The University of Reading

THE BUSINESS SCHOOL
FOR FINANCIAL MARKETS

Normal Mixture GARCH(1,1)

Applications to Exchange Rate Modelling

ISMA Centre Discussion Papers in Finance 2004-06
March 2004

Carol Alexander
ISMA Centre, University of Reading, UK

Emese Lazar
ISMA Centre, University of Reading, UK

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

Some recent specifications for GARCH error processes explicitly assume a conditional variance that is generated by a mixture of normal components, albeit with some parameter restrictions. This paper analyses the general normal mixture GARCH(1,1) model which can capture time-variation in both conditional skewness and kurtosis. A main focus of the paper is to provide conclusive evidence that, for modelling exchange rates, generalized two component normal mixture GARCH(1,1) models perform better than those with three or more components, and better than symmetric and skewed Student's *t*-GARCH models. In addition to the extensive empirical results based on simulation and on historical data on three US dollar foreign exchange rates (British pound, Euro and Japanese yen) we derive: expressions for the conditional and unconditional moments of all models; parameter conditions to ensure that the second and fourth conditional and unconditional moments are positive and finite; and analytic derivatives for the maximum likelihood estimation of the model parameters and standard errors of the estimates.

Author Contacts:

Prof. Carol Alexander (corresponding author)
Chair of Risk Management and Director of Research
ISMA Centre, School of Business,
University of Reading, Reading RG6 6BA
Email: c.alexander@ismacentre.rdg.ac.uk

Emese Lazar
ISMA Centre, School of Business
University of Reading, Reading RG6 6BA
Email: e.lazar@ismacentre.rdg.ac.uk

JEL classification: C22, C51, C52
Keywords: Volatility regimes, conditional excess kurtosis, normal mixture, heavy tails, exchange rates, conditional heteroscedasticity, GARCH models.

Acknowledgement:
We would like to thank Chris Brooks, Tim Bollerslev, two anonymous referees, Simon Burke, Franz Palm and Charles Bos for insightful comments and very helpful suggestions on an earlier draft of this paper.

1. Introduction

One of the main lines of research in finance is focused on finding an appropriate quantitative description for market returns – and here the unconditional features as well as the time evolution of these characteristics are of interest. Already forty years ago Mandelbrot (1963) and Fama (1965), followed by Westerfield (1977), McFarland, Petit and Sung (1982) and many others, showed that given the excess kurtosis and volatility clustering that characterize returns in financial markets time invariant normal distributions do not offer an appropriate framework. Also, many studies showed that financial returns are characterized by skewness (Bakshi, Kapadia and Madan, 2003).

Consequently there has been a keen interest in developing tractable non-normal models that can be used for option pricing and risk analysis. In particular, there is a large growing literature on the class of hyperbolic distributions, pioneered by Barndorff-Nielsen (1977) and lately developed by Eberlein and Keller (1995) and Barndorff-Nielsen and Shephard (2001, 2002).

One of the simplest and most tractable hyperbolic distributions is the mixture of normal densities, introduced to the financial community by Ball and Torous (1983) and Kon (1984).[1] Normal mixture (NM) densities are weighted sums of normal densities:

$$\eta(x) = \sum_{i=1}^{K} p_i \varphi_i(x) \tag{1}$$

where $[p_1, p_2, \dots, p_K]$ is the positive *mixing law*, with $\sum_{i=1}^{K} p_i = 1$, and $\varphi_i(x) = \varphi(x; \mu_i, \sigma_i^2)$ are normal density functions. We use the notation $X \sim NM(p_1, \dots, p_K; \mu_1, \dots, \mu_K; \sigma_1^2, \dots, \sigma_K^2)$ for a random variable whose distribution is characterised by a density function of this form. Normal mixture densities accommodate for skewness and can be mesokurtic (thin tailed) when the constituent means are different, but are always symmetric leptokurtic (heavy tailed) when the constituent means are equal.

One advantage of the normal mixture over the Student's *t* model for conditional returns distributions is that intuitive interpretations can be placed in the normal mixture framework. For example, Ball and Torous (1983) applied normal mixture models to risk analysis, where the individual distributions in the mixture represent different market states and the mixing law gives the probabilities of these states. In the case of a mixture of two normal densities they differentiate between 'normal' and 'unusual' market conditions, depending on the arrival of new relevant information. Such a distribution may also be supported by the seasonal changes in volatility, for instance the day-of-the week effect as in McFarland, Pettit and Sung (1982). There are also behavioural models to support the use of normal mixtures on market data. For example, the normal densities in the mixture may arise from the different types of

[1] As early as forty years ago, Fama (1965) already discussed a simple form of the normal mixture distribution for returns.

traders in the market, having different expectations regarding returns and volatilities according to which they form their own prices and trade. In this context it is the proportions of the different types of traders that determine the mixing law (Epps & Epps, 1976). [2]

Kon (1984) argued that a mixture of normal distributions fits stock returns distributions better than the Student's *t* distribution. For exchange rate returns, as shown for example by Boothe and Glassman (1987), both the Student's *t* and the normal mixture distribution offer a better description of the data than the normal model. Studying the unconditional fit of stock and index returns, comparing the Student's *t*, the asymmetric stable paretian, the normal mixture and the mixed diffusion-jump distributions, Tucker (1992) showed that it was the class of normal mixture densities that offered the best fit. Also, he showed that in most cases using only two normals in the mixture is sufficient.

In this paper we focus on the classic econometric approach to volatility modelling – as opposed to the option theoretic approach that treats volatility as a continuous-time process.[3] The milestone model in this context is the generalised autoregressive conditional heteroscedasticity (GARCH) model that was pioneered by Engle (1982) and Bollerslev (1986). For example in the univariate *symmetric normal* GARCH(1,1) model the regression equation for the return y_t has conditionally normal errors:

$$y_t = \mathbf{X_t}'\boldsymbol{\gamma} + \varepsilon_t, \qquad \varepsilon_t \mid I_{t-1} \sim N(0, \sigma_t^2)$$

where I_{t-1} represents the information set available at time $t-1$. The conditional variance of these errors is given by: [4]

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad \text{where} \quad \omega > 0, \quad \alpha, \beta \geq 0, \quad \alpha + \beta < 1, \tag{2}$$

The conditional excess kurtosis is zero in the normal GARCH(1,1) model but some degree of leptokurtosis in the unconditional returns distribution can be captured. However, Bollerslev (1987), Baillie and Bollerslev (1989), Hsieh (1989a and 1989b), Johnston and Scott (2000) and many others have all concluded that, in daily or higher frequency data, the observed leptokurtosis in both conditional and unconditional returns is often higher than predicted by the normal GARCH(1,1) model.

Consequently several heavy-tailed conditional densities have been considered in the GARCH framework, including the Student's *t*-GARCH model introduced by Bollerslev (1987) – but this model is characterized by constant conditional skewness and kurtosis, unless these higher moments are explicitly modelled as in

---

[2] The Student's *t* distribution can be represented as a mixture of normal distributions with a continuous mixing variable following an inverted Gamma distribution, so in principle the Student's t allows for a similar interpretation.

[3] The volatility process can be deterministic or stochastic. Deterministic models assume that volatility changes in time according to a predetermined function; popular approaches are the implied tree local volatility approach introduced by Dupire (1994) and the CEV model developed by Cox and Ross (1976). Stochastic volatility models assume that volatility follows a Brownian diffusion process, as in the models developed by Hull & While (1987), Heston (1993), Stein & Stein (1991) and many others. There are also several approaches that combine the option-pricing and statistical volatility schools, one of the most important being the GARCH option pricing model developed by Duan (1995) and Heston and Nandi (2000).

[4] Parameter conditions are needed in order to ensure positivity for the second moment.

Hansen (1994) and Harvey and Siddique (1999), modelling conditional skewness, and Brooks, Burke and Persand (2002), modelling kurtosis. Still, these don't model time-varying conditional skewness and kurtosis taken altogether.

As noted by Bates (1991), Hansen (1994) and Nelson (1996), there are good reasons to favour GARCH models that capture time-variability in higher moments. It is well known that the existence of the smile and/or skew in implied volatility surfaces is largely due to excess kurtosis and/or skewness in the conditional densities of the underlying returns (for instance, see Alexander, 2001, p. 30-31). The assumption that conditional kurtosis and/or skewness are constant over time implies that the implied volatility surface should also be fixed over time. However this is clearly not the case as shown by several studies of the dynamics of implied volatility surfaces, for instance by Derman (1999) and many others. Bakshi, Kapadia and Madan (2003) also emphasise the importance of capturing time-variation in skewness for equity option pricing. Thus statistical volatility models with constant conditional higher moments clearly place unrealistic restrictions on the underlying price process.

The class of GARCH models with a flexible parametric error distribution based on the exponential generalized beta (developed by Wang, Fawson, Barrett and McDonald, 2001), allowing for conditional skewness and kurtosis, was shown to outperform *t*-GARCH models for FX data. A discussion of other non-normal distributions that were used in the context of ARCH models can be found in Bollerslev, Chou and Kroner (1992).

As an alternative to heavy-tailed distributions, the inclusion of a trend for the long-term volatility increases the amount of unconditional leptokurtosis captured by a GARCH(1,1) model , as in the component model of Engle and Lee (1999) – but this model cannot account for the time-variability of the conditional higher moments. Non-distributional models, like the semi-parametric ARCH model of Engle and Gonzalez-Rivera (1991) have also been considered.

Also, squared financial returns are highly autocorrelated – as a study by Taylor (1986) shows. GARCH models can partly account for this autocorrelation – but cannot explain the full extent of it, as found by Ding, Granger and Engle (1993). To overcome this problem, the Long Memory ARCH model was proposed that mimics the empirical autocorrelations quite well (Ding and Granger, 1996). Unfortunately this model does not imply time-varying higher moments.

In the financial risk analysis framework, these models yield returns distributions that are more realistic than the simple normal GARCH returns distributions. However, these models are not very tractable. Analytic derivatives are too complex to derive for the model parameter estimates and their standard errors, necessitating the use of numerical methods.[5] By contrast with normal GARCH, the connection with stochastic volatility option pricing models of the *t*-GARCH is opaque, due to the complexity of its diffusion limit. The Student's *t*-GARCH price density will not have simple analytic properties and consequently there is no clear relationship between Student's *t*-GARCH option prices and Black-Scholes prices.

The above observations indicate another advantage of the normal mixture over the Student's *t* framework for modelling conditional returns distributions. The analytic results on option prices for normal GARCH models may be easily translated into the normal mixture setting. Considering first the local volatility framework, Brigo and Mercurio (2002) prove that if the underlying returns have a conditional normal variance mixture representation (equivalently, if the risk-neutral density of the log price is a normal mixture) then, under certain regularity conditions, the local volatility function has a simple analytic form. The consequence is that unique risk-neutral European option prices and hedge ratios exist under the normal mixture diffusion and these will be averages of Black-Scholes prices and hedge ratios weighted by the mixing law.[6] Now, although the diffusion limit of the normal mixture GARCH will be a stochastic volatility model, recent research on the normal mixture diffusion also encompasses this interpretation, where the mixing law gives the probabilities that the volatility takes a specific value (Mercurio, 2002). Hence a normal mixture GARCH process implies a normal mixture log price density.[7] It follows that closed-form normal mixture GARCH European option prices will be weighted sums of the closed-form normal GARCH option prices derived by Heston and Nandi (2000), where the weights are those given in the mixing law. Finally it is worth mentioning that when the implied volatility smile/skew is decomposed into short-term and long-term effects, the normal mixture diffusion local volatility model can capture the decreasing term structure of excess kurtosis/skewness that is implied by the central limit theorem (Alexander 2004). Clearly the excess kurtosis and skewness forecasts from a statistical volatility model should also display such a term structure property. However, although this is implicit in many normal mixture GARCH models, there is no such term structure in the Student's *t*-GARCH model.

Recently, several authors have examined the class of 'NM-GARCH models', i.e. models where errors have a normal mixture conditional distribution with GARCH variance components. These models, besides having a skewed leptokurtic conditional density, can also account for time-variation in the conditional skewness and kurtosis. Several authors have already studied restricted versions of such models. The simplest model of this form, treated by Roberts (2001), has error conditional densities that are a mixture of two normal densities where one of the variance components is constant. Earlier, Vlaar

---

[5] Of course, the lack of analytical framework can be compensated by improved computational power, and this will facilitate the increasing use of numerical methods in the future.

[6] Melick and Thomas (1997) also assumed lognormal mixture stock prices and applied European option prices as weighted averages of Black-Scholes prices.

[7] In fact, normal mixture log price densities can result from other stochastic volatility models: see Andersen, Bollerslev and Diebold (2002); Barndorff-Nielsen and Shephard (2002).

and Palm (1993) considered another restricted form of NM-GARCH, assuming a mixture of two normal distributions where the difference between the conditional variances of the components is constant, this way incorporating only constant jumps in the level of the variance. The simplicity of the model is appealing. When applied to weekly exchange rate data, the model dramatically improves the fit compared to the normal GARCH(1,1), also accounting for almost all the skewness and kurtosis in the data. Another restricted NM-GARCH model is that of Bauwens, Bos and van Dijk (1999) and Bai, Russell and Tiao (2001, 2003) where the ratio of the two conditional variances is constant, so the conditional kurtosis is constrained to be constant. Also, one of the models proposed by Ding and Granger (1996, p.200-203) to capture volatility components of differing persistence can be viewed as a restricted version of the NM(*K*)-GARCH(1,1) model. In this case, however, the individual densities in the mixture have zero means and all individual variances are restricted to have the same long-term expectations, so the unconditional and conditional skewness are zero. [8]

In a recent discussion paper, Haas, Mittnik and Paolella (2002) specified the general framework for NM(*K*)-GARCH(*p*, *q*) models, assuming an inter-dependent autoregressive evolution for the variance series:

$$y_t = \mathbf{X_t}'\boldsymbol{\gamma} + \varepsilon_t \qquad \varepsilon_t \big| I_{t-1} \sim NM(p_1,\ldots,p_K;\mu_1,\ldots,\mu_K;\sigma_{1t}^2,\ldots,\sigma_{Kt}^2), \qquad \sum_{i=1}^{K} p_i = 1, \tag{3}$$

$$\sigma_{it}^2 = \omega_i + \sum_{j=1}^{q}\alpha_{ij}\varepsilon_{t-j}^2 + \sum_{k=1}^{K}\sum_{j=1}^{p}\beta_{ikj}\sigma_{k,t-j}^2 \qquad \text{for } i = 1,\ldots,K$$

The individual variances are related through their common dependence on ε_t and also through cross-equation effects where lagged values of the k^{th} variance component affect the current value of each variance component. However Haas, Mittnik and Paolella observed that the cross dependence of component variances does not appear to lead to significant improvements of the model, and neither does the inclusion of more than one lag in the conditional variance equations.

These models are related to another important GARCH model with non-normal error distributions, the Markov Switching (MS) GARCH model – both of them assume more than one volatility regime and both have *K* individual conditional variance equations. There are two differences between the two models: (1) whilst a MS-GARCH model estimates the (time-varying) probability that each observation belongs to a given volatility regime, for the NM-GARCH model what is important is the overall (time-invariant) probability that a given regime occurs over the entire sample; and (2) the conditional variance of the MS-GARCH model is equal with the variance of the prevailing regime, but for the NM-GARCH the conditional variance is equal with the weighted average of the conditional regime variances. Hamilton and

[8] Under certain conditions, the limiting case of this model when the α's and β's follow a Beta distribution and the number of components goes to infinity is the Long memory ARCH model (Ding and Granger, 1996). This model, although very good in capturing the ACF pattern of the absolute returns, has only 4 variables in the (unique) variance equation and is not able to model time-variation in higher moments.

Susmel (1994) and Cai (1994) introduced the MS-ARCH model, but also concluded that GARCH models with regimes are impossible to estimate due to the dependence of the conditional volatility on the ruling regime. Later, a tractable MS-GARCH model was presented by Gray (1996) and a modification of this was suggested by Klaassen (1998). Here the variance, instead of being equal to the variance of the existing regime (as it would be in a pure MS model), is the weighted average of all the regime-specific variances, no matter what the ruling regime is. The model is basically a combination of normal mixture and Markov switching GARCH models. It is similar to a normal mixture specification in the sense that the conditional distribution of the error term is a normal mixture, but it is a Markov switching model in the sense that, having time-varying probabilities, the model estimates the regime for each time step.

The purpose of this paper is to extend the literature on symmetric normal mixture GARCH models in the following ways: (1) to derive analytic expressions for the derivatives in the maximum likelihood optimisation and standard error computation, thus avoiding the need for time-consuming and imprecise numerical methods; (2) to derive analytic expressions for the conditional and unconditional moments of the error term and hence state explicit parameter conditions for positive, finite second and fourth moments of the error term and to derive the autocorrelation function of squared returns; (3) to assess the potential bias and inefficiency in parameter estimates, as a function of the mixing law parameters; (4) to determine the optimal number of normal densities in the mixture when the model is applied to historical exchange rate returns and (5) to compare the model with the standardized and skewed *t*-GARCH(1,1) models using several selection criteria. To achieve this last aim we apply NM(*K*)-GARCH(1,1) models to three US dollar exchange rates: the British pound, the Euro and the Japanese yen, with one, two and three normal densities in the mixture.

The structure of the paper is the following: the next section uses the conditional and unconditional moments of normal mixture GARCH models to define parameter restrictions for positivity and finiteness of the conditional and unconditional even moments. Section 3 examines the bias and efficiency of the model parameter estimates using Monte Carlo simulations. These results highlight some pitfalls when using normal mixture GARCH models with more than two components, namely that estimation becomes difficult and can lead to biased estimates because one of the components will, most likely, have a low weight in the mixture. Section 4 surveys the existing literature on exchange rate modelling and investigates the fit of these models when applied to the three major exchange rates, comparing GARCH(1,1) models with normal, normal mixture (with two and three components) and Student's *t* densities, with and without specific parameter restrictions. Our comparisons of a total of fifteen different models are based on eight types of model selection criteria, including likelihood tests, conditional moment tests, conditional and unconditional moment analysis, density fitting based on simulated histograms, the comparison of empirical with theoretical autocorrelation functions of the squared residuals and Value-at-Risk (VaR) estimations. Normal mixture GARCH model estimations are here based on the use of the analytic derivatives of the likelihood function and analytic standard errors that are stated in the appendix. Our main results and conclusions are summarized in Section 5.

2. The NM(*K*)-GARCH(1,1) Models

The general model (3) as formulated by Haas, Mittnik and Paolella (2002) has a very large number of parameters. Since these authors found no demonstrated advantage of allowing for cross-equation effects, or of using more than one lag in each of the individual conditional variance equations, we shall examine only this form of NM-GARCH model, which we label the NM(*K*)-GARCH(1, 1) models. Also, since the focus of the GARCH is a volatility model and not a returns model, we shall assume that the conditional mean equation contains no explanatory variables, not even a constant, so that after de-meaning the series we have $y_t = \varepsilon_t$:

$$\varepsilon_t \left| I_{t-1} \sim NM(p_1,\ldots,p_K;\mu_1,\ldots,\mu_K;\sigma_{1t}^2,\ldots,\sigma_{Kt}^2)\right., \quad \sum_{i=1}^{K} p_i = 1 \text{ and } \sum_{i=1}^{K} p_i\mu_i = 0 \tag{4}$$

where the conditional density $\eta(\varepsilon_t) = \sum_{i=1}^{K} p_i \varphi_i(\varepsilon_t)$ is the normal mixture density with φ_i, for $i = 1,\ldots, K$ representing K (≥ 2) normal density functions with the conditional component variance at time *t* given by σ_{it}^2.

A mixture of normal densities generally has non-zero excess kurtosis and skewness. Models characterized by zero-mean densities in the mixture will be called *symmetric* NM(*K*)-GARCH(1, 1) models and these will exhibit zero conditional skewness. Allowing for non-zero means in the normal components we have the general NM(*K*)-GARCH(1, 1) models that can be applied to markets where the underlying returns densities are expected to be skewed and heavy-tailed: for example, the foreign exchange markets.

The NM(*K*)-GARCH(1,1) model requires *K* equations to specify the conditional variance with the variance of each normal in the mixture assumed to follow a GARCH(1,1) process:

$$\sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2 \qquad i = 1,\ldots,K \tag{5}$$

Since the *K* component variances of a NM(*K*)−GARCH(1,1) model have a GARCH(1,1) specification, the overall variance can be expressed as a GARCH(*K*,*K*) variance process (see Appendix A). Thus, according to Bollerslev (1986) the autocorrelations of the squared residuals can be written as an AR(*K*) process. Alternatively, knowing the moments the autocorrelations can be calculated directly (also see Appendix B). The representation as a GARCH(*K*,*K*) process does not imply equivalence, but that the GARCH(*K*,*K*) model (having only 2*K*+1 parameters) is a restricted form of the NM(*K*)−GARCH(1,1) model. This way, all models which can be expressed as restricted GARCH(*K*,*K*) models are also restricted forms of the NM(*K*)-GARCH(1,1) model.

The conditions given in (4) imply that the last mixing parameter and the mean of the last density in the mixture can be expressed as a function of the other parameters, so the NM(K)-GARCH(1, 1) model that we study here has $5K$-2 parameters:

$$\boldsymbol{\theta} = (p_1,\ldots, p_{K-1}, \mu_1,\ldots,\mu_{K-1}, \omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2,\ldots, \omega_K, \alpha_K, \beta_K)'.$$

To obtain the optimal estimates, we maximize $\sum_{t=1}^{T} ln\left[\eta(\varepsilon_t)\right]$ using a gradient method.[9,10] In the earlier studies on normal mixture GARCH models, numerical approximations were implemented for the gradient vectors used to derive the optimal estimates and the Hessian matrices that are required to estimate the standard errors of the parameters. Here, in Appendix C, we derive analytic expressions for the first and second order derivatives of the likelihood function with respect to the parameters for the NM(K)-GARCH(1,1) model. Consequently, these have been used in a more efficient implementation of the optimisation algorithm to obtain the simulation and historical estimation results reported in this paper.

The conditional and unconditional variance, skewness and kurtosis of the NM(K)−GARCH(1,1) model are given in Appendix B. Clearly parameter conditions are required to ensure these exist and the even moments are positive. The following parameters restrictions are necessary and sufficient for all the component unconditional variances, the overall unconditional variance and the unconditional kurtosis to be finite and positive:

R1. The mixture parameters must be positive and their sum must be less than one. Also, to ensure that the variance processes are not explosive, the α's must be positive and each β_i must be positive and less than one. It is not necessary to set all ω's positive; some of them may be negative but with small absolute value. Thus we have the following first set of restrictions:

$$0 < p_i < 1, \quad i = 1,\ldots, K-1, \quad \sum_{i=1}^{K-1} p_i < 1, \quad 0 < \alpha_i, \quad 0 \leq \beta_i < 1, \quad i = 1,\ldots, K \tag{R1}$$

R2. All individual long-term variances must exist and be finite and positive. In fact, a finite and positive overall long-term variance and a small ω in absolute value (if negative) are sufficient to ensure that all individual long-term variances are finite and positive. This observation follows from restriction (R1) and the following relationship between the individual and overall long-term variances:

$$E(\sigma_{it}^2)(1-\beta_i) = \omega_i + \alpha_i E(\varepsilon_t^2) \tag{6}$$

---

[9] Common algorithms used for ML optimisation are the Berndt-Hall-Hall-Hausmann (BHHH) algorithm – see Bollerslev (1986) – which approximates the Hessian with the first derivatives and the class of quasi-Newton methods. The most important one of these is the Broyden-Fletcher-Goldfarb-Shanno algorithm, based on an approximating and updating method for the information matrix.

[10] One major problem in any type of optimisation is the search for appropriate starting values, to ensure that the optimisation process leads to the global optimum, instead of a local one. To overcome this problem, as suggested by Doornik and Ooms (2000), an initial grid search can be performed. However, the difficulty of optimisation increases with the number of parameters, thus with the number of components in the mixture.

The overall long-term variance in the case of a NM(K)-GARCH(1,1) model is given by:

$$E(\varepsilon_t^2) = E(\sigma_t^2) = \frac{\sum_{i=1}^{K} p_i \mu_i^2 + \sum_{i=1}^{K} \frac{p_i \omega_i}{(1-\beta_i)}}{\sum_{i=1}^{K} \frac{p_i (1-\alpha_i-\beta_i)}{(1-\beta_i)}} \tag{7}$$

From this and (6), we get that a (necessary and sufficient) additional condition for a finite positive overall and component unconditional second moment is:

$$m = \sum_{i=1}^{K} p_i \mu_i^2 + \sum_{i=1}^{K} \frac{p_i \omega_i}{(1-\beta_i)} > 0, \ n = \sum_{i=1}^{K} \frac{p_i (1-\alpha_i-\beta_i)}{(1-\beta_i)} > 0 \text{ and } \omega_i + \alpha_i \frac{m}{n} > 0, \ i = 1,\ldots,K \qquad \text{(R2)}$$

It can be noticed that if the sum $\alpha_i + \beta_i < 1$ for all i, then the $n > 0$ part of the condition is met. However, this is a sufficient, and not a necessary condition and the condition above is more exact. Actually it is too strict a condition to force $\alpha_i + \beta_i < 1$ for all individual variance components. To see this, note that $E(\varepsilon_t^2)$ is the weighted average of the individual long-term variances, which means that at least one of these variances is higher than $E(\varepsilon_t^2)$. Thus there exists at least one i, $1 \le i \le K$ such that:

$$E(\sigma_{it}^2) < \omega_i + (\alpha_i + \beta_i) E(\sigma_{it}^2) \tag{8}$$

which can be written as:

$$(1-\alpha_i-\beta_i) < \frac{\omega_i}{E(\sigma_{it}^2)} \tag{9}$$

and it can happen that the left hand side (and maybe the right one as well) is negative.

In fact, our own empirical results (see column (15) in Tables 4 – 9) and the results of Haas, Mittnik and Paolella (2002) have shown that the α parameter of the highest variance component may take unusually high values in order to capture a high level for the conditional variance when there are large values or outliers in the data. However, it should also be noted that the α parameter estimate could also be subject to a considerable upwards bias; and this is shown by our simulation results in the next section.

R3. Estimated parameters must also satisfy conditions for the third and fourth moments to be finite and for the fourth moment to be positive.[11] Note that the skewness is finite if each component variance is finite, so R1 and R2 are already sufficient for this. The expression for the fourth moment (given in Appendix B) is complex, and it is not possible to relate existence conditions to simple conditions on estimated parameters. However R1 and R2 allow some of the components to have $\alpha_i + \beta_i > 1$ and in this case the fourth moment might not exist or be negative for certain values of the parameters. For instance, for the symmetric NM(2)-GARCH(1,1) model such a region for the mixing parameter p (keeping the

[11] Amongst others, the studies of Huisman, Koedijk, Kool and Palm (2001 and 2002) support the existence of the fourth moment for the major exchange rates vis-à-vis the US dollar.

other parameters constant) is shown in Fig. 1. The figure also shows that the positivity of the second moment does not ensure the positivity and existence of the fourth moment. This way, our third condition that the parameters must satisfy is that the estimated fourth moment is positive.[12]

[Fig. 1 about here]

3. Simulation Results

From henceforth, for brevity, we use an abbreviated notation 'NM(K)' to denote the class of NM(K)-GARCH(1,1) models. The model parameter estimates are obtained by maximizing the log-likelihood using a gradient method. Earlier studies on normal mixture GARCH models used numerical approximations for the gradient vectors used to derive the optimal estimates and the Hessian matrices that are required to estimate the standard errors of the parameters. Therefore, the analytic forms of the derivatives of the log-likelihood function have been used in a more efficient implementation of the optimisation algorithm to obtain the simulation and historical estimation results reported in this paper.

We are interested in three issues: the *bias* and the *efficiency* of the estimation and the *stability* of our results. In this section we show that the variance parameter estimates for any component(s) having low weight in the mixture will be subject to significant bias. One of the main problems with NM(K) models for $K > 2$ is that often at least one of the mixing parameters takes very low values; so our results in this section indicate that NM(K) models for $K > 2$ are very likely to have estimation problems.

