The University of Reading

THE BUSINESS SCHOOL
FOR FINANCIAL MARKETS

Autoregressive Conditional Kurtosis

ISMA Centre Discussion Papers In Finance 2002-05
This Version: 22 February 2002

Chris Brooks, Simon P. Burke and Gita Persand

The University of Reading • ISMA Centre • Whiteknights • PO Box 242 • Reading RG6 6BA • UK
Tel: +44 (0)118 931 8239 • Fax: +44 (0)118 931 4741
Email: research@ismacentre.rdg.ac.uk • Web: www.ismacentre.rdg.ac.uk

Director: Professor Brian Scott-Quinn, ISMA Chair in Investment Banking
The ISMA Centre is supported by the International Securities Market Association

Abstract[*]

This paper proposes a new model for autoregressive conditional heteroscedasticity and kurtosis. Via a time-varying degrees of freedom parameter, the conditional variance and conditional kurtosis are permitted to evolve separately. The model uses only the standard Student's *t* density and consequently can be estimated simply using maximum likelihood. The method is applied to a set of four daily financial asset return series comprising US and UK stocks and bonds, and significant evidence in favour of the presence of autoregressive conditional kurtosis is observed. Various extensions to the basic model are examined, and show that conditional kurtosis appears to be positively but not significantly related to returns, and that the response of kurtosis to good and bad news is not significantly asymmetric. A multivariate model for conditional heteroscedasticity and conditional kurtosis, which can provide useful information on the co-movements between the higher moments of series, is also proposed.

Contacting Author:
Chris Brooks
ISMA Centre,
University of Reading
Box 242, Whiteknights, Reading, UK RG6 6BA
Tel: (+44) 118 931 67 68
Fax: (+44) 118 931 47 41
E-Mail: C.Brooks@reading.ac.uk

J.E.L. Classifications: C22, C51, C52
Keywords: conditional kurtosis, GARCH, fourth moment, fat tails, Student's t distribution.

[*] The authors' affiliations are respectively: ISMA Centre, University of Reading; School of Business, University of Reading; Department of Economics, University of Bristol.

1. Introduction

It is an almost universally accepted stylised fact that asset returns are not normally distributed, following early research by Mandelbrot (1963). It is further typical to observe that the non-normality of return distributions arises predominantly as a result of excess kurtosis rather than asymmetry (skewness). This property implies that extreme market movements, in either direction, will occur with greater frequency in practice than would be predicted by the normal distribution. For example, a 5% daily loss is observed to occur in equity markets approximately once every two years, while if returns were normally distributed, such a change would be expected only once in every one thousand years (Johansen and Sornette, 1999), given the estimated return variances. Clearly this is an important observation in finance since, under the normality assumption for returns, variance is widely used as a proxy for market risk. If however, asset returns are fat-tailed, this will lead to a systematic underestimate of the true riskiness of a portfolio, where risk is measured as the likelihood of achieving a loss greater than some threshold.

The standard ARCH and GARCH models introduced by Engle (1982) and Bollerslev (1986) respectively allow normally distributed disturbances to have time varying (conditional) variance. Such models are able to generate data with unconditionally fat tails, but not sufficiently fat to capture all of the observed unconditional leptokurtosis in returns series. Engle and Bollerslev (1986) explore the Gaussian model further, and although conditional kurtosis is not of direct interest to their study, they derive the conditional kurtosis forecasts from a GARCH(1,1), process as a function of the conditional variance. Since the variance is dynamic (time-varying), so is the kurtosis. Their derivation is sufficient to illustrate a key point relevant to this paper: that since the normal distribution is characterised entirely by its first two moments, the behaviour of the kurtosis is entirely determined by that of the variance.

The observation that GARCH models with normal disturbances cannot generate sufficient leptokurtosis to replicate that observed in actual data was in part the motivation for the study of Bollerslev (1987). He developed a more general model that allowed the disturbances to have a transformed *t*-distribution so that extreme values, occurring more commonly than under a normal distribution, may be accommodated. However, whilst such a model can lead to sufficiently fat tails to provide a realistic model for asset returns, the conditional kurtosis from such a model is tied to the conditional variance via a time-invariant degrees of freedom parameter. To introduce some notation, suppose that the conditional distribution of the series of interest, y_t , $t = 1, 2, \ldots, T$, is a transformed central t with conditional mean $y_{t \mid t-1}$, variance $h_{t \mid t-1}$ and degrees of freedom v, and let $f_v(e_t \mid y_{t-1})$ denote the density function for e_t conditional upon all information available to time t-1. A GARCH-t model may thus be written

$$y_t = g_0 + e_t \quad , \quad e_t \mid y_{t-1} \sim f_v(e_t \mid y_{t-1})$$
$$h_{t \mid t-1} = a_0 + a_1 e_{t-1}^2 + a_2\, h_{t-1 \mid t-2} \tag{1}$$

where g_0, a_0, a_1, a_2 are parameters to be estimated.

From a finance perspective, consideration of the higher moments of portfolio return distributions is important to ensure that investors make optimal decisions given their tolerances for risk and, for example, so that fund management or trading rule performance is correctly appraised. Research by Chunhachinda *et al*. (1997), for instance, suggests that the incorporation of moments higher than the second into the investor's portfolio decision causes a major change in the construction of the optimal combination of risky assets. Since higher moment deviations from normality are agreed to be non-negligible, there is no reason to suppose that they should be time-invariant, other than for simplicity, and allowing them to be time-varying may improve their approximation to the actual return distributions. Failure to consider moments higher than the second or assuming that those moments are time-invariant could also lead to avoidably high approximation error. Nelson (1996), for example, plots the standardised residuals exceeding 4 in absolute value from an EGARCH fit to daily S&P 500 data, and finds that large residuals of either sign tend to bunch together through time. He argues that this finding implies evidence for time-varying kurtosis.

Recent research by Harvey and Siddique (1999, 2000) has proposed and employed a model that allows for time-varying conditional skewness, based upon a non-central Student's t-distribution. However, their approach does not model time-varying conditional kurtosis. We would argue that an examination the conditional fourth moment is of more importance than the third given the leptokurtosis that is almost universally observed in financial asset returns, irrespective of the frequency of observation (although excess kurtosis is seen to diminish somewhat with temporal and cross-sectional aggregation of returns). On the other hand, some asset return distributions appear to be negatively skewed (implying a higher probability of negative returns than positive returns of the same order), while others are positively skewed. The degree and sign of the conditional skewness appears also to vary with sampling frequency and for given assets over time (see Harvey and Siddique, 1999, Figure 3B, for an illustration of the latter). Fernandez and Steel (1998) introduce a model for conditional skewness that can be applied to any continuous, unimodal and symmetric distribution. The method is then applied to a set of stock returns, although they examine only unconditional moments and an extension to allow for conditional GARCH-type dependence for the third moment would be non-trivial. Lambert and Laurent (2001) employ the Fernandez and Steel approach in the context of maximum likelihood estimation of a GARCH model. They do not, however, consider conditional third or fourth moments.

Hansen (1994) develops a general model for autoregressive conditional density estimation, centred on a skewed version of the Student's *t* density. The resulting formulation is rather complex and, whilst it allows for conditional skewness, the conditional kurtosis is not explicitly parameterised. Premaratne and Bera (2001) propose the use of a Pearson type IV distribution for the unconditional returns data, which is able to account for both asymmetry and extreme fat tails simultaneously, although they do not fully examine these in a conditional setting that allows the third and fourth moments to be time-varying. The various types of asymmetric conditional density function that are available for GARCH-type modelling are reviewed in Bond (2001).

In this paper, we develop a model for autoregressive conditional kurtosis that permits the kurtosis to develop over time in a fashion that is not fixed with respect to the variance. This is achieved via a central *t*-distribution with time-varying degrees of freedom, so that the variance, degrees of freedom and hence kurtosis all vary over time, in a manner determined by the data. It is this allowance of the degrees of freedom parameter to vary over time that permits the relaxation of the relationship between the variance and kurtosis. It is worth noting at the outset that such a set up with a central *t* cannot produce asymmetries in the return distribution, although incorporation of asymmetry terms in the conditional variance and conditional kurtosis equations can generate asymmetries in the unconditional return distribution. For the reasons above, and for its tractability and ease of subsequent model estimation, the restriction implied by the use of a central rather than a non-central *t* may not be undesirable. As Premaratne and Bera (2001) note, the non-central *t* involves the sum of an infinite series which enters the log-likelihood function and which makes computation exceedingly difficult. The Hansen (1994) approach allows for time-varying skewness but requires the imposition of several *ad hoc* restrictions on the parameter values to permit estimation. These parameter restrictions are lifted in work by Jondeu and Rockinger (2000).

We also propose a multivariate extension to the model that is based on a combination of transformed Student *t* variates. To our knowledge, this is the first study that allows for such cross-dependencies in fourth moments other than recent papers employing copulas (see, for example, Rockinger and Jondeu, 2001).

The remainder of the paper is laid out as follows. Section 2 develops the univariate model for conditional kurtosis and discusses estimation issues, while Section 3 describes the data that we employ to illustrate the model's applicability. Section 4 discusses the univariate results and their interpretation and Section 5 extends the model to a multivariate setting. Finally, Section 6 concludes and offers suggestions for further research.

2.1 Univariate Model Development

In order to obtain a model with freely varying conditional kurtosis as well as conditional variance, an approach based on that of Bollerslev (1987), but extended, is employed. Let e_t, $t=1,2,\ldots,T$, be independently distributed as central Student's *t* variates with n_t degrees of freedom. Extending the work of Bollerslev (1987), consider a time varying transformation of e_t, to result in a new process that may have any desired variance h_t and kurtosis k_t. Let this be given by

$$e_t^* = l_t e_t . \quad (2)$$

The e_t^* are the analogues of the disturbances of a *t*-GARCH model. Their conditional variance, $h_t = m_{2,t}$ and conditional fourth moment, $m_{4,t}$, are given respectively by

$$m_{2,t} = l_t^2 \frac{n_t}{n_t - 2};\tag{3a}$$

$$m_{4,t} = l_t^4 \frac{3n_t^2}{(n_t - 2)(n_t - 4)}.\tag{3b}$$

Equations (3) arise from the moment generating function for a central *t*, where all odd moments are by definition zero. Rearranging equation (3a) gives the time-varying transformation as a function of the conditional variance and the time-varying degrees of freedom

$$l_t = \left(\frac{h_t (v_t - 2)}{v_t} \right)^{1/2}.\tag{4}$$

Defining $k_t = m_{4,t} / h_t^2$, and substituting (4) into (3b) gives the conditional kurtosis as a function of the degrees of freedom at time *t* as

$$k_t = \frac{3(n_t - 2)}{(n_t - 4)}.\tag{5}$$

Equation (5) can be rearranged to give the degrees of freedom as a function of the conditional kurtosis

$$n_t = \frac{2(2k_t - 3)}{k_t - 3}.\tag{6}$$

and (6) can be substituted into (4) to give

$$l_t = \left(\frac{k_t h_t}{2k_t - 3} \right)^{1/2}.\tag{7}$$

Equations (5) to (7) show that the conditional variance and conditional kurtosis are not tied together in a fixed fashion and may thus vary freely over time, since k_t is a function of v_t only while h_t also depends on l_t. Since they are not functionally related, the terms h_t and k_t may be parameterised individually as desired. In order to estimate the parameters of these processes, note that the Jacobian of the transformation $e_t^* / l_t = e_t$ is

$$J = \frac{\partial e_t}{\partial e_t *} = \frac{1}{l_t}\tag{8}$$

Taking the Student's *t* density for e_t, substituting $e_t^* / l_t = e_t$ and multiplying by the Jacobian gives the density of e_t^*:

$$f\left(e_t^*\right) = \frac{1}{l_t} \frac{\Gamma\left[\left(n_t + 1\right)/2\right]}{p^{1/2} n_t^{1/2} \Gamma\left[n_t/2\right]\left(1 + e_t^{*2} / l_t^2 n_t\right)^{\left(n_t+1\right)/2}} . \tag{9}$$

Substituting for l_t in (9) and taking logarithms yields the log-likelihood for the t^{th} observation (dropping constant terms):

$$\begin{aligned} l_t = \log\left(\Gamma\left[\left(n_t + 1\right)/2\right]\right) - \log\left(\Gamma\left[n_t/2\right]\right) - \frac{1}{2}\log\left(h_t\right) - \frac{1}{2}\log\left(n_t - 2\right) \\ - \frac{n_t + 1}{2}\log\left(1 + \frac{e_t^{*2}}{h_t\left(n_t - 2\right)}\right) \end{aligned} \tag{10}$$

where v_t is a function of k_t, as given by (6). Maximisation of the log-likelihood function, $\sum_{t=1}^{T} l_t$, yields the maximum likelihood estimates of all the parameters of the model. From (5), there exists a degrees of freedom restriction,

$$n_t > 4$$

generated by the requirement for the existence of a fourth moment and hence equation (5) also implies that $k_t \rightarrow 3$ as $v_t \rightarrow \infty$ and $k_t \rightarrow \infty$ as $v_t \rightarrow 4$ from above.

Many specifications of the variance and kurtosis equations are conceivable, most obviously they may be of the typical GARCH type, namely,

$$h_t = a_0 + a_1 e_{t-1}^{*\ 2} + a_2 h_{t-1} , \tag{11a}$$

$$k_t = b_0 + b_1 \frac{e_{t-1}^{*\ 4}}{h_{t-1}^2} + b_2 k_{t-1} , \tag{11b}$$

all coefficients being non-negative.

