leigh (1997) "On the covariance matrices used in Value-at-Risk models" *Journal of Derivatives* 4:3 pp50-62

Alexander, C.O. (1998) "Volatility and correlation: Methods, models and applications" in *Risk Management and Analysis*: *Measuring and Modelling Financial Risk* (C. O. Alexander, Ed) Wileys

Alexander, C.O. (1999) "Correlation and cointegration in energy markets" in *Managing Energy Price Risk* (2nd Edition) RISK Publications pp291-304

Alexander, C.O. (2000) "A primer on the orthogonal GARCH model" Draft in pdf on www.ismacentre.reading.ac.uk

Basle Committee on Banking Supervision (1995) "An internal model based approach to market risk capital requirements"

Bollerslev, T. (1986) "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics* 31 pp 307-327.

Brailsford, T.J. and Faff, R.W. (1996) “An evaluation of volatility forecasting techniques” *Journal of Banking and Finance* 20:3 pp419-438

Derman, E. (1999) "Volatility regimes" *RISK Magazine* 12:4 pp55-59

Derman, E. and M. Kamal (1997) "The patterns of change in implied index volatilities" *Quantitative Strategies Research Notes*, Goldman Sachs

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

Duan, JC (1995) "The GARCH option pricing model" *Mathematical Finance* 5:1 pp13-32.

Duan, JC (1996) ‘Cracking the smile’ *RISK Magazine* 9:12 pp55- 59

Engle, R.F. (1982) "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". *Econometrica* 50:4 pp987-1007.

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 pp213-238

Engle, R.F. and K.F. Kroner (1993) "Multivariate simultaneous generalized ARCH" *Econometric Theory* 11 pp22-150

Engle, R.F. and J. Rosenberg (1995) "GARCH gamma" *Journal of Derivatives* 2 pp47-59

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

Jamshidian, F. and Y. Zhu (1996) "Scenario simulation: Theory and methodology" *Finance and Stochastics* 1:1 pp43-67