A symmetric NM(2) model with a weight of p on the first variance component is sufficient to illustrate the results. Indeed, it is the best framework to use for this exercise since it has only seven parameters, so the likelihood surface is better conditioned than it is for higher order normal mixture GARCH models. The base parameters for our simulations were chosen to be both realistic (i.e. close to the empirical estimates obtained when implementing the model on historical daily exchange rates) and useful (i.e. they should help us answer questions about bias and efficiency).[13] Thus the base parameter values chosen for these simulations were:

$$\omega_1 = 0.0001,\ \alpha_1 = 0.05,\ \beta_1 = 0.85,\ \omega_2 = 0.01,\ \alpha_2 = 0.1,\ \beta_2 = 0.8$$

The simulation has the following steps: first, the two individual variance processes are simulated and combined to obtain a single time series for the error term. Secondly, the NM(2) parameters of this simulated process are estimated using the time series for the error term as input. The estimated means and standard errors of the parameter estimates are sensitive to the values chosen for the model parameters in the simulation, and to the value of the mixing parameter p in particular.

---

[12] It is common practice to impose restrictions by re-parameterisation, but in this case it would over-complicate the model.

[13] For instance, slightly higher ω_2 than the values estimated from real data may be chosen if estimated values are very low (e.g. of the order of 10^{-4}).

Consequently, to investigate the bias and efficiency of the estimation as a function of the mixing parameter, we performed 9 different simulations for mixing parameter values of 0.1 up to 0.9 (with a step of 0.1) but with the above fixed values for the other parameters. These 9 simulations were performed with different sample sizes: 1000, 2000 and 4000. Fig. 2 summarizes our preliminary classification of estimation results into 'good' and 'unrealistic' estimates.[14] For non-extreme values of the mixing parameters a high percentage of the simulated time series lead to realistic estimates; but the more extreme the mixing parameters the more difficult the estimation becomes because only few observations are drawn from some of the normal distributions, making it difficult to estimate the parameters of the associated variance processes. For example, when p = 0.1 and the sample size is 1000 there are only 100 observations drawn from the first subordinate distribution.

[Fig. 2 about here]

3.1 Bias

With zero bias the histograms of the estimated parameters should be centred on their true values. Fig. 3 presents the estimation results, illustrating the bias of the estimation as a function of the mixing parameter. Clearly ω and α have a *positive bias* whilst β is *biased downwards.* The size of this bias decreases as the sample size increases. The size is also inversely proportional to the mixing parameter associated with the variance component. That is, the bias on the parameters of the first variance component becomes greater as p decreases and that of the second variance component becomes greater as p increases. Even the mixing parameter has a small upward bias, but only when it takes very low values. Interestingly, the overall long-term volatility and the kurtosis are both estimated without bias or with very small bias, even for sample sizes of only 1000. Even the individual component long-term volatilities have no bias, except for very low values of p.

3.2 Efficiency

If the estimation is efficient, then the standard errors should decrease as the sample size increases and the standard error of the estimated parameters should be around the Cramér-Rao lower bound.[15] Fig. 4 illustrates the efficiency of the estimation by comparing the estimated standard errors on the parameter estimates with their Cramér-Rao bound for different sample sizes. Our results show that the estimation of the variance parameters for a component of the mixture becomes more exact as the mixing parameter associated with this component increases and the sample size increases. Intuitively, when the mixing parameter associated with a component is high or when the sample size is high we have many observations drawn from this density leading to a better precision of the estimation of the associated

[14] An estimation is considered *un*realistic in the following three cases: (1) when convergence is not achieved, (2) when boundary values are obtained for one of the α's or β's; or (3) when the main diagonal of the information matrix contains a negative value – all these indicating that a local and not the global optimum was obtained.

[15] The Cramér-Rao bound constitutes the main diagonal element of the inverse of the Information Matrix. These values represent a lower limit for the variance of certain unbiased estimators. Given the complexity of an analytical formula for the Cramér-Rao bound, simulations were used to find an approximation for it.

parameters. The standard errors of the estimation match the Cramér-Rao bound very closely, indicating the efficiency of the estimation, except for small values of p. It is important to note that in all cases the bias is less than the standard error of the estimation, as can be seen comparing the two graphs.

[Fig. 3 and 4 about here]

3.3 Stability

The size of the variance parameters themselves also affects their bias. To see this we individually increase each of the parameters by 10% of their base value, but with a fixed value for the mixing parameter (p = 0.8) using a sample size of 2000. This way the effect of the individual changes in the simulated values of the other parameters is investigated. Fig. 3 shows that the largest change in the bias (expressed as percentage of the standard errors) – and this affects all parameters – is obtained when increasing β_1 by 10%. This is expected because β_1 is the largest parameter in the more important component and with an increase of 10% it is changed from 0.85 to 0.935. However, the change in the bias is still less than 25% of the standard errors of the estimates. When the other parameters are increased by 10% the changes in the bias are less than 5% of the standard error.

[Fig. 5 about here]

3.4 Conclusion

In summary, even with a sample size of 4000 there will be a clear positive bias on the ω and α parameter estimates and a clear negative bias on the β estimate in the variance components having a low weight in the mixture. These biases decrease as sample size increases and as the weight on the component increases. Although the bias is less than the standard error of the estimate, the standard errors increase as the weight on the component decreases (also, the efficiency of the estimation deteriorates). We conclude that if one of the variance components has a very low weight, as is often the case when $K > 2$, precise parameter estimation will be difficult because of the non-linearities of the high dimensional likelihood surface.

4. Empirical Application

In this section we present the literature on exchange rate modelling and estimate the parameters of NM(K) models for different values of K, and some of their restricted forms, using historical data on major exchange rates. The main aim is to test the specifications of these models and to use many model selection criteria to compare them with the fitted normal, and standardized symmetric and standardized skewed t-GARCH(1,1) models.

4.1 Literature Survey

The use of a GARCH framework for exchange rate modelling is an established approach in the literature. Applications arguably began with Engle and Bollerslev (1986) and Hsieh (1988), who examined the statistical properties of daily exchange rates and concluded that exchange rate returns have a distribution that varies over time (a similar result was obtained by Zhou, 1996) and rejected the hypothesis that the data has a heavy-tailed distribution with fixed parameters over time. In a subsequent paper, Hsieh (1989b) showed that a GARCH model can explain a significant proportion of the observed non-linearities for five major exchange rates, but it cannot account for the entire leptokurtosis in the data. Different non-normal distributions – Student's *t*, GED, normal-Poisson mixture and normal-lognormal distributions – combined with the exponential GARCH model can fit exchange rate returns quite well, as obtained by Hsieh (1989a). Similar results were found by Johnston and Scott (2000) and Huisman, Koedijk, Kool and Palm (2002). Other studies using a GARCH framework to model the statistical properties of exchange rates include Baillie and Bollerslev (1989, 1991), Engle, Ito and Lin (1990) and Engle and Gau (1997).

The empirical evidence on the leptokurtosis and volatility clustering found in the distribution of high frequency data makes GARCH techniques preferable and their superiority for modelling exchange rates in particular has been stressed by many authors: Mckenzie (1987), West and Cho (1994), Christoffersen (1998) to mention but a few. The use of asymmetry in the unconditional distribution of exchange rates has been stressed by de Vries (1994). Nevertheless, other related lines of research on exchange rate behaviour include the use of a mixed jump diffusion – see Jorion (1988) – and, more recently, Markov switching models, as in Engel and Hamilton (1990) and Engel (1994). Another alternative to analyse exchange rate return variability, proposed by Andersen, Bollerslev, Diebold and Labys (2000, 2001, 2003), is to estimate the realized volatility directly by simply summing the intra-day squared returns.

We are currently witnessing a debate concerning the performance of these models. The results of Tucker and Pond (1988) and Akgiray and Booth (1988) favour the mixed jump model. On the other hand, a study of Johnston and Scott (1999) concludes that none of these models consistently dominates the others. In a recent study, Hansen and Lunde (2001), comparing 330 GARCH models for volatility forecasting, cannot point out one single model that outperforms the others in the case of exchange rates, although they found that models based on leptokurtic distributions do better than those based on normal distributions. They conclude that the best models do not provide a significantly better forecast than the normal GARCH(1,1) model, and that none of the models that they considered capture totally the behaviour of exchange rates: even the *t*-GARCH models were not able to explain fully the excess kurtosis in the data. A possible explanation for their results is that symmetric Student's *t* densities have only three parameters, so they are not flexible enough to capture the heavy tails of the empirical distributions. On the other hand, normal mixture distributions have more parameters (for example, a zero-mean mixture of just two normal densities already has four parameters), so they are more flexible in capturing leptokurtosis.

Normal mixture distributions have been applied to the unconditional distribution of exchange rate returns by many authors: Boothe and Glassman (1987), Zangari (1996) and Hull and White (1998) to mention but a few. However, the empirical literature on NM-GARCH modelling of exchange rate returns is still in its infancy. The study of Vlaar and Palm (1993) on weekly European exchange rates shows that their model is capable, in most cases, of accounting for the excess conditional kurtosis present in the data. Also, Bai, Russell and Tiao (2001, 2003) using intra-day data on the Deutsche mark, French franc and Japanese yen exchange rates, obtained significant improvements on the unconditional kurtosis estimates, compared to the GARCH(1,1) model.

4.2 The Data and its Properties

The data consists of daily prices of three foreign currencies (British pound, Euro and Japanese yen) in terms of US dollar, covering a fourteen-year period from 2nd January 1989 to 31st December 2002 (a total of 3652 observations), provided by Datastream. We denote these exchange rates GBP, EUR and JPY respectively. Daily returns are computed as the (annualised) difference in the logarithm of daily closing prices.[16] The time evolution of the returns is presented in Fig. 6, and Table 1 reports some statistical properties of the data.[17] From the first four moments of the unconditional distributions, we see that the mean returns are not significantly different from zero for any of the three exchange rates. The skewness is significant for the GBP and JPY rates, while the significance of the excess kurtosis for all three rates is high, especially for JPY. Also, the Ljung-Box statistic shows that the data provide no evidence of autocorrelation.

[Fig. 6 and Table 1 about here]

Since the focus of this section is to estimate the parameters of NM(*K*) models and some of its restricted forms, for different values of *K*, and to compare them using model selection criteria with the symmetric and skewed *t*-GARCH(1,1) models, and since these models are designed to capture, specifically, the time-variation in the higher moments of conditional distributions of returns, we need to test for the existence of these moments in the sample data used. Following Hill (1975), we shall estimate the *maximal moment exponent* defined as

$$a = sup\left\{b > 0 : E\,|\,\varepsilon_t\,|^b < \infty\right\}$$

To this end, let

$$\hat{a}_s = \left(s^{-1}\left(\sum_{j=0}^{s-1} ln\,\varepsilon_{n-j} \right) - ln\,\varepsilon_{n-s} \right)^{-1} \tag{10}$$

---

16 Zero returns have been removed as they most often indicate missing data and distort the likelihood surface, namely 78, 138 and 168 data points are missing for the GBP, EUR and JPY series, respectively.

17 Based on the BIC criteria, the following AR(1) model is chosen for the daily returns on the British pound : $r_t = \varepsilon_t + 0.06061 * \varepsilon_{t-1}$. Similarly, the following AR(1) model describes the EUR/USD rate: $r_t = \varepsilon_t - 0.03115 * \varepsilon_{t-1}$. Subsequently the terms GBP and EUR will signify the residuals from these regressions. No autoregressive effects were found necessary in the conditional mean dollar returns for the Japanese Yen.

where s is a positive integer and $\varepsilon_1 \leq \ldots \leq \varepsilon_n$ are the ordered returns (with n being the sample size). Hall (1982) showed that under certain conditions, $\sqrt{s}(\hat{a}_s - a)$ is asymptotically normally distributed with standard deviation a.[18] Thus a large sample fourth moment test becomes one of testing the hypothesis that $a \geq 4$ using this as test statistic. Table 2 presents the estimates for the maximal moment exponent and the test results for the existence of the fourth moment, according to which we cannot reject the existence of the fourth moment.

[Table 2 about here]

We also performed the small sample 'HKKP' tests for the existence of the fourth moment that have been derived by Huisman, Koedijk, Kool and Palm (2001, 2002), based on the same exchange rate data but now divided into the three sub-periods defined below. The starting point is the dependence of Hill's estimate on s and the goal is to obtain an estimate for $s = 0$. The HKKP method computes Hill's estimate for values of s smaller than a threshold value and then regresses the obtained estimates on s. The correct estimate for the maximal moment exponent is the intercept in this regression.

We already have a strong prior that the fourth moment exists because Huisman, Koedijk, Kool and Palm's own results support the existence of the fourth moment for all major US dollar exchange rates, based on daily data from 1979 to 1996. The results of our HKKP tests led to the same conclusion, for all exchange rates.[19]

4.3 Description of the Models

Thirteen NM(K) models, for $K = 1$, 2 and 3 respectively, are estimated for each of the three exchange rate series, using the whole fourteen years of data.[20] Of course, the NM(1) model is equivalent to the normal GARCH(1,1). In addition to the general normal mixture GARCH models several restricted models were also estimated, restrictions being based on: *(i)* zero-mean normals in the mixture, *(ii)* similar GARCH processes for the individual variances (inspired by the Vlaar and Palm model), and *(iii)* a constant variance for one of the components of the mixture density (as in Roberts, 2001). These models all have time-varying conditional kurtosis and, except for the symmetric model *(i)* without any other restriction, their conditional skewness also displays time-variation. Additionally, the GARCH(1,1) model with Student's *t*-distributed errors (Bollerslev, 1987) and its skewed version (Fernandez and Steel, 1998 and Lambert and Laurent, 2001a and 2001b) were also fitted to the data. Although the Student's *t*-GARCH models do not

---

[18] The conditions are that n is large enough, that $s = s(n)$ is a function of n of a pre-specified order, that s/n is small enough and that the tails of the distribution have the asymptotic Pareto-Lévy form. Also, the heteroscedasticity found in the data is not affecting the results since Hill's method excludes the analysis of the dynamics of the series.

[19] Results available from the authors on request. Only in one particular sub-period (2nd Jan 1998 – 31st Dec 2002) did the HKKP test for JPY rejected the existence of the fourth moment.

[20] The study of Tucker (1992) showed in most cases using only two normals in the mixture is sufficient to model stock returns. Although exchange rates behave differently from stock prices, we shall show that Tucker's findings are supported by our own.

exhibit time-varying conditional skewness and kurtosis, they might offer a better unconditional fit than normal mixture GARCH models. The fifteen models are summarized in Table 3. [21]

[Table 3 about here]

4.4 Estimation Results

Tables 4, 5 and 6 present the estimation results for the three exchange rates for the entire period.[22] Standard errors are computed as the square root of the diagonal elements of the information matrix (see Appendix C for its derivation).

[Tables 4, 5 and 6 about here]

When fitting the NM(2) models (4) – (9) the components of the mixture distributions can easily be differentiated: in all cases the lower long-term volatility component has the higher value for the mixing parameter. Thus the model captures two distinct 'regimes' in exchange rate volatility, a 'usual market circumstance' volatility which occurs most of the time, and an 'extreme market circumstance' volatility which occurs rarely, but which is higher than the 'usual' one. The estimated weights in the mixing law may be interpreted as the frequencies with which these two states occurred during the sample period.

When fitting the NM(3) models (10) – (15) the components of the mixture distributions are more difficult to differentiate. In most cases the component with the highest mixing parameter has an average long-term volatility, and the other two components (with lower and higher long-term volatilities) each have a smaller mixing parameter. In this case the model is capturing two 'exceptional circumstances' in volatility – one corresponding to unusually tranquil markets, and the other corresponding to unusually volatile markets. However it can happen that the highest weight in the mixing law is associated with the lowest long-term volatility. Thus adding an extra component to the mixture of two distributions leads to a decrease in the weights on both volatility components – the NM(3) model captures a 'medium' volatility or 'low' volatility for usual market circumstances.

All parameter estimates in Tables 3 – 5 are in line with the stylised facts of GARCH models on major exchange rates. In fact, the pattern in the estimates is remarkably similar across all three exchange rates: except for the unrestricted symmetric and asymmetric NM(3) models (12) and (15), all β parameter estimates lie in the range [0.93, 0.95] and all α parameter estimates lie in the range [0.02, 0.12]. However, models (12) and (15) have a larger α and/or a smaller β for all three exchange rates.[23] There are two reasons for this effect: (1) the third (lowest weight) component is most sensitive to large absolute returns, leading to a less persistent individual GARCH process; and (2) as shown by our simulation results, there

---

[21] Recall that, for brevity, we refer to the NM(2)-GARCH(1,1) models as the NM(2) models and the NM(3)-GARCH(1,1) as the NM(3) models.

[22] The results were generated using C++ and Ox version 3.30 (Doornik, 2002) and the G@rch package version 3.0 (Laurent, S. and Peters, J.-P., 2002)

[23] For the GBP and JPY rates this is also observed in the restricted models (11) and (14).

can be a large upwards bias on α and a large downwards bias on the β parameter estimate of the third component. We can already conclude that the general symmetric and asymmetric NM(3) models (12) and (15) have estimation pitfalls that users should be aware of. In fact, several further empirical results in this section will corroborate this view.

4.4.1 Robustness Check 1

In order to study the effect of extreme events on the estimation we repeated all the estimations after excluding the most extreme returns.[24] Namely, an (annualized) return of -51% on the 16th Sept 1992 for the GBP rate, one of 67% on the 26th May 1995 for EUR, and a return of 121% on the 7th of Oct 1998 for the JPY rate. Tables 7, 8 and 9 present the results. Clearly the model estimations pass this robustness check because although these returns were each very extreme, no significant changes in the parameter estimates were found. Which parameters are most affected by the exclusion of the most extreme return? One would expect a decrease in the mixing parameter of the component with the highest long-term volatility and a slight increase in the β parameters of the individual variances. Generally, the results are indeed as expected. Also, although NM(3) parameter estimates are more sensitive to extreme observations than NM(2) models, the problems with the unrestricted models (12) and (15) outlined above remain. Thus the very large bias in parameter estimates for these two models is not caused by extreme values in the data.

[Tables 7, 8 and 9 about here]

4.4.2 Robustness Check 2

Though useful for estimating models with many parameters – and the general NM(3)-GARCH(1,1) has thirteen – some GARCH empiricists would consider fourteen years too long a sample to obtain meaningful parameter estimates from the GARCH estimation. Therefore, in addition to the entire period under study, the basic models

- Model (1): Normal GARCH(1,1)
- Model (3): Skewed *t*-GARCH
- Model (9): Unrestricted NM(2)-GARCH(1,1)

were re-estimated on three sub-periods chosen to have roughly equal length: the first four years (2nd Jan 1989 – 31st Dec 1992), the middle five years (2nd Jan 1993 – 31st Dec 1997) and the last five years of the sample (2nd Jan 1998 – 31st Dec 2002). The NM(3) model estimations over the sub-periods are not reported because, although each period contained several years of data, substantial convergence problems arose in many instances from the ill-conditioned likelihood surface.

---

[24] Since normal mixture distributions capture heavy tails, the concept of outlier has to be used carefully in our situation. Instead, we prefer the expression 'extreme returns' and – since no outlier detection method exists for NM-GARCH models (although they might be constructed) – we simply choose the highest return in absolute value.

This split of the sample into three parts is useful to check the robustness of our parameter estimates, but one has to bear in mind that the three sub-periods have quite different characteristics. For the British pound, the first sub-period is characterized by an average volatility of 11.8% and an excess kurtosis of 1.60; the second sub-period is more stable, with a lower average volatility of 8.4%, but with a higher excess kurtosis, of 2.89, and the last sub-period has a very low average volatility, only 7.2%, and the excess kurtosis decreases to 1.14. Similarly, in the case of the Euro, the first period is the most volatile, and the excess kurtosis, probably due to extreme returns, is highest for the second period, having a value of 3.8. The last period is characterized by a very low excess kurtosis of only 1.24. For the Japanese yen, the last period is the most volatile and also has an exceptionally high sample excess kurtosis, of 8.97 (and this was the period where the HKKP fourth moment test failed, for the Japanese yen). Given the different characteristics of each of the sub-periods, we should not expect the model parameter estimates to be identical in all three periods. The results in Tables 10, 11 and 12 show that, generally speaking, the parameter estimates do indeed change in line with the particular qualities of the data during the period.

[Tables 10 – 12 about here]

In each exchange rate and for each sub-period the simple GARCH(1,1) model has a much lower in-sample likelihood than the other two models. However there are many other and more important criteria that should be taken into account when classifying models and we shall now consider these.

4.5 Model Selection

To decide which model has the best fit, we have applied a number of model selection criteria. The results are given in the lower portion of Tables 4 – 9. It is notable that the removal of the largest extreme value from the sample had little effect on the parameter estimates and consequently did not affect the qualitative nature of our results. Thus similar inference can be made from both sets of table: Tables 4 – 6, and Tables 7 – 9.

4.5.1 Likelihood-based criteria

First the values of the log-likelihood, the Akaike Information Criterion (AIC) and Schwartz's Bayesian Information Criterion (BIC) are examined. As expected the highest likelihood is obtained from the model with the most parameters, the unrestricted NM(3). The information criteria, which take into account the parsimony of the parameterisation, generally favour the *t*-GARCH models. The simple GARCH(1,1) model performs worst under all three criteria, for all three exchange rates.

4.5.2 Likelihood ratio test

To complete the likelihood-based selection criteria we also compare the restricted NM(*K*) models with their simplest generalizations. That is, models (4) and (5) are compared with model (6), models (6), (7) and (8) are compared with model (9), models (10) and (11) are compared with model (12) and models (12), (13) and (14) are compared with model (15). The Likelihood ratio statistic for this test is:

$$LR = -2(L_R - L_U) \sim \chi^2(r)$$

where L represents the log-likelihood for the restricted and unrestricted models, and r is the number of restrictions. These tests favour few, if any restrictions on the parameters. For the GBP rate the symmetric normal mixture models (with no other restrictions) are preferred, whilst for the other two exchange rates the criterion selects the general unrestricted asymmetric models.

4.5.3 Moment specification tests

To check the adequacy of each model to capture the higher moments of the conditional returns densities, the tables also report results of moment specification tests. In order to test whether the moments of the error densities match the ones specified by the estimated distributions, the errors must be transformed into a series that has a standard normal distribution under the null hypothesis that the estimated model is valid. Thus for each realization of the error term ε_t, the cumulative normal mixture distribution is computed:

$$P_t = \mathrm{N}(\varepsilon_t) = \sum_{k=1}^{K} p_k \Phi_k(\varepsilon_t) \tag{11}$$

where K is the number of normal densities in the mixture, and Φ_k is the cumulative normal distribution function of the k^{th} element of the mixture. Under the null hypothesis, P_t will be independently and uniformly distributed and then the inverse cumulative standard normal distribution of P_t gives a series $u_t = \Phi^{-1}(P_t)$ which should be independent standard normally distributed. As in Harvey and Siddique (1999), this is verified by checking its moments for the following conditions:

$$\begin{aligned} &E(u_t) = 0 \\ &E(u_t^2 - 1) = 0 \\ &E(u_t^3) = 0 \\ &E(u_t^4 - 3) = 0 \end{aligned} \tag{12}$$

Since the transformed errors should not exhibit any autocorrelation in the powers, we also should have that:

$$\begin{aligned} &E(u_t u_{t-j}) = 0 \\ &E(u_t^2 u_{t-j}^2) = 0 \\ &E(u_t^3 u_{t-j}^3) = 0 \\ &E(u_t^4 u_{t-j}^4) = 0 \end{aligned} \qquad j = 1, \ldots, m \tag{13}$$

A cumulative test is also carried out that the all of the conditions for the powers of the error term mentioned above are jointly true – we take m = 4 – although it is well-known that the cumulative test is too stringent for most practical applications. Following Greene (2000, chapter 11.6.2), a Wald test approach is used and the derivation of the test statistic is given in Appendix F.

Our results show that the GARCH(1,1) model, the restricted versions of the NM(2) model and some of the NM(3) restrictions have severe rejections of model appropriateness, especially when testing the third and fourth moments. Again it is the two *t*-GARCH and the NM(3) models that are favoured by the moment tests. These models can account for almost all the unconditional skewness and excess kurtosis in the data, since the residuals show no clear-cut evidence of non-normality.

4.5.4 Histogram-based criterion

One of the more difficult unconditional distribution tests for GARCH models to pass is based on the comparison of the empirical returns density with a simulated returns density generated by the estimated GARCH model. That is, we simulate returns based on the estimated parameters (and to make sure the simulation is not affected by small sample size we use 50000 replications) where each simulation is 1000 steps ahead in time (in order to avoid any influence of the starting values) but we only use the last simulated return. Subsequently we estimate the theoretical (simulated) histogram using a nonparametric kernel approach, approximating the density with the following expression:

$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{y_i - y}{h}\right)$$

where n is the size of simulation, h is the smoothing parameter (window width), y_i are the simulated returns, and k is a smooth weighting function satisfying certain conditions. Several alternatives are available for the kernel k, our chosen function being the *Epanechnikov optimal kernel* (Epanechnikov, 1969):

$$k(x) = \begin{cases} \frac{3}{4}(1 - x^2), & |x| \le 1 \\ 0 & \textit{otherwise} \end{cases}$$

The estimator can be sensitive to the choice of the window width so we used the following two bandwidths:

$$h = 1.06\sigma_x n^{-1/5} \text{ and } \qquad h = 2.34\sigma_x n^{-1/5}$$

but our results were the same in each case. Fig. 10 presents the approximated histograms for the data and the most important models: GARCH(1,1), skewed *t*-GARCH, and the unrestricted NM(2) and NM(3). These show how bad a histogram fit is offered by basic normal GARCH(1,1) models.

[Fig. 10 about here]

The model selection criteria is based on the *Kolmogorov-Smirnov statistic* (Kolmogoroff, 1933, Smirnov, 1939 and Massey, 1951) $KS = D\sqrt{n}$ where

$$D = \max_{1 \le i \le n}\left|F(y_i) - \frac{i}{n}\right|,$$

and $F(y_i)$ denotes the cumulative density function. In fact the two-stage delta-corrected version of (Khamis, 2000) is a more powerful tool; it modifies the above expression to:

$$D = \max_{\substack{1 \le i \le n \\ \delta = 0 \, or \, \delta = 1}} \left| F(y_i) - \pi_i(\delta) \right|, \qquad \text{where } \pi_i(\delta) = \frac{i - \delta}{n - 2\delta + 1}, \qquad 0 \le \delta \le 1$$

The model with the minimum *KS* (or modified *KS*) generates the histogram that is closest to the empirical histogram of returns.[25] In Tables 4 – 9 we only report the modified *KS* statistic as similar results are obtained for the basic *KS* criteria. For all three currencies, including or excluding the most extreme return, it is always the unrestricted NM(2) model (9) which minimizes both the *KS* statistic and its modified form.[26] This result is not marginal: the *KS* statistics are *substantially* lower for the unrestricted NM(2) model than for most other models.

4.5.5 ACF-based criterion

The ability of the unrestricted NM(2) model to fit the unconditional returns density is a very important result. However, this does not imply that the same model best captures the *dynamic* properties of the returns – namely, the empirical autocorrelations of the squared residuals. We therefore use the theoretical autocorrelation functions of the different models (see Appendix B, D and E) to estimate the autocorrelations of the squared residuals, and apply the Mean Squared Error (MSE) criterion to compare how different models fit the empirical autocorrelations. The MSE statistics are reported in the second part of Tables 4 – 9.[27] Overall, again the NM(2) models perform best. In general they are better than the NM(3) models and both normal mixture models do better than the *t*-GARCH models. For the GBP and JPY rates the *t*-GARCH specifications perform very badly indeed, having an MSE around 10 times higher than that of the NM(2) general model. On the other hand, for the EUR series the *t*-GARCH models give a reasonably good fit. It is worth mentioning that the GARCH(1,1) model and the restricted normal mixture models also perform quite well according to this criterion.

4.5.6 Conditional moment comparison

Fig. 6 above compares the time series for conditional volatility obtained from the normal GARCH(1,1), skewed *t*-GARCH, and the two unrestricted normal mixture GARCH(1,1) models, denoted NM(2) and NM(3) for short, as before. From this perspective there is no significant difference between the more advanced models and the simple normal GARCH(1,1) model. Fig. 7 plots the evolution of individual volatilities of the unrestricted NM(2) and NM(3) models, compared with the overall conditional volatilities. The shapes are similar (not same!) with drifts differentiating the individual volatilities. The highest volatilities in the NM(3) model are far too unstable. Important differences arise when the higher

---

[25] This statistic has a *KS* or modified *KS* distribution if the theoretical distribution is specified in advance. Since our simulations assume an estimated and not a predefined set of parameters, the KS tests have lower significance level, invalidating the test results – this is why we don't perform the *KS* tests, we only report the test statistics. However, it is correct to compare the test statistics across different models.