To summarise, the model may be termed generalised autoregressive conditional heteroscedasticity and kurtosis (GARCHK), and may be described by the following equations.

$$y_t = g_0 + e_t^* ; \tag{12a}$$

$$e_t^* = l_t e_t , \; e_t \sim t_{n_t} ; \tag{12b}$$

$$h_t = a_0 + a_1 e_{t-1}^{*\ 2} + a_2 h_{t-1} ; \tag{12c}$$

$$k_t = b_0 + b_1 \frac{{e^*_{t-1}}^4}{h^2_{t-1}} + b_2 k_{t-1}; \tag{12d}$$

$$n_t = \frac{2(2k_t - 3)}{k_t - 3}. \tag{12e}$$

$$l_t = \left(\frac{h_t (v_t - 2)}{v_t} \right)^{1/2}; \tag{12f}$$

All parameters are estimated using quasi-maximum likelihood with the BFGS algorithm.

2.2 Development of a Test for Autoregressive Conditional Kurtosis

At first blush, it may appear sensible to attempt the formulation of test for autoregressive conditional kurtosis in a similar vein to Engle's TR^2 Lagrange Multiplier test for conditional heteroscedasticity, defining, $y_t = \frac{\hat{e}_t^4}{\hat{h}_t^2}$, and regressing it on p of its lagged values to test for autoregressive conditional kurtosis of order p. However, such an approach is inappropriate in the context of a student-t density since the TR^2 approximate form for the LM statistic depends crucially on an assumption of conditionally normal disturbances. Therefore, in order to test whether there is evidence of autoregressive conditional kurtosis in the data, we revert to the application of a likelihood ratio test to the relevant estimated model parameters (b_1 and b_2 in (12d)). The development of a form for the test that does not require estimation of the model is left to future research.

2.3 Extensions of the Basic GARCHK Model

There are several natural extensions of the model given by equations (12) that arise from an examination of the comparable GARCH literature. The simplest of these would be to increase the number of lags of the fourth power of the standardised error, ${e^*_{t-1}}^4 / h^2_{t-1}$, and of the conditional kurtosis in the same way that a GARCH(1,1) can be extended to a GARCH(p,q). We experimented with models of a higher order in the conditional kurtosis equation but all additional terms could be restricted to zero under a likelihood ratio test. This indicates that one lag of each of the fourth power of the standardised error and of the conditional kurtosis is sufficient to capture all of the autoregressive conditional kurtosis in the data.

The second obvious extension to make to the GARCHK model is to add additional terms to the conditional variance and conditional kurtosis equations that permit the next period values of these quantities to have asymmetric responses to the signs of the innovations, in the style of Glosten *et al*. (1993). This could be viewed as an alternative parameterisation for the skewness in the unconditional return distribution. It would also, of course, be plausible to specify logarithmic formulations for the conditional variance and kurtosis equations in the manner of Nelson (1991), although this possibility is not pursued here. Equations (12c) and (12d) thus become

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^{*\,2} + \alpha_2 h_{t-1} + \alpha_3 I_{t-1} e_{t-1}^{*\,2}; \tag{12c′}$$

$$k_t = \beta_0 + \beta_1 \frac{e_{t-1}^{*\,4}}{h_t^2} + \beta_2 k_{t-1} + \beta_3 I_{t-1} \frac{e_{t-1}^{*\,4}}{h_t^2}; \tag{12d′}$$

where $I_{t\text{-}1}$ is an indicator function taking the value 1 if $e_{t\text{-}1}^*$ is negative and zero otherwise, with all other parts of equations (12) remaining unchanged. Clearly, it would be possible to include the asymmetry term in either the conditional variance equation or the conditional kurtosis equation or both, and we opt for the latter.

Engle *et al*. (1987) suggested that investors usually require compensation in the form of additional returns for taking on additional risk. Thus, a third intuitive extension of the GARCHK model would therefore be to allow the current return to depend on the current value of the conditional kurtosis as well as on the conditional standard deviation. Thus, equation (12a) becomes

$$y_t = \gamma_0 + \gamma_3 \frac{h_t^{1/2}}{\bar{y}} + \gamma_4 k_t + e_t^*; \tag{12a′}$$

where $\bar{y}$ is the mean value of y_t. Standard GARCH-M formulations use $h_t^{1/2}$ in the conditional mean. This has units of returns but k_t is unitless, so the conditional standard deviation is divided by the mean of the returns in order to ensure that the conditional standard deviation and conditional kurtosis terms in mean have the same scale. The scaled conditional standard deviation could also be viewed as the conditional coefficient of variation. We employ the mean value of y as the divisor rather than the current value, y_t, since the latter takes on zero values for some of the sample observations. Overall, therefore, both variables in (12a′) are unitless so that the coefficients g_3 and g_4 both have units of returns.

2.4 Residual Standardisation and Moment Specification Testing

Given the model, if it is capturing all of the relevant features of the data, certain moment relationships should hold on an appropriately standardised measure of the residuals. These can be tested on their sample moments. Effectively, the tests are of non-linear restrictions on the parameters of the model given the data, and can be regarded as mis-specification tests. The relevant forms for the tests will now be presented.

Let q denote a $k \times 1$ parameter vector (containing all model parameters) and $r(q)$ denote a $J \times 1$ restrictions function. The form of test used is the Wald test. That is, the unrestricted model is estimated, the moment restrictions not being imposed, and the (joint) significance of departures from the restrictions calculated using the freely estimated parameters, tested. The null hypothesis of the test is:

$$H_0: \quad r(q) = 0. \tag{13}$$

The test statistic is of the usual form. Let

$$\Omega = \mathrm{var}\left(r\left(\hat{q}\right)\right). \tag{14}$$

Then the test statistic is given by

$$W = r\left(\hat{q}\right)^T \hat{\Omega}^{-1} r\left(\hat{q}\right). \tag{15}$$

Under the null, $W \overset{a}{\sim} c_J^2$, and so the null hypothesis is rejected if $W > c_{J,a}^2$, where a is the size of the test.

Consider first the case where a single (J=1) moment restriction is applied. This would take the form

$$r(q) = \frac{1}{T}\sum_{t=1}^{T} m(y_t), \tag{16}$$

where y_t is the t^{th} observation. There will thus be a contribution from each observation (excluding the division by the sample size), of $m(y_t)$. Substituting the estimated parameters:

$$r(\hat{q}) = \frac{1}{T}\sum_{t=1}^{T}\hat{m}(y_t) \tag{17}$$

It remains to obtain the variance of $r(\hat{q})$ so that the test statistic (15) may be constructed. Greene (2000) shows that this may be calculated using the derivatives of the log-likelihood function. Let the t^{th} contribution to the log-likelihood function be $l_t(q)$, so that the log-likelihood itself is

$$l(q) = \sum_{t=1}^{T} l_t(q)\,. \tag{18}$$

Now let q_i be the i^{th} element of the parameter vector, q. Denote the derivative of $l_t(q)$ with respect to q_i by

$$\frac{\P l_t(q)}{\P q_i} = d_{t,i}(q)\,. \tag{19}$$

The estimated analogues are

$$\left.\frac{\P l_t(q)}{\P q_i}\right|_{q=\hat{q}} = d_{t,i}(\hat{q})\,. \tag{20}$$

These derivatives, one for each observation, may be stacked into a column,

$$\hat{d}_i = \begin{pmatrix} d_{1,i}(\hat{q}) \\ \vdots \\ d_{T,i}(\hat{q}) \end{pmatrix}. \tag{21}$$

The columns for each *i*, *i*=1,2,...,*k* may then be put together into a $T \times k$ matrix of derivatives,

$$D = \begin{bmatrix} \hat{d}_1 & \cdots & \hat{d}_k \end{bmatrix} = \begin{bmatrix} \hat{d}_{1,1} & \cdots & \hat{d}_{1,k} \\ \vdots & \vdots & \vdots \\ \hat{d}_{T,1} & \cdots & \hat{d}_{T,k} \end{bmatrix}, \tag{22}$$

where the $(t,i)^{th}$ element, $\hat{d}_{t,i}$, is $d_{t,i}(\hat{q})$, the derivative of the t^{th} contribution to the log-likelihood with respect to the i^{th} element of the parameter vector. This is like a time series regressor data matrix and is used as such below. The variance-covariance matrix in (15), $\hat{\Omega}$ (a scalar in this one moment restriction case), may then be calculated from the residual sum of squares of the regression of $\hat{m}$ (the values of elements of the moment restriction function by observation) on $\hat{d}_1$, $\hat{d}_2$, ..., $\hat{d}_k$, the values of each of the derivatives of the log-likelihood, observation by observation. This residual sum of squares of this regression is given by

$$R = \hat{m}^T \hat{m} - \hat{m}^T D\left(D^T D\right)^{-1} D^T \hat{m}. \tag{23}$$

The required variance is then

$$\hat{\Omega} = \frac{R}{T^2}. \tag{24}$$

Returning now to the multiple moment restriction case, generalisation is achieved by replacing a single column of observation by observation values of a moment restriction by columns generated using all of the moment restrictions. The $J \times 1$ restrictions function $r(q)$, might be written,

$$r(q) = \begin{bmatrix} r_1(q) \\ \vdots \\ r_J(q) \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} m_1(y_t) \\ \vdots \\ \frac{1}{T}\sum_{t=1}^{T} m_J(y_t) \end{bmatrix}. \tag{25}$$

Substituting estimated parameters gives

$$r(\hat{q}) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} \hat{m}_1(y_t) \\ \vdots \\ \frac{1}{T}\sum_{t=1}^{T} \hat{m}_J(y_t) \end{bmatrix}. \tag{26}$$

The estimated moment values are given by a matrix whose j^{th} column is

$$\hat{m}_j = \begin{pmatrix} \hat{m}_j(y_1) \\ \vdots \\ \hat{m}_j(y_T) \end{pmatrix}, \tag{27}$$

This is the vector of observation by observation contributions to the j^{th} moment restriction function before division by sample size. Let M be the following $T \times J$ matrix

$$M = \left[\hat{m}_1 \quad \cdots \quad \hat{m}_J\right] = \begin{bmatrix} \hat{m}_1(y_1) & \cdots & \hat{m}_J(y_1) \\ \vdots & \vdots & \vdots \\ \hat{m}_1(y_T) & \cdots & \hat{m}_J(y_T) \end{bmatrix}. \tag{28}$$

For convenience, write the $(t,j)^{th}$ element of this matrix as

$$\hat{m}_{t,j} = \hat{m}_j(y_t). \tag{29}$$

Then $\hat{m}_{t,j}$ is the contribution from the t^{th} observation to the j^{th} moment restriction function, before division by the sample size. As already stated, these elements are not all required to be zero, but the j^{th} moment restriction states that the sum of the $\hat{m}_{t,j}$ across observations, t, should be zero.

The variance-covariance matrix in (24), $\hat{\Omega}$ (now a $J \times J$ matrix), is a function of what is in effect, the residual sum of squares matrix (residual sums of squares and cross products of residuals) from the set of regressions of each of the J ($T \times 1$) columns $\hat{m}_j$, j =1, 2, ..., J on all of the log-likelihood derivative columns, $\hat{d}_1$, ..., $\hat{d}_k$. This residual sum of squares matrix can be written, analogously to equation (13) as

$$R = M^T M - M^T D\left(D^T D\right)^{-1} D^T M\,, \tag{30}$$

and the required variance-covariance matrix as

$$\hat{\Omega} = \frac{1}{T^2} R\,. \tag{31}$$

The intuition that can be gleaned from the test is that the values of the moment functions, $r_j\left(\hat{q}\right)$, are well determined if their observation by observation contributions are well explained by those of the derivatives of the log-likelihood function. That is, atypical restriction values must tend to be accompanied by atypical score values, in some direction. For the value of the restrictions functions to be well determined, deviations of individual observation contributions (around their average values) must correspond to deviations of the contributions to the derivatives of the log-likelihood function (the score) around their average of zero. This will lead to a small value of the variance-covariance matrix. The value of the test statistic will then be larger, and hence more likely to lead to rejection of the moment restrictions. Low correlation between moment restriction values and the score constitutes poorer determination of the value of the moment restrictions, whatever that value, and thus a reduced probability of rejection of the moment restrictions given a fixed value of the restrictions function.

The construction of analytical derivatives for a complex model such as that outlined above has not yet been achieved, so instead, all are based on the numerical approximation

$$\frac{\partial g(y)}{\partial y_i} \approx \frac{g(y_1,\ldots,y_i+d,\ldots,y_n)-g(y_1,\ldots,y_i-d,\ldots,y_n)}{2d}, \qquad (32)$$

for sufficiently small d. The latter parameter is chosen following Greene (2000) as max[10^{-5},abs(q_i)$\times 10^{-5}$].

The only remaining issue is to determine the appropriate standardised form of the residuals that should be used as the basis for the test. It should be evident from the specification of the model of (12) that taking the residuals from the estimated GARCHK model and dividing them by the square root of the conditional variance, i.e. forming $\hat{e}_t^* \big/ \hat{h}_t^{1/2}$, will in general not provide an independently distributed series. Such a procedure would be inappropriate since it would ignore any conditional kurtosis that was present in the data. Instead, a more appropriate approach would be to take the estimated residual from the model and to divide it by the contemporaneous estimated value of the transforming variable: $\hat{e}_t^* \big/ \hat{l}_t$. Under correct model specification, the result will be a standardised measure that will be an independently distributed *t*-variate with v_t degrees of freedom. From (5), it should also be clear that as the degrees of freedom increase towards infinity, the conditional kurtosis reduces to three and from (12e), $l_t \to h_t^{1/2}$, and hence this new standardised residual becomes the usual measure employed to test GARCH model effectiveness, $\hat{e}_t^* \big/ \hat{h}_t^{1/2}$.