[26] The one exception being in the EUR data set over all 14 years with the most extreme observation removed, where the unrestricted NM(2) model is second to the skewed *t*-GARCH.

[27] Our criterion is based on the first 250 autocorrelations, the reason being twofold: first, lower order autocorrelations bear higher significance, and secondly the length of our data period does not allow us to take too high orders for the ACF.

conditional moments are inspected. One major disadvantage of the GARCH(1,1) and the two *t*-GARCH models is that they have no time-varying conditional moments – although these could be explicitly modelled as in Hansen(1994), Harvey and Siddique (1999) and Brooks, Burke and Persand (2002). Thus, in Fig. 8 and 9, we compare only the conditional evolution of skewness and kurtosis of the NM(2) and NM(3) models.

[Fig. 6, 7, 8 and 9 about here]

(1) Unrestricted – restricted models comparison

Immediately obvious is that the restricted models which have a time-varying conditional skewness (models 7, 8, 13 and 14) result in a too volatile process for the conditional skewness – whilst the levels are similar in most cases. The evolution of the conditional kurtosis for the different models also emphasizes that the conditional estimates of the restricted models are far too volatile. The only exception is the restriction resulting in a symmetric NM(K) model (without any other restriction), case in which the conditional skewness is zero, but the conditional kurtosis estimate closely matches that of the unrestricted model.

(2) NM(2) – NM(3) comparison

The graphs show that whilst the NM(2) and the NM(3) models have a similar level for the conditional skewness, the NM(3) models have a too volatile evolution. Whilst the two classes of models also agree on the general level of the conditional kurtosis, the NM(3) model gives conditional kurtosis estimates that are excessively jumpy, fluctuating between very large and very small values.

4.5.7 Unconditional moment comparison

Tables 13 and 14 present the unconditional 'realised' skewness and excess kurtosis in the data and compare this with the long-term values derived from the GARCH(1,1), the skewed *t*-GARCH, and the unrestricted NM(2) and NM(3) models (the derivations are presented in Appendix B, C and D). These are estimated over the whole sample period and, except for the NM(3) models which could not be reliably estimated, over the three sub-periods defined in section 4.4.2. The NM(2) model gave unconditional moment estimates that were closest to the realized moments in almost all cases, as shown in the tables with bold type.

[Table 13 about here]

4.5.8 VaR-based criterion

When a statistical volatility model is applied to Value-at-Risk (VaR) estimation the accuracy of quantile predictions is a very important property. However, there are many possible VaR-based model selection criteria and, despite much research, there is no consensus about which method is best. Here we use a criterion which counts the percentage of the returns less than the α% forecasted VaR, and then compares this with its theoretical value, α%. We shall do this for α% = 1%, 5% and 10%, for the entire period. We use the parameters estimated for the entire dataset and in-sample VaR forecasts are computed for the entire period.

First, variances are forecasted for each model for every non-overlapping 1-day period (using the parameters estimated for the entire period). In the case of the NM(*K*) models, we forecast the individual variances of the components of the mixture and combine these using the estimated weight parameters. The forecasted VaR values corresponding to each model are then compared with the realized returns, counting the number of times the VaR was exceeded and denoting the proportion of these 'exceptions' by $\beta\%_{\alpha\%}$.

To accumulate the information we chose the following methodology: first, in order to avoid underestimating errors in the 1% VaR we translate each exceptions proportion into an absolute percentage, relative to the significance level:

$$\gamma_{\alpha\%} = \left| \frac{\beta\%_{\alpha\%} - \alpha\%}{\alpha\%} \right|$$

Then we average the *γ*'s across the different significance levels to obtain two exceptions percentages:

$$\gamma = \left(\gamma_{1\%} + \gamma_{5\%} + \gamma_{10\%} \right) / 3$$

This γ gives our criterion for the VaR ranking of the models, reported in the last rows of Tables 4 – 9. According to this criterion the simple GARCH(1,1) and the *t*-GARCH models do not perform well at all. Clearly there is insufficient flexibility in the tails of the conditional densities in these models. In some cases the best models are the NM(3) specifications and in 5 cases out of 6, the general NM(2) model performs better than the skewed *t*-GARCH specification.

4.5.9 Conclusion of model selection criteria

The model selection results based on the eight criteria described above can be summarized as:

(1) Likelihood-based criteria prefer the skewed *t*-GARCH or NM(3) specifications.

(2) Likelihood ratio tests prefer unrestricted asymmetric parameterisation for the normal mixture models, except that the GBP rate favours symmetric normal mixture models.

(3) Moment specification tests also prefer the *t*-GARCH and NM(3) models.

(4) The criterion based on density fitting concludes that the unrestricted NM(2) model is clearly the best.

(5) The ACF-based criterion has mixed results.

(6) Conditional moment comparisons clearly prefer the unrestricted asymmetric and symmetric NM(2) models.

(7) Unconditional moment comparisons also favour the NM(2) specifications.

(8) VaR tests show that all normal mixture models forecast tails much better than *t*-GARCH or simple normal GARCH models.

We believe that the most important criteria are those based on histogram fit, moment tests and VaR rankings. Based on these we conclude that the unrestricted NM(2) model offers the best fit for our data.

5. Summary and Conclusions

This paper has examined the properties of the general class of 'normal mixture GARCH(1,1) models'. These are GARCH(1,1) models where the error term follows a normal mixture distribution with two or more GARCH(1,1) variance components. We have derived many theoretical results and have compared the implementation of these models with implementations of the normal and (symmetric and skewed) Student's *t*-GARCH models. Our contributions may be summarized as follows:

1. *The analytic framework for the estimation* of NM(*K*)-GARCH(1,1) models (or 'NM(*K*)' models for short) has been developed: we have derived expressions for the derivatives of the likelihood function and for the standard errors of the estimates; we have also set specific parameter conditions for the existence of finite and positive second and fourth moments;
2. *Explicit formulae for the higher moments* (both conditional and unconditional skewness and excess kurtosis) of the general class of normal mixture GARCH models, and the unconditional higher moments of the normal, standardized and skewed Student's *t*-GARCH have also been derived; furthermore, the autocorrelation function for the squared returns was developped;
3. *The bias, efficiency and stability* of parameter estimates of normal mixture GARCH models was investigated using Monte Carlo simulations and some important conclusions were derived from this exercise;
4. *Fifteen models* (i.e. normal, standardized and skewed Student's *t*-GARCH, six NM(2) and six NM(3) models) have been estimated, using daily data for major US dollar exchange rates covering a variety of sample periods;
5. *Eight different types of model selection criteria* were applied to determine the best fitting model(s) for these data. These criteria are based on likelihood, moment specification tests, likelihood ratio tests, histogram fitting, unconditional and conditional moment analysis, autocorrelation function (ACF) and Value-at-Risk (VaR) criteria. All analytic results that were necessary to apply these criteria to our models were derived, including the framework for VaR 'ranking' tests and the conditional and unconditional moment comparisons.

5.1 Analytic Framework for Estimation

The analytic framework greatly improves the efficiency of estimation algorithms. The specific parameters conditions we derive in fact allow for $\alpha + \beta > 1$ for some components. We have seen that there is a region for the parameters where the fourth moment does not exist and around this region the fourth moment is highly non-linear, providing intuition for the too high values of the unconditional excess kurtosis in some NM(3) implementations.

5.2 Bias, Efficiency and Stability

Simulations indicate potential convergence problems and a possible bias on parameter estimates for any variance components having low weights in the mixture. In particular, the ω and α parameters will have a large upward bias and the β parameter will have a large downward bias in variance components that have a low weight in the mixture. Although this bias decreases as the sample size increases, and the bias on parameter estimates is generally less than the estimated standard error, the standard errors also become large for parameters in low weight components. Nevertheless, the standard errors are close to their Cramér-Rao lower bound so they are 'efficient' in this sense. We conclude that the estimation of NM(3) models, where one of the components is likely to carry low weight in the mixture, can be fraught with difficulties because parameter estimates are likely to have a serious bias and to lack of robustness.

5.3 Model Estimations

Empirical results on fifteen models (normal, standardized symmetric and skewed Student's *t*-GARCH, six NM(2) and six NM(3) models) were based on daily data for US dollar exchange rates with the British pound, Euro and Japanese yen. All estimations were robust to the removal of extreme returns. We find that the NM(2) estimates are highly intuitive and repeating the exercise for three sub-periods also gave intuitive results, with model parameters adjusting to reflect the prevalent circumstances in each market. However, four substantial problems arose with the estimation of NM(3) models: *(i)* ill-conditioning of the likelihood surface induces convergence problems so that a very long data period was necessary to estimate the parameters; *(ii)* a serious potential bias was observed in the parameter estimates of the third component; *(iii)* the conditional kurtosis estimates fluctuated excessively between unrealistic extreme values; and *(iv)* the parameter estimates on the third variance component were apparently close to the non-existence boundary for the fourth moment and for this reason the unconditional kurtosis estimate may be biased upwards.

5.4 Model Selection

Model selection criteria were applied to determine the best fitting model(s) for these data. These are based on likelihood, moment specification tests, density fitting, unconditional and conditional moment analysis, autocorrelation function (ACF) of squared residuals and Value-at-Risk estimation. The NM(3) models offer a better fit than the NM(2) models according to some of our criteria – e.g. the in-sample likelihood. But this is to be expected because it has up to 11 parameters, depending on the number of restrictions applied. However, having so many parameters is not always desirable: as seen above, it can lead to estimation problems and it should also be borne in mind that *over* fitting the data can itself lead to specification problems. Certainly the NM(3) models did not perform well according to the density fitting, the unconditional moment fitting or the VaR criteria. Likelihood ratio tests and the analysis of conditional moment graphs generally rejected the use of restricted forms of normal mixture GARCH models, in favour of their unrestricted symmetric or asymmetric alternatives. The *t*-GARCH models performed well in moment specification tests, and the skewed *t*-GARCH model obviously improved on the symmetric *t*-GARCH according to almost all of the criteria. However, they were found to be inferior to NM(2) models for unconditional density and moment fitting and they also performed very poorly under the VaR criteria. Density fitting clearly points to the unrestricted NM(2) model as the best fit. Also, the analysis of the conditional and unconditional moments of the models, as well as the VaR based criterion all favour the unrestricted NM(2) model.

5.5 General Conclusions

In summary, whilst the standardized skewed Student's *t*-GARCH model is a strong candidate for modelling time-varying conditional volatility for returns that are skewed and leptokurtic, it has a number of limitations compared with normal mixture GARCH models. Perhaps the most significant of these is that it fails to model time-variation in the conditional skewness and excess kurtosis, unless an explicit augmentation of the model is defined. Moreover, when applied to daily exchange rate data both Student's *t*-GARCH models were out-performed by the NM(2) models, based on several important model selection criteria.

Although the class of NM(3) models scored highly on some model specification tests, we do not recommend these model be applied, at least to data on major exchange rates. A failure of the NM(3) models is likely to arise from over-fitting the data, and/or because the mixing parameter on the third component is so small that the likelihood surface becomes ill-conditioned, leading to biased parameter and conditional and unconditional kurtosis estimates. On the other hand, the NM(2) implementations for each of the data series, and over all sample periods considered, had intuitive and appealing interpretations. The estimated parameters of each conditional variance equation were easily distinguished, in that the higher weight component had much lower variance than the lower weight component. This interpretation supports the intuition provided by behavioural models where two different volatility 'states' can exist in the market. Clearly NM(2) models are able to capture two distinct components in exchange rate volatility: indeed they are closely related to Markov Switching GARCH models, but they share none of their estimation problems.

Our empirical results were conclusive and lead to the recommendation of the general unrestricted NM(2) model for exchange rate modelling in preference to Student's *t*-GARCH models, the restricted NM(2) models that have been previously studied, or normal mixture GARCH(1,1) models with more than two variance components.

In contrast to some of the restricted versions of normal mixture GARCH models that have been studied previously, and the Student's *t*-GARCH models, a key feature of unrestricted normal mixture GARCH models is that they capture time-variation in conditional skewness and kurtosis. It is well known that the implied volatility smile/skew surface, a major part of which is caused by skewed and leptokurtic returns in the underlying asset, also exhibits substantial variation over time. Thus smile/skew dynamics are consistent with the general version of the normal mixture GARCH models that are studied in this paper. For these reasons normal mixture GARCH models, and the NM(2) model in particular, should prove useful for smile consistent option pricing and hedging. An avenue of theoretical interest will be the derivation of the continuous time limit of normal mixture GARCH models and their relation to the class of normal mixture diffusion models with stochastic volatility. Another application of the NM(2) model will be to generate analytic term structure forecasts for excess kurtosis, as these play an important role in the dynamic delta hedging of options portfolios.

Among the possible extensions of general normal mixture GARCH(1,1) models, different asymmetric GARCH and multivariate parameterisations would be of interest. The extension of our theoretical results in this respect would prove interesting, as would empirical applications where equity returns are assumed to follow a normal mixture process with two asymmetric GARCH variance components, and empirical applications where term structures of commodity futures are modelled in a multivariate normal mixture GARCH framework.

References

Akgiray, V. and Booth, G.G. (1988): 'Mixed Diffusion-Jump Process Modeling of Exchange Rate Movements', *The Review of Economics and Statistics*, Vol. 70, Issue 4, 631-637

Alexander, C. (2001): 'Market Models: a Guide to Financial Data Analysis', John Wiley & Sons, Chicester

Alexander, C. (2004): 'Normal Mixture Diffusion with Uncertain Volatility: Modelling Short and Long Term Smile Effects' *Journal of Banking and Finance*, forthcoming.

Andersen, T.G., Bollerslev, T., Diebold, F.X. (2002): 'Parametric and Nonparametric Volatility Measurement', in Aït-Sahalia, Y. and Hansen, L. P.: 'Handbook of Financial Econometrics', Amsterdam: North Holland

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2000): 'Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian', *Multinational Finance Journal*, Vol. 4, No. 3&4, 159-179

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001): 'The Distribution of Realized Exchange Rate Volatility', *Journal of the American Statistical Association*, Vol. 96, 42-55

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003): 'Modelling and Forecasting Realized Volatility', *Econometrica*, Vol. 71, No. 2, 579-625

Bai, X., Russell, J.R. and Tiao, G.C. (2001): 'Beyond Merton's Utopia (I): Effects of Non-normality and Dependence on the Precision of Variance Using High-frequency Financial Data', *University of Chicago, GSB Working Paper* July 2001

Bai, X., Russell, J.R. and Tiao, G.C. (2003): 'Kurtosis of GARCH and Stochastic Volatility Models with Non-normal Innovations' *Journal of Econometrics*, Vol. 114, Issue 2, 349-360

Baillie, R.T. and Bollerslev, T. (1989): 'The Message in Daily Exchange Rates: A Conditional-Variance Tale', *Journal of Business and Economic Statistics*, Vol. 7, No. 3, 297-305

Baillie, R.T. and Bollerslev, T. (1991): 'Intra-Day and Inter-Market Volatility in Foreign Exchange Rates', *The Review of Economic Studies*, Vol. 58, Issue 3, 565-585

Bakshi, G., Kapadia, N. and Madan, D. (2003): 'Stock Return Characteristics, Skew Laws, and the Differential Pricing for Individual Equity Options', *The Review of Financial Studies*, Vol. 16, Issue 1, 101-143

Ball, C.A. and Torous, W.N. (1983): 'A Simplified Jump Process for Common Stock Returns', *Journal of Financial and Quantitative Analysis*, Vol. 18, No. 1, 53-65

Barndorff-Nielsen, O.E. (1977): 'Exponentially Decreasing Distributions for the Logarithmic of Particle Size', *Proceedings of the Royal Society of London A*, Vol. 353, 401-409

Barndorff-Nielsen, O.E. and Shephard, N. (2001): 'Non-Gaussian Ornstein-Uhlenbeck-based Models and Some of Their Uses in Financial Economics (with discussion)', *Journal of the Royal Statistical Society Series B*, Vol. 63, 167-241

Barndorff-Nielsen, O.E. and Shephard, N. (2002): 'Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models', *Journal of the Royal Statistical Society Series B*, Vol. 64, No. 2, 253-280

Bates, D.S. (1991): 'The Crash of '87: Was it expected? The Evidence from Options Markets', *Journal of Finance*, Vol. 46, 1009-1044

Bauwens, L., Bos, C.S. and van Dijk, H.K. (1999): 'Adaptive Polar Sampling with an Application to a Bayes measure of Value-at-Risk', *Tinbergen Institute Discussion Paper* TI 99-082/4

Bollerslev, T. (1986): 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics*, Vol. 31, 309-328

Bollerslev, T. (1987): 'A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return', *Review of Economics and Statistics*, Vol. 69, 542-547

Bollerslev, T., Chou, R.Y. and Kroner, K.F.(1992): 'ARCH Modeling in Finance', *Journal of Economics*, Vol. 52, 5-59

Boothe, P. and Glassman, D. (1987): 'The Statistical Distribution of Exchange Rates: Empirical Evidence and Economic Implications', *Journal of International Economics*, Vol. 22, 297-319

Brigo, D. and Mercurio F. (2000): 'A Mixed up Smile', *RISK*, Vol. 13, No. 9, 123-126

Brigo, D. and Mercurio, F. (2001): 'Interest Rate Models: Theory and Practice', Springer-Verlag Berlin Heidelberg

Brooks, C., Burke, S.P. and Persand, G. (2002): 'Autoregressive Conditional Kurtosis', *ISMA Centre Discussion Papers in Finance* Series 2002-5

Cai, J. (1994): 'A Markov Model of Switching-Regime ARCH', *Journal of Business & Economic Statistics*, Vol. 12, 309-316

Christoffersen, P.F. (1998): 'Evaluating Interval Forecasts', *International Economic Review*, Vol. 39, No. 4, 841-862

Cox, J.C. and Ross, S.A. (1976): 'The Valuation of Options for Alternative Stochastic Processes', *Journal of Financial Economics*, Vol. 3, 145-166

Derman, E. (1999): 'Regimes of Volatility' *Risk Magazine*, Vol. 12, No. 4, 55-59.

de Vries, C.G. (1994): 'Stylized Facts of Nominal Exchange Rate Returns', in F. van der Ploeg (ed.), Handbook of International Macroeconomics (Basil Blackwell), 348-389

Ding, Z. and Granger, C. W. J. (1996): 'Modeling Volatility Persistence of Speculative Returns: A New Approach', *Journal of Econometrics*, Vol. 73, 185-215

Ding, Z., Granger, C.W.J. and Engle, R.F. (1993): 'A Long Memory Property of Stock Market Returns and a New Model', *Journal of Empirical Finance*, Vol. 1, 83-106

Doornik, J.A. (2002): 'Object-Oriented Matrix Programming Using Ox', 3rd ed. London, *Timberlake Consultants Press and Oxford*: www.nuff.ox.ac.uk/Users/Doornik.

Doornik, J.A. and Ooms, M. (2000): 'Multimodality and the GARCH Likelihood', mimeo, Nuffield College

Duan, J.-C. (1995): 'The GARCH Option Pricing Model', *Mathematical Finance*, Vol. 5, No. 1, 13-32

Dupire, B. (1994): 'Pricing with a Smile', *RISK*, Vol. 7, No. 1, 18-20

Eberlein, E. and Keller, U. (1995): 'Hyperbolic distributions in Finance', *Bernoulli*, Vol. 1, 281-299

Engel, C. (1994): 'Can the Markov Switching Model Forecast Exchange Rates?', *Journal of International Economics*, Vol. 36, 151-165

Engel, C. and Hamilton, J.D. (1990): 'Long Swings in the Dollar: Are They in the Data and Do Markets Know It?', *American Economic Review*, Vol. 80, 689-713

Engle, R.F. (1982): 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica*, Vol. 50, No. 4, 987-1007

Engle, R.F. and Bollerslev, T. (1986): 'Modelling the Persistence of Conditional Variances', *Econometrics Review*, Vol. 5, 1-50

Engle, R.F. and Gau, Y.-F. (1997): 'Conditional Volatility of Exchange Rates Under a Target Zone', *University of California, San Diego, Department of Economics Discussion Paper* Series 06

Engle, R.F. and Gonzalez-Rivera, G. (1991): 'Semiparametric GARCH Models', *Journal of Business & Economic Statistics*, Vol. 9, No. 4, 345-359

Engle, R.F., Ito, T. and Lin, W.-L. (1990): 'Meteor Showers or Heat Waves? Heteroskedastic Intra-daily Volatility in the Foreign Exchange Market', *Econometrica*, Vol. 58, No. 3, 525-542

Engle, R. F. and Lee, G.. J. (1999): 'A permanent and Transitory Component Model of Stock Return Volatility', in ed. R.F. Engle and H. White, Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger, (Oxford University Press), 475-497

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

Epanechnikov, V. (1969): 'Nonparametric estimation of a multidimensional probability density', *Teoriya Veroyatnostej i Ee Primeneniya*, Vol. 14, 156-162

Epps, T.W., Epps, M.L. (1976): 'The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-distributions Hypothesis', *Econometrica*, Vol. 44, No. 2, 305-321

Fama, E. (1965): 'The Behavior of Stock-Market Prices', *Journal of Business*, Vol. 38, 34-105

Fernandez, C. and Steel, M. (1998): 'On Bayesian Modelling of Fat Tails and Skewness', *Journal of the American Statistical Association*, Vol. 93, 359-371

Gray, S.F. (1996): 'Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process', *Journal of Financial Economics*, Vol. 42, 27-62

Greene, W.H. (2000): 'Econometric Analysis', Prentice Hall International, Fourth Edition

Haas, M., Mittnik, S. and Paolella, M.S. (2002): 'Mixed Normal Conditional Heteroskedasticity', *Center for Financial Studies Working Paper* No. 2002/10

Hall, P. (1982): 'On Some Simple Estimates of an Exponent of Regular Variation', *Journal of the Royal Statistical Society, Series B (Methodological)*', Vol. 44, No. 1, 37-42

Hamilton, J.D. and Susmel, R. (1994): 'Autoregressive Conditional Heteroscedasticity and Changes in Regime', *Journal of Econometrics*, Vol. 64, 307-333

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

Hansen, P.R. and Lunde, A. (2001): 'A Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?' *Centre for Analytical Finance Working Paper Series* No. 84

Harvey, C.R. and Siddique, A. (1999): 'Autoregressive Conditional Skewness', *Journal of Financial and Quantitative Studies*, Vol. 34, No. 4, 465-487

Heston, S.L. (1993): 'A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options', *Review of Financial Studies*, Vol. 6, 327-343

Heston, S.L., and S. Nandi (2000): 'A Closed-Form GARCH Option Valuation Model', *Review of Financial Studies*, Vol. 13, 585–625.

Hill, B.M. (1975): 'A Simple General Approach to Inference About the Tail of a Distribution', *The Annals of Statistics*, Vol. 3, No. 5, 1163-1174

Hsieh, D.A. (1988): 'The Statistical Properties of Daily Foreign Exchange Rates: 1974 – 1983', *Journal of International Economics*, Vol. 24, 129-145

Hsieh, D.A.(1989a): 'Modeling Heteroscedasticity in Daily Foreign-Exchange Rates', *Journal of Business and Economic Statistics*, Vol. 7, No. 3, 307-317

Hsieh, D.A. (1989b): 'Testing for Nonlinear Dependence in Daily Foreign Exchange Rates', *The Journal of Business*, Vol. 62, Issue 3, 339-368

Huisman, R., Koedijk, K.G., Kool, C.J.M. and Palm, F. (2001): 'Tail-Index Estimator in Small Sample', *Journal of Business and Economic Statistics*, Vol. 19, 208-216

Huisman, R., Koedijk, K.G., Kool, C.J.M. and Palm, F. (2002): 'The Tail-Fatness of FX Returns Reconsidered', *De Economist*, Vol. 150, 299-312

Hull, J. and White, A. (1987): 'The Pricing of Options on Assets with Stochastic Volatilities', *The Journal of Finance*, Vol. 42, 281-300

Hull, J. and White, A. (1998): 'Value at Risk when Daily Changes in Market Variables are not Normally Distributed', *Journal of Derivatives*, Vol. 5, No. 3, 9-19

Johnston, K. and Scott, E. (1999): 'The Statistical Distribution of Daily Exchange Rate Price Changes: Dependent vs. Independent Models', *Journal of Financial and Strategic Decisions*, Vol.12, No. 2, 39-49

Johnston, K. and Scott, E. (2000): 'GARCH Models and the Stochastic Process Underlying Exchange Rate Price Changes', *Journal of Financial and Strategic Decisions*, Vol.13, No. 2, 13-24

Jorion, P. (1988): 'On Jump Processes in the Foreign Exchange and Stock Markets', *The Review of Financial Studies*, Vol. 1, Issue 4, 427-445

Khamis, H. J. (2000): 'The Two-Stage δ-Corrected Kolmogorov-Smirnov Test', *Journal of Applied Statistics*, Vol. 27, No. 4, 439-450

Klaassen, F. (1998): 'Improving GARCH Volatility Forecasts', *SSRN Discussion Paper* No. 1998-52

Kolmogoroff, A. (1933): 'Sulla determinazione empirica di una legge di distribuzione', *Giornale dell'Istituto Italiano degh Attuari*, Vol. 4, 83-91

Kon, S.J. (1984): 'Models of Stock Returns – A Comparison', *The Journal of Finance*, Vol. 39, Issue 1, 147-165

Lambert, P. and Laurent, S. (2001a): 'Modelling Skewness Dynamics in Series of Financial Data Using Skewed Location-Scale Distributions', *Institut de Statistique, Louvain-la-Neuve Discussion Paper* 0119,

Lambert, P. and Laurent, S. (2001b): 'Modelling Financial Time Series Using GARCH-Type Models with a Skewed Student Distribution for the Innovations', *Institut de Statistique, Louvain-la-Neuve Discussion Paper* 0125,

Laurent, S. and Peters, J.-P. (2002): 'G@RCH 2.2: an Ox Package for Estimating and Forecasting Various ARCH Models', *Journal of Economic Surveys*, Vol. 16, No. 3, 447-485

Mandelbrot, B. (1963): 'The Variation of Certain Speculative Prices', *Journal of Business*, Vol. 36, 394-419

Massey, F. J. Jr. (1951): 'The Kolmogorov-Smirnov Test for Goodness of Fit', *Journal of the American Statistical Association*, Vol. 46, No. 253, 68-78

McFarland, J.W., Pettit, R.R. and Sung, S.K. (1982): 'The Distribution of Foreign Exchange Price Changes: Trading Day Effects and Risk Measurement', *The Journal of Finance*, Vol. 37, No. 3, 693-715

Mckenzie, M.D. (1997): Forecasting Australian Exchange Rate Volatility: A Comparative Study of Alternate Modelling Techniques And the Impact of Power Transformations', *RMIT University, School of Economics and Finance Working Paper* Series no. 10

Melick, W.R. and Thomas C.P. (1997): 'Recovering an Asset's Implied PDF from Option Prices: An Application to Crude Oil during the Gulf Crisis', *The Journal of Financial and Quantitative Analysis*, Vol. 32, No. 1, 91-115

Mercurio, F. (2002): 'A Multi-stage Uncertian Volatility Model' (available from www.fabiomercurio.it)

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

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

Newey, W.K. (1985): 'Generalized Method of Moments Specification Testing', *Journal of Econometrics*, Vol. 29, 229-256

Roberts, M.C. (2001): 'Commodities, Options, and Volatility: Modelling Agricultural Futures Prices', *Working Paper, The Ohio State University*

Smirnov, N. (1939): 'Sur les ecarts de la courbe de distribution empirique' (Russian, French summary), *Matematicheskii Sbornik*, Vol, 48, No. 6, 3-26

Stein, E.M. and Stein, J.C. (1991): 'Stock Price Distributions with Stochastic Volatility: an Analytic Approach', *Review of Financial Studies*, Vol. 4, 727-752

Taylor, S. (1986): 'Modeling Financial Time Series', *Wiley, New York, NY*

Tucker, A.L. (1992): 'A Reexamination of Finite- and Infinite-Variance Distributions as Models of Daily Stock Returns', *Journal of Business & Economic Statistics*, Vol. 10, No. 1, 73-81

Tucker, A.L. and Pond, L. (1988): 'The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes', *The Review of Economics and Statistics*, Vol. 70, Issue 4, 638-647

Vlaar, P.J.G. and Palm, F.C. (1993): 'The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps', *Journal of Business & Economic Statistics*, Vol. 11, No. 3, 351-60

Wang, K.-L., Fawson, C., Barrett, C. B. and McDonald, J. B. (2001): 'A Flexible Parametric GARCH Model with an Application to Exchange Rates', *Journal of Applied Econometrics*, Vol. 16, 521-536

West, K.D. and Cho, D. (1994): 'The Predictive Ability of Several Models of Exchange Rate Volatility', *NBER Technical Working Paper* No. 152

Westerfield, J.M. (1977): 'An Examination of Foreign Exchange Risk Under Fixed and Floating Rate Regimes', *Journal of International Economics*, Vol. 7, 181-200

Zangari, P. (1996): 'An Improved Methodology for Measuring VaR', *RiskMetrics Monitor*, JPMorgan, Second quarter 1996

Zhou, S. (1996): 'High-Frequency Data and Volatility in Foreign-Exchange Rates', *Journal of Business & Economic Statistics*, Vol. 14, No. 1, 45-52

Appendix A. The Relationship between NM(*K*)-GARCH(1,1) and GARCH(*K*,*K*) models

For the NM(*K*)-GARCH(1,1) model the variances are expressed as:

$$\sigma_t^2 = \sum_{i=1}^{K} p_i \sigma_{it}^2 + \sum_{i=1}^{K} p_i \mu_i^2$$

$$\sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2 \qquad \text{i} = 1,\ldots, K$$

In the following we use the following notations for simplicity:

$$A_n = L + \sum_{j=1}^{n} \varphi_j \varepsilon_{t-j}^2 + c_j + d_j L \quad L = \sum_{i=1}^{K} p_i \mu_i^2 \qquad \varphi_j = (-1)^{j+1} \sum_{\substack{i_1=1 \\ \ldots \\ i_j=1 \\ i_1 \neq \ldots \neq i_j}}^{K} p_{i_1} \alpha_{i_1} \prod_{m=2}^{j} \beta_{i_m}$$

$$c_j = (-1)^{j+1} \sum_{\substack{i_1=1 \\ \ldots \\ i_j=1 \\ i_1 \neq \ldots \neq i_j}}^{K} p_{i_1} \omega_{i_1} \prod_{m=2}^{j} \beta_{i_m} \qquad d_j = (-1)^{j} \sum_{\substack{i_1=1 \\ \ldots \\ i_j=1 \\ i_1 \neq \ldots \neq i_j}}^{K} \prod_{m=1}^{j} \beta_{i_m}$$

$$B_n = \sum_{j=1}^{n} \theta_j \sigma_{t-j}^2 \qquad \theta_j = (-1)^{j+1} \sum_{\substack{i_1=1 \\ \ldots \\ i_j=1 \\ i_1 \neq \ldots \neq i_j}}^{K} \prod_{m=1}^{j} \beta_{i_m}$$

We have that:

$$\sigma_t^2 = L + \sum_{i=1}^{K} p_i \sigma_{it}^2 = A_1 + \sum_{i_1=1}^{K} p_{i_1} \beta_{i_1} \sigma_{i_1 t-1}^2 = A_1 + \sum_{i_1=1}^{K} p_{i_1} \sum_{i_2=1}^{K} \beta_{i_2} \sigma_{i_1 t-1}^2 + (-1) \sum_{i_1=1}^{K} p_{i_1} \sum_{\substack{i_2=1 \\ i_1 \neq i_2}}^{K} \beta_{i_2} \sigma_{i_1 t-1}^2 =$$

$$= A_1 + B_1 - \sum_{i_1=1}^{K} \beta_{i_1} L + (-1) \sum_{i_1=1}^{K} p_{i_1} \sum_{\substack{i_2=1 \\ i_1 \neq i_2}}^{K} \beta_{i_2} \sigma_{i_1 t-1}^2 = A_2 + B_1 + (-1) \sum_{i_1=1}^{K} p_{i_1} \sum_{\substack{i_2=1 \\ i_1 \neq i_2}}^{K} \beta_{i_1} \beta_{i_2} \sigma_{i_1 t-2}^2 =$$

$$= A_2 + B_1 + (-1) \sum_{i_1=1}^{K} p_{i_1} \sum_{\substack{i_2=1 \\ i_3=1 \\ i_2 \neq i_3}}^{K} \beta_{i_2} \beta_{i_3} \sigma_{i_1 t-2}^2 + (-1)^2 \sum_{i_1=1}^{K} p_{i_1} \sum_{\substack{i_2=1 \\ i_3=1 \\ i_1 \neq i_2 \neq i_3}}^{K} \beta_{i_2} \beta_{i_3} \sigma_{i_1 t-2}^2 =$$

$$= A_2 + B_2 + \sum_{\substack{i_1=1 \\ i_2=1 \\ i_1 \neq i_2}}^{K} \beta_{i_1} \beta_{i_2} L + (-1)^2 \sum_{i_1=1}^{K} p_{i_1} \sum_{\substack{i_2=1 \\ i_3=1 \\ i_1 \neq i_2 \neq i_3}}^{K} \beta_{i_2} \beta_{i_3} \sigma_{i_1 t-2}^2$$

$$= A_3 + B_2 + (-1)^2 \sum_{i_1=1}^{K} p_{i_1} \sum_{\substack{i_2=1 \\ i_3=1 \\ i_1 \neq i_2 \neq i_3}}^{K} \beta_{i_1} \beta_{i_2} \beta_{i_3} \sigma_{i_1 t-3}^2 =$$

$$\ldots$$

$$= A_K + B_{K-1} + (-1)^{n-1} \sum_{i_1=1}^{K} p_{i_1} \sum_{\substack{i_2=1 \\ \ldots \\ i_n=1 \\ i_1 \neq \ldots \neq i_n}}^{K} \beta_{i_1} \beta_{i_2} \ldots \beta_{i_n} \sigma_{i_1 t-n}^2 = A_K + B_K$$

So the NM(*K*)-GARCH(1,1) model can be expressed as a GARCH(*K*,*K*) process for the variance.

For instance, in the NM(2)-GARCH(1,1) model we have:

$$\begin{aligned}
\sigma_t^2 &= \left(p\mu_1^2 + (1-p)\mu_2^2\right) + p\sigma_{1t}^2 + (1-p)\sigma_{2t}^2 = \\
&= \left(p\mu_1^2 + (1-p)\mu_2^2\right) + \left(p\omega_1 + (1-p)\omega_2\right) + \left(p\alpha_1 + (1-p)\alpha_2\right)\varepsilon_{t-1}^2 + p\beta_1\sigma_{1t-1}^2 + (1-p)\beta_2\sigma_{2t-1}^2 = \\
&= \left(p\mu_1^2 + (1-p)\mu_2^2\right) + \left(p\omega_1 + (1-p)\omega_2\right) + \left(p\alpha_1 + (1-p)\alpha_2\right)\varepsilon_{t-1}^2 + (\beta_1 + \beta_2)\left(p\sigma_{1t-1}^2 + (1-p)\sigma_{2t-1}^2\right) \\
&\quad - \left(p\beta_2\sigma_{1t-1}^2 + (1-p)\beta_1\sigma_{2t-1}^2\right) = \\
&= \left(p\mu_1^2 + (1-p)\mu_2^2\right)\left(1 - (\beta_1 + \beta_2) + \beta_1\beta_2\right) + \left(p\omega_1 + (1-p)\omega_2\right) - \left(p\omega_1\beta_2 + (1-p)\omega_2\beta_1\right) + \left(p\alpha_1 + (1-p)\alpha_2\right)\varepsilon_{t-1}^2 \\
&\quad - \left(p\alpha_1\beta_2 + (1-p)\alpha_2\beta_1\right)\varepsilon_{t-2}^2 + (\beta_1 + \beta_2)\sigma_{t-1}^2 - \beta_1\beta_2\sigma_{t-2}^2
\end{aligned}$$

Appendix B. Derivation of the Conditional and Unconditional Moments of the Error Term and the Autocorrelation of the Squared Residuals of the NM(K)-GARCH(1,1) Model

The model is specified as:

$$y_t = \varepsilon_t \qquad \varepsilon_t \big| I_{t-1} \sim NM(p_1,..,p_K;\mu_1,...,\mu_K;\sigma^2_{1t},...,\sigma^2_{Kt}) \qquad \sum_{i=1}^{K} p_i = 1$$

$$\sigma^2_{it} = \omega_i + \alpha_i \varepsilon^2_{t-1} + \beta_i \sigma^2_{it-1} \qquad \text{i} = 1,\ldots, K \tag{1}$$

where the conditional variance is:

$$\sigma^2_t = \sum_{i=1}^{K} p_i \sigma^2_{it} + \sum_{i=1}^{K} p_i \mu^2_i \tag{2}$$

We use the following notations: $x = E(\varepsilon^2_t) = E(\sigma^2_t)$ and $y_i = E(\sigma^2_{it})$ i = 1,…, K.

Taking expectations of (1) and (2) gives:

$$x = \sum_{i=1}^{K} p_i y_i + \sum_{i=1}^{K} p_i \mu^2_i \tag{3}$$

$$y_i = \omega_i + \alpha_i x + \beta_i y_i \qquad \text{i} = 1,\ldots, K \tag{4}$$

The second equations are equivalent to:

$$y_i = \frac{\omega_i + \alpha_i x}{1-\beta_i} \qquad \text{i} = 1,\ldots, K \tag{5}$$

using (5) in (3) we obtain:

$$x = E(\varepsilon^2_t) = E(\sigma^2_t) = \frac{\sum_{i=1}^{K} p_i \mu^2_i + \sum_{i=1}^{K} \frac{p_i \omega_i}{(1-\beta_i)}}{\left(1 - \sum_{i=1}^{K} \frac{p_i \alpha_i}{(1-\beta_i)}\right)} \tag{6}$$

The conditional skewness and excess kurtosis are given by:

$$\tau_t = \frac{E_{t-1}\left(\varepsilon^3_t\right)}{\left(\sigma^2_t\right)^{3/2}} = \frac{3\sum_{i=1}^{K} p_i \mu_i \sigma^2_{it} + \sum_{i=1}^{K} p_i \mu^3_i}{\left(\sigma^2_t\right)^{3/2}}$$

$$k_t = \frac{E_{t-1}\left(\varepsilon^4_t\right)}{\left(\sigma^2_t\right)^2} - 3 = \frac{3\sum_{i=1}^{K} p_i \left(\sigma^2_{it}\right)^2 + 6\sum_{i=1}^{K} p_i \mu^2_i \sigma^2_{it} + \sum_{i=1}^{K} p_i \mu^4_i}{\left(\sigma^2_t\right)^2} - 3$$

The unconditional third moment is:

$$h = E(\varepsilon^3_t) = \sum_{i=1}^{K} p_i E_i(\varepsilon^3_t) = \sum_{i=1}^{K} p_i \left(3 y_i \mu_i + \mu^3_i\right)$$

and the unconditional skewness can be expressed as

$$s = \frac{h}{x^{3/2}}$$

Now let:

$$z = E(\varepsilon^4_t) = \sum_{i=1}^{K} p_i \left(3E(\sigma^4_{it}) + 6\mu^2_i y^2_i + \mu^4_i\right) \tag{7}$$

Using the notations $v_i = E(\sigma_{it}^4)$, $\mathbf{v} = (v_1, \ldots, v_K)'$, $\mathbf{p} = (p_1, \ldots, p_K)'$ and $s = \sum_{i=1}^{K} p_i \left(6\mu_i^2 y_i^2 + \mu_i^4\right)$

we obtain:

$$z = 3\mathbf{p}'\mathbf{v} + s \tag{8}$$

Computing v_i from the individual conditional variance equation (1) we obtain:

$$v_i(1-\beta_i^2) = \omega_i^2 + 2\omega_i\alpha_i x + 2\omega_i\beta_i y_i + \alpha_i^2 z + 2\alpha_i\beta_i E(\varepsilon_t^2\sigma_{it}^2) \tag{9}$$

We denote the first (known) part of the RHS by w_i:

$$w_i = \omega_i^2 + 2\omega_i\alpha_i x + 2\omega_i\beta_i y_i$$

Also, let m_i denote the last term: $m_i = E(\varepsilon_t^2\sigma_{it}^2)$ so (9) becomes:

$$v_i(1-\beta_i^2) = w_i + \alpha_i^2 z + 2\alpha_i\beta_i m_i \tag{10}$$

Note that $m_i = p_i v_i + \sum_{\substack{k=1\\k\neq i}}^{K} p_k E(\sigma_{it}^2\sigma_{kt}^2) + y_i \sum_{k=1}^{K} p_k\mu_k^2 = p_i v_i + \sum_{\substack{k=1\\k\neq i}}^{K} p_k l_{ik} + y_i q$ where $q = \sum_{k=1}^{K} p_k\mu_k^2$ and

$l_{ik} = E(\sigma_{it}^2\sigma_{kt}^2)$.

Now by (1):

$$l_{ik}\left(1-\beta_i\beta_k\right) = \omega_i\omega_k + x\left(\omega_i\alpha_k + \omega_k\alpha_i\right) + \beta_i y_i\omega_k + \beta_k y_k\omega_i + \alpha_i\alpha_k z + \alpha_i\beta_k m_k + \alpha_k\beta_i m_i$$

Using the notation

$r_{ik} = \omega_i\omega_k + x\left(\omega_i\alpha_k + \omega_k\alpha_i\right) + \beta_i y_i\omega_k + \beta_k y_k\omega_i$ we have:

$$l_{ik} = \frac{r_{ik} + \alpha_i\alpha_k z + \alpha_i\beta_k m_k + \alpha_k\beta_i m_i}{1-\beta_i\beta_k}$$

so (10) becomes:

$$m_i\left(1 - \sum_{\substack{k=1\\k\neq i}}^{K}\frac{p_k\alpha_k\beta_i}{1-\beta_i\beta_k}\right) = p_i v_i + \left(\sum_{\substack{k=1\\k\neq i}}^{K}\frac{p_k r_{ik}}{1-\beta_i\beta_k}\right) + \left(\sum_{\substack{k=1\\k\neq i}}^{K}\frac{p_k\alpha_i\alpha_k}{1-\beta_i\beta_k}\right)z + \left(\sum_{\substack{k=1\\k\neq i}}^{K}\frac{p_k\alpha_i\beta_k}{1-\beta_i\beta_k}m_k\right) + y_i q \tag{11}$$

Let $\mathbf{m} = (m_1, \ldots, m_K)$ and

$$\mathbf{A} = \begin{bmatrix} 1 - \sum_{\substack{k=1\\k\neq 1}}^{K}\frac{p_k\beta_1\alpha_k}{1-\beta_1\beta_k} & -\frac{p_2\alpha_1\beta_2}{1-\beta_1\beta_2} & \cdots & -\frac{p_K\alpha_1\beta_K}{1-\beta_1\beta_K} \\ -\frac{p_1\alpha_2\beta_1}{1-\beta_2\beta_1} & 1 - \sum_{\substack{k=1\\k\neq 2}}^{K}\frac{p_k\beta_2\alpha_k}{1-\beta_2\beta_k} & \cdots & -\frac{p_K\alpha_2\beta_K}{1-\beta_2\beta_K} \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{p_1\alpha_K\beta_1}{1-\beta_K\beta_1} & -\frac{p_2\alpha_K\beta_2}{1-\beta_K\beta_2} & \cdots & 1 - \sum_{\substack{k=1\\k\neq K}}^{K}\frac{p_k\beta_K\alpha_k}{1-\beta_K\beta_k} \end{bmatrix}$$

So (11) may be written in matrix form as:

$$\mathbf{m}=\mathbf{A}^{-1}\begin{pmatrix} p_1v_1+\left(\sum_{\substack{k=1\\k\neq 1}}^{K}\frac{p_k r_{1k}}{1-\beta_1\beta_k}\right)+\left(\sum_{\substack{k=1\\k\neq 1}}^{K}\frac{p_k\alpha_1\alpha_k}{1-\beta_1\beta_k}\right)z+y_1q \\ \vdots \\ p_Kv_K+\left(\sum_{\substack{k=1\\k\neq K}}^{K}\frac{p_k r_{Kk}}{1-\beta_K\beta_k}\right)+\left(\sum_{\substack{k=1\\k\neq K}}^{K}\frac{p_k\alpha_K\alpha_k}{1-\beta_K\beta_k}\right)z+y_Kq \end{pmatrix}$$

Denoting the elements of $\mathbf{A}^{-1}$ by (a_{ij}), the individual elements of the vector **m** can be expressed as:

$$m_i=\sum_{j=1}^{K}a_{ij}\left[p_jv_j+\left(\sum_{\substack{k=1\\k\neq j}}^{K}\frac{p_k r_{jk}}{1-\beta_j\beta_k}\right)+\left(\sum_{\substack{k=1\\k\neq j}}^{K}\frac{p_k\alpha_j\alpha_k}{1-\beta_j\beta_k}\right)z+y_jq\right]$$

Let c_i denote the known part of the above expression: $c_i=\sum_{j=1}^{K}a_{ij}\left[\left(\sum_{\substack{k=1\\k\neq j}}^{K}\frac{p_k r_{jk}}{1-\beta_j\beta_k}\right)+y_jq\right]$

Also, using the notations $d_i=\sum_{j=1}^{K}a_{ij}\left(\sum_{\substack{k=1\\k\neq j}}^{K}\frac{p_k\alpha_j\alpha_k}{1-\beta_j\beta_k}\right)$ and $e_{ij}=a_{ij}p_j$ we obtain the following result:

$$m_i=c_i+d_iz+\sum_{j=1}^{K}e_{ij}v_j$$

Substituting this expression into (10) yields:

$$v_i(1-\beta_i^2)=w_i+\alpha_i^2z+2\alpha_i\beta_i\left(c_i+d_iz+\sum_{j=1}^{K}e_{ij}v_j\right) \tag{12}$$

Let **B** denote the matrix:

$$\mathbf{B}=\begin{bmatrix} 1-\beta_1^2-2\alpha_1\beta_1e_{11} & -2\alpha_1\beta_1e_{12} & \dots & -2\alpha_1\beta_1e_{1K} \\ -2\alpha_2\beta_2e_{21} & 1-\beta_2^2-2\alpha_2\beta_2e_{22} & \dots & -2\alpha_K\beta_Ke_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ -2\alpha_K\beta_Ke_{K1} & -2\alpha_K\beta_Ke_{K2} & \dots & 1-\beta_K^2-2\alpha_K\beta_Ke_{KK} \end{bmatrix}$$

and introduce the vectors: $\mathbf{f}=\begin{pmatrix} w_1+2\alpha_1\beta_1c_1 \\ \vdots \\ w_K+2\alpha_K\beta_Kc_K \end{pmatrix}$ and $\mathbf{g}=\begin{pmatrix} \alpha_1^2+2\alpha_1\beta_1d_1 \\ \vdots \\ \alpha_K^2+2\alpha_K\beta_Kd_K \end{pmatrix}$ so that (12) may be written as:

$$\boldsymbol{v}=\boldsymbol{B}^{-1}(\boldsymbol{f}+\boldsymbol{g}z)$$

Substituting this into (8) yields $z=3\mathbf{p}'\mathbf{B}^{-1}(\mathbf{f}+\mathbf{g}z)+s$ which leads to the following result:

$$z=E(\varepsilon_t^4)=\frac{3\mathbf{p}'\mathbf{B}^{-1}\mathbf{f}+s}{1-3\mathbf{p}'\mathbf{B}^{-1}\mathbf{g}} \tag{13}$$

hence the excess kurtosis is:

$$K=\frac{E(\varepsilon_t^4)}{E(\varepsilon_t^2)^2}-3=\frac{z}{x^2}-3$$

The autocorrelations of the squared errors can be expressed as:[28]

$$\varrho_k = Corr\left(\varepsilon_t^2, \varepsilon_{t-k}^2\right) = \frac{Cov\left(\varepsilon_t^2, \varepsilon_{t-k}^2\right)}{Var\left(\varepsilon_t^2\right)} = \frac{E[\varepsilon_t^2 \varepsilon_{t-k}^2] - x^2}{E[\varepsilon_t^4] - x^2} = \frac{c_k - x^2}{z - x^2},$$

$$c_k = E\left[\varepsilon_t^2 \varepsilon_{t-k}^2\right] = x \sum_{i=1}^{K} p_i \mu_i^2 + \sum_{i=1}^{K} p_i E\left[\sigma_{it}^2 \varepsilon_{t-k}^2\right] = x \sum_{i=1}^{K} p_i \mu_i^2 + \sum_{i=1}^{K} p_i b_{ik}$$

$$b_{ik} = \omega_i x + \alpha_i c_{k-1} + \beta_i b_{ik-1}$$

The starting values are: $c_0 = z$ and $b_{i0} = c_i + d_i z + \mathbf{e}_i{}' \mathbf{B}^{-1} (\mathbf{f} + \mathbf{g} z)$.

In the case of a mixture of two zero-mean normal densities, equations (6), (5) and (13) become:

$$E(\varepsilon_t^2) = E(\sigma_t^2) = \frac{\dfrac{p\omega_1}{(1-\beta_1)} + \dfrac{(1-p)\omega_2}{(1-\beta_2)}}{\left(1 - \dfrac{p\alpha_1}{(1-\beta_1)} - \dfrac{(1-p)\alpha_2}{(1-\beta_2)}\right)}$$

$$E(\sigma_{it}^2) = \frac{\dfrac{\omega_i}{(1-\beta_i)} + p_{3-i} \dfrac{(\alpha_i \omega_{3-i} - \alpha_{3-i}\omega_i)}{(1-\beta_1)(1-\beta_2)}}{\left(1 - \dfrac{p_1\alpha_1}{(1-\beta_1)} - \dfrac{p_2\alpha_2}{(1-\beta_2)}\right)}$$

$$E(\varepsilon_t^4) = 3 \frac{pA_1B_1 + (1-p)A_2B_2 + 2p(1-p)A_3B_3}{T_1 + T_2 + T_3 + T_4 + T_5}$$

where: $A_1 = (1-\beta_2^2)\left[1 - \beta_1\beta_2 - p\alpha_1\beta_2 - (1-p)\alpha_2\beta_1\right] - 2\alpha_2\beta_2(1-p)\left[1 - \beta_1\beta_2 - \alpha_2\beta_1\right]$

$$B_1 = \omega_1^2 + 2\omega_1\alpha_1 x + 2\omega_1\beta_1 y_1$$

$$A_2 = (1-\beta_1^2)\left[1 - \beta_1\beta_2 - p\alpha_1\beta_2 - (1-p)\alpha_2\beta_1\right] - 2\alpha_1\beta_1 p\left[1 - \beta_1\beta_2 - \alpha_1\beta_2\right]$$

$$B_2 = \omega_2^2 + 2\omega_2\alpha_2 x + 2\omega_2\beta_2 y_2$$

$$A_3 = \alpha_1\beta_1(1-\beta_2^2) + \alpha_2\beta_2(1-\beta_1^2) - 2\alpha_1\alpha_2\beta_1\beta_2$$

$$B_3 = \omega_1\omega_2 + \left(\omega_1\alpha_2 + \omega_2\alpha_1\right)x + \omega_1\beta_2 y_2 + \omega_2\beta_1 y_1$$

$$T_1 = \left[1 - \beta_1\beta_2 - p\alpha_1\beta_2 - (1-p)\alpha_2\beta_1\right]\left\{(1-\beta_1^2)(1-\beta_2^2) - 3(1-\beta_2^2)\alpha_1^2 p - 3(1-\beta_1^2)\alpha_2^2(1-p)\right\}$$

$$T_2 = -2(1-\beta_2^2)p\alpha_1\beta_1\left(1 - \beta_1\beta_2 - p\alpha_1\beta_2\right)$$

$$T_3 = -2(1-\beta_1^2)(1-p)\alpha_2\beta_2\left(1 - \beta_1\beta_2 - (1-p)\alpha_2\beta_1\right)$$

$$T_4 = 4\alpha_1\alpha_2\beta_1\beta_2 p(1-p)\left(1 - \beta_1\beta_2\right)$$

$$T_5 = -6p(1-p)\alpha_1\alpha_2\left(\alpha_1 - \alpha_2\right)\left(\beta_1 - \beta_2\right)$$

[28] Since the variance of the NM(*K*)-GARCH(1,1) model can be expressed as a GARCH(*K*,*K*) variance, according to Bollerslev (1986) the autocorrelations can also be written as an AR(*K*) process.

Appendix C. Estimation with Analytical Derivatives of the NM(K)-GARCH(1,1) Model

The specification of the model comprises $K + 1$ equations: the first one for the conditional mean and the next K for the variance behaviour. The conditional mean equation of the model is $y_t = \varepsilon_t$. For simplicity it contains no explanatory variables (these can be estimated separately). There are K conditional variance equations:

$$\sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2 \qquad i = 1,\ldots,K$$

The error term ε_t is assumed to have a conditional normal mixture density with zero mean:

$$\varepsilon_t \middle| I_{t-1} \sim NM(p_1,\ldots,p_K;\mu_1,\ldots,\mu_K;\sigma_{1t}^2,\ldots,\sigma_{Kt}^2), \quad \sum_{i=1}^{K} p_i = 1, \quad \sum_{i=1}^{K} p_i \mu_i = 1$$

The density of the error term is represented as a weighted average of K normal density functions φ_i:

$$\eta(\varepsilon_t) = \sum_{i=1}^{K} p_i \varphi_i(\varepsilon_t)$$

The source of asymmetry is that a mixture of normals with different means leads to a non-zero skewness. Since the sum of the probabilities is one, it is only necessary to use (K-1) parameters for the mixing law. Similarly, because the error term is centred on zero, the mean value of the last component of the mixture can be expressed with the help of the other parameters in the following way:

$$\mu_K = -\frac{1}{p_K} \sum_{i=1}^{K-1} p_i \mu_i$$

This way we use the following notations for the parameters:

$$\mathbf{p} = (p_1,\ldots,p_{K-1})', \qquad \boldsymbol{\mu} = (\mu_1,\ldots,\mu_{K-1})', \qquad \boldsymbol{\gamma_i} = (\omega_i,\alpha_i,\beta_i)' \quad i = 1,\ldots,K$$

$$\boldsymbol{\theta} = (p_1,\ldots,p_{K-1},\mu_1,\ldots,\mu_{K-1},\omega_1,\alpha_1,\beta_1,\omega_2,\alpha_2,\beta_2,\ldots,\omega_K,\alpha_K,\beta_K)' = (\mathbf{p}',\boldsymbol{\mu}',\boldsymbol{\gamma_1}',\boldsymbol{\gamma_2}'\ldots,\boldsymbol{\gamma_K}')'$$

Maximizing the likelihood, or equivalently, maximizing $\sum_{t=1}^{T} ln[\eta(\varepsilon_t)] + \frac{T}{2} ln(2\pi)$ gives the optimal parameter values, given the data $(y_1,\ldots,y_T)$.[29]

To ease the analysis, in the following let g_i denote $p_i \frac{1}{\sigma_{it}} e^{-\frac{1}{2}\frac{(\varepsilon_t-\mu_i)^2}{\sigma_{it}^2}}$, $i = 1,\ldots, K$. It can be easily seen that g_i is a function of p_i, μ_i and $\boldsymbol{\gamma_i}$ only, for $i = 1,\ldots, K$-1 and g_K is a function of $\mathbf{p}$, $\boldsymbol{\mu}$ and $\boldsymbol{\gamma_K}$. Also, we denote $ln\left(\sum_{i=1}^{K} g_i\right)$ by $m_t(\boldsymbol{\theta})$. Using the new notations, our objective is to maximize $M(\boldsymbol{\theta}) = \sum_{t=1}^{T} m_t(\boldsymbol{\theta})$. Newton's gradient method is used to obtain the optimal parameters.

The non-negativity of variance and the positivity of the fourth moment are assured by imposing the restrictions:

---

[29] Adding a constant to the function being maximized does not alter the optimal values.