The quantity $\hat{z}_t = \hat{e}_t^* \big/ \hat{l}_t$ can then be used to apply conditional moment-based specification tests of the form proposed by Newey (1985) and Nelson (1991), and in a similar vein to those examined by Harvey and Siddique (1999); see also Greene (2000, pp493-496). The orthogonality conditions examined in this paper are as follows

$E[z_t] = 0$

$E[z_t \cdot z_{t-j}] = 0$ for $j = 1, 2, 3, 4$

$E[\{z_t^2 - (v_t / (v_t - 2))\} \cdot \{(z_{t-j}^2 - (v_{t-j} / (v_{t-j} - 2))\}] = 0$ for $j = 1, 2, 3, 4$

$E[z_t^3 \cdot z_{t-j}^3] = 0$ for $j = 1, 2, 3, 4$

$E[\{z_t^4 - (3v_t^2/((v_t-2)(v_t-4)))\} \cdot \{z_{t-j}^4 - (3v_{t-j}^2/((v_{t-j}-2)(v_{t-j}-4)))\}] = 0$ for $j = 1, 2, 3, 4$

Thus there are a total of 17 conditions, which are evaluated separately and via a joint test.

3. Data

The data employed in this study comprise four financial time series – two equity indices (US Standard and Poor's 500 and UK FTSE 100), and two bond indices (UK, US). The equity indices are total return indices where dividends have been added back to calculate the index values, while the bonds are both 10-year maturity benchmark bond indices. The daily sample spans the period 2 January 1990 to 14 June 2000, implying a total of 2727 observations. The series are transformed into continuously compounded returns by taking the natural logarithms of the price relatives in the usual fashion.

Panel A of Table 1 presents summary statistics for the 4 series. Clearly, whilst all of the returns series show statistically significant evidence of leptokurtosis, the degree of unconditional skewness varies from one series to another. The FTSE 100 index returns are positively skewed but not significantly so. On the other hand, the S&P500 returns and the US bond returns are significantly negatively skewed while the UK bond returns are negatively skewed but not significantly so. In all cases, the null hypothesis that the unconditional return distributions are normal is rejected convincingly. An application of the Ljung-Box Q* test using 5 lags of the returns suggest reasonable evidence of autocorrelation in the conditional mean. Closer examination of the autocorrelation and partial autocorrelation functions (not shown but available from the authors on request) suggested a first order AR (for UK equities, and both bond series) or a first-order MA (for the US equities) model as sufficient to capture this linear dependence. Engle's (1982) LM test for ARCH is suggestive of highly significant volatility clustering effects in each case.

4. Results

Panel B of Table 4 gives results for estimation of Bollerslev's GARCH-*t* model with time-invariant kurtosis and degrees of freedom. The estimated degrees of freedom in each case can be calculated from a time-invariant version of (12e), and are 5.36, 6.45, 6.15, and 5.56 for the US equities, UK equities, US bonds and UK bonds respectively. These values together with those for the conditional variance equations are plausible and in line with the conclusions of previous research and are again indicative of the fatness of the tails of the return distributions in each case. All series except the US bond returns demonstrate strong persistence of shocks to the conditional variance, as demonstrated by the closeness of the sum of a_1 and a_2 to unity.

4.1 Results for Basic Model

Table 2 presents the results from estimating the GARCHK model separately for each of the 4 series. First, it is evident that all of the conditional variance and conditional kurtosis coefficient values are positive as required. The parameter estimates concerning the conditional variance equation are entirely as expected: the persistence of shocks to volatility is high in most cases, with a higher coefficient value on the lagged conditional variance than the lagged squared error. The lagged kurtosis coefficient values are significant for all series, while the coefficients on the lagged fourth powers of the standardised error are only significant for the two equity series at the 5% level, and for the US bond returns at the 10% level. The a_1 coefficient has often been termed the "volatility of variance" parameter in the GARCH literature, since it measures how much the conditional variance will move around over time in response to innovations. The corresponding coefficient in the conditional kurtosis equation, b_1, could usefully be termed the "volatility of kurtosis" and would similarly measure the stability of the fitted conditional kurtosis in response to innovations. It is clear, for all four series, that the volatility of kurtosis is considerably greater than the corresponding volatility of variance parameters, suggesting that the fitted kurtoses are likely to be far less smooth than the fitted variances. The conditional kurtosis coefficients also show a high degree of persistence for the US bond series, although the persistence is overall far less apparent than in the corresponding conditional variance equations. In the case of the kurtosis equations, much of the persistence results from the lagged standardised innovation term than the lagged conditional kurtosis, while almost all of the persistence in the conditional variance comes from the lagged conditional variance terms.

The penultimate column of Table 2 presents the maximal value of the log-likelihood function for the model with GARCH and GARCHK. This can be compared with the last column of Panel B in Table 1 which gives the LLF values for the restricted model where $b_1 = 0$ and $b_2 = 0$. The final column presents the implied likelihood ratio test value for the restriction of the GARCHK model to have fixed kurtosis over time. This statistic will follows a $\chi^2(2)$ under the null hypothesis, with critical value 5.99 at the 5% level. Restrictions of the conditional kurtosis to be fixed are resoundingly rejected for all series except the UK bond returns, suggesting that there seems to be less evidence of conditional kurtosis in the latter, in contrast with the highly significant lagged conditional kurtosis term. This may be indicative, in the context of the complex non-linear optimisation that is required, that the overall model fit has only improved marginally upon inclusion of the b_1 and b_2 parameters as a result of a reduction in the goodness of fit of other parts of the model (although it is not apparent from the results which parts).

For illustration, Figure 1 plots the fitted conditional variance and conditional kurtosis obtained for the FTSE returns by estimating equations (12). Although there are periods when the both the conditional variance and conditional kurtosis are both high, it is evident that the model has succeeded in enabling the kurtosis to develop a different dynamic pattern. In general, the conditional kurtosis is much more stable for most of the time than the variance, but has periods where it rises very substantially. Thus the conditional kurtosis appears to be fitting to the extreme events, as one may expect. Both the conditional variance and conditional kurtosis take on high values in the early 1990's, but have historically low values for the 1993-96 period, before rising again towards the end of the sample.

Figure 2 graphs the estimated degrees of freedom over time for the FTSE returns. The degrees of freedom never rises above 15 for the whole sample period, with an average of around 8, further indicating the fatness of the unconditional return distribution. Mirroring the conditional kurtosis, the degrees of freedom rises during the mid-1990's as the return distribution's tails thin, before falling considerably in the late 1990's, when it spends most of its time in the 4-8 range. It is also worth noting, however, that the typical range of values that the degree of freedom parameter takes over time are higher than the fixed unconditional estimate of 6.4 from the Bollerslev *t* model. This is suggestive that forcing the degree of freedom parameter to be fixed over time will lead to a measured tail fatness that is greater than the values it would take for most of the sample period if it were allowed to vary. Figure 3 plots the time-varying transformation parameter for the UK stock returns. The extent to which this varies over time is a measure of the extent to which the relationship between the conditional variance and the degrees of freedom varies over time. From (12f), it is evident that the transformation series must be non-negative. The value of l_t is also less than unity for the majority of the sample period, and so from equation (12b), the transformation results in a shrinking of the *t*-variate relative to one that follows an untransformed *t* with v_t degrees of freedom.

For comparison, Figures 4 to 6 replicate Figures 1 to 3, but for the US Treasury bond series. First, considering Figures 1 and 4, it is evident by examining the ranges of values that the conditional variance and conditional kurtosis take, that the conditional variance is typically lower for the US bond series than for the UK equities, although the higher conditional kurtosis has peaks that are higher in the bond case. Unlike the equity series, the bond returns were relatively stable in the early 1990's and at the end of our sample, while the conditional variance and conditional kurtosis rose substantially in the mid 1990's. Figure 4 scales very differently from Figure 1. In the former case, one very large value of the kurtosis relative to its typical values has considerably changed the scaling, while there are no such extreme estimates for the US bonds. The fitted degree of freedom parameter for the US bonds (Figure 5) remains within a narrow range of values (4,7.2) for the whole sample period, leading the time-varying transformation parameter to also lie within a narrower range of values than was the case for UK equities. No restrictions are placed on the model that constrain the value of the degree of freedom parameter, other than that it must be greater than 4 at all times.

4.2 Results from Extensions of the Basic GARCHK Model

The results of estimating equations (12a), (12b), (12c′), (12d′) and (12e) are presented in Table 3. This is the GARCHK model with asymmetries in the conditional mean and conditional kurtosis equations. Considering first the conditional variance asymmetry parameters, a_3, they are statistically significant at the 1% level for all four series. Such results are consistent with previous studies of asymmetry in volatility. The asymmetries in conditional kurtosis appear to be, from a consideration of the *t*-ratios, considerably less significant than those in conditional variance in all cases, although they do always have the expected positive sign. Thus, it can be said that negative innovations lead to higher (although not significantly so) future values of kurtosis than positive innovations of the same magnitude. The coefficient values on the asymmetry terms are bigger in the conditional kurtosis equation, ranging from 0.01 to 0.6, indicating that the asymmetric effects are larger in magnitude, although not significant, compared with those in the conditional variance equation. The last two columns of Table 3 give the maximal values of the log-likelihood function for the GARCHK model with asymmetries in the conditional variance and conditional kurtosis equations, and the results of a likelihood ratio test of the restriction that the asymmetry terms in both equations are jointly zero. Such a restriction is rejected marginally for the US equity and UK bond series but not for the UK equities or US bonds. It is clear that asymmetries are very much stronger in conditional variance than conditional kurtosis, and that the results of the latter dilute those of the former in the joint test. The asymmetry terms in the conditional variance can capture the skewness in the unconditional distribution of returns, while it is less clear what the impact of asymmetry terms in the kurtosis equation would be on the unconditional distribution.

In the context of equity markets, such asymmetries have been attributed to leverage effects – see, for example, Black (1976) or Christie (1982). The argument goes that as equity values fall, the relative weight attached to debt in a firm’s capital structure rises, *ceteris paribus*. This induces equity holders, who bear the residual risk of the firm, to perceive the stream of future income accruing to their portfolios as being relatively more risky. An alternative view is provided by the 'volatility-feedback' hypothesis. Assuming constant dividends, if expected returns increase when stock price volatility increases, then stock prices should fall when volatility rises. Whilst one cannot appeal to such explanations of leverage effects in the context of asymmetries in other financial asset return time series, there is equally no reason to suppose that such asymmetries do not exist. In the context of exchange rates, for example, it is possible that good news for one currency will have a differential impact on the perceptions of investors compared with an equivalent amount of good news for the other.

The results from estimating equations (12a′) together with (12b)-(12e) are presented in Table 4. This is the model with conditional variance and conditional kurtosis terms in the mean equation and asymmetries in the conditional variance and conditional kurtosis equations. The signs of the coefficients on both the conditional standard deviation and the conditional kurtosis terms in the mean equations are positive in all cases, suggesting that investors do require additional returns for accepting additional risk. Risk in this sense comprises both standard deviation and kurtosis, neither of which would be considered desirable by any rational investor if some fairly weak assumptions are made concerning the shape of the utility function (see Scott and Hovarth, 1980). Considering the conditional standard deviation terms in the mean equation however, none of these are significant, even at the 10% level. The conditional kurtosis terms in the mean equations are also insignificant, although the parameter estimates for the conditional kurtosis in mean are of the same order of magnitude as those of the conditional standard deviation in mean. It can also be seen from a comparison of Table 4 with Table 3 that the intercept coefficients in the conditional kurtosis equations in most cases rise, and the values of the asymmetry coefficients in the conditional kurtosis equation decrease upon adding the mean terms. This may be taken as evidence that the models excluding such terms in the mean are mis-specified, and that this failure to adequately parameterise the mean is resulting in misleading estimates of the higher moment dynamics. The penultimate column of Table 4 presents the values of the log-likelihood function for the model including conditional mean and asymmetry terms. Likelihood ratio tests of restrictions that neither the scaled conditional standard deviation nor the conditional kurtosis terms appear in the conditional mean are given in the last column of Table 4. The result for both tests is a rejection at the 1% level for the S&P returns but non-rejections for the other three series. It turns out that a re-estimation of the model of Table 4 but excluding the AR or MA terms in the conditional mean results in statistically significant (and very slightly larger) coefficient values on the conditional standard deviation and conditional kurtosis terms in the mean equation. Since, in order for these coefficients to be interpretable as risk premia, there should not be lagged returns or an MA term in the mean, it should not be concluded that conditional kurtosis is not priced in the market. Collectively, these results suggest that it is not clear whether a better measure of total risk may be obtained by using a composite statistic including standard deviation and kurtosis, than by using standard deviation alone.

4.2 Results of Moment-Based Specification Tests

The results from an application of the moment-based specification tests described in Section 2.4 are given in Table 5 for the most general of the models that we consider, whose coefficient values were described in Table 4. There are 17 conditions in all. The first five conditions in Table 5 examine the specification of the conditional mean of the standardised residuals, the next 4 examine the conditional variance, the next 4 the conditional third moment, and the last 4 the conditional fourth moment. If the model has captured all of the dynamic features of the first four moments of the returns series, none of the test statistics should be significantly different from zero. Table 5 suggests some evidence of further structure in the conditional mean for the S&P and for both of the bond series. The covariance of the standardised residuals with their first lags are statistically significant only for the two bond series, while the covariance with the third lag is significant for the US equities. This evidence of further linear structure is not consistent with the acf and pacf results obtained from the raw (i.e. the original unstandardised) returns, but may be symptomatic of the difficulty in estimating the model also containing conditional standard deviation and conditional kurtosis terms in the mean equation. There is evidence that the dependence in the conditional variance has not been fully captured in the cases of the S&P 500, and the US and UK bond series. However, there is virtually no evidence for temporal dependencies in the conditional third moment that would give motivation for the consideration of a conditional skewness model. There is also no evidence of remaining unparameterised temporal structure in the conditional fourth moment. The joint test of significance for all 17 conditions shows a rejection in all four cases as a result of the small numbers of significant individual moment statistics. Thus, overall, there still remain some features of the data that have apparently not been fully captured by the model, consistent with the results of Harvey and Siddique (1999). This should not, however, be taken as evidence that the model proposed here is less well specified than those described in previous studies. Rather, since virtually none of the studies in the GARCH literature have employed specification tests, our results may be indicative that standard GARCH(1,1,) models are not sufficient to fully capture the temporal dependencies in financial asset return series. Obtaining a model that is able to capture such features in their entirety is an open question for future research.

5. Development of a Multivariate Generalised Autoregressive Conditionally Heteroscedastic and Leptokurtic Model using the Student's *t*-distribution

The objective of this section is to obtain a multivariate *t*-distribution with separate time-varying degrees of freedom for every element of the random variable vector. This is essentially achieved by multiplying together the densities of independent *t*-variates and then applying a time-varying linear transformation to the vector of random variables to achieve the desired variance-covariance matrix and kurtosis. The details of this derivation are now presented.

5.1 General Transformation of Independent Scalar Student-*t*'s

Let x_t denote a $n \times 1$ vector of independent Student's *t* variates, each with degrees of freedom $v_{i,t}$. Note that, to avoid confusion, the notation below differs slightly from that employed above in the univariate case. This random vector may be transformed to have whatever variance-covariance structure is required. Let Λ_t be a non-singular $n \times n$ matrix, and define the transformed random vector

$$z_t = \Lambda_t x_t \,, \tag{33}$$

where

$$\Lambda_t = \{l_{i,j,t}\}. \tag{34}$$

Let $f(x_{i,t} \mid v_{i,t})$ be the probability density function of $x_{i,t}$. Its parameters are the vector of n degrees of freedom and the n^2 elements of the transformation matrix, Λ_t. These parameters will be used model both the variance-covariance structure of z_t, and its kurtosis. The density of $x_{i,t}$ is given by

$$f\left(x_{i,t} \middle| v_{i,t}\right) = \frac{\Gamma\left[\left(n_{i,t}+1\right)/2\right]}{p^{1/2} n_{i,t}^{1/2} \Gamma\left[n_{i,t}/2\right]\left(1 + x_{i,t}^2 / n_{i,t}\right)^{\left(n_{i,t}+1\right)/2}} . \tag{35}$$

Denoting the (joint) density of x_t as $f(x_t \mid v_t)$,

$$f\left(x_t \middle| v_t\right) = \prod_{i=1}^{n} \frac{\Gamma\left[\left(n_{i,t}+1\right)/2\right]}{p^{1/2} n_{i,t}^{1/2} \Gamma\left[n_{i,t}/2\right]\left(1 + x_{i,t}^2 / n_{i,t}\right)^{\left(n_{i,t}+1\right)/2}} . \tag{36}$$

The density of the transformed t, z_t, will be given by the standard Jacobian. The Jacobian of the transformation in equation (33) is

$$J_t = abs\left|\Lambda_t^{-1}\right| = \frac{1}{abs\left|\Lambda_t\right|} \tag{37}$$

The density of z_t is then found by replacing x_t by $\Lambda_t^{-1} z_t$ (element by element) in (36) and multiplying by the Jacobian. Thus, the density of z_t will be given by

$$f\left(z_t \middle| v_t, \Lambda_t\right) = \frac{1}{abs\left|\Lambda_t\right|} \prod_{i=1}^{n} \frac{\Gamma\left[\left(n_{i,t}+1\right)/2\right]}{p^{1/2} n_{i,t}^{1/2} \Gamma\left[n_{i,t}/2\right] \left(1 + \left[\left(\sum_{j=1}^{n} \Lambda_t^{i,j} z_{j,t}\right)^2 \middle/ n_{i,t}\right]\right)^{\left(n_{i,t}+1\right)/2}} \tag{38}$$

where $\Lambda_t^{i,j}$ is the *i*, *j*th element of Λ_t^{-1}.

5.2 Constraints on Modelling Variance-Covariance and Kurtosis

There are $n(n+1)/2$ variance-covariance parameters, leaving only $n(n+1)/2$ to explicitly model the kurtosis. In order to ensure identification, typically, using integer powers, there would be $5n!/(2!(n-2)!)$ fourth moments of the form $E\left(z_{i,t}^r z_{j,t}^s\right)$, where $r, s \geq 0$ and $r + s = 4$. Considering products of powers of three or more variables (with powers summing to 4) generates even more moments to account for. It is therefore necessary to make a selection of fourth order moments to model (either as the moments themselves or standardised to give the kurtosis).

One appealing way of restricting the quantities modelled so as to number $n(n+1)/2$ is to restrict attention to the variance-covariance matrix of the squares of the $z_{i,t}$. Since this is an $n \times n$ variance-covariance matrix, it has the required number of parameters. They are all of the form $E\left(z_{i,t}^r z_{j,t}^s\right)$, where $r, s = 0,2,4$ and $r + s = 4$. To illustrate, in the bivariate case, the moments modelled would be $E\left(z_{1,t}^4\right)$, $E\left(z_{1,t}^2 z_{2,t}^2\right)$, and $E\left(z_{2,t}^4\right)$. For convenience, the covariances of the squares will be referred to as cokurtoses (if they are in a standardised form, or cross-fourth moments when unstandardised). Apart from all those fourth order moments involving products of three or more random variables, the pairwise moments not modelled are $E\left(z_{1,t}^3 z_{2,t}\right)$ and $E\left(z_{1,t} z_{2,t}^3\right)$. The values of these ‘missing’ moments will be functions of those modelled.

Clearly, the choice made here, to model those fourth order moments that are the elements of the variance-covariance matrix of the squares of the process, is not the only choice that could be made. It is however appealing in the context of GARCH models, and is structurally useful in the context of financial applications, as discussed below.

5.3 Modelling Equations

The density is in terms of the elements of the transformation matrix and the degrees of freedom. The modelling will be in terms of the variance-covariance and the kurtosis of the process. It is therefore necessary to establish the mapping between the parameters of interest and those of the density.

Variance-Covariance Equations

The variance-covariance matrix of the transformed process is

$$\Sigma_{z,t} = \Lambda_t \Sigma_{x,t} \Lambda^T{}_t \, , \tag{39}$$

where the i, j^{th} element of $\Sigma_{x,t}$ is $\nu_{i,t} / (\nu_{i,t} - 2)$ for $i=j$ and zero otherwise. Denoting the i, j^{th} element of $\Sigma_{z,t}$ as $h_{i,j,t}$, these elements are given by

$$h_{i,i,t} = \sum_{k=1}^{n} l_{i,k,t}^2 \frac{\nu_{k,t}}{(\nu_{k,t} - 2)} \, , \tag{40a}$$

$$h_{i,j,t} = \sum_{k=1}^{n} l_{i,k,t} \, l_{j,k,t} \frac{\nu_{k,t}}{(\nu_{k,t} - 2)} \, . \tag{40b}$$

Kurtosis Equations

Exploiting the fact that the odd moments of the underlying *t*-variates, $x_{i,t}$, are zero, that they are independent, and that $E\left(x_{i,t}^4\right) = 3\nu_{i,t}^2 / (\nu_{i,t} - 2)(\nu_{i,t} - 4)$, leads to the fourth moment equations as follows:

$$m_{4,i,i,t} = E\left(z_{i,t}^4\right) = E\left(\left(\sum_{j=1}^{n} l_{i,j,t} x_{j,t}\right)^4\right) = 3\sum_{j=1}^{n} l_{i,j,t}^4 \frac{\nu_{j,t}^2}{(\nu_{j,t}-2)(\nu_{j,t}-4)} + 3\sum_{\substack{j,k=1\\ j\neq k}}^{n} l_{i,j,t}^2 l_{i,k,t}^2 \frac{\nu_{j,t}\nu_{k,t}}{(\nu_{j,t}-2)(\nu_{k,t}-2)}$$

i=1,...,n, (41a)

and

$$m_{4,i,m,t} = E\left(z_{i,t}^2 z_{m,t}^2\right) = E\left(\left(\sum_{j=1}^{n} l_{i,j,t} x_{j,t}\right)^2 \left(\sum_{k=1}^{n} l_{i,k,t} x_{k,t}\right)^2\right)$$

$$= 3\sum_{j=1}^{n} l_{i,j,t}^2 l_{m,j,t}^2 \frac{\nu_{j,t}^2}{(\nu_{j,t}-2)(\nu_{j,t}-4)} + \sum_{\substack{j,k=1\\ j\neq k}}^{n} l_{i,j,t}^2 l_{m,k,t}^2 \frac{\nu_{j,t}\nu_{k,t}}{(\nu_{j,t}-2)(\nu_{k,t}-2)} \quad (41b)$$

$$+ \sum_{\substack{j,k=1\\ j\neq k}}^{n} \sum_{\substack{r,s=1\\ r\neq s}}^{n} l_{i,j,t} l_{i,k,t} l_{m,r,t} l_{m,s,t} \frac{\nu_{j,t}\nu_{k,t}}{(\nu_{j,t}-2)(\nu_{k,t}-2)}$$

$$j=r,k=s \quad j=s,k=r$$

for i,m=1,...,n, $i \neq m$.

The Bivariate Case as an Illustration

For simplicity, the ensuing analysis will restrict attention to the bivariate case (*n*=2). Then the equations to solve - (40a), (40b), (41a), and (41b), become:

Variances

$$h_{1,1,t} = \sum_{k=1}^{2} l_{1,k,t}^2 \frac{\nu_{k,t}}{(\nu_{k,t}-2)} \quad (42a)$$

$$h_{2,2,t} = \sum_{k=1}^{2} l_{2,k,t}^2 \frac{\nu_{k,t}}{(\nu_{k,t}-2)} \quad (42b)$$

Covariance

$$h_{1,2,t} = \sum_{k=1}^{n} l_{1,k,t} l_{2,k,t} \frac{\nu_{k,t}}{(\nu_{k,t}-2)} \quad (42c)$$

Fourth moment

$$m_{4,1,1,t} = 3\sum_{j=1}^{2} l_{1,j,t}^4 \frac{\nu_{j,t}^2}{(\nu_{j,t}-2)(\nu_{j,t}-4)} + 6\, l_{1,1,t}^2 l_{1,2,t}^2 \frac{\nu_{1,t}\nu_{2,t}}{(\nu_{1,t}-2)(\nu_{2,t}-2)} \quad (42d)$$

$$m_{4,2,2,t} = 3\sum_{j=1}^{2} l_{2,j,t}^4 \frac{\nu_{j,t}^2}{(\nu_{j,t}-2)(\nu_{j,t}-4)} + 6\, l_{2,1,t}^2 l_{2,2,t}^2 \frac{\nu_{1,t}\nu_{2,t}}{(\nu_{1,t}-2)(\nu_{2,t}-2)} \quad (42e)$$

Co-fourth moment

$$m_{4,1,2,t} = 3\sum_{j=1}^{2} l_{1,j,t}^{2}\, l_{2,j,t}^{2} \frac{n_{j,t}^{2}}{(n_{j,t}-2)(n_{j,t}-4)} + \sum_{\substack{j,k=1\\ j\neq k}}^{2} l_{1,j,t}^{2}\, l_{2,k,t}^{2} \frac{n_{j,t} n_{k,t}}{(n_{j,t}-2)(n_{k,t}-2)}$$
$$+ 4\, l_{1,1,t}\, l_{1,2,t}\, l_{2,1,t}\, l_{2,2,t} \frac{n_{1,t} n_{2,t}}{(n_{1,t}-2)(n_{2,t}-2)} \qquad (42f)$$

5.4 Solving the Equations

It is required to solve (42a)-(42f), for the degrees of freedom and transformation parameters in terms of the moments. While in principle, this is possible, numerically at least, it is likely to be computationally burdensome. In a maximum likelihood setting, the equations have to be solved for each observation and for each evaluation of the log-likelihood required in the process of optimisation. For this reason, a further simplification is suggested.

To reduce the size of the problem, restrictions may be imposed on the transformation matrix. This means that fewer moments may be modelled - one fewer for each free transformation parameter lost. Even so, in order to make the problem more tractable, the requirement to directly model the co-fourth moment may be dropped, so that its value will follow from those of other parameters calculated from the relationships generated by the remaining moment equations. Relaxing the requirement to model the co-fourth moment reduces the number moment equations by $n(n-1)/2$, requiring the same reduction in the number of free parameters of the transformation matrix. Two obvious possible choices are to make the matrix symmetric, or to make it triangular. In the former case, the equations remain complex. However, in the triangular case, considerable simplifications are available, resulting in analytical solutions for the density parameters in terms of the moments. In all that follows, we employ the triangular approach, and therefore the transformation matrix is upper triangular, so that

$$l_{i,j,t} = 0 \ \ for \ \ i > j \, . \qquad (43)$$

It is easiest to illustrate the impact this has in the bivariate case, and the remainder of this paper focuses on the bivariate case only.

The Bivariate Case with Upper Triangular Transformation Matrix

The equations to be solved are (42a)-(42d). The co-fourth moment equation, (42e) will be returned to later, once the other parameters have been solved for. Imposing an upper triangular Λ_t in this case simply means setting $l_{2,1,t} = 0$. Substituting this into (42a)-(42d) results in the following equations.