R1. $0 < p_i < 1,\quad i = 1,\ldots, K-1,\quad \sum_{i=1}^{K-1} p_i < 1,\quad 0 \le \alpha_i,\quad 0 \le \beta_i < 1,\quad i = 1,\ldots K$

R2. $m = \sum_{i=1}^{K} p_i \mu_i^2 + \sum_{i=1}^{K} \frac{p_i \omega_i}{(1-\beta_i)} > 0$, $n = \sum_{i=1}^{K} \frac{p_i(1-\alpha_i-\beta_i)}{(1-\beta_i)} > 0$ and $\omega_i + \alpha_i \frac{m}{n} > 0$, $i = 1,\ldots,K$

R3. $0 < E(\varepsilon_t^4) < \infty$

The updating formula has the following form, where $\boldsymbol{g}$ is the gradient vector, $\boldsymbol{H}$ the Hessian matrix and s represents the step-length:

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} - s[\mathbf{H}(\boldsymbol{\theta}^{(m)})]^{-1}\mathbf{g}(\boldsymbol{\theta}^{(m)})$$

To compute the Hessian matrix and the gradient vector, we need to compute the first and second order derivatives of $m_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. The first order derivatives are:

$$\frac{\partial m_t(\boldsymbol{\theta})}{\partial p_i} = \left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left(\frac{g_i}{p_i} + \frac{\partial g_K}{\partial p_i}\right)$$

where

$$\frac{\partial g_K}{\partial p_i} = g_K\left(-\frac{1}{p_K} + \left(\frac{\varepsilon_t - \mu_K}{\sigma_{Kt}^2}\right)\left(\frac{\partial \mu_K}{\partial p_i}\right)\right) \quad \text{and} \quad \frac{\partial \mu_K}{\partial p_i} = -\frac{1}{p_K}\left(\mu_i + \sum_{k=1}^{K-1}\frac{p_k}{p_K}\mu_k\right)$$

$$\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_i} = \left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left(\frac{\partial g_i}{\partial \mu_i} + \frac{\partial g_K}{\partial \mu_i}\right)$$

After some calculations we obtain:

$$\frac{\partial g_i}{\partial \mu_i} = \frac{g_i(\varepsilon_t - \mu_i)}{\sigma_{it}^2} \quad \text{and} \quad \frac{\partial g_K}{\partial \mu_i} = -\left(\frac{p_i}{p_K}\right)\left(\frac{g_K(\varepsilon_t - \mu_K)}{\sigma_{Kt}^2}\right)$$

$$\frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i}} = \left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left(\frac{\partial g_i}{\partial \boldsymbol{\gamma_i}}\right)$$

Also, we have that $\frac{\partial \ln g_{\boldsymbol{i}}}{\partial \boldsymbol{\gamma_i}} = \left(\frac{1}{g_i}\right)\left(\frac{\partial g_{\boldsymbol{i}}}{\partial \boldsymbol{\gamma_i}}\right)$, so (1)

$$\frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i}} = \left(\frac{g_i}{\sum_{k=1}^{K} g_k}\right)\left(\frac{\partial \ln g_i}{\partial \boldsymbol{\gamma_i}}\right) \tag{2}$$

After further calculations we obtain:

$$\frac{\partial \ln g_{\boldsymbol{i}}}{\partial \boldsymbol{\gamma_i}} = \left(\frac{1}{2\sigma_{it}^2}\right)\left(\left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}}\right)^2 - 1\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right) \tag{3}$$

This leads to:

$$\frac{\partial g_{\boldsymbol{i}}}{\partial \boldsymbol{\gamma_i}} = \left(\frac{g_{\boldsymbol{i}}}{2\sigma_{it}^2}\right)\left(\left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}}\right)^2 - 1\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right)$$

From (2) and (3) we obtain the following result:

$$\frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i}} = \left(\frac{1}{2\sigma_{it}^2}\right)\left(\frac{g_i}{\sum_{k=1}^{K} g_k}\right)\left(\left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}}\right)^2 - 1\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right) \tag{4}$$

The second order derivatives are:

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial p_i \partial p_j} = -\left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial p_i}\right)\left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial p_j}\right) + \left(\frac{1}{g_K \sum_{k=1}^{K} g_k}\right)\left(\frac{\partial g_K}{\partial p_i}\right)\left(\frac{\partial g_K}{\partial p_j}\right) -$$

$$-\left(\frac{g_K}{\sum_{k=1}^{K} g_k}\right)\left[\left(\frac{1}{\sigma_{Kt}^2}\right)\left(\frac{\partial \mu_K}{\partial p_i}\right)\left(\frac{\partial \mu_K}{\partial pj}\right) - \left(\frac{1}{p_K}\right)\left(\frac{\varepsilon_t - \mu_K}{\sigma_{Kt}^2}\right)\left(\left(\frac{\partial \mu_K}{\partial p_i}\right) + \left(\frac{\partial \mu_K}{\partial p_j}\right)\right) + \left(\frac{1}{p_K^2}\right)\right]$$

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial p_{\boldsymbol{i}} \partial \mu_i} = -\left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial p_i}\right)\left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_i}\right) + \left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left[\left(\frac{1}{p_i}\right)\left(\frac{\partial g_i}{\partial \mu_i}\right) + \left(\frac{1}{g_K}\right)\left(\frac{\partial g_K}{\partial p_i}\right)\left(\frac{\partial g_K}{\partial \mu_i}\right)\right] +$$

$$+\left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left(\frac{g_K}{p_K}\right)\left(\frac{1}{\sigma_{Kt}^2}\right)\left[p_i\left(\frac{\partial \mu_K}{\partial p_i}\right) - (\varepsilon_t - \mu_K)\left(1 + \frac{p_i}{p_K}\right)\right]$$

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial p_{\boldsymbol{i}} \partial \mu_j} = -\left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial p_i}\right)\left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_j}\right) + \left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left(\frac{1}{g_K}\right)\left(\frac{\partial g_K}{\partial p_i}\right)\left(\frac{\partial g_K}{\partial \mu_j}\right) +$$

$$+\left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left(\frac{g_K}{p_K}\right)\left(\frac{1}{\sigma_{Kt}^2}\right)\left[p_j\left(\frac{\partial \mu_K}{\partial p_i}\right) - (\varepsilon_t - \mu_K)\left(\frac{p_j}{p_K}\right)\right], \qquad j \neq i$$

Based on (4), we have:

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial p_i \partial \boldsymbol{\gamma_j}} = \left(\frac{1}{2\sigma_{jt}^2}\right)\left(\left(\frac{\varepsilon_t - \mu_j}{\sigma_{jt}}\right)^2 - 1\right)\frac{\partial\left(\frac{g_j}{\sum_{k=1}^{K} g_k}\right)}{\partial p_i}\left(\frac{\partial \sigma_{jt}^2}{\partial \boldsymbol{\gamma_j}}\right), \qquad j \neq K$$

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial p_i \partial \boldsymbol{\gamma_K}} = \left(\frac{1}{2\sigma_{Kt}^2}\right)\left[\left(\left(\left(\frac{\varepsilon_t - \mu_K}{\sigma_{Kt}}\right)^2 - 1\right)\frac{\partial\left(\dfrac{g_K}{\sum_{k=1}^{K} g_k}\right)}{\partial p_i} - 2\left(\frac{g_K}{\sum_{k=1}^{K} g_k}\right)\left(\frac{\varepsilon_t - \mu_K}{\sigma_{Kt}^2}\right)\left(\frac{\partial \mu_K}{\partial p_i}\right)\right)\right]\left(\frac{\partial \sigma_{Kt}^2}{\partial \boldsymbol{\gamma_K}}\right)$$

After some calculations we can see that:

$$\frac{\partial\left(\dfrac{g_i}{\sum_{k=1}^{K} g_k}\right)}{\partial p_i} = \frac{\left(\dfrac{g_i}{p_i}\right)\left(\sum_{k=1}^{K} g_k\right) - \left(\dfrac{g_i}{p_i} + \dfrac{\partial g_K}{\partial p_i}\right) g_i}{\left(\sum_{k=1}^{K} g_k\right)^2}$$

$$\frac{\partial\left(\dfrac{g_j}{\sum_{k=1}^{K} g_k}\right)}{\partial p_i} = \frac{-\left(\dfrac{g_i}{p_i} + \dfrac{\partial g_K}{\partial p_j}\right) g_j}{\left(\sum_{k=1}^{K} g_k\right)^2}, \qquad j \neq K \text{ and } j \neq i$$

$$\frac{\partial\left(\dfrac{g_K}{\sum_{k=1}^{K} g_k}\right)}{\partial p_i} = \frac{\left(\dfrac{\partial g_K}{\partial p_K}\right)\left(\sum_{k=1}^{K} g_k\right) - \left(\dfrac{g_i}{p_i} + \dfrac{\partial g_K}{\partial p_i}\right) g_K}{\left(\sum_{k=1}^{K} g_k\right)^2}$$

Furthermore, we have that:

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \mu_i^2} = -\left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_i}\right)^2 + \left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left[\left(\frac{1}{g_i}\right)\left(\frac{\partial g_i}{\partial \mu_i}\right)^2 + \left(\frac{1}{g_K}\right)\left(\frac{\partial g_K}{\partial \mu_i}\right)^2 - \left(\frac{g_i}{\sigma_{it}^2}\right) - \left(\frac{p_i}{p_K}\right)^2\left(\frac{g_K}{\sigma_{Kt}^2}\right)\right]$$

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \mu_i \partial \mu_j} = -\left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_i}\right)\left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_j}\right) + \left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left[\left(\frac{1}{g_K}\right)\left(\frac{\partial g_K}{\partial \mu_i}\right)\left(\frac{\partial g_K}{\partial \mu_j}\right) - \left(\frac{p_i}{p_K}\right)\left(\frac{p_j}{p_K}\right)\left(\frac{g_K}{\sigma_{Kt}^2}\right)\right], \quad j \neq i$$

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \mu_i \partial \boldsymbol{\gamma}_i} = \left(\frac{1}{\sum_{k=1}^{K} g_k}\right)\left\{\left[\left(\frac{1}{g_i}\right)\left(\frac{\partial g_i}{\partial \mu_i}\right) - \left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_i}\right)\right]\left(\frac{\partial g_i}{\partial \boldsymbol{\gamma}_i}\right) - \left(\frac{1}{\sigma_{it}^2}\right)\left(\frac{\partial g_i}{\partial \mu_i}\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma}_i}\right)\right\}$$

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \mu_i \partial \boldsymbol{\gamma}_j} = \left(\frac{-1}{\sum_{k=1}^{K} g_k} \right) \left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_i} \right) \left(\frac{\partial g_i}{\partial \boldsymbol{\gamma}_j} \right), \qquad j \neq K \text{ and } j \neq i$$

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \mu_i \partial \boldsymbol{\gamma}_K} = \left(\frac{1}{\sum_{k=1}^{K} g_k} \right) \left\{ \left[\left(\frac{1}{g_K} \right) \left(\frac{\partial g_K}{\partial \mu_i} \right) - \left(\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_i} \right) \right] \left(\frac{\partial g_K}{\partial \boldsymbol{\gamma}_K} \right) - \left(\frac{1}{\sigma_{Kt}^2} \right) \left(\frac{\partial g_K}{\partial \mu_i} \right) \left(\frac{\partial \sigma_{Kt}^2}{\partial \boldsymbol{\gamma}_K} \right) \right\}$$

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} = \left(\frac{\partial \, ln \, g_i}{\partial \boldsymbol{\gamma_i}} \right) \frac{\partial \left(\frac{g_i}{\sum_{k=1}^{K} g_k} \right)}{\partial \boldsymbol{\gamma_i}'} + \left(\frac{g_i}{\sum_{k=1}^{K} g_k} \right) \left(\frac{\partial^2 \, ln \, g_i}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} \right) \tag{5}$$

We need to compute the partial derivatives on the right hand side of (5). Using (1) the following expression results:

$$\frac{\partial \left(\frac{g_i}{\sum_{k=1}^{K} g_k} \right)}{\partial \boldsymbol{\gamma_i}'} = \left(\frac{g_i \left(\sum_{k=1}^{K} g_k - g_i \right)}{\left(\sum_{k=1}^{K} g_k \right)^2} \right) \left(\frac{\partial \, ln \, g_i}{\partial \boldsymbol{\gamma_i}} \right)'$$

which, based on (3), simplifies to:

$$\frac{\partial \left(\frac{g_i}{\sum_{k=1}^{K} g_k} \right)}{\partial \boldsymbol{\gamma_i}'} = \left(\frac{1}{2\sigma_{it}^2} \right) \left(\frac{g_i \left(\sum_{k=1}^{K} g_k - g_i \right)}{\left(\sum_{k=1}^{K} g_k \right)^2} \right) \left(\left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 - 1 \right) \left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}} \right)' \tag{6}$$

Using equation (3) again we obtain:

$$\frac{\partial^2 \, ln \, g_i}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} = \left(\frac{1}{2\sigma_{it}^2} \right) \left[\left(\frac{1}{\sigma_{it}^2} \right) \left(1 - 2 \left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 \right) \left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}} \right) \left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}} \right)' + \left(\left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 - 1 \right) \left(\frac{\partial^2 \sigma_{it}^2}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} \right) \right] \tag{7}$$

Combining and grouping (3), (5), (6) and (7) we have the following result:

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} = \left(\frac{1}{\sigma_{it}^2} \right)^2 \left[\left(\frac{g_i \left(\sum_{k=1}^{K} g_k - g_i \right)}{4 \left(\sum_{k=1}^{K} g_k \right)^2} \right) \left(1 - \left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 \right)^2 + \left(\frac{g_i}{2 \left(\sum_{k=1}^{K} g_k \right)} \right) \left(1 - 2 \left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 \right) \right] \left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}} \right) \left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}} \right)'$$

$$+ \left(\frac{1}{2\sigma_{it}^2} \right) \left(\frac{g_i}{\sum_{k=1}^{K} g_k} \right) \left(\left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 - 1 \right) \left(\frac{\partial^2 \sigma_{it}^2}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} \right)$$

Based on (4), we can write the cross-derivative with respect to $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_j$:

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_j}'} = \left(\frac{1}{2\sigma_{it}^2}\right)\left(\frac{g_i}{\left(\sum_{k=1}^{K} g_k\right)^2}\right)\left(1-\left(\frac{\varepsilon_t-\mu_i}{\sigma_{it}}\right)^2\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right)\left(\frac{\partial g_j}{\partial \boldsymbol{\gamma_j}}\right)', \quad j \neq i$$

Using (1) and (3) the above implies:

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_j}'} = \left(\frac{-1}{4\sigma_{it}^2\sigma_{jt}^2}\right)\left(\frac{g_i g_j}{\left(\sum_{k=1}^{K} g_k\right)^2}\right)\left(1-\left(\frac{\varepsilon_t-\mu_i}{\sigma_{it}}\right)^2\right)\left(1-\left(\frac{\varepsilon_t-\mu_j}{\sigma_{jt}}\right)^2\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right)\left(\frac{\partial \sigma_{jt}^2}{\partial \boldsymbol{\gamma_j}}\right)'$$

The first and second order derivatives of σ_{it}^2 with respect to $\boldsymbol{\gamma_i}$ still need to be computed:

$$\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}} = z_{it} + \beta_i \frac{\partial \sigma_{it-1}^2}{\partial \boldsymbol{\gamma_i}}$$

where $z_{it} = (1, \varepsilon_{t-1}^2, \sigma_{it-1}^2)'$. The starting values for this expression (for $t = 0$) are:

$$\frac{\partial \sigma_{i0}^2}{\partial \boldsymbol{\gamma_i}} = (1, s^2, s^2)', \text{ where } s^2 = \frac{\sum_{t=1}^{T} \varepsilon_t^2}{T}$$

The second order derivatives are:

$$\frac{\partial^2 \sigma_{it}^2}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} = w_{it} + \beta_i \frac{\partial^2 \sigma_{it-1}^2}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} \qquad w_{it} = A_{it} + A_{it}^T \qquad A_{it} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ & \left(\frac{\partial \sigma_{it-1}^2}{\partial \boldsymbol{\gamma_i}}\right)' & \end{bmatrix}$$

The starting values for this computation are $\frac{\partial^2 \sigma_{i0}^2}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} = \frac{1}{(1-\beta_i)} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & s^2 \\ 1 & s^2 & 2s^2 \end{bmatrix}$

Appendix D. Derivation of the Conditional and Unconditional Moments of the Error Term and the Autocorrelation of the Squared Residuals of the normal GARCH(1,1) Model

The model is specified as:

$$y_t = \varepsilon_t, \qquad \varepsilon_t \big| I_{t-1} \sim N(0, \sigma_t^2)$$

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

Using the notation $x = E(\varepsilon_t^2) = E(\sigma_t^2)$ we have:

$$x = E(\varepsilon_t^2) = E(\sigma_t^2) = \frac{\omega}{1-\alpha-\beta}$$

This model has no conditional skewness and excess kurtosis, and the unconditional third moment is:

$$h = E\left(\varepsilon_t^3\right) = E\left(E_{t-1}\left(\varepsilon_t^3\right)\right) = E(0) = 0$$

and the unconditional skewness can be expressed as:

$$s = \frac{h}{x^{3/2}} = 0$$

The derivation of the unconditional fourth moment is the following:

$$z = E(\varepsilon_t^4) = 3v \text{, where } v = E(\sigma_t^4)$$

We can write:

$$v(1-\beta^2 - 3\alpha^2 - 2\alpha\beta) = \omega^2 + 2\omega x(\alpha+\beta)$$

which implies:

$$v = \frac{\omega^2 + 2\omega x(\alpha+\beta)}{1-\beta^2-3\alpha^2-2\alpha\beta}$$

and the unconditional excess kurtosis is:

$$K = \frac{E(\varepsilon_t^4)}{E(\varepsilon_t^2)^2} - 3 = \frac{z}{x^2} - 3$$

The autocorrelations of the squared residuals are:

$$\varrho_k = Corr\left(\varepsilon_t^2, \varepsilon_{t-k}^2\right) = \frac{Cov\left(\varepsilon_t^2, \varepsilon_{t-k}^2\right)}{Var\left(\varepsilon_t^2\right)} = \frac{E[\varepsilon_t^2\varepsilon_{t-k}^2] - x^2}{E[\varepsilon_t^4] - x^2} = \frac{c_k - x^2}{z - x^2},$$

where

$$c_k = \omega x + (\alpha+\beta)c_{k-1} \qquad k > 1$$
$$c_1 = \omega x + \alpha c_0 + \beta z / 3 \qquad k = 1$$
$$c_0 = z$$

This reduces to:

$$\varrho_k = \left(\alpha + \frac{\alpha^2\beta}{1-2\alpha\beta-\beta^2}\right)(\alpha+\beta)^{k-1}$$

Appendix E. Derivation of the Conditional and Unconditional Moments of the Error Term and the Autocorrelation of the Squared Residuals of the Standardized and Standardized Skewed t-GARCH(1,1) Models

(a) Standardized t-GARCH(1,1) Model

The model is given by:

$$y_t = \varepsilon_t \qquad \varepsilon_t = z_t \sigma_t \qquad z_t \big| I_{t-1} \sim \text{Standardized Student's } t(v)$$

The density function is:

$$g(z \mid v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{(v-2)\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{z^2}{v-2}\right)^{-\frac{v+1}{2}}$$

The variance of the model follows a GARCH(1,1) process:

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

The unconditional second moment is:

$$x = E(\varepsilon_t^2) = E(\sigma_t^2) = \frac{\omega}{1-\alpha-\beta}$$

The conditional and unconditional third moment and skewness are zero, and the conditional kurtosis is:

$$k = \frac{E_{t-1}(\varepsilon_t^4)}{\sigma_t^4} = \frac{\sigma_t^4 E_{t-1}(z_t^4)}{\sigma_t^4} = E_{t-1}(z_t^4) = \left(\frac{3v-6}{v-4}\right)\left(E_{t-1}(z_t^2)\right)^2 = \frac{3v-6}{v-4} \tag{1}$$

To derive the unconditional kurtosis we start with the unconditional fourth moment:

$$z = E(\varepsilon_t^4) = E\left(E_{t-1}(z_t^4\sigma_t^4)\right) = E\left(\sigma_t^4 E_{t-1}(z_t^4)\right) = \left(\frac{3v-6}{v-4}\right)E(\sigma_t^4) = kE(\sigma_t^4)$$

Also, we have that:

$$E(\sigma_t^4) = \left(\frac{\omega^2 + 2\omega x(\alpha+\beta)}{1-\beta^2-k\alpha^2-2\alpha\beta}\right) \tag{2}$$

So the unconditional kurtosis is:

$$K = \frac{E(\varepsilon_t^4)}{x^2} = \left(\frac{k}{x^2}\right)\left(\frac{\omega^2 + 2\omega x(\alpha+\beta)}{1-\beta^2-k\alpha^2-2\alpha\beta}\right) \tag{3}$$

(b) Standardized Skewed t-GARCH(1,1) Model

The model is specified as:

$$y_t = \varepsilon_t \qquad \varepsilon_t = z_t \sigma_t \qquad z_t \big| I_{t-1} \sim \text{Standardized Skewed Student's } t(v,\gamma)$$

The density is – as introduced by Fernandez and Steel (1998) and Lambert and Laurent (2001a and 2001b):

$$f(z\mid v,\gamma)=\left(\frac{2}{\gamma+\frac{1}{\gamma}}\right)s\left\{g\left[\gamma(sz+e_1)\mid v\right]I_{(-\infty,0)}(z+e_1/s)+g\left[(sz+e_1)/\gamma\mid v\right]I_{[0,\infty)}(z+e_1/s)\right\}$$

where I is an indicator function such that: $I_t(z)=\begin{cases}1\ if\ z_t\geq -e_1/s\\ -1\ if z_t<-e_1/s\end{cases}$

and we use the following notation for the r[th] moment:[30]

$$e_r=e_r(v,\gamma)=E(y^r\mid v,\gamma)=M_r(v)\left(\frac{\gamma^{r+1}+\frac{(-1)^r}{\gamma^{r+1}}}{\gamma+\frac{1}{\gamma}}\right)\qquad y\sim\text{Skewed Student's } t(v,\gamma).$$

where $M_r(v)$ is the r^{th} order moment of the Student's $t(v)$ distribution, truncated to the positive real values:

$$M_r(v)=\int_0^\infty 2w^r g(w\mid v)dw=\frac{\Gamma\left(\frac{v-r}{2}\right)\Gamma\left(\frac{1+r}{2}\right)(v-2)^{r/2}}{\sqrt{\pi}\Gamma\left(\frac{v}{2}\right)}$$

and $\qquad s=s(v,\gamma)=\sqrt{Var(y\mid v,\gamma)}$

We let $\qquad V=Var(y\mid v,\gamma)=e_2-e_1^2$

The variance is a GARCH(1,1) process:

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

The unconditional second moment is:

$$x=E(\varepsilon_t^2)=E(\sigma_t^2)=\frac{\omega}{1-\alpha-\beta}$$

The conditional skewness is:

$$\tau=\frac{E_{t-1}(\varepsilon_t^3)}{\sigma_t^3}=\frac{\sigma_t^3E_{t-1}(z_t^3)}{\sigma_t^3}=E_{t-1}(z_t^3)=\frac{e_3-3e_1e_2+2e_1^3}{V^{3/2}}\tag{4}$$

The unconditional skewness can be expressed as:

$$S=\frac{E(\varepsilon_t^3)}{x^{3/2}}$$

The third moment is:

$$E(\varepsilon_t^3)=E(z_t^3\sigma_t^3)=E\left[\sigma_t^3E_{t-1}(z_t^3)\right]=\tau E(\sigma_t^3)$$

By taking limits in $E(\sigma_t^3)=\omega E(\sigma_t)+(\alpha+\beta)E(\sigma_t\sigma_{t-1}^2)$ we obtain that

$$E(\sigma_t^3)=E(\sigma_t^2)*E(\sigma_t)$$

---

[30] Adopted from Fernandez and Steel (1998) and Lambert and Laurent (2001a and 2001b).

$E(\sigma_t)$ has no exact formula, an approximation available for it being $E(\sigma_t) \cong \sqrt{E(\sigma_t^2)}$, leading to:

$$E(\sigma_t^3) = x^{3/2}$$

Hence the conditional and unconditional skewness is $S = \tau$ given by (4).

The conditional kurtosis is expressed as:

$$k = \frac{E_{t-1}(\varepsilon_t^4)}{\sigma_t^4} = \frac{\sigma_t^4 E_{t-1}(z_t^4)}{\sigma_t^4} = E_{t-1}(z_t^4) = \left(\frac{3v-6}{v-4}\right)\left(E_{t-1}(z_t^2)\right)^2 = \frac{e_4 - 4e_1e_3 + 6e_2e_1^2 - 3e_1^4}{V^2} \quad (5)$$

The unconditional fourth moment is:

$$z = E(\varepsilon_t^4) = E\left(E_{t-1}(z_t^4\sigma_t^4)\right) = E\left(\sigma_t^4 E_{t-1}(z_t^4)\right) = kE(\sigma_t^4)$$

It can be shown that:

$$E(\sigma_t^4) = \left(\frac{\omega^2 + 2\omega x(\alpha+\beta)}{1-\beta^2 - k\alpha^2 - 2\alpha\beta}\right) \quad (6)$$

So the unconditional kurtosis is:

$$K = \frac{E(\varepsilon_t^4)}{x^2} = \left(\frac{k}{x^2}\right)\left(\frac{\omega^2 + 2\omega x(\alpha+\beta)}{1-\beta^2 - k\alpha^2 - 2\alpha\beta}\right) \quad (7)$$

Unfortunately this framework does not allow time variation in conditional skewness and excess kurtosis. For both models, the autocorrelation of the squared residuals has the form:

$$\varrho_k = Corr\left(\varepsilon_t^2, \varepsilon_{t-k}^2\right) = \frac{Cov\left(\varepsilon_t^2, \varepsilon_{t-k}^2\right)}{Var\left(\varepsilon_t^2\right)} = \frac{E[\varepsilon_t^2\varepsilon_{t-k}^2] - x^2}{E[\varepsilon_t^4] - x^2} = \frac{c_k - x^2}{z - x^2}$$

$$c_k = \omega x + (\alpha+\beta)c_{k-1} \qquad k > 1$$
$$c_1 = \omega x + \alpha c_0 + \beta z / 3 \qquad k = 1$$
$$c_0 = z$$

which reduces to:

$$\varrho_k = \left(\alpha + \frac{\alpha^2\beta}{1 - 2\alpha\beta - \beta^2}\right)(\alpha+\beta)^{k-1} \quad (8)$$

Appendix F. Moment Specification Tests

The specifications of the model present an error term that has a normal mixture conditional distribution. Given $\boldsymbol{\theta}$ the k-dimensional vector of parameters, $\hat{\boldsymbol{\theta}}$ the parameter estimates and $\hat{u}_t$ the standardized residuals, the estimated vector of restrictions, $\mathbf{r}(\hat{\boldsymbol{\theta}})$ can be written as:

$$\mathbf{r}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} r_1(\hat{\boldsymbol{\theta}}) \\ \vdots \\ r_J(\hat{\boldsymbol{\theta}}) \end{pmatrix} = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} m_1(\hat{u}_t) \\ \vdots \\ \frac{1}{T}\sum_{t=1}^{T} m_J(\hat{u}_t) \end{pmatrix} \tag{1}$$

This is a J-dimensional vector, where J represents the number of restrictions to be tested. For example, when verifying the first moment, we have that J =1 and $m_1(\hat{u}_t) = \hat{u}_t$. Let $\hat{\mathbf{M}}$ denote the following $T \times J$ matrix:

$$\hat{\mathbf{M}} = \begin{bmatrix} m_1(\hat{u}_1) & \dots & m_J(\hat{u}_1) \\ \vdots & \vdots & \vdots \\ m_1(\hat{u}_T) & \dots & m_J(\hat{u}_T) \end{bmatrix} \tag{2}$$

The log-likelihood function is $l(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta})$. The partial derivative of $l_t(\boldsymbol{\theta})$ with respect to θ_i, evaluated at the estimated parameter values is denoted by:

$$d_{t,i}(\hat{\boldsymbol{\theta}}) = \left.\frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

Let $\mathbf{D}$ be the first derivatives matrix of the realizations of the log-likelihood function with respect to the parameters, evaluated at the estimated values:

$$\hat{\mathbf{D}} = \begin{bmatrix} d_{1,1}(\hat{\boldsymbol{\theta}}) & \dots & d_{1,k}(\hat{\boldsymbol{\theta}}) \\ \vdots & \vdots & \vdots \\ d_{T,1}(\hat{\boldsymbol{\theta}}) & \dots & d_{T,k}(\hat{\boldsymbol{\theta}}) \end{bmatrix} \tag{3}$$

The null hypothesis is that the restrictions are zero:

$$H_0\colon \boldsymbol{r}(\boldsymbol{\theta}) = 0 \tag{4}$$

Let $\hat{\boldsymbol{\Omega}}$ denote the variance-covariance matrix of $\mathbf{r}(\hat{\boldsymbol{\theta}})$. The test statistic is:

$$W = T\mathbf{r}(\hat{\boldsymbol{\theta}})^T \hat{\mathbf{\Omega}}^{-1}\mathbf{r}(\hat{\boldsymbol{\theta}}), \tag{5}$$

which has a χ^2_J distribution under the null.