Variances

$$h_{1,1,t} = \sum_{k=1}^{2} l_{1,k,t}^2 \frac{n_{k,t}}{(n_{k,t} - 2)} \tag{44a}$$

$$h_{2,2,t} = l_{2,2,t}^2 \frac{n_{2,t}}{(n_{2,t} - 2)} \tag{44b}$$

Covariance

$$h_{1,2,t} = l_{1,2,t}\, l_{2,2,t} \frac{n_{2,t}}{(n_{2,t} - 2)} \tag{44c}$$

Fourth moment

$$m_{4,1,1,t} = 3\sum_{j=1}^{2} l_{1,j,t}^4 \frac{n_{j,t}^2}{(n_{j,t} - 2)(n_{j,t} - 4)} + 6\, l_{1,1,t}^2\, l_{1,2,t}^2 \frac{n_{1,t} n_{2,t}}{(n_{1,t} - 2)(n_{2,t} - 2)} \tag{44d}$$

$$m_{4,2,2,t} = 3 l_{2,2,t}^4 \frac{n_{2,t}^2}{(n_{2,t} - 2)(n_{2,t} - 4)} \tag{44e}$$

Since this paper considers models for conditional kurtosis rather than the conditional fourth moment, equations for the former are obtained by taking (44d) and (44e) and dividing them by the squares of their respective conditional variances from (44a) and (44b):

$$k_{1,1,t} = \frac{3\sum_{j=1}^{2} l_{1,j,t}^4 \frac{n_{j,t}^2}{(n_{j,t} - 2)(n_{j,t} - 4)} + 6\, l_{1,1,t}^2\, l_{1,2,t}^2 \frac{n_{1,t} n_{2,t}}{(n_{1,t} - 2)(n_{2,t} - 2)}}{\left(l_{1,1,t}^2 \frac{v_{1,t}}{v_{1,t} - 2} + l_{1,2,t}^2 \frac{v_{2,t}}{v_{2,t} - 2} \right)^2} \tag{44d′}$$

and

$$k_{2,2,t} = \frac{3 l_{2,2,t}^4 \frac{n_{2,t}^2}{(n_{2,t} - 2)(n_{2,t} - 4)}}{l_{2,2,t}^4 \frac{n_{2,t}^2}{(v_{2,t} - 2)^2}} = \frac{3(v_{2,t} - 2)}{(v_{2,t} - 4)} \tag{44e′}$$

Equations (44) can now be solved - that is, expressions may be obtained for the degrees of freedom parameters and the transformation parameters in terms of the moments. The solutions are obtained as follows.

$$l_{2,2,t} = \sqrt{h_{2,2,t}\frac{(n_{2,t}-2)}{n_{2,t}}} \tag{45a}$$

$$l_{1,2,t} = h_{1,2,t}\sqrt{\frac{(n_{2,t}-2)}{n_{2,t}h_{2,2,t}}} \tag{45b}$$

$$l_{1,1,t} = \sqrt{\left(h_{1,1,t}-\frac{h_{1,2,t}^2}{h_{2,2,t}}\right)\frac{(n_{1,t}-2)}{n_{1,t}}} \tag{45c}$$

It follows directly by rearranging (44e′) that

$$n_{2,t} = \frac{4k_{2,2,t}-6}{k_{2,2,t}-3} \tag{45d}$$

From (45d′), and substituting in for the transformation parameters,

$$k_{1,1,t} = \frac{3}{h_{1,1,t}^2}\left[\left(h_{1,1,t}-\frac{h_{1,2,t}^2}{h_{2,2,t}}\right)^2\frac{(v_{1,t}-2)}{(v_{1,t}-4)}+\frac{h_{1,2,t}^4(v_{2,t}-2)}{h_{2,2,t}^2(v_{2,t}-4)}+2\left(h_{1,1,t}-\frac{h_{1,2,t}^2}{h_{2,2,t}}\right)\frac{h_{1,2,t}^2}{h_{2,2,t}}\right] \tag{44d′′}$$

Rearranging (44d′′) and substituting for $v_{2,t}$ from (45d),

$$n_{1,t} = 4+\frac{6\left(h_{1,1,t}-\frac{h_{1,2,t}^2}{h_{2,2,t}}\right)^2 h_{2,2,t}^2}{\left(k_{1,1,t}h_{1,1,t}^2h_{2,2,t}^2-3h_{2,2,t}^2\left(h_{1,1,t}-\frac{h_{1,2,t}^2}{h_{2,2,t}}\right)^2-h_{1,2,t}^4k_{2,2,t}-6h_{1,2,t}^2h_{2,2,t}\left(h_{1,1,t}-\frac{h_{1,2,t}^2}{h_{2,2,t}}\right)\right)} \tag{45e}$$

Equations (45d) and (45e) can then be substituted back into (45a), (45b), (45c) to obtain the solutions for the non-zero transformation matrix parameters in terms of the conditional kurtoses and conditional variances if desired. Finally, using these solutions and the restriction that $l_{2,1,t}=0$, equation (42f) can be used to obtain a value for the co-fourth moment

$$m_{4,1,2,t} = \frac{3h_{1,2,t}^2(v_{2,t}-2)}{(v_{2,t}-4)}+h_{1,1,t}h_{2,2,t}-h_{1,2,t}^2 \tag{46}$$

The co-fourth moment is clearly related to the covariance between the sample variances of the two series in the bivariate system[1]. As such, this series is could be of particular interest in finance, where it is important to know whether variances move together over time or can be viewed as operating relatively independently of one another. Multivariate GARCH models, by contrast, consider only co-movements between the levels of the series; it appears to be a common misconception that multivariate GARCH models consider relationships between the variances of the processes.

In the bivariate case with the restriction imposed, the log-likelihood function is obtained by taking the natural logarithm of (38) and can be written as

$$\begin{aligned} LLF(z_t|v_t,\Lambda_t) = &-\frac{1}{2}\log\left(h_{1,1,t}h_{2,2,t}-h_{1,2,t}^2\right)-\frac{1}{2}\log\left(v_{1,t}-2\right)-\frac{1}{2}\log\left(v_{2,t}-2\right)+\log\left[\Gamma\left(\frac{v_{1,t}+1}{2}\right)\right] \\ &-\log\left[\Gamma\left(\frac{v_{1,t}}{2}\right)\right]+\log\left[\Gamma\left(\frac{v_{2,t}+1}{2}\right)\right]-\log\left[\Gamma\left(\frac{v_{2,t}}{2}\right)\right]-\left(\frac{v_{2,t}+1}{2}\right)\log\left[1+\left(\frac{z_{2,t}^2}{(v_{2,t}-2)h_{2,2,t}}\right)\right] \\ &-\left(\frac{v_{1,t}+1}{2}\right)\log\left[1+\left(\frac{(z_{1,t}h_{2,2,t}-z_{2,t}h_{1,2,t})^2}{(v_{1,t}-2)(h_{1,1,t}h_{2,2,t}-h_{1,2,t}^2)h_{2,2,t}}\right)\right] \end{aligned} \quad (47)$$

5.5 Results

The results obtained from estimating the bivariate autoregressive conditional heteroscedastic and conditionally leptokurtic model (which will be referred to using the acronym MGARCHK) are presented in Table 6. Two separate estimations are conducted: one for the two equity series in a system and one for the bond series. The univariate coefficient estimates described above are used as starting values for optimisation of the multivariate model. The coefficient estimates in the conditional variance and conditional kurtosis equations are broadly similar to those obtained from univariate model estimation, but a set of coefficients for the conditional covariances are also estimated. The last two rows of Panels A and B show the conditional covariance parameter estimates between the two stock series and between the two bond series respectively. The conditional covariances for the equities are slightly smoother and less persistent, shown by the smaller values of the d_1 and d_2 coefficients

[1] Specifically, co-fourth moment measures the relationship between the squares of the residuals, and

respectively, than the conditional variances are, while the conditional covariance estimates for the bond system are similar to those of the variances.

Figure 7 plots the conditional covariances between the two sets of bond returns. The estimated covariance is fairly smooth over time, but is larger on average and more volatile during 1990 and mid 1994, while it reaches a peak of over 5 times its average value in October 1998. Figure 8 plots the co-fourth moment between the equities, together with the two fitted conditional variance series. There is a very strong correlation between the variances, which move closely together over time, although the UK equity market appears to be on average more volatile overall from the plot (although its estimated unconditional variance is slightly lower). The fitted UK conditional variance is particularly high in September 1990, April 1992 and September 1992, while both markets are volatile in August-September 1998 and the US is more volatile in April 2000. The conditional co-fourth moment series, however, takes on values close to unity most of the time but also comprises a small number of very large values which are orders of magnitude greater than the average and make it harder to visualise the day to day variation. It is evident from the equation for the co-fourth moment, $m_{4,1,2,t}$ that its values will be large when (assuming all else is constant) either $h_{1,1,t}$ or $h_{2,2,t}$ are large, or when $v_{2,t} \downarrow 4$. If the equation is suitably rearranged, it is also possible to show that $m_{4,1,2,t}$ also varies positively with $h_{1,2,t}$. Over the 10-year sample period, a total of 5 daily observations have co-fourth moment values larger than 10: 28 October 1997, 28 August 1998, 31 August 1998, 5 January 2000, and 17 April 2000. All of these days witnessed extreme movements in both equity markets, except for 28 August 1998 where the UK return was zero since the market was closed due to a public holiday. These 5 observations are therefore removed and the 3 series re-plotted in Figure 9[2]. The high degree of association between the conditional variances and the conditional co-fourth moment can now be seen. In general, extreme negative returns appear to result in increased variances, increased covariance, increased co-fourth moment and a fall in the degree of freedom parameters to values very close to 4. However, the co-fourth moment interestingly appears to have increased by more during the last three and a half years of the sample than either of

---

thus the analogy with the co-movement between the variances is not perfect.

[2] Note that these points are still included in all of the estimations – they are simply removed from the plots to make the day to day variations easier to see.

the two variances, while the degree of freedom estimates for both the US and the UK fell relative to their historical averages. This suggests that the conditional variance is not sufficient to describe the dynamics of the series in the late 1990's, and that the series have become more fat-tailed and therefore further from being normally distributed over time. This fact is also evident from the relatively small and stable values that the conditional co-fourth moment takes on (typically in the range 0.3-0.7), while it jumps occasionally to values that are twenty times this – jumps that are far greater than those of the conditional variances.

Finally, Figure 10 presents a scatter plot of the conditional co-fourth moment against the conditional co-variance for the equity returns (after the five largest values of the co-fourth moment are removed). The positive relationship between the two series is clearly evident and indeed, the correlation between them (after the 5 extreme observations are removed) is 71.2%, while the correlations of $m_{4,1,2,t}$ with the US and UK conditional variances are 76.3% and 78% respectively. Further, a quadratic relationship between the co-fourth moment and the conditional covariance is evident from the lower part of the "frontier" that appears. A plot of the co-fourth moment against the square of the conditional covariance (not shown here) demonstrates a more linear relationship.

It would be possible to further extend the model to allow for asymmetries in the conditional variance, covariance, or kurtosis equations, or to allow for feedback between these series and the conditional mean. These generalisations, together with an application of the specification testing methodology outlined above, are not pursued further here in the context of the multivariate model and are left for future research. It would also be of interest to explicitly examine the co-moments that were left unmodelled above of the form $E(z_i^3 z_j)$, and so on, although this would considerably increase the complexity of the model and would require numerically solving for the time-varying transformation parameters at each iteration in the optimisation. Clearly, then, there is a trade-off between flexibility and simplicity and we have leaned towards the latter, with the result that our model is estimable using standard maximum likelihood methods.

6. Conclusions

This paper has proposed and estimated a model for conditional kurtosis. The model is based on the approach of Bollerslev (1987), but its novelty lies in its ability to allow the conditional kurtosis to develop in a fashion that is not fixed to the conditional variance. This occurs via a time-varying degree of freedom parameter. The model was applied to a set of four financial asset return time series, and the results indicate strong evidence for the presence of “GARCH-style” dependence in the conditional kurtosis, suggesting the presence of a hitherto unexplored phenomenon in such series.

Several extensions to the basic model were proposed, including considerations of asymmetries in the relationship between the sign of the innovations and the size of the next period conditional kurtosis, and of a possible feedback from the kurtosis to the returns. Evidence for these relationships appeared to be weaker in the context of conditional kurtosis than was the case for the conditional variance. Finally, a multivariate version of the model was described, which was constructed from a set of independent *t*-variates that were multiplied together and then subjected to a time-varying linear transformation to achieve the desired variance-covariance matrix and kurtosis. This approach not only allows for the conditional co-variance to be time-varying as a standard multivariate GARCH approach would, but also permits a consideration of the co-fourth moment, which is the covariance between the variances of the series.

The research described above could be usefully extended in a number of different directions. First, the models proposed could be used to produce conditional kurtosis forecasts. These predictions may be useful – for example in the pricing of some classes of financial asset, such as options on options that require estimates of the “variance of a variance”. This quantity could be obtained from the forecasts of conditional kurtosis, and the option prices obtained compared with those from simpler approaches. Second, it may be the case that models allowing for dynamic higher moments can better describe the distributional properties of financial asset returns, especially when measured at high frequency, than less complex models that do not. Third, models containing feedback terms from the conditional variance and kurtosis to the conditional mean (“GARCHK-M” models) may, when appropriately formulated, be used to obtain separate estimates of the market-required risk premia for accepting variance (or standard deviation) and kurtosis risk. GARCHK-M models could then be used in portfolio construction or in investment performance appraisal to evaluate whether the trade-off between mean, variance, and kurtosis that is implicit from the series of returns to the chosen portfolio was an optimal one given the market-required returns for each type of risk. Finally, further exploration of the multivariate version of the model should shed light on the co-relationships between the moments of each series employed in the system, which is likely to be of relevance in the context of portfolio construction or financial risk management.