Greene (2000) has shown that the variance-covariance matrix of $\mathbf{r}(\hat{\boldsymbol{\theta}})$ can be calculated as:

$$\hat{\mathbf{\Omega}} = \frac{1}{T}\left[\hat{\mathbf{M}}^T\hat{\mathbf{M}} - \hat{\mathbf{M}}^T\hat{\mathbf{D}}\left(\hat{\mathbf{D}}^T\hat{\mathbf{D}}\right)^{-1}\hat{\mathbf{D}}^T\hat{\mathbf{M}}\right] \tag{6}$$

Legends to Figures and Tables:

Fig. 1. Non-existence of the fourth moment

The unconditional volatility, the fourth moment and the excess kurtosis (as derived in Appendix A) as a function of the first mixing parameter for the following NM(2)-GARCH(1,1) model:

$$y_t = \varepsilon_t, \quad \varepsilon_t \mid I_{t-1} \sim NM(p, 1-p; 0, 0; \sigma_{1t}^2, \sigma_{2t}^2)$$

$$\sigma_{1t}^2 = 0.00001 + 0.03\varepsilon_{t-1}^2 + 0.9\sigma_{1,t-1}^2 \quad \sigma_{2t}^2 = 0.0001 + 0.05\varepsilon_{t-1}^2 + 0.96\sigma_{2,t-1}^2$$

Fig. 2. Classification of estimation results

Estimation results are considered unrealistic when convergence is not achieved, boundary values are obtained or the long-term variance obtained is negative.

Fig. 3. Bias in parameter estimates

The estimated values of the NM(2)-GARCH(1,1) model parameters, and the unconditional volatilities and excess kurtosis estimated by the model, averaged over all the simulations for which realistic estimates were obtained. 5000 simulations were based on 2000 realisations of the same data generating process as in Fig. 1. Note that the estimates of the mixing parameter and the two component long term volatilities (and the overall long term volatility) have little or no large sample bias.

Fig. 4. Efficiency of parameter estimates

The efficiency of the estimators is given by a comparison of the Cramér-Rao bounds (where available) and the standard errors of the estimates of the parameters. The Cramér-Rao bound is computed as the main diagonal elements of the inverse of the information matrix. The data generating process is as in Fig. 1.

Fig. 5. Stability of the estimation

The effect of increasing each variable by 10% of its values is presented, expressed as the size of the bias of the parameter estimates as percentage of their standard errors.

Fig. 6. The returns on the three exchange rates and their conditional volatilities

Daily returns are computed as the first difference of the logarithm of the exchange rates over the period 3rd January 1989 to 31st December 2002 (a total of 3651 observations). Subsequently this 14 year period is divided into three sub-periods of roughly equal length. The conditional volatilities of the GARCH(1,1), Skewed t-GARCH(1,1), NM(2)-GARCH(1,1) and NM(3)-GARCH(1,1) models have the same shape.

Fig. 7. The unrestricted NM(2) and NM(3) conditional volatilities of the three exchange rates

The individual conditional volatilities of the NM(2)-GARCH(1,1) and NM(3)-GARCH(1,1) models have similar shape to the overall volatilities, shifted upwards or downwards.

Fig. 8. Estimated conditional skewness

Conditional skewness's for the three exchange rates estimated via the NM(2)- and NM(3)-GARCH(1,1) models respectively. The conditional skewness's are too jumpy except for the unrestricted models.

Fig. 9. Estimated conditional excess kurtosis

Conditional excess kurtosis's for the three exchange rates estimated via the NM(2)- and NM(3)-GARCH(1,1) model respectively. The conditional excess kurtosis's are too jumpy except for the NM(2) unrestricted (symmetric or asymmetric) models.

Fig. 10. Histogram fitting

Using a nonparametric kernel approach, histograms were fitted to the real data and simulated returns for the GARCH(1,1), Skewed t-GARCH(1,1) and the unrestricted NM(2) and NM(3)-GARCH(1,1) models.

Table 1. Statistical description of the data

For a return X the first four moments of its distributions are the mean $\mu = E(X)$, *variance* $\sigma^2 = E[(X-\mu)^2]$, *skewness,* $\tau = E[(X-\mu)^3] / \sigma^3$ *and excess kurtosis,* $k = E[(X-\mu)^4] / \sigma^4 - 3$. *The standard error (s.e.) of the sample estimates of these parameters are as follows: s.e. sample mean =* $\sigma/\sqrt{T}$, *s.e. sample variance =* $\sqrt{2}\ \sigma^2/T$, *s.e. sample skewness* $\approx \sqrt{6/T}$, *s.e. sample excess kurtosis* $\approx \sqrt{24/T}$, *where T represents the total number of observations. In the table where s.e. of the higher sample moments have been estimated, *, ** and *** represent results significantly different from zero at the 5%, 1% and 0.1% level, respectively.*

Table 2. Estimates of the maximal moment exponent and tests for the existence of the fourth moment

The maximal moment exponent is expressed as $\hat{a}_s = \left(s^{-1} \left(\sum_{j=0}^{s-1} ln\, \varepsilon_{n-j} \right) - ln\, \varepsilon_{n-s} \right)^{-1}$ *where* $\varepsilon_1 \leq \ldots \leq \varepsilon_n$ *are the ordered returns and s is defined as the percentage of number of the observations in the right (left) tail. The fourth moment test is based on the fact that, approximately* $\sqrt{s}\,(\hat{a}_s - a) \sim N(0, a)$.

Table 3.The Models

Table 4. Estimation results for NM(K)-GARCH(1,1) and t-GARCH(1,1) models for GBP

Parameters are estimated by MLE. Numbers in parenthesis represent t-ratios. The cumulative test is a joint test that the moment and AC conditions for all moments are met (a total of 20 conditions). Test statistics for the moment tests have a $\chi^2(1)$ *distribution and for the cumulative test have a* $\chi^2(20)$ *distribution.* and ** signify significance at 5% and 1% significance level, respectively. Histogram results show the fit of simulated data with the original data, and the Likelihood Ratio is based on the comparison with the simplest extension of the model. ACF results show the fit of the autocorrelation function of the squared residuals of the model to the autocorrelation of the squared returns, whilst VaR rankings are based on a counting method of the losses higher than the estimated VaR.*

Table 5. Estimation results for NM(K)-GARCH(1,1) and t-GARCH(1,1) models for EUR

Parameters are estimated by MLE. Numbers in parenthesis represent t-ratios. The cumulative test is a joint test that the moment and AC conditions for all moments are met (a total of 20 conditions). Test statistics for the moment tests have a $\chi^2(1)$ *distribution and for the cumulative test have a* $\chi^2(20)$ *distribution. * and ** signify significance at 5% and 1% significance level, respectively. Histogram results show the fit of simulated data with the original data, and the Likelihood Ratio is based on the comparison with the simplest extension of the model. ACF results show the fit of the autocorrelation function of the squared residuals of the model to the autocorrelation of the squared returns, whilst VaR rankings are based on a counting method of the losses higher than the estimated VaR..*

Table 6. Estimation results for NM(K)-GARCH(1,1) and t-GARCH(1,1) models for JPY

Parameters are estimated by MLE. Numbers in parenthesis represent t-ratios. The cumulative test is a joint test that the moment and AC conditions for all moments are met (a total of 20 conditions). Test statistics for the moment tests have a $\chi^2(1)$ *distribution and for the cumulative test have a* $\chi^2(20)$ *distribution. * and ** signify significance at 5% and 1% significance level, respectively. Histogram results show the fit of simulated data with the original data, and the Likelihood Ratio is based on the comparison with the simplest extension of the model. ACF results show the fit of the autocorrelation function of the squared residuals of the model to the autocorrelation of the squared returns, whilst VaR rankings are based on a counting method of the losses higher than the estimated VaR.*

Table 7. Estimation results for NM(K)-GARCH(1,1) and t-GARCH(1,1) models for GBP (excluding the most extreme observation)

Parameters are estimated by MLE. Numbers in parenthesis represent t-ratios. The cumulative test is a joint test that the moment and AC conditions for all moments are met (a total of 20 conditions). Test statistics for the moment tests have a $\chi^2(1)$ *distribution and for the cumulative test have a* $\chi^2(20)$ *distribution. * and ** signify significance at 5% and 1% significance level, respectively. Histogram results show the fit of simulated data with the original data, and the Likelihood Ratio is based on the comparison with the simplest extension of the model. ACF results show the fit of the autocorrelation function of the squared residuals of the model to the autocorrelation of the squared returns, whilst VaR rankings are based on a counting method of the losses higher than the estimated VaR.*

Table 8. Estimation results for NM(K)-GARCH(1,1) and t-GARCH(1,1) models for EUR (excluding the most extreme observation)

Parameters are estimated by MLE. Numbers in parenthesis represent t-ratios. The cumulative test is a joint test that the moment and AC conditions for all moments are met (a total of 20 conditions). Test statistics for the moment tests have a $\chi^2(1)$ *distribution and for the cumulative test have a* $\chi^2(20)$ *distribution. * and ** signify significance at 5% and 1% significance level, respectively. Histogram results show the fit of simulated data with the original data, and the Likelihood Ratio is based on the comparison with the simplest extension of the model. ACF results show the fit of the autocorrelation function of the squared residuals of the model to the autocorrelation of the squared returns, whilst VaR rankings are based on a counting method of the losses higher than the estimated VaR.*

Table 9. Estimation results for NM(K)-GARCH(1,1) and t-GARCH(1,1) models for JPY (excluding the most extreme observation)

Parameters are estimated by MLE. Numbers in parenthesis represent t-ratios. The cumulative test is a joint test that the moment and AC conditions for all moments are met (a total of 20 conditions). Test statistics for the moment tests have a $\chi^2(1)$ *distribution and for the cumulative test have a* $\chi^2(20)$ *distribution. * and ** signify significance at 5% and 1% significance level, respectively. Histogram results show the fit of simulated data with the original data, and the Likelihood Ratio is based on the comparison with the simplest* *extension of the model. ACF results show the fit of the autocorrelation function of the squared residuals of the model to the autocorrelation of the squared returns, whilst VaR rankings are based on a counting method of the losses higher than the estimated VaR.*

Table 10. Estimation results for GARCH(1,1), the skewed t-GARCH(1,1) and NM(K)-GARCH(1,1) models for GBP for different subperiods

Parameters are estimated by MLE. Numbers in parenthesis represent t-ratios.

Table 11. Estimation results for GARCH(1,1), the skewed t-GARCH(1,1) and NM(K)-GARCH(1,1) models for EUR for different subperiods

Parameters are estimated by MLE. Numbers in parenthesis represent t-ratios.

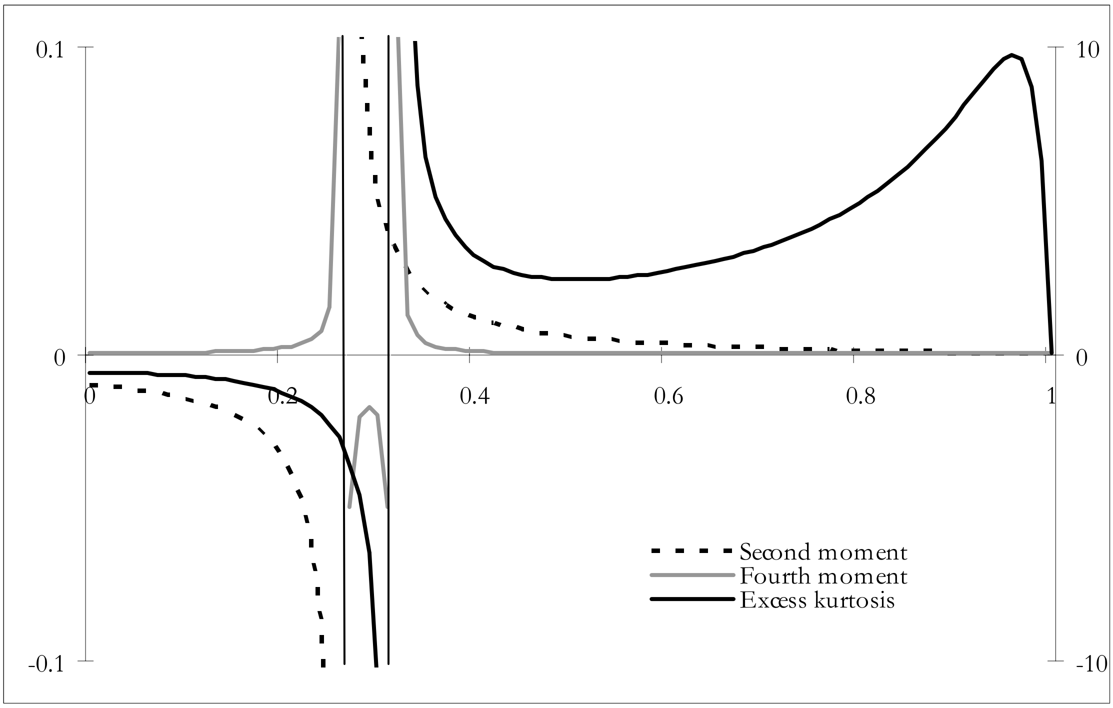
Table 12. Estimation results for GARCH(1,1), the skewed t-GARCH(1,1) and NM(K)-GARCH(1,1) models for JPY for different subperiods

Parameters are estimated by MLE. Numbers in parenthesis represent t-ratios.

Table 13. The realized and the modelled unconditional skewness and excess kurtosis for the three exchange rates for different periods

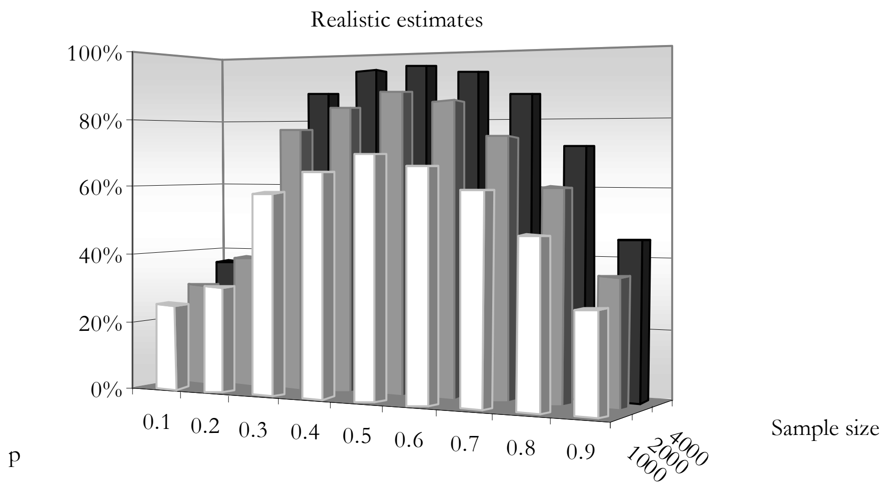
The unconditional realized skewness is the sample skewness, also shown in Table 1 for the entire period. The unconditional realized excess kurtosis is the sample kurtosis, estimated as in Table 1. The formulae for the unconditional skewness and excess kurtosis in the NM(K)-GARCH(1,1) models are given in Appendix B. Numbers in bold represent values that are closest to the realized values.

Fig. 1.



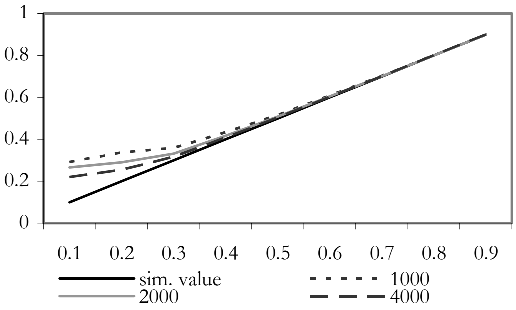
0.1
0
-0.1
10
0
-10
0
0.2
0.4
0.6
0.8
1
Second moment
Fourth moment
Excess kurtosis

Fig. 2.

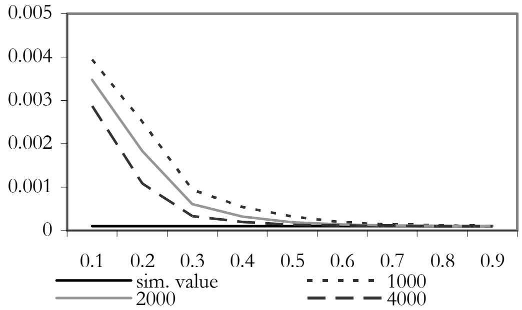


Realistic estimates
100%
80%
60%
40%
20%
0%
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
p
4000
2000
1000
Sample size

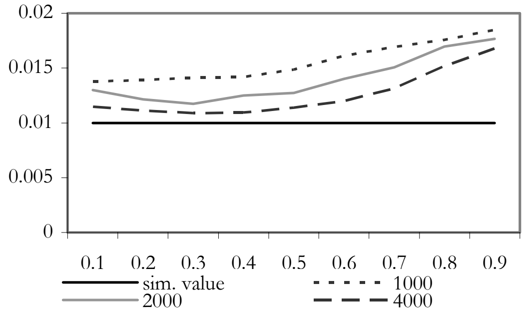
Fig. 3.



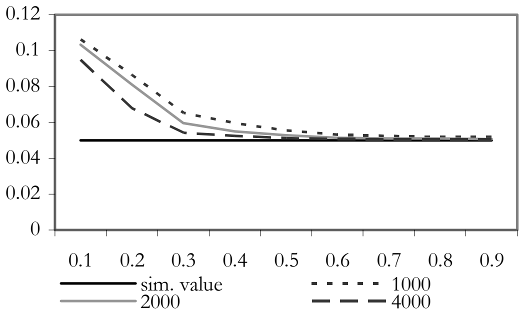
(a) p



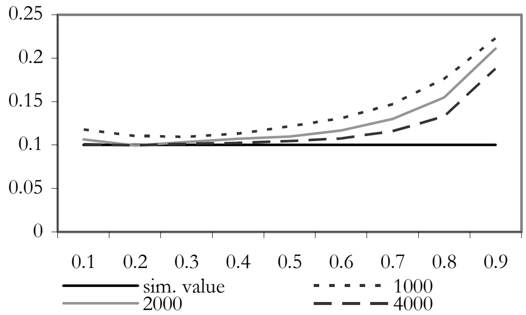
(b) omega1



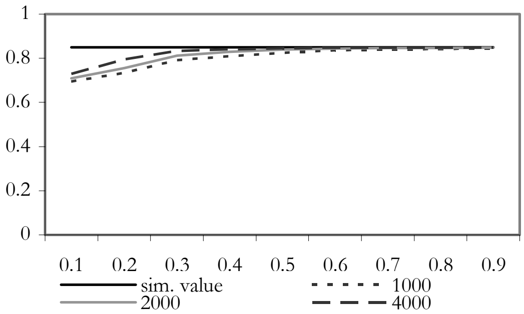
(c)omega2



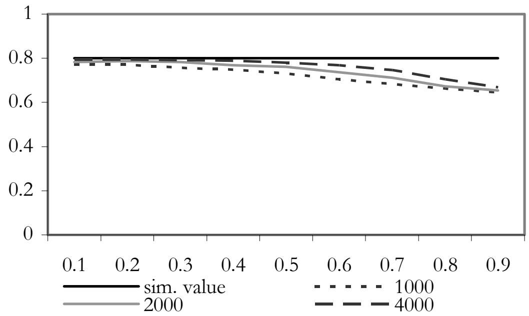
(d) alpha1



(e) alpha2

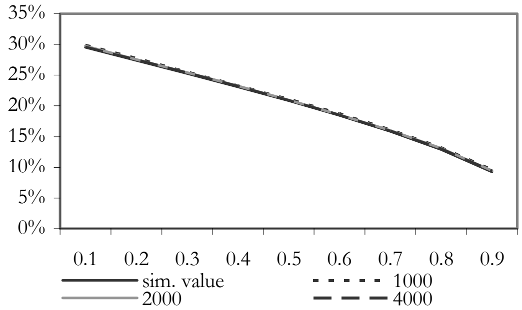


(f) beta1

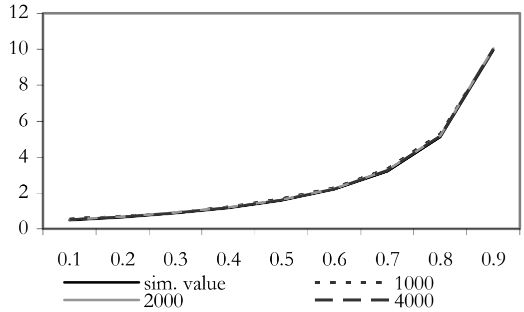


(g) beta2

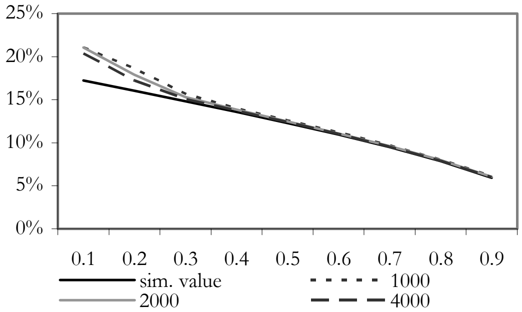
(continuation of Fig. 3.)



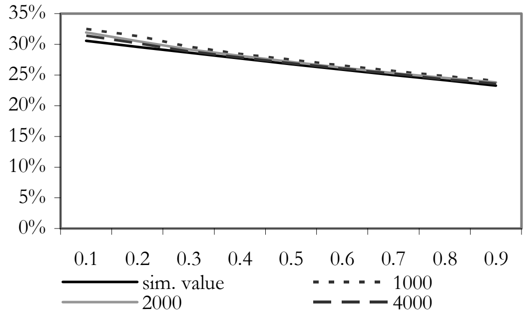
(h) long-term volatililty



(i) kurtosis

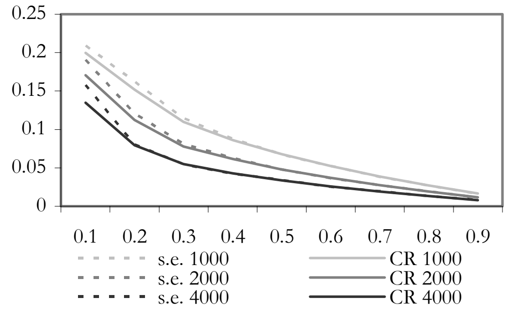


(j) long-term volatility 1

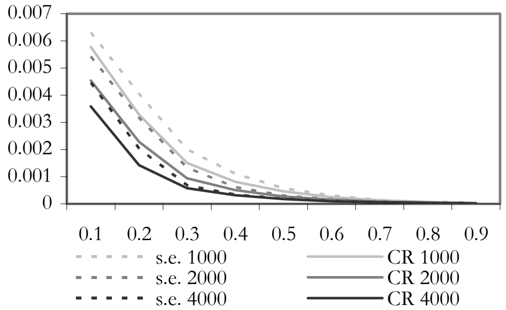


(k) long-term volatility 2

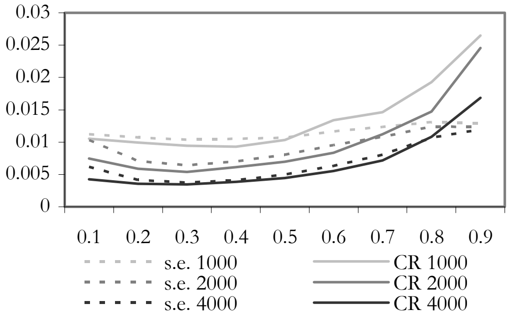
Fig. 4.



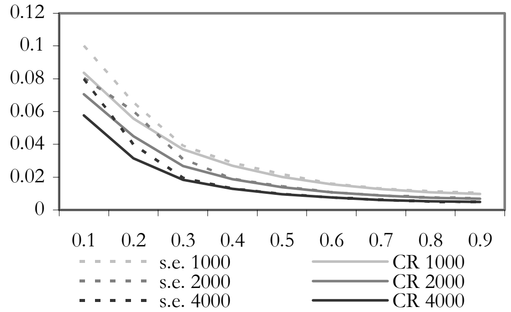
(a) p



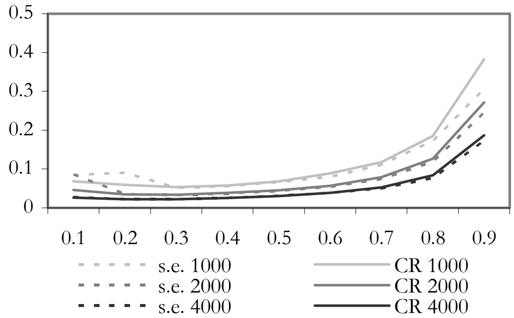
(b) omega1



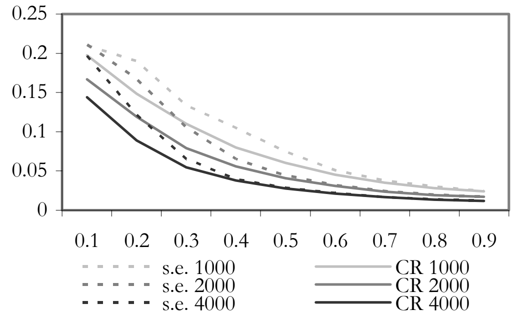
(c)omega2



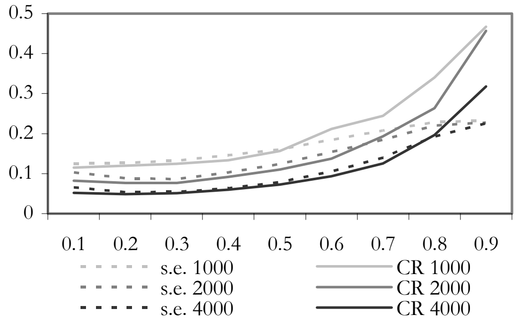
(d) alpha1



(e) alpha2

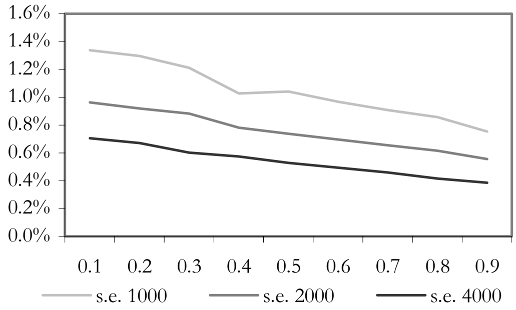


(f) beta1

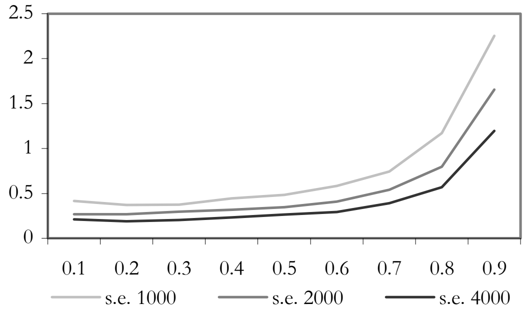


(g) beta2

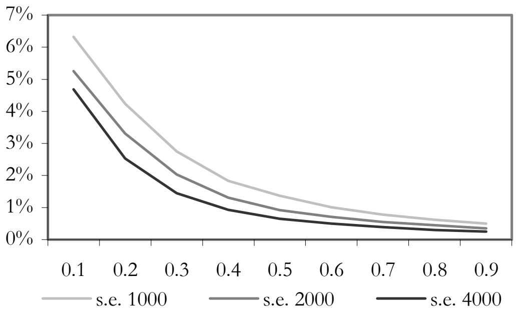
(continuation of Fig. 4.)