References

Black, F. (1976) Studies in price volatility changes, Proceedings of the 1976 Meeting of the Business and Economics Statistics Section *American Statistical Association*, 177-181.

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

Bollerslev, T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return *Review of Economics and Statistics*, 69, 542 – 547.

Bond, S.A. (2001) A review of asymmetric conditional density functions in autoregressive conditional heteroscedasticity models in Knight, J. and Satchell, S. (eds.) *Return Distributions in Finance* Butterworth-Heinemann, Oxford, UK.

Christie, A. (1982) The stochastic behaviour of common stock variance: Value, leverage and interest rate effects *Journal of Financial Economics*, 10, 407-432.

Chunhachinda, P., Dandapani, K., Hamid, S., and Prakash, A. (1997) Portfolio Selection and Skewness: Evidence from International Stock Markets *Journal of Banking and Finance* 21, 143-167.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1006.

Engle, R.F. and Bollerslev, T. (1986) Modelling the persistence of conditional variances. *Econometric Reviews* 5, 1-50.

Engle, R.F., Lilien, D.M., Robins, R.P. (1987) Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model *Econometrica* 55(2), 391-407.

Fernandez, C. and Steel, M.F.J. (2998) On Bayesian Modelling of Fat Tails and Skewness *Journal of the American Statistical Association* 93, 359-371.

Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993) On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks *The Journal of Finance* 48(5), 1779-1801.

Greene, W.H. (2000) *Econometric Analysis* Prentice Hall, Upper Saddle River, New Jersey.

Hansen, B.E. (1994) Autoregressive Conditional Density Estimation *International Economic Review* 35, 705-730.

Harvey, C.R. and Siddique, A. (2000) Conditional Skewness in Asset Pricing Tests *Journal of Finance* 55, 1263-1295.

Harvey, C.R. and Siddique, A. (1999) Autoregressive Conditional Skewness *Journal of Financial and Quantitative Analysis* 34, 465-487.

Johansen, A., and Sornette, D. (1999) Critical Crashes *Risk* 12(1), 91-95.

Jondeau, E. and Rockinger, M. (2000) Conditional Volatility, Skewness and Kurtosis: Existence and Persistence *Mimeo.* Banque de France and HEC School of Management.

Lambert, P. and Laurent, S. (2001) Modelling Financial Time Series using GARCH-type Models and a Skewed Student Density *Mimeo*. Université Catholique de Lovain.

Nelson, D. (1996) Asymptotic Filtering Theory for Multivariate ARCH Models *Journal of Econometrics* 71, 1-47.

Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach *Econometrica* 59(2), 347-370

Newey, W. (1985) Generalised Method of Moments Specification Testing *Journal of Econometrics* 29, 229-256.

Premaratne, G. and Bera, A.K. (2001) Modelling Asymmetry and Excess Kurtosis in Stock Return Data *Mimeo*. Department of Economics, University of Illinois.

Rockinger, M and Jondeau, E. (2001) Conditional Dependency of Financial Series: An Application of Copulas Persistence *Mimeo.* Banque de France and HEC School of Management.

Scott, R. and P. Horvath (1980), On the Direction of Preference for Moments of Higher Order Than the Variance, *Journal of Finance*, 35, 915-919.

Table 1: Summary Statistics and Model Estimates for GARCH with fixed degrees of freedom

Panel A: Summary Statistics

	Mean	Variance	Skewness	Kurtosis (excess)	BJ♣ Normality	Ljung-Box Q*(5)	ARCH(5)
Equities							
S&P500	0.052	0.842	-0.345**	5.415**	3385.130**	12.754*	255.253**
FTSE 100	0.034	0.663	0.014	2.826**	906.954**	29.024**	274.823**
Bonds							
US	0.005	0.279	-0.305**	1.957**	477.062**	19.961**	16.649**
UK	0.016	0.267	-0.020	3.736**	1585.630**	12.226*	78.74**

Panel B: Estimates for GARCH with fixed degrees of freedom

$$y_t = g_0 + g_1 y_{t-1} + g_2 e_{t-1}^{*} + e_t^{*} \quad (12a)$$

$$e_t^{*} = l_t e_t \,,\; e_t \sim t_{n_t} \quad (12b)$$

$$h_t = a_0 + a_1 e_{t-1}^{*\,2} + a_2 h_{t-1} \quad (12c)$$

$$k_t = b_0 \quad (12d)$$

	g_0	g_1	g_2	a_0	a_1	a_2	b_0	LLF
Equities								
S&P500	0.060	-	0.034	0.043	0.286	0.730	7.404	-1736.96
	(0.014)**		(0.024)	(0.005)**	(0.015)**	(0.008)**	(2.126)**	
FTSE 100	0.031	0.063	-	0.057	0.273	0.711	5.449	-1517.63
	(0.012)*	(0.022)**		(0.004)**	(0.014)**	(0.009)**	(1.319)**	
Bonds								
US	0.0116	0.036	-	0.172	0.019	0.384	5.791	-495.74
	(0.009)	(0.019)		(0.006)**	(0.012)	(0.020)**	(0.799)**	
UK	0.0249	0.043	-	0.004	0.045	0.938	6.846	-279.36
	(0.008)**	(0.019)*		(0.001)**	(0.002)**	(0.002)**	(0.683)**	

Notes: Standard errors are shown in parentheses. ♣ The Bera-Jarque (BJ) normality test is asymptotically distributed as a $\chi^2(2)$ variate under the null of normality while the Ljung-Box and ARCH tests are asymptotically distributed as $\chi^2(5)$ variates under their respective null hypotheses. The 5% $\chi^2(2)$ and $\chi^2(5)$ critical values are respectively 5.991 and 11.071, and at the 1% level, the critical values are 9.210 and 15.086 respectively. * and ** denote significance at the 5% and 1% levels respectively; LLF denotes the maximal value of the log-likelihood function.

Table 2: Conditional Variance & Kurtosis

Model: $y_t = g_0 + g_1 y_{t-1} + g_2 e_{t-1}^{*} + e_t^{*}$, $h_t = a_0 + a_1 e_{t-1}^{*\,2} + a_2 h_{t-1}$, $e_t = l\, e_t \quad e \quad {}_{n_t}$ $k_t = b_0 + b_1 \dfrac{e_{t-1}^{*\,4}}{h_{t-1}^2} + b_2 k_{t-1}$

	g_0	g_1	g_2	a_0	a_1	a_2	b_0	b_1	b_2	LLF	LR $b_1 = b_2 = 0$
Equities											
S&P500	0.060	-	0.021	0.098	0.023	0.872	5.041	0.412	0.171	-1689.16	95.60**
	(0.014)**		(0.020)	(0.005)**	(0.002)**	(0.003)**	(0.509)**	(0.176)*	(0.090)*		
FTSE 100	0.036	0.081	-	0.097	0.023	0.865	2.662	0.309	0.285	-1471.18	92.90**
	(0.014)**	(0.020)**		(0.005)**	(0.002)**	(0.004)**	(0.164)**	(0.156)*	(0.039)**		
Bonds											
US	0.015	0.040	-	0.044	0.057	0.786	3.314	0.545	0.324	-485.96	19.56**
	(0.009)**	(0.018)*		(0.001)**	(0.005)**	(0.004)**	(0.110)**	(0.286)	(0.027)**		
UK	0.023	0.043	-	0.005	0.050	0.932	4.667	0.222	0.379	-277.72	3.28
	(0.008)**	(0.019)*		(0.001)**	(0.002)**	(0.002)**	(1.225)**	(0.404)	(0.132)**		

Notes: Standard errors are shown in parentheses. * and ** denote significance at the 5% and 1% levels respectively. LLF denotes the maximal value of the log-likelihood function, while LR denotes the value of the likelihood ratio test statistic. The $\chi^2(2)$ critical values are 5.99 and 9.21 at the 5% and 1% levels respectively.

Table 3: Conditional Variance & Kurtosis – Asymmetric

Model: $y_t = g_0 + g_1 y_{t-1} + g_2 e_{t-1}^{*} + e_t^{*},\ e_t^{*} = l_t e_t,\ e_t \sim t_{n_t},\ h_t = a_0 + a_1 e_{t-1}^{*\,2} + a_2 h_{t-1} + a_3 I_{t-1} e_{t-1}^{*\,2},\ k_t = b_0 + b_1 \frac{e_{t-1}^{*\,4}}{h_{t-1}^2} + b_2 k_{t-1} + b_3 I_{t-1} \frac{e_{t-1}^{*\,4}}{h_{t-1}^2}$

	g_0	g_1	g_2	a_0	a_1	a_2	a_3	b_0	b_1	b_2	b_3	LLF	LR: $a_3 = 0$ and $b_3 = 0$
Equities													
S&P500	0.048	-	0.031	0.025	0.081	0.816	0.164	5.348	0.419	0.209	0.606	-1685.66	7.00**
	(0.013)**		(0.023)	(0.003)**	(0.008)**	(0.005)**	(0.016)**	(0.566)**	(0.151)**	(0.100)*	(0.318)		
FTSE 100	0.027	0.079	-	0.026	0.080	0.819	0.140	2.239	0.444	0.370	0.500	-1469.76	2.84
	(0.012)*	(0.020)**		(0.002)**	(0.006)**	(0.005)**	(0.014)**	(0.161)**	(0.136)	(0.036)**	(0.330)		
Bonds													
US	0.008	0.039	-	0.062	0.013	0.729	0.098	4.640	0.356	0.139	0.291	-484.35	3.22
	(0.009)	(0.019)**		(0.002)**	(0.009)	(0.009)**	(0.018)**	(0.912)**	(0.425)	(0.125)	(0.865)		
UK	0.017	0.054	-	0.005	0.024	0.937	0.045	4.610	0.021	0.519	0.012	-274.61	6.22*
	(0.008)*	(0.018)**		(0.001)**	(0.002)**	(0.002)	(0.005)**	(2.000)*	(0.767)	(0.149)**	(1.890)		

Notes: Standard errors are shown in parentheses. * and ** denote significance at the 5% and 1% levels respectively. LLF denotes the maximal value of the log-likelihood function, while LR denotes the value of the likelihood ratio test statistic. The $\chi^2(2)$ critical values are 5.99 and 9.21 at the 5% and 1% levels respectively.

Table 4: Asymmetric Conditional Variance & Kurtosis – GARCH and Kurtosis in mean

Model: $y_t = g_0 + g_1 y_{t-1} + g_2 e_{t-1}^{*} + g_3 \frac{h_t^{1/2}}{\bar{y}} + g_4 k_t + e_t^{*}, \quad e_t^{*} = l_t e_t,\ e_t \sim t_{n_t},$

$$h_t = a_0 + a_1 e_{t-1}^{*\,2} + a_2 h_{t-1} + a_3 I_{t-1} e_{t-1}^{*\,2}, \quad k_t = b_0 + b_1 \frac{e_{t-1}^{*\,4}}{h_{t-1}^2} + b_2 k_{t-1} + b_3 I_{t-1} \frac{e_{t-1}^{*\,4}}{h_{t-1}^2}$$

	g_0	g_1	g_2	g_3	g_4	a_0	a_1	a_2	a_3	b_0	b_1	b_2	b_3	LLF	LR: $g_3 = 0$ and $g_4 = 0$
Equities															
S&P500	0.028	-	0.020	0.0010	0.0003	0.025	0.079	0.083	0.147	5.331	0.702	0.281	0.170	-1676.9	17.44**
	(0.013)**		(0.021)	(0.0009)	(0.0012)	(0.003)**	(0.007)**	(0.005)**	(0.014)**	(0.458)**	(0.375)	(0.124)*	(0.289)		
FTSE	0.015	0.089	-	0.0003	0.0001	0.024	0.078	0.826	0.136	2.021	0.285	0.437	0.176	-1469.3	1.00
100	(0.012)	(0.020)**		(0.0006)	(0.0011)	(0.002)**	(0.006)**	(0.005)**	(0.013)	(0.105)**	(0.117)**	(0.030)**	(0.172)		
Bonds															
US	0.014	0.052	-	0.0001	0.0003	0.053	0.007	0.766	0.078	4.294	0.531	0.268	0.027	-482.3	4.18
	(0.009)	(0.018)**		(0.0001)	(0.0009)	(0.002)**	(0.008)	(0.007)**	(0.015)**	(0.734)**	(0.529)	(0.106)*	(0.058)		
UK	0.002	0.044	-	0.0004	0.0006	0.005	0.026	0.932	0.052	6.799	0.161	0.244	0.074	-274.4	0.44
	(0.008)	(0.020)*		(0.0003)	(0.0006)	(0.001)**	(0.002)**	(0.002)**	(0.005)**	(0.127)**	(0.100)	(0.053)**	(0.779)		

Notes: Standard errors are shown in parentheses. * and ** denote significance at the 5% and 1% levels respectively. LLF denotes the maximal value of the log-likelihood function, while LR denotes the value of the likelihood ratio test statistic. The $\chi^2(2)$ critical values are 5.99 and 9.21 at the 5% and 1% levels respectively.

Table 5: Moment Specification Tests

Orthogonality Conditions	US S&P500	UK FTSE100	US T-Bond	UK T-Bond
1. $E[z_t] = 0$	-0.012	0.010	-0.016	-0.004
	(1.621)	(1.910)	(3.825)	(0.194)
2. $E[z_t \cdot z_{t-1}] = 0$	0.076**	0.018	0.028**	0.038**
	(35.045)	(2.381)	(7.457)	(7.249)
3. $E[z_t \cdot z_{t-2}] = 0$	0.037	0.026	0.029	0.038
	(1.580)	(0.098)	(0.965)	(2.00)
4. $E[z_t \cdot z_{t-3}] = 0$	-0.087**	-0.014	-0.052	-0.029
	(9.218)	(0.281)	(3.264)	(1.022)
5. $E[z_t \cdot z_{t-4}] = 0$	-0.044	0.023	-0.053	0.033
	(2.282)	(0.837)	(3.173)	(1.267)
6. $E[(z_t^2 - (v_t / (v_t - 2)))(z_{t-1}^2 - (v_{t-1} / (v_{t-1} - 2)))] = 0$	-0.643**	-0.277	-0.443**	-0.146
	(19.298)	(3.317)	(21.188)	(0.419)
7. $E[(z_t^2 - (v_t / (v_t - 2)))(z_{t-2}^2 - (v_{t-2} / (v_{t-2} - 2)))] = 0$	-0.108	-0.231*	-0.066	-0.336**
	(0.222)	(4.761)	(0.312)	(8.739)
8. $E[(z_t^2 - (v_t / (v_t - 2)))(z_{t-3}^2 - (v_{t-3} / (v_{t-3} - 2)))] = 0$	-0.453**	-0.111	-0.160	-0.066
	(9.421)	(0.924)	(1.169)	(0.176)
9. $E[(z_t^2 - (v_t / (v_t - 2)))(z_{t-4}^2 - (v_{t-4} / (v_{t-4} - 2)))] = 0$	-0.321*	-0.338**	0.099	-0.006
	(4.522)	(12.636)	(0.580)	(0.001)
10. $E[z_t^3 \cdot z_{t-1}^3] = 0$	4.449*	3.514	2.442	8.981
	(4.339)	(1.276)	(2.524)	(2.774)
11. $E[z_t^3 \cdot z_{t-2}^3] = 0$	10.659*	0.915	0.986	2.002
	(4.467)	(0.617)	(0.306)	(1.948)
12. $E[z_t^3 \cdot z_{t-3}^3] = 0$	-2.702	-0.862	-1.661	2.576
	(1.195)	(0.220)	(0.308)	(1.276)
13. $E[z_t^3 \cdot z_{t-4}^3] = 0$	-2.632	0.990	-3.041	5.239
	(1.634)	(1.616)	(2.101)	(1.769)
14. $E[(z_t^4-(3v_t^2/((v_t-2)(v_t-4))))\cdot(z_{t-1}^4-(3v_{t-1}^2/((v_{t-1}-2)(v_{t-1}-4))))] = 0$	-17237	-4721.708	-4595.259	-14829*
	(1.011)	(1.096)	(3.497)	(4.323)
15. $E[(z_t^4-(3v_t^2/((v_t-2)(v_t-4))))\cdot(z_{t-2}^4-(3v_{t-2}^2/((v_{t-2}-2)(v_{t-2}-4))))] = 0$	-3879.381	-2090.880	-1045.669	-540.994
	(0.633)	(1.071)	(1.184)	(2.833)
16. $E[(z_t^4-(3v_t^2/((v_t-2)(v_t-4))))\cdot(z_{t-3}^4-(3v_{t-3}^2/((v_{t-3}-2)(v_{t-3}-4))))] = 0$	-739.302	-808.454	-65.157	-28.075
	(0.204)	(0.854)	(0.057)	(0.061)
17. $E[(z_t^4-(3v_t^2/((v_t-2)(v_t-4))))\cdot(z_{t-4}^4-(3v_{t-4}^2/((v_{t-4}-2)(v_{t-4}-4))))] = 0$	291.682	-397.981	144.793	19.461
	(0.313)	(1.015)	(1.887)	(0.034)
Test statistic for joint test of all moment restrictions 1 to 17	171.981**	63.393**	79.599**	48.396**

Notes: Orthogonality conditions are based on $\hat{z}_t = \hat{e}_t^* / \hat{l}_t$. * and ** denote significance at the 5% and 1% levels respectively. Cell entries refer to sample averages with test statistics in parentheses. The latter follow a $\chi^2(1)$ distribution under the null hypothesis in each case. The joint test follows a $\chi^2(17)$ under the null that all four moments are correctly specified. The critical values for the $\chi^2(20)$ are 27.58 and 33.41 at the 5% and 1% levels respectively.

Table 6: Multivariate Conditional Heteroscedasticity and Kurtosis

Model (i =1,2; 1 = US, 2 = UK): $y_{i,t} = g_{i,0} + g_{i,1} y_{i,t-1} + g_{i,2} z_{i,t-1} + z_{i,t}$, $\quad _{,} = a_0 \quad a_1 z \;_{-1} + a_{,} \quad _{,\, t-1}$,

$$h_{1,2,t} = d_0 + d_1 z_{1,t-1} z_{2,t-1} + d_2 h_{1,2,t-1},\; z_t = \Lambda_t x_t,\; x_{i,t} \sim t_{n_{i,t}} \quad k_{i,i,t} = b_{i,0} + b_{i,1} \frac{z^{4}_{i,t-1}}{h^2_{i,t-1}} + b_{i,2} k_{i,i,t-1}$$

Panel A: US and UK Equity Series: LLF = -3206.93

	g_0	g_1	g_2	a_0	a_1	a_2	b_0	b_1	b_2
S&P500	0.062	-	0.006	0.031	0.090	0.823	5.248	0.053	0.144
Equation	(0.013)**		(0.014)	(0.003)**	(0.003)**	(0.005**)	(0.884)**	(0.642)	(0.117)
FTSE 100	0.033	0.086	-	0.062	0.080	0.800	2.333	0.506	0.318
Equation	(0.015)*	(0.018)**		(0.005)**	(0.0037)	(0.005)**	(0.114)**	(0.063)**	(0.233)

	d_0	d_1	d_2
Covariance	0.022	0.046	0.695
Equation	(0.003)**	(0.004)**	(0.037)**

Panel B: US and UK Bond Series: LLF = -578.97

	g_0	g_1	a_0	a_1	a_2	b_0	b_1	b_2
US Bonds	0.006	0.024	0.043	0.059	0.790	3.247	0.269	0.321
Equation	(0.009)	(0.017)	(0.001)**	(0.005)**	(0.004)**	(0.065)**	(0.464)	(0.062)**
UK Bonds	0.0186	0.047	0.005	0.050	0.933	4.837	0.013	0.315
Equation	(0.007)*	(0.017)**	(0.001)**	(0.002)**	(0.002)**	(0.215)	(0.216)	(0.089)**

	d_0	d_1	d_2
Covariance	0.010	0.051	0.820
Equation	(0.001)**	(0.002)**	(0.005)**

Notes: Standard errors are shown in parentheses. * and ** denote significance at the 5% and 1% levels respectively. LLF denotes the maximal value of the log-likelihood function. The $\chi^2(2)$ critical values are 5.99 and 9.21 at the 5% and 1% levels respectively.

Figure 1: Fitted Conditional Heteroscedasticity and Conditional Kurtosis over Time for UK Stock Returns

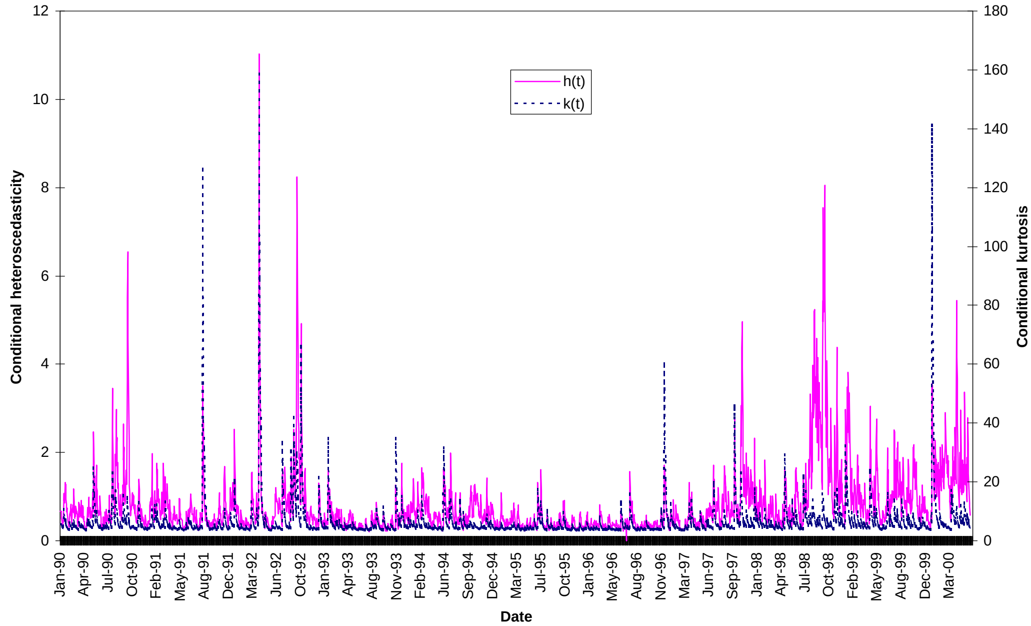

Figure 2: Estimated Degrees of Freedom over Time for UK Stock Returns

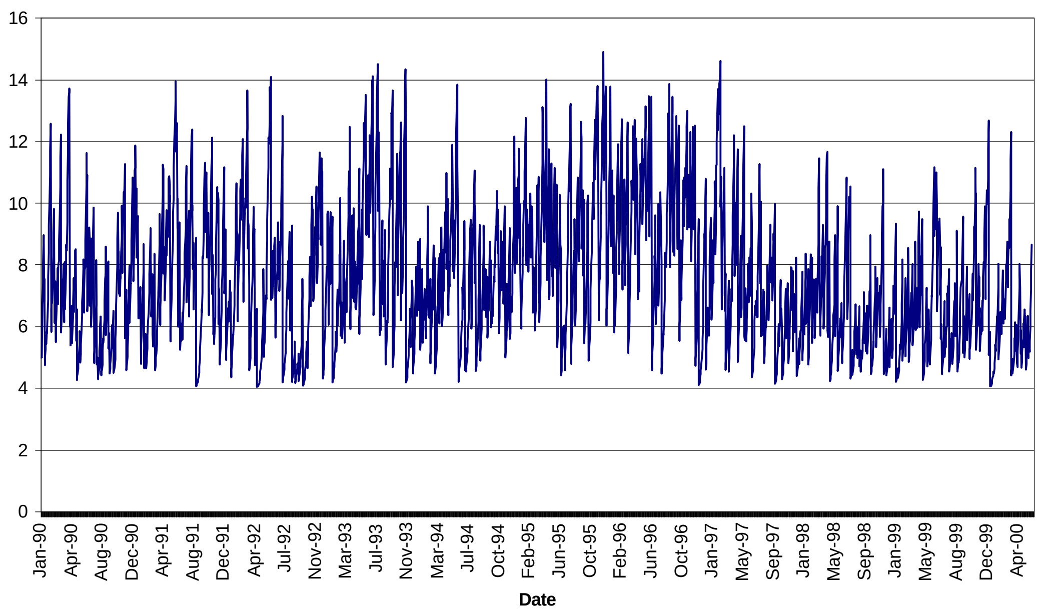

Figure 3: Plot of Transformation Series, l_t for UK Stock Returns

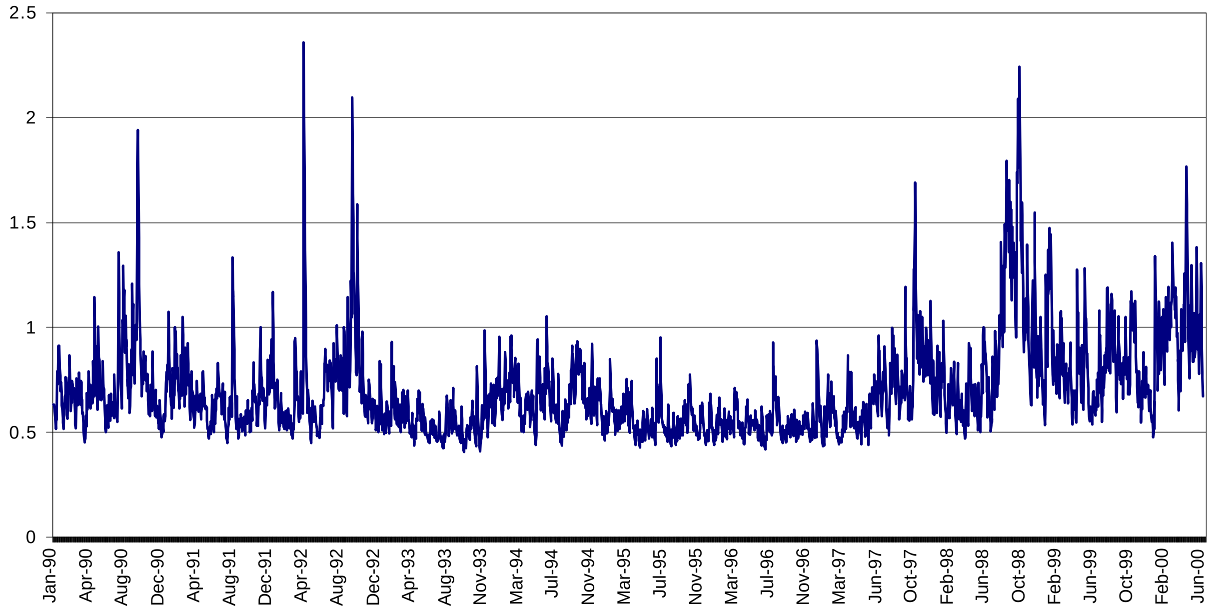

Figure 4: Fitted Conditional Heteroscedasticity and Conditional Kurtosis over Time for US T-Bonds