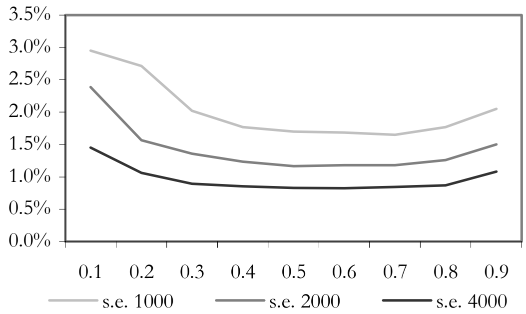
(h) long-term volatililty



(i) kurtosis

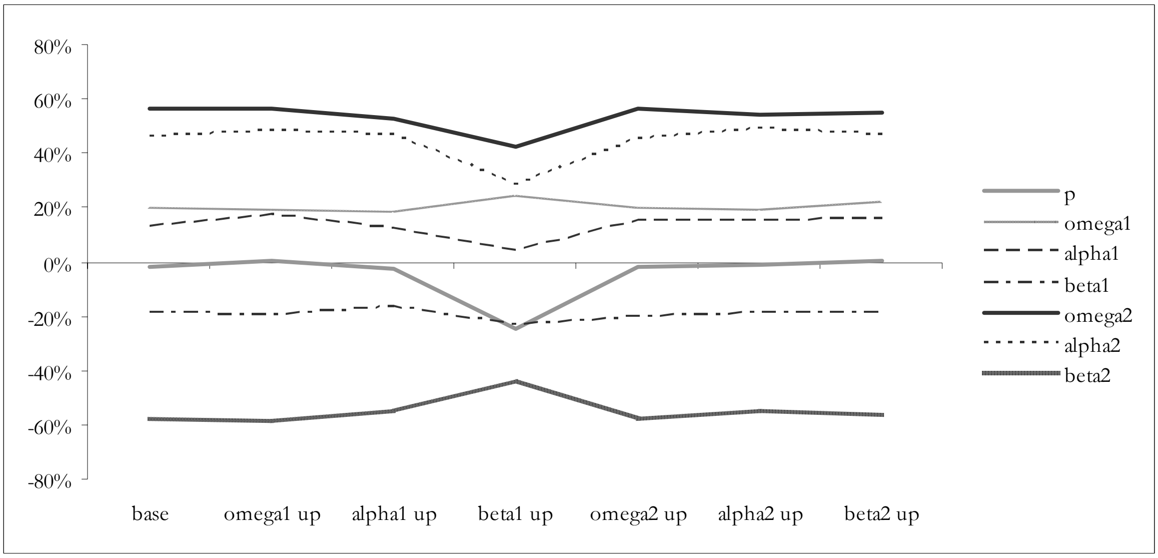


(j) long-term volatility 1



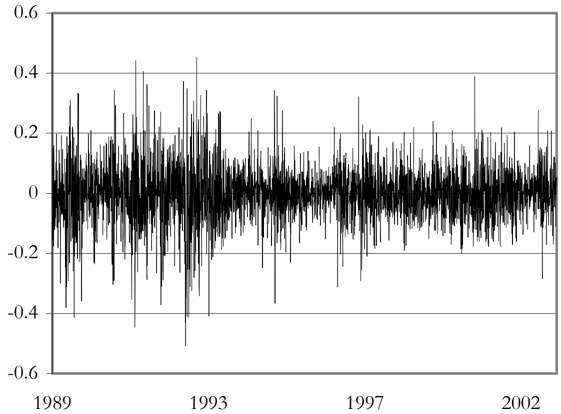
(k) long-term volatility 2

Fig. 5.

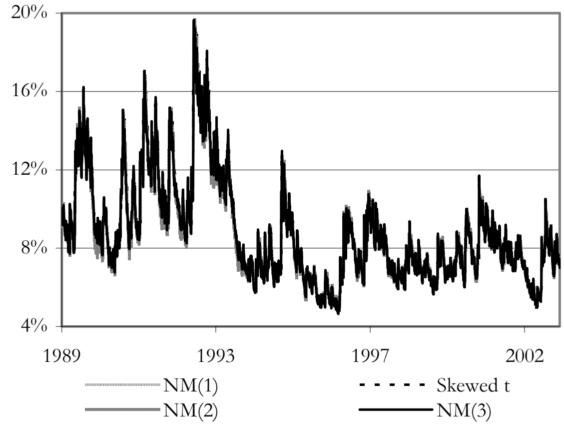


80%
60%
40%
20%
0%
-20%
-40%
-60%
-80%
base
omega1 up
alpha1 up
beta1 up
omega2 up
alpha2 up
beta2 up
p
omega1
alpha1
beta1
omega2
alpha2
beta2

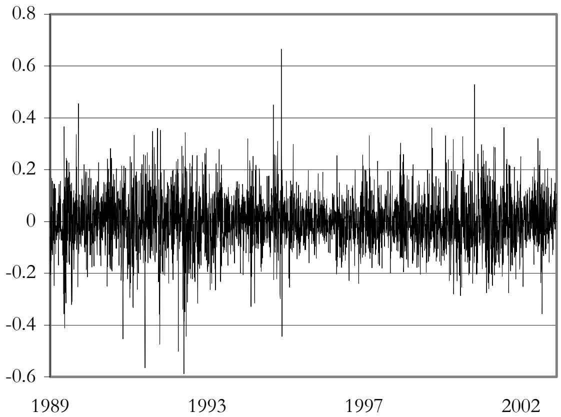
Fig. 6.



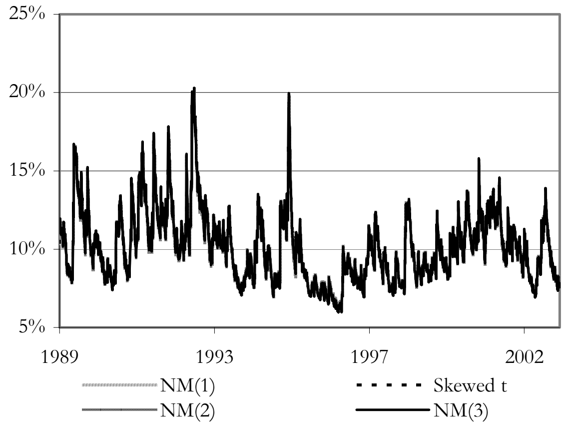
(a) GBP returns



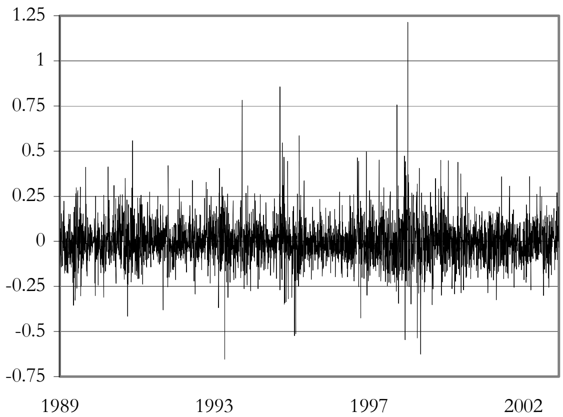
(b) Conditional volatility for GBP



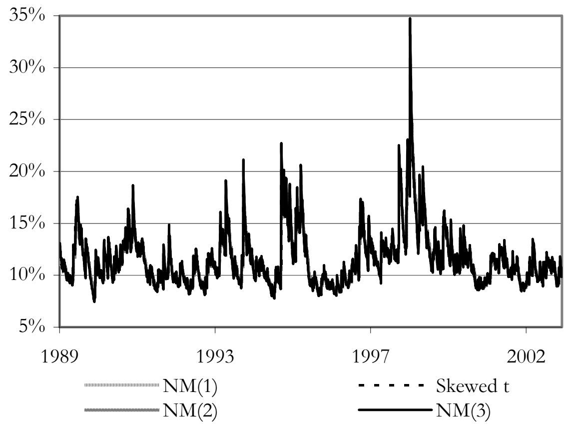
(c) EUR returns



(d) Conditional volatility for EUR

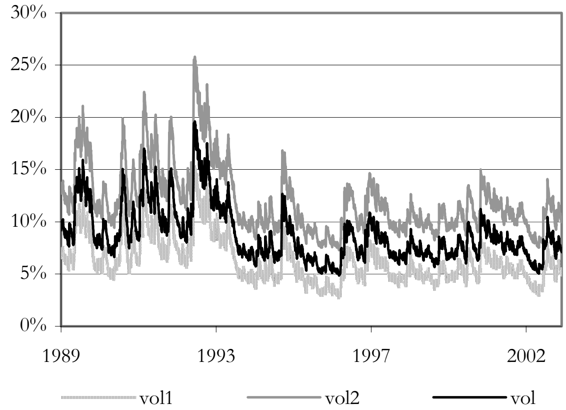


(e) JPY returns

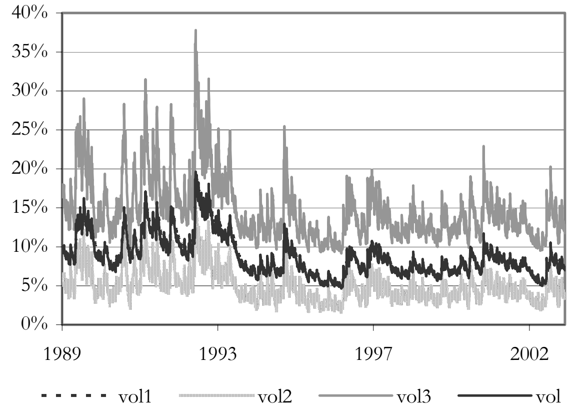


(f) Conditional volatility for JPY

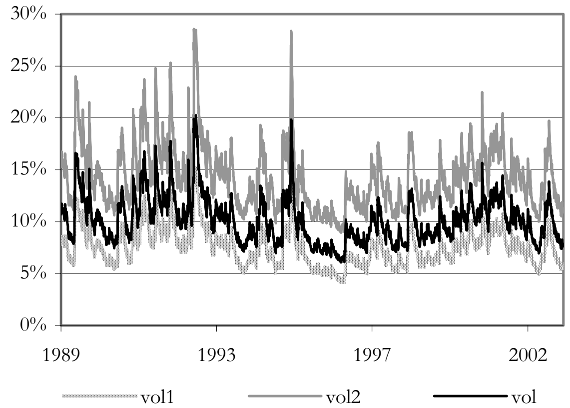
Fig. 7.



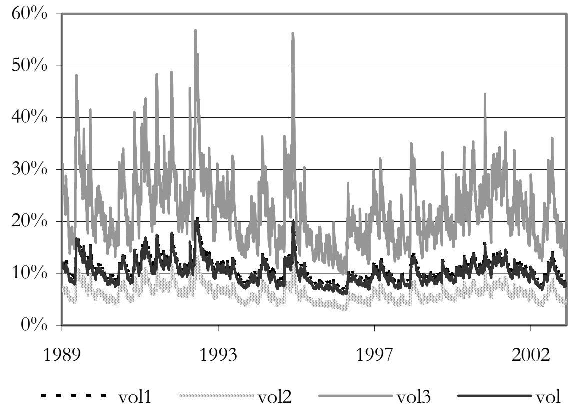
(a) Conditional NM(2) volatilities for GBP



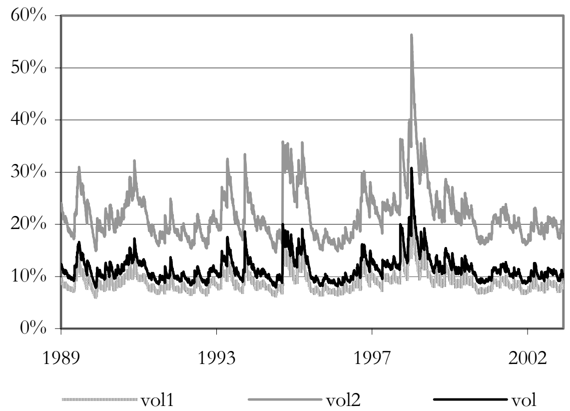
(b) Conditional NM(3) volatilites for GBP



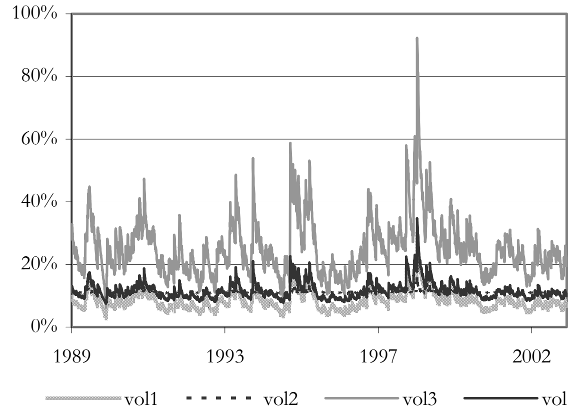
(c) Conditional NM(2) volatilities for EUR



(d) Conditional NM(3) volatilites for EUR

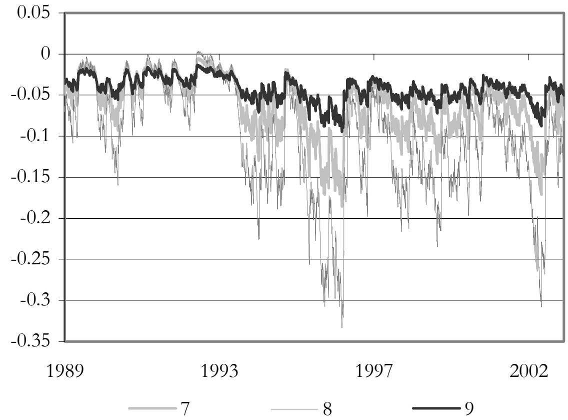


(e) Conditional NM(2) volatilities for JPY

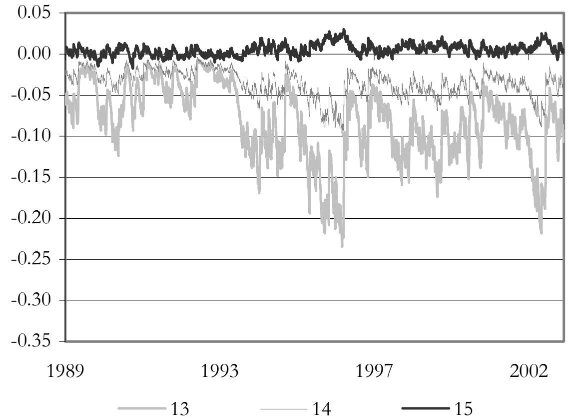


(f) Conditional NM(3) volatilites for JPY

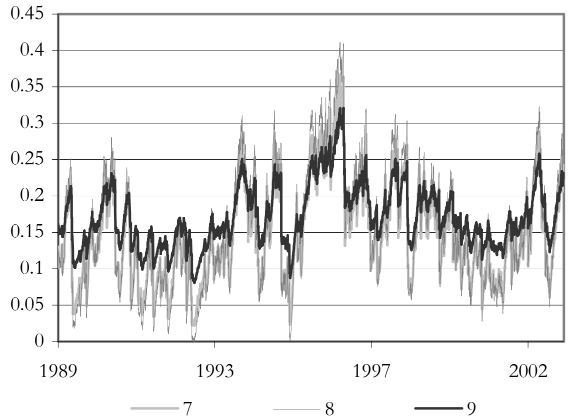
Fig. 8.



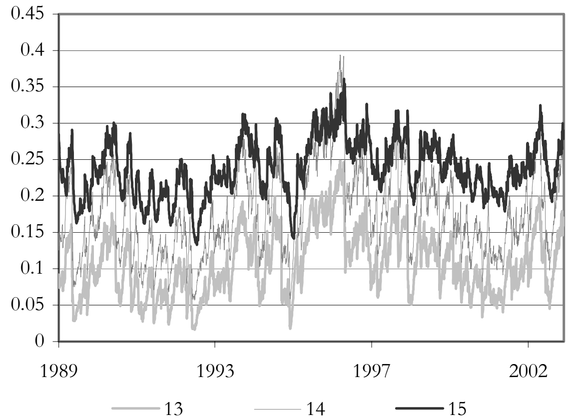
(a) Conditional NM(2) skewness for GBP



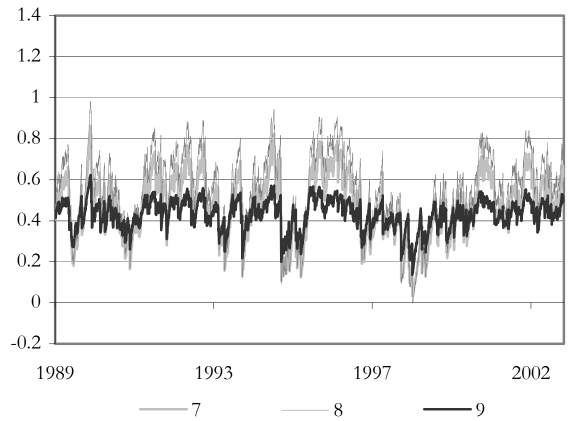
(b) Conditional NM(3) skewness for GBP



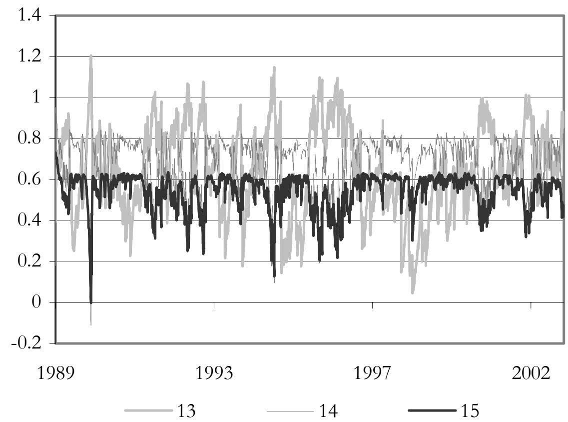
(c) Conditional NM(2) skewness for EUR



(d) Conditional NM(3) skewness for EUR

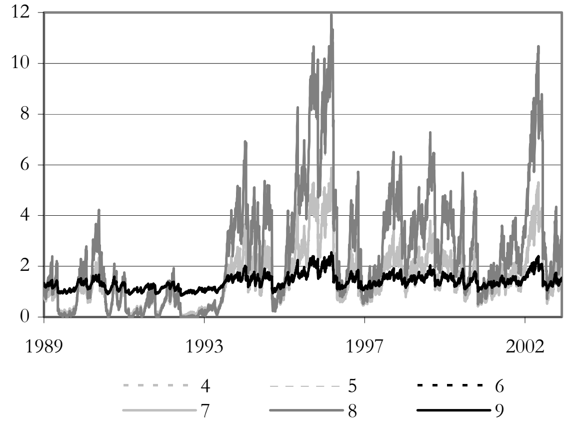


(e) Conditional NM(2) skewness for JPY

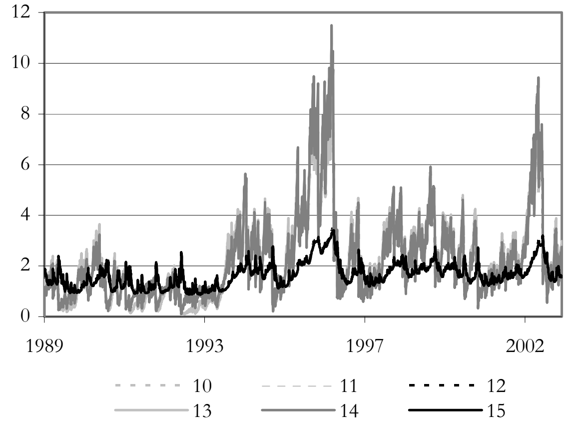


(f) Conditional NM(3) skewness for JPY

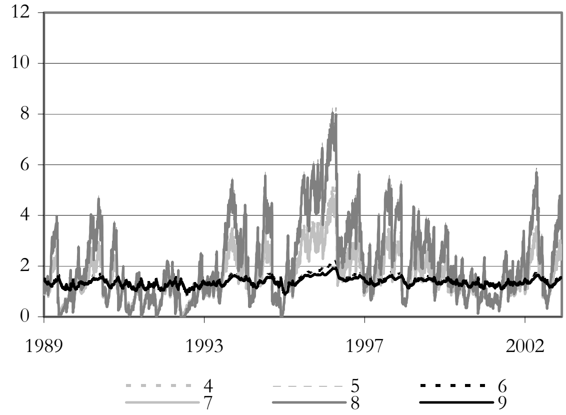
Fig. 9.



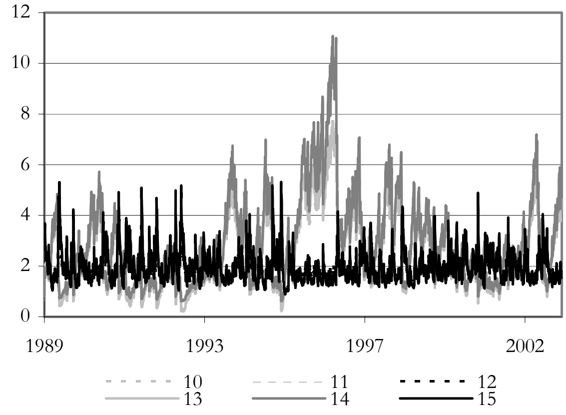
(a) Conditional NM(2) excess kurtosis for GBP



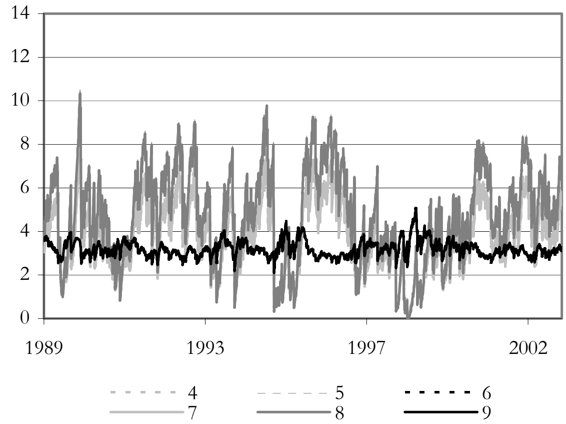
(b) Conditional NM(3) excess kurtosis for GBP



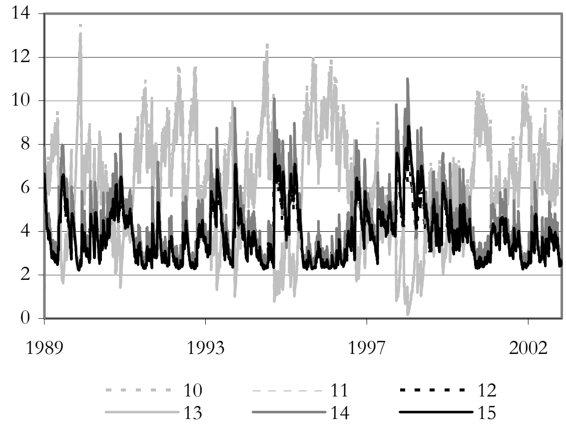
(c) Conditional NM(2) excess kurtosis for EUR



(d) Conditional NM(3) excess kurtosis for EUR

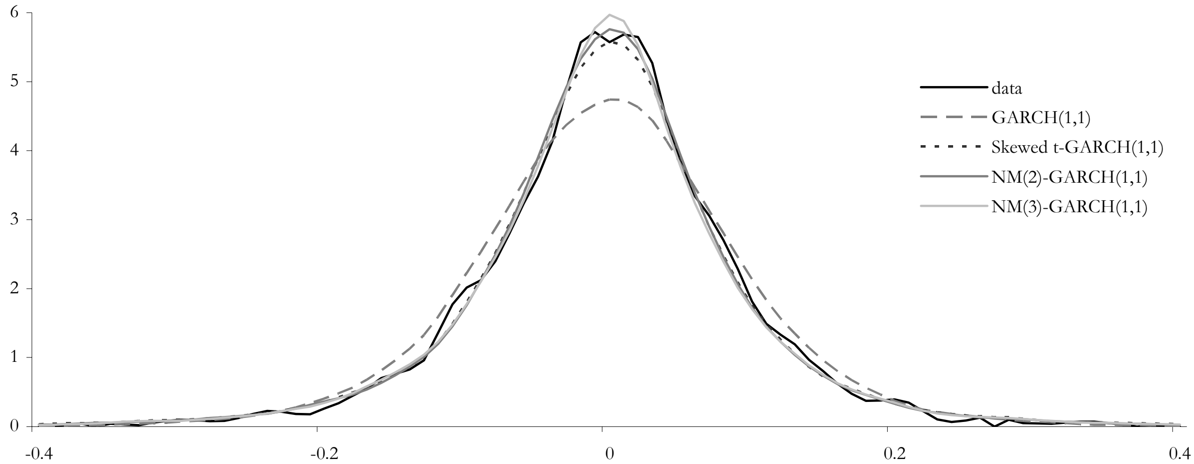


(e) Conditional NM(2) excess kurtosis for JPY

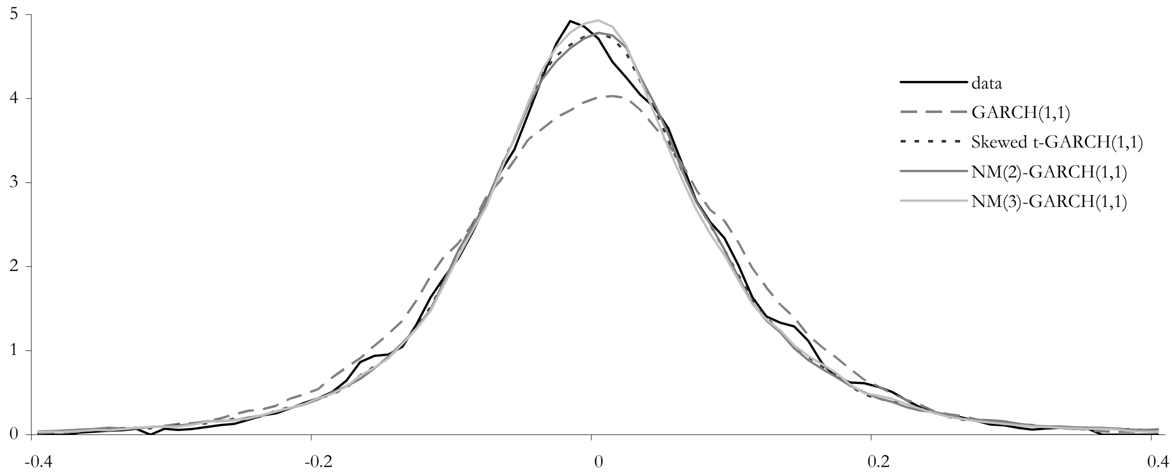


(f) Conditional NM(3) excess kurtosis for JPY

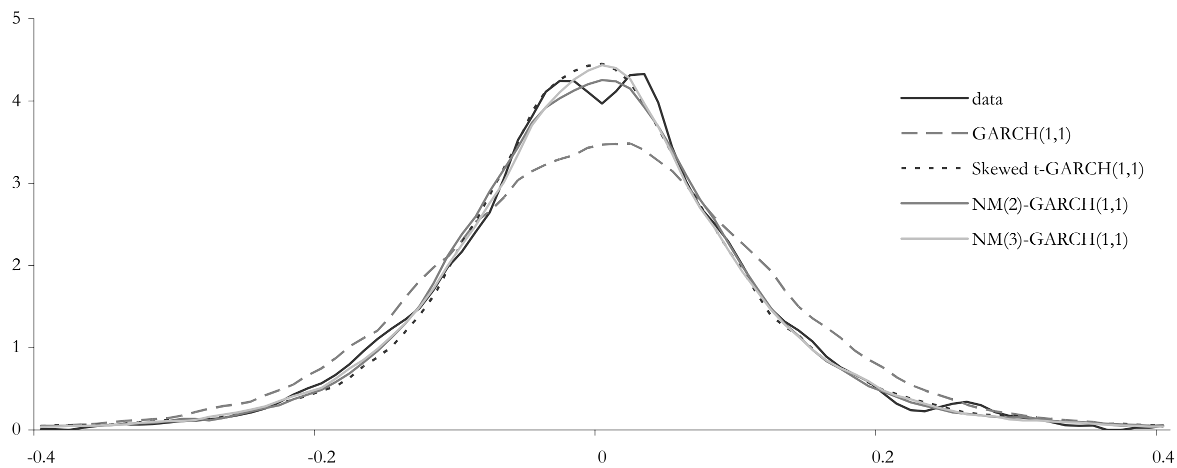
Fig. 10.