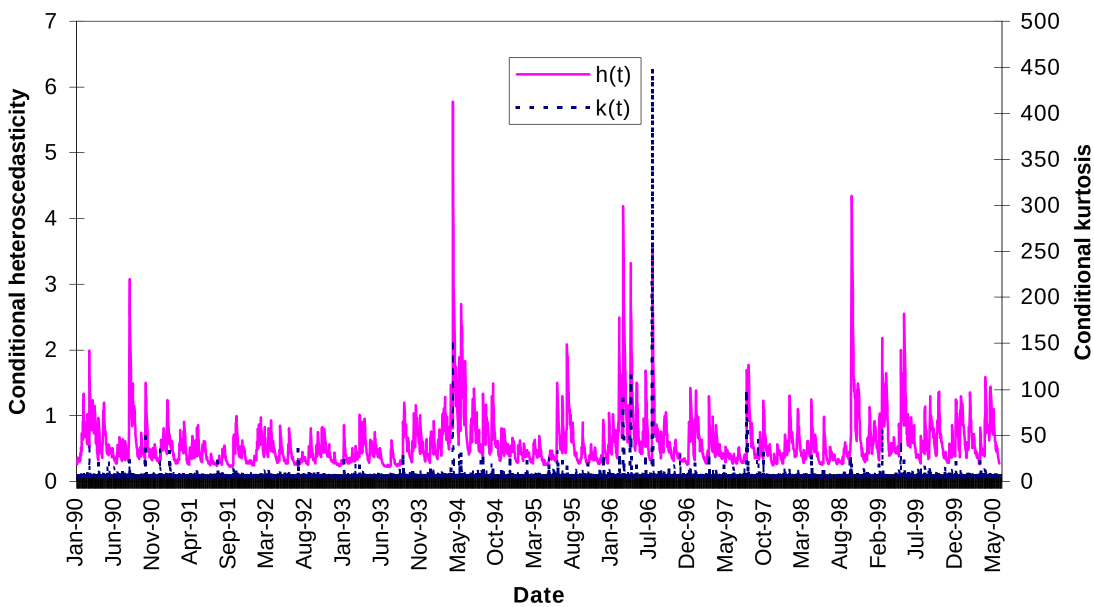

Figure 5: Estimated Degrees of Freedom over Time for US T-bonds

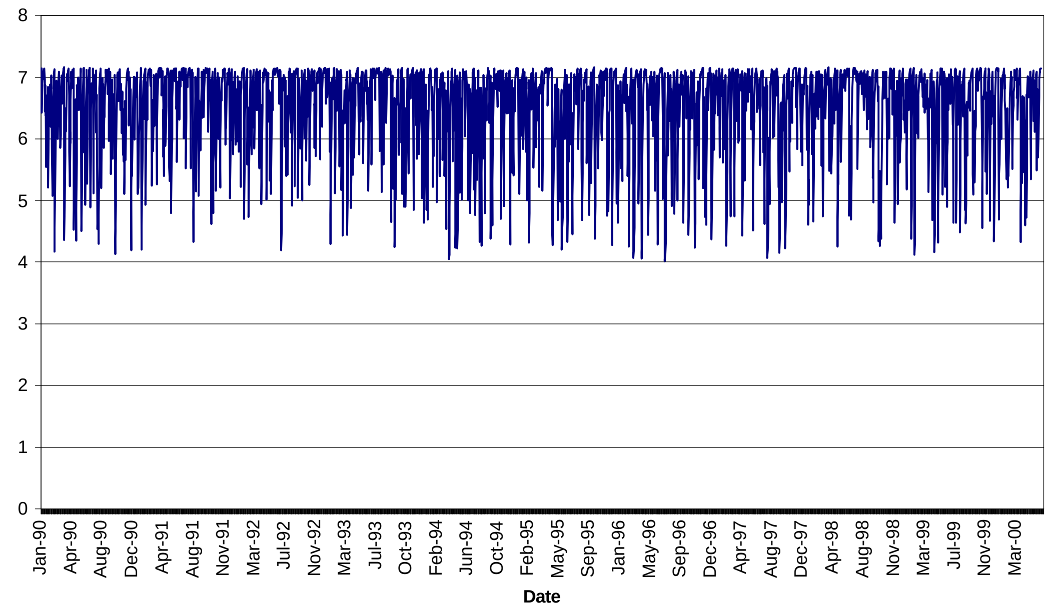

Figure 6: Plot of Transformation Series, l_t for US T-Bonds

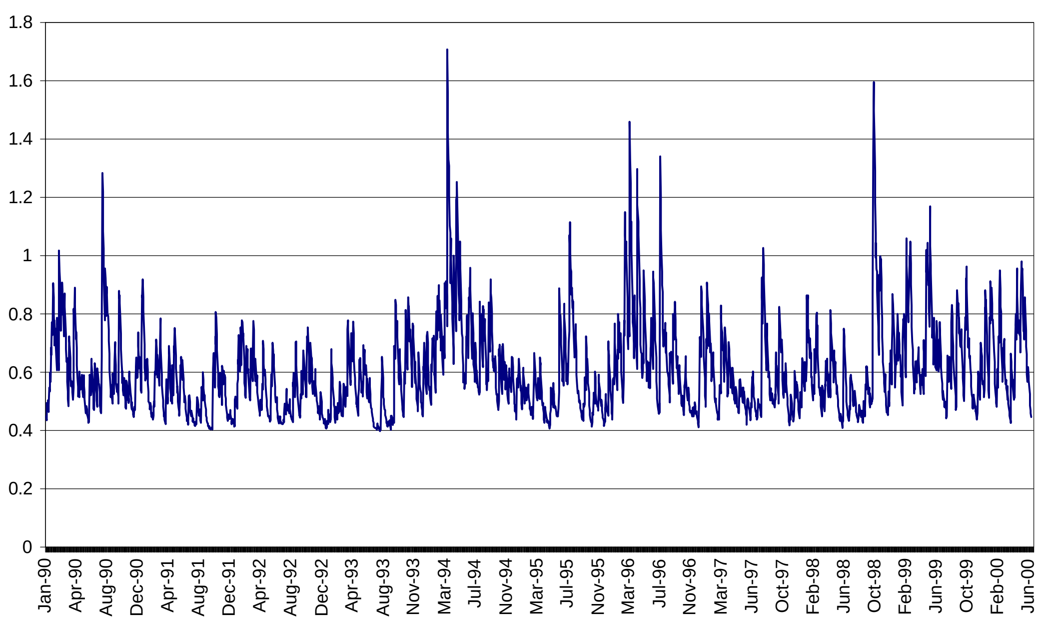

Figure 7: Conditional Covariance between US and UK Bond Returns

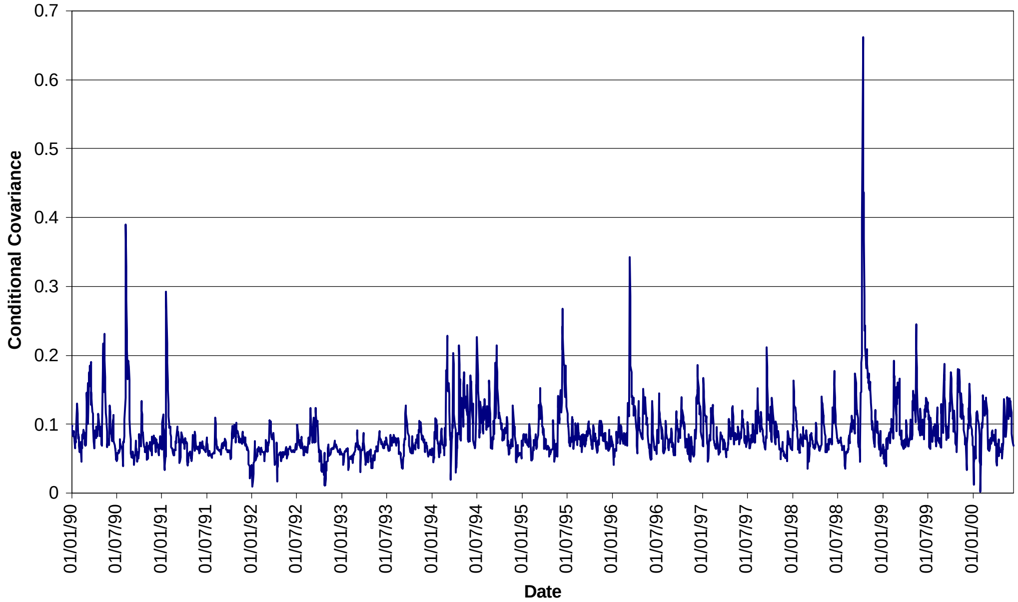

Figure 8: Conditional Co-fourth Moment and Conditional Variances for US and UK Stock Returns

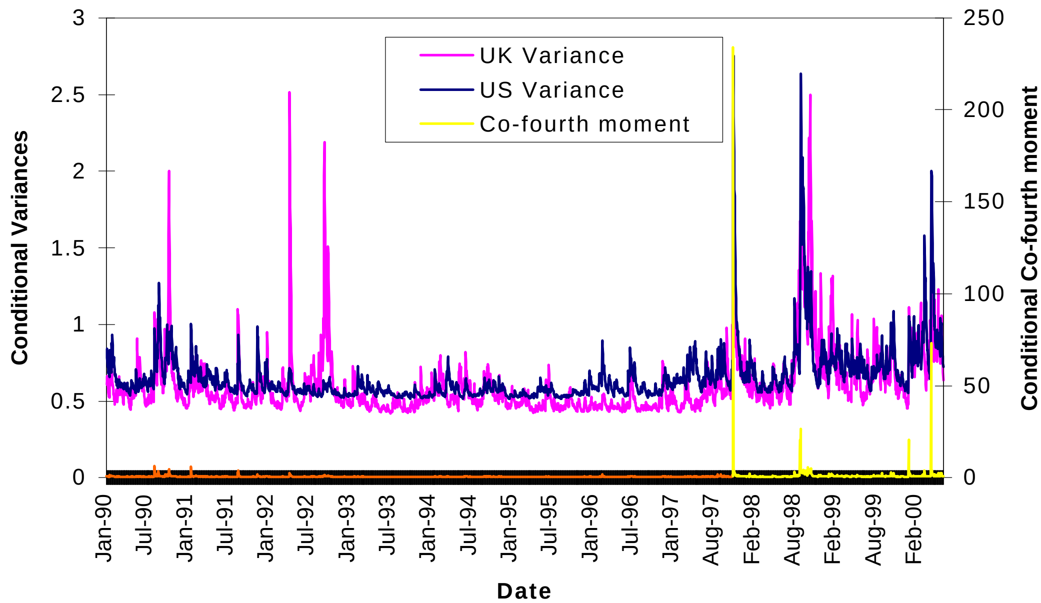

Figure 9: Conditional Co-fourth Moment and Conditional Variances for US and UK Stock Returns (co-fourth moment values above 10 removed)

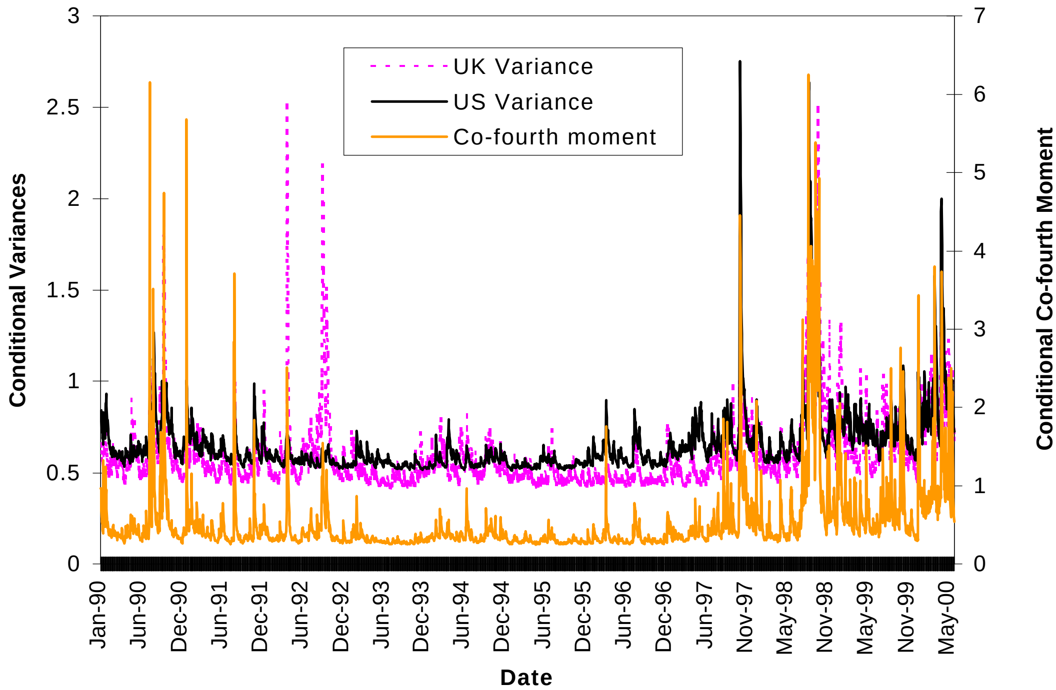

Figure 10: Conditional Co-fourth Moment against Conditional Covariance for US and UK Stock Returns (co-fourth moment values above 10 removed)