(a) Histogram fitting for GBP



(b) Histogram fitting for EUR



(c) Histogram fitting for JPY

Table 1.

Returns on exchange rate	**GBP**	**EUR**	**JPY**
Mean return	-0.05%	-0.05%	0.02%
Standard deviation	9.12%	10.35%	11.79%
Skewness	-0.1533***	0.0256	0.7711***
Excess kurtosis	2.7285***	2.4232***	7.0853***
Minimum	-50.89%	*-58.85%*	-65.36%
Maximum	45.35%	66.52%	121.39%
Ljung-Box statistic (4 lags)	2.5708	3.5032	1.1121

Table 2.

Exchange Rate			**GBP**	**EUR**	**JPY**
5%	Right tail	$\hat{a}_s$	3.83	4.45	3.34
		4^{th} moment test	*-0.37*	*1.02*	*-1.5*
	Left tail	$\hat{a}_s$	3.17	3.69	3.76
		4^{th} moment test	*-1.93*	*-0.73*	*-0.57*
1%	Right tail	$\hat{a}_s$	5.52	5.85	4.24
		4^{th} moment test	*1.62*	*1.91*	*0.25*
	Left tail	$\hat{a}_s$	5.37	4.23	3.94
		4^{th} moment test	*1.45*	*0.24*	*-0.06*

Table 3.

(1) Normal GARCH(1,1)

(2) GARCH(1,1) with Student's *t* distributed errors

(3) GARCH(1,1) with Skewed Student's *t* distributed errors

(4) NM(2)-GARCH(1,1) with restrictions:

$\mu_1 = \mu_2 = 0$

$a_1 = a_2$

$\beta_1 = \beta_2$

(5) NM(2)-GARCH(1,1) with restrictions:

$\mu_1 = \mu_2 = 0$

$a_2 = \beta_2 = 0$

(6) NM(2)-GARCH(1,1) with restrictions:

$\mu_1 = \mu_2 = 0$

(7) NM(2)-GARCH(1,1) with restrictions:

$a_1 = a_2$

$\beta_1 = \beta_2$

(8) NM(2)-GARCH(1,1) with restriction:

$a_2 = \beta_2 = 0$

(9) NM(2)-GARCH(1,1) without restrictions

(10) NM(3)-GARCH(1,1) with restrictions:

$\mu_1 = \mu_2 = \mu_3 = 0$

$a_1 = a_2 = a_3$

$\beta_1 = \beta_2 = \beta_3$

(11) NM(3)-GARCH(1,1) with restrictions:

$\mu_1 = \mu_2 = \mu_3 = 0$

$a_3 = \beta_3 = 0$

(12) NM(3)-GARCH(1,1) with restrictions:

$\mu_1 = \mu_2 = \mu_3 = 0$

(13) NM(3)-GARCH(1,1) with restrictions:

$a_1 = a_2 = a_3$

$\beta_1 = \beta_2 = \beta_3$

(14) NM(3)-GARCH(1,1) with restriction:

$a_3 = \beta_3 = 0$

(15) NM(3)-GARCH(1,1) without restriction

Table 4.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
p1				0.7747	0.9093	0.6595	0.7673	0.9088	0.6509	0.5068	0.6007	0.6036	0.5132	0.5954	0.5995
				(18.32)	*(45.50)*	*(12.92)*	*(17.93)*	*(45.56)*	*(12.77)*	*(3.81)*	*(8.98)*	*(7.45)*	*(3.91)*	*(8.76)*	*(7.34)*
μ1				0	0	0	0.0013	0.0010	0.0012	0	0	0	0.0021	-0.0008	-0.0010
							(0.92)	*(0.93)*	*(0.74)*				*(0.89)*	*(-0.37)*	*(-0.37)*
ω1	7.0E-5	4.8E-5	4.8E-5	-8.0E-6	1.1E-5	1.0E-5	-1.0E-5	1.0E-5	9.0E-6	-3.4E-5	2.8E-5	1.4E-5	-3.4E-5	2.8E-5	1.5E-5
	(5.36)	*(2.61)*	*(2.61)*	*(-0.71)*	*(1.01)*	*(0.71)*	*(-0.90)*	*(0.94)*	*(0.64)*	*(-1.98)*	*(1.25)*	*(0.79)*	*(-2.03)*	*(1.23)*	*(0.84)*
α1	0.0439	0.0440	0.0439	0.0373	0.0346	0.0292	0.0372	0.0342	0.0295	0.0374	0.0408	0.0262	0.0371	0.0408	0.0261
	(10.33)	*(6.71)*	*(6.69)*	*(6.03)*	*(7.72)*	*(4.55)*	*(5.98)*	*(7.66)*	*(4.51)*	*(3.16)*	*(5.46)*	*(3.70)*	*(3.21)*	*(5.40)*	*(3.69)*
β1	0.9476	0.9513	0.9514	0.9479	0.9548	0.9434	0.9479	0.9554	0.9424	0.9461	0.9650	0.9722	0.9464	0.9651	0.9724
	(191.59)	*(134.49)*	*(134.39)*	*(116.75)*	*(166.91)*	*(82.65)*	*(115.84)*	*(168.91)*	*(80.62)*	*(63.10)*	*(151.85)*	*(147.49)*	*(64.08)*	*(150.96)*	*(147.74)*
p2 (d.f. for (2),(3))		6.0897	6.0784	0.2253	0.0907	0.3405	0.2327	0.0912	0.3491	0.4376	0.3716	0.2936	0.4329	0.3772	0.2988
		(9.23)	*(9.23)*							*(3.79)*	*(5.53)*	*(3.75)*	*(3.77)*	*(5.53)*	*(3.80)*
μ2 (Asym. for (3))			-0.0113	0	0	0	-0.0044	-0.0102	-0.0023	0	0	0	-0.0014	0.0016	0.0014
			(-0.51)										*(-0.34)*	*(0.66)*	*(0.46)*
ω2				0.0005	0.0245	0.0002	0.0005	0.0245	0.0002	1.7E-4	1.4E-5	2.0E-6	1.7E-4	1.5E-5	4.0E-6
				(1.75)	*(0.02)*	*(2.46)*	*(1.79)*	*(0.02)*	*(2.52)*	*(1.15)*	*(0.46)*	*(0.06)*	*(1.14)*	*(0.49)*	*(0.11)*
α2				0.0373	0	0.0786	0.0372	0	0.0766	0.0374	0.0425	0.0487	0.0371	0.0422	0.0488
				(6.03)		*(3.77)*	*(5.98)*		*(3.81)*	*(3.16)*	*(3.16)*	*(2.68)*	*(3.21)*	*(3.17)*	*(2.69)*
β2				0.9479	0	0.9465	0.9479	0	0.9471	0.9461	0.8779	0.8575	0.9464	0.8792	0.8575
				(116.75)		*(68.06)*	*(115.84)*	0.9088	*(69.59)*	*(63.10)*	*(23.54)*	*(16.38)*	*(64.08)*	*(23.79)*	*(16.42)*
p3										0.0555	0.0277	0.1029	0.0540	0.0275	0.1017
μ3										0	0	0	-0.0091	-0.0037	0.0021
ω 3										0.0012	0.0305	0.0008	0.0013	0.0307	0.0008
										(0.41)	*(0.01)*	*(1.02)*	*(0.40)*	*(0.01)*	*(1.02)*
α3										0.0374	0	0.2265	0.0371	0	0.2310
										(3.16)		*(1.13)*	*(3.21)*		*(1.13)*
β3										0.9461	0	0.8861	0.9464	0	0.8852
										(63.10)		*(11.92)*	*(64.08)*		*(11.86)*
Unconditional σ	9.05%	10.13%	10.13%	8.67%	8.97%	9.13%	8.69%	9.00%	9.13%	8.72%	8.96%	9.03%	8.75%	8.96%	9.03%
Unconditional σ1				7.23%	8.01%	6.69%	7.22%	8.03%	6.65%	6.81%	10.08%	9.06%	6.83%	10.11%	9.08%
Unconditional σ2				12.41%	15.64%	12.56%	12.34%	15.64%	12.49%	9.19%	5.39%	5.29%	9.25%	5.41%	5.31%
Unconditional σ3										16.80%	17.47%	15.22%	16.96%	17.53%	15.25%
Unconditional *τ*	0	0	-0.0276	0	0	0	-0.0472	-0.0693	-0.0348	0	0	0	-0.0634	-0.0267	0.0043
Unconditional *k*	0.8763	-1031.52	1.79	1.4234	1.6686	3.5680	1.3978	1.6415	3.4614	1.9519	2.4837	4.3112	1.9557	2.4784	4.3903

(continuation of Table 4.)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Loglikelihood	3740.6	3823.2	3823.3	3812.4	3801.9	3824.3	3812.9	3802.3	3824.6	3817.8	3830.9	3835.9	3818.4	3831.2	3836.0
AIC	-2.106	-2.152	-2.151	-2.145	-2.139	-2.151	-2.145	-2.139	-2.150	-2.147	-2.153	-2.155	-2.146	-2.152	-2.154
BIC	-2.100	-2.145	-2.142	-2.136	-2.130	-2.138	-2.134	-2.128	-2.136	-2.135	-2.138	-2.136	-2.130	-2.133	-2.131
Likelihood Ratio				23.83**	44.90**	0.55	23.49**	44.55**		36.28**	9.96**	0.22	35.31**	9.72**	
Moment specification tests:															
1st moment	0.51	0.54	0.47	0.48	0.42	0.59	0.33	0.25	0.47	0.48	0.49	0.58	0.31	0.39	0.56
2nd moment	0.80	1.55	1.12	0.00	14.94**	0.20	0.18	17.11**	0.20	1.32	0.74	1.87	0.59	0.61	2.26
3rd moment	0.67	0.30	0.82	0.02	0.00	0.51	0.40	0.18	1.21	0.02	0.01	0.46	0.47	0.17	0.51
4th moment	43.07**	1.18	1.34	140.5**	22.46**	7.49**	121.8**	21.47**	7.64**	156.3**	4.13*	0.09	136.0**	3.36	0.07
1st moment AC(1)	0.01	0.20	0.21	0.13	0.10	0.21	0.14	0.12	0.22	0.15	0.17	0.23	0.18	0.18	0.23
1st moment AC(2)	0.48	0.33	0.30	0.39	0.39	0.36	0.35	0.34	0.33	0.37	0.37	0.34	0.33	0.34	0.33
1st moment AC(3)	1.20	2.18	2.21	2.09	2.10	2.08	2.13	2.14	2.11	2.24	2.04	2.29	2.30	2.07	2.31
1st moment AC(4)	0.79	0.36	0.36	0.44	0.50	0.35	0.42	0.46	0.34	0.43	0.43	0.41	0.41	0.43	0.43
2ndmoment AC(1)	0.55	1.60	1.57	1.49	4.16*	1.01	1.45	4.08*	0.96	1.45	1.44	0.46	1.42	1.40	0.43
2ndmoment AC(2)	0.72	0.05	0.06	0.00	0.14	0.29	0.01	0.15	0.31	0.00	0.10	0.54	0.00	0.10	0.56
2ndmoment AC(3)	0.62	0.08	0.06	0.71	0.87	0.04	0.59	0.74	0.03	0.65	0.46	0.04	0.50	0.43	0.03
2ndmoment AC(4)	1.27	0.58	0.62	0.16	0.02	0.96	0.21	0.04	1.05	0.11	0.52	1.45	0.15	0.55	1.45
3rd moment AC(1)	5.75*	4.28*	4.23*	5.05*	6.35*	4.62*	4.97*	5.90*	4.61*	4.69*	6.26*	4.05*	4.55*	6.16*	4.13*
3rd moment AC(2)	0.75	1.70	1.65	1.21	1.41	1.33	1.15	1.35	1.29	1.27	1.69	1.21	1.19	1.67	1.23
3rd moment AC(3)	1.73	0.77	0.75	0.73	0.64	0.85	0.70	0.61	0.86	0.70	0.89	0.40	0.62	0.88	0.40
3rd moment AC(4)	3.64	3.52	3.49	3.73	3.88*	3.64	3.57	3.62	3.55	3.72	4.02*	3.92*	3.54	3.99*	4.02*
4th moment AC(1)	0.98	0.00	0.01	0.04	0.63	0.23	0.01	0.38	0.31	0.11	0.06	0.30	0.06	0.03	0.31
4th moment AC(2)	4.60*	1.96	2.06	0.04	0.25	2.86	0.06	0.24	3.03	0.00	0.27	2.81	0.01	0.29	2.86
4th moment AC(3)	1.45	1.11	1.07	2.19	1.59	0.90	2.14	1.59	0.88	3.63	1.74	0.56	3.55	1.73	0.56
4th moment AC(4)	4.02*	2.41	2.74	0.40	0.10	3.56	0.64	0.24	4.11*	0.23	0.82	4.78*	0.41	0.98	4.72*
Cumulative test	76.67**	71.34**	74.46**	196.9**	83.29	61.11**	175.8**	81.29	61.52**	207.1**	49.12**	56.62**	196.3**	48.36**	57.79**
Rejections at 5%	4	1	1	2	5	2	2	4	3	2	3	3	2	2	3
Rejections at 1%	1	0	0	1	2	1	1	2	1	1	0	0	1	0	0
Density	2.6186	1.1047	1.0312	1.3112	1.9210	0.9672	1.1873	1.7971	0.8075	1.2421	1.1715	1.1786	1.1289	1.0059	1.0962
ACF	0.0833	2.5289	0.7558	0.1545	0.2035	0.1081	0.1568	0.2042	0.0999	0.1883	0.0437	0.1222	0.1910	0.0438	0.1267
VaR ranking	14	13	12	7	8	10	3	9	11	4	15	4	6	1	2

Table 5.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
p1				0.8076	0.8904	0.7299	0.7855	0.8831	0.6927	0.5527	0.5287	0.5635	0.5438	0.5553	0.5881
				(19.88)	*(37.68)*	*(15.07)*	*(18.56)*	*(36.02)*	*(14.36)*	*(4.22)*	*(5.05)*	*(4.48)*	*(5.01)*	*(6.27)*	*(5.73)*
μ1				0	0	0	-0.0036	-0.0023	-0.0055	0	0	0	-0.0070	0.0064	0.0039
							(-1.98)	*(-1.56)*	*(-2.66)*				*(-2.38)*	*(1.87)*	*(0.91)*
ω1	1.5E-4	1.3E-4	1.3E-4	5.0E-5	7.3E-5	4.7E-5	4.4E-5	7.1E-5	4.6E-5	-8.0E-6	2.3E-4	1.7E-4	-8.0E-6	2.0E-4	1.6E-4
	(4.84)	*(3.03)*	*(2.99)*	*(1.73)*	*(2.50)*	*(1.69)*	*(1.53)*	*(2.43)*	*(1.61)*	*(-0.24)*	*(2.15)*	*(1.92)*	*(-0.27)*	*(2.27)*	*(2.10)*
α1	0.0473	0.0460	0.0463	0.0366	0.0364	0.0256	0.0371	0.0370	0.0248	0.0362	0.0578	0.0440	0.0358	0.0560	0.0452
	(9.69)	*(6.41)*	*(6.47)*	*(5.57)*	*(6.68)*	*(4.43)*	*(5.47)*	*(6.67)*	*(4.20)*	*(3.50)*	*(4.12)*	*(3.17)*	*(3.53)*	*(4.44)*	*(3.51)*
β1	0.9393	0.9427	0.9429	0.9399	0.9427	0.9495	0.9385	0.9418	0.9481	0.9386	0.9415	0.9511	0.9378	0.9437	0.9499
	(145.25)	*(105.36)*	*(106.48)*	*(88.60)*	*(108.40)*	*(87.24)*	*(84.65)*	*(106.25)*	*(79.80)*	*(58.45)*	*(70.84)*	*(77.08)*	*(57.87)*	*(78.92)*	*(81.79)*
p2 (d.f. for (2),(3))		6.5384	6.5239	0.1924	0.1096	0.2701	0.2145	0.1169	0.3073	0.4264	0.4587	0.4035	0.4362	0.4318	0.3818
		(9.14)	*(9.07)*							*(3.63)*	*(4.19)*	*(2.72)*	*(4.45)*	*(4.65)*	*(3.21)*
μ2 (Asym. for (3))			0.0641	0	0	0	0.0132	0.0174	0.0124	0	0	0	0.0086	-0.0089	-0.0096
			(2.82)										*(1.69)*	*(-2.88)*	*(-2.73)*
ω2				0.0011	0.0322	0.0005	0.0010	0.0306	0.0004	4.6E-4	2.7E-5	2.7E-5	4.7E-4	2.7E-5	2.4E-5
				(0.84)	*(0.02)*	*(1.75)*	*(0.94)*	*(0.02)*	*(1.88)*	*(1.40)*	*(0.79)*	*(0.67)*	*(1.48)*	*(0.79)*	*(0.64)*
α2				0.0366	0	0.1124	0.0371	0	0.1066	0.0362	0.0219	0.0236	0.0358	0.0217	0.0224
				(5.57)		*(3.04)*	*(5.47)*		*(3.46)*	*(3.50)*	*(2.64)*	*(2.39)*	*(3.53)*	*(2.61)*	*(2.37)*
β2				0.9399	0	0.9241	0.9385	0	0.9272	0.9386	0.9430	0.9359	0.9378	0.9392	0.9348
				(88.60)		*(35.81)*	*(84.65)*		*(42.64)*	*(58.45)*	*(48.35)*	*(37.68)*	*(57.87)*	*(43.99)*	*(36.61)*
p3										0.0209	0.0126	0.0330	0.0200	0.0129	0.0301
μ3										0	0	0	0.0031	0.0243	0.0444
ω 3										0.0035	0.0768	0.0009	0.0036	0.0751	0.0004
										(0.05)	*(0.00)*	*(0.39)*	*(0.05)*	*(0.00)*	*(0.20)*
α3										0.0362	0	0.3958	0.0358	0	0.5421
										(3.50)		*(0.41)*	*(3.53)*		*(0.45)*
β3										0.9386	0	0.9000	0.9378	0	0.8848
										(58.45)		*(4.75)*	*(57.87)*		*(4.76)*
Unconditional σ	10.60%	10.76%	10.81%	10.22%	10.36%	10.54%	10.18%	10.33%	10.57%	10.23%	10.44%	10.56%	10.19%	10.44%	10.68%
Unconditional σ1				8.48%	8.99%	8.11%	8.35%	8.94%	7.88%	7.77%	12.14%	11.59%	7.64%	11.99%	11.62%
Unconditional σ2				15.53%	17.95%	15.29%	15.02%	17.50%	14.87%	11.69%	6.84%	6.73%	11.61%	6.59%	6.55%
Unconditional σ3										25.00%	27.72%	23.14%	25.31%	27.41%	23.94%
Unconditional *τ*	0	0	0.1394	0	0	0	0.1259	0.1257	0.1548	0	0	0	0.0918	0.1521	0.2290
Unconditional *k*	0.6081	6.0879	-0.0153	1.5166	1.7421	2.8294	1.3976	1.6227	2.8857	2.2959	2.8074	4.7447	2.3026	2.7501	6.808

(continuation of Table 5.)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Loglikelihood	3098.1	3175.3	3179.5	3165.8	3160.5	3172.8	3168.0	3161.9	3176.6	3172.6	3177.4	3179.6	3175.9	3182.1	3185.2
AIC	-1.774	-1.817	-1.819	-1.811	-1.808	-1.814	-1.812	-1.809	-1.816	-1.814	-1.816	-1.816	-1.815	-1.817	-1.818
BIC	-1.768	-1.810	-1.810	-1.803	-1.799	-1.802	-1.801	-1.798	-1.802	-1.802	-1.800	-1.796	-1.799	-1.798	-1.795
Likelihood Ratio				13.98**	24.56**	7.51**	17.20**	29.40**		13.96**	4.35	11.23**	18.51**	6.24*	
Moment specification tests:															
1st moment	0.08	3.5E-5	0.09	1.4E-3	0.00	7.2E-4	0.05	0.04	0.10	2.8E-3	2.3E-3	8.4E-6	0.03	0.07	0.15
2nd moment	0.46	1.82	2.27	2.05	16.28**	0.38	0.62	13.37**	0.00	0.98	0.75	2.46	2.50	0.98	2.64
3rd moment	3.25	2.70	0.05	1.98	1.45	4.19*	0.07	0.06	0.26	2.08	2.72	4.28*	0.08	0.01	0.03
4th moment	27.05**	0.01	2.4E-3	27.93**	19.44**	9.05**	33.08**	24.05**	9.60**	60.17**	1.53	3.01	40.81**	0.02	0.49
1st moment AC(1)	0.86	0.47	0.58	0.41	0.40	0.48	0.48	0.45	0.58	0.42	0.39	0.44	0.48	0.48	0.58
1st moment AC(2)	2.03	1.19	1.17	1.23	1.27	1.20	1.21	1.26	1.19	1.16	1.20	1.14	1.14	1.19	1.11
1st moment AC(3)	0.00	0.07	0.11	0.04	0.03	0.06	0.06	0.05	0.09	0.05	0.06	0.05	0.08	0.09	0.09
1st moment AC(4)	0.17	0.08	0.08	0.09	0.11	0.07	0.10	0.12	0.08	0.09	0.07	0.07	0.11	0.07	0.08
2ndmoment AC(1)	0.53	0.36	0.29	0.22	0.11	0.72	0.19	0.07	0.73	0.20	0.35	0.35	0.16	0.27	0.39
2ndmoment AC(2)	0.33	0.87	0.67	0.27	0.13	1.23	0.15	0.05	1.02	0.28	0.43	0.82	0.23	0.32	0.77
2ndmoment AC(3)	0.27	0.12	0.19	0.68	1.02	0.10	0.80	1.15	0.15	0.67	0.31	0.20	0.86	0.41	0.21
2ndmoment AC(4)	0.86	1.61	1.46	1.92	2.13	1.39	1.68	1.87	1.17	2.34	1.73	1.99	2.38	1.61	1.64
3rd moment AC(1)	1.31	2.07	2.18	1.11	0.89	3.09	1.11	0.88	3.04	1.03	1.31	2.26	1.04	1.28	2.45
3rd moment AC(2)	3.99*	4.82*	4.63*	4.09*	4.03*	4.36*	3.84*	3.76	4.14*	5.11*	5.37*	5.07*	5.04*	5.15*	4.67*
3rd moment AC(3)	2.52	1.42	1.21	1.72	1.66	1.69	1.65	1.63	1.57	1.55	1.59	1.69	1.47	1.50	1.45
3rd moment AC(4)	0.05	0.69	0.60	0.32	0.46	0.44	0.24	0.40	0.32	0.52	0.57	0.70	0.44	0.46	0.71
4th moment AC(1)	2.29	0.42	0.31	0.11	0.02	1.85	0.08	0.10	1.98	0.00	0.22	0.49	0.00	0.12	0.57
4th moment AC(2)	0.01	0.00	0.11	0.40	0.69	0.23	0.67	0.97	0.03	0.63	0.34	0.00	0.72	0.53	0.07
4th moment AC(3)	0.58	0.46	0.62	1.01	1.20	0.36	1.07	1.29	0.45	1.24	0.84	0.61	1.41	0.98	0.68
4th moment AC(4)	0.30	0.40	0.34	0.21	0.16	0.13	0.13	0.07	0.09	0.60	0.30	0.55	0.62	0.28	0.37
Cumulative test	49.22**	29.44	21.80	57.02**	54.62	47.41**	62.80**	45.10**	30.60	124.5**	78.81**	155.6*	78.43**	22.71	30.29
Rejections at 5%	2	1	1	2	3	3	2	2	2	2	1	2	2	1	1
Rejections at 1%	1	0	0	1	2	1	1	2	1	1	0	0	1	0	0
Density	2.1097	1.0569	0.6976	0.9966	0.9624	1.0435	0.8685	1.0883	0.6759	1.0238	1.0947	1.1037	0.7397	0.8696	1.0024
ACF	0.1741	1.5414	0.0514	0.0428	0.0525	0.1455	0.0436	0.0535	0.1813	0.0464	0.0475	0.1744	0.0492	0.0504	0.2549
VaR ranking	15	11	3	9	6	12	4	7	5	10	13	14	8	1	2

Table 6.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
p1				0.8927 *(52.02)*	0.9156 *(67.62)*	0.8734 *(42.16)*	0.8930 *(53.83)*	0.9142 *(68.53)*	0.8766 *(44.59)*	0.6353 *(6.23)*	0.5992 *(8.49)*	0.5469 *(4.83)*	0.6402 *(6.30)*	0.6068 *(8.81)*	0.5608 *(6.06)*
μ1				0	0	0	-0.0057 *(-3.87)*	-0.0054 *(-3.87)*	-0.0060 *(-3.95)*	0	0	0	-0.0069 *(-2.70)*	-0.0047 *(-1.54)*	-0.0045 *(-1.37)*
ω1	2.4E-4 *(7.55)*	2.6E-4 *(3.34)*	2.5E-4 *(3.34)*	1.5E-4 *(3.25)*	1.8E-4 *(3.60)*	1.7E-4 *(3.32)*	1.5E-4 *(3.29)*	1.8E-4 *(3.61)*	1.7E-4 *(3.35)*	4.6E-5 *(0.92)*	-3.1E-5 *(-0.63)*	3.0E-6 *(0.05)*	4.6E-5 *(0.91)*	-3.6E-5 *(-0.76)*	-4.4E-5 *(-0.88)*
α1	0.0387 *(13.07)*	0.0390 *(5.17)*	0.0386 *(5.19)*	0.0287 *(5.75)*	0.0277 *(5.75)*	0.0271 *(5.36)*	0.0287 *(5.72)*	0.0276 *(5.71)*	0.0270 *(5.34)*	0.0292 *(3.97)*	0.0414 *(5.12)*	0.0405 *(4.12)*	0.0295 *(4.00)*	0.0411 *(5.35)*	0.0403 *(4.39)*
β1	0.9448 *(220.46)*	0.9426 *(86.88)*	0.9431 *(88.07)*	0.9384 *(88.95)*	0.9385 *(88.81)*	0.9360 *(79.96)*	0.9380 *(88.05)*	0.9380 *(87.75)*	0.9358 *(79.52)*	0.9368 *(62.16)*	0.9341 *(81.47)*	0.9226 *(52.05)*	0.9365 *(62.03)*	0.9365 *(87.25)*	0.9366 *(72.84)*
p2 (d.f. for (2),(3))		4.8394 *(12.52)*	4.9109 *(12.10)*	0.1073	0.0844	0.1266	0.1070	0.0858	0.1234	0.3405 *(3.58)*	0.0607 *(3.37)*	0.3991 *(3.63)*	0.3338 *(3.51)*	0.0523 *(4.22)*	0.3778 *(4.23)*
μ2 (Asym. for (3))			0.0737 *(3.41)*	0	0	0	0.0474	0.0571	0.0425	0	0	0	0.0047 *(0.68)*	0.1193 *(4.25)*	-0.0075 *(-1.07)*
ω2				0.0029 *(0.46)*	0.0657 *(0.02)*	0.0005 *(1.32)*	0.0028 *(0.48)*	0.0617 *(0.01)*	0.0004 *(1.23)*	7.8E-4 *(1.19)*	5.3E-5 *(0.04)*	2.4E-4 *(1.00)*	7.7E-4 *(1.18)*	-7.7E-4 *(-2.96)*	3.8E-3 *(0.50)*
α2				0.0287 *(5.75)*	0	0.1141 *(2.55)*	0.0287 *(5.72)*	0	0.1186 *(2.49)*	0.0292 *(3.97)*	0.3208 *(2.36)*	0.0078 *(0.80)*	0.0295 *(4.00)*	0.4821 *(3.16)*	0.0206 *(0.61)*
β2				0.9384 *(88.95)*	0	0.9599 *(56.05)*	0.9380 *(88.05)*	0	0.9587 *(55.46)*	0.9368 *(62.16)*	0.9417 *(38.23)*	0.9756 *(39.64)*	0.9365 *(62.03)*	0.9237 *(33.42)*	0.6746 *(1.10)*
p3										0.0242	0.3400	0.0540	0.0260	0.3409	0.0613
μ3										0	0	0	0.1103	-0.0099	0.0875
ω 3										0.0073 *(0.05)*	0.0131 *(0.01)*	0.0003 *(0.14)*	0.0062 *(0.07)*	0.0129 *(0.01)*	-0.0005 *(-0.36)*
α3										0.0292 *(3.97)*	0	0.3952 *(2.27)*	0.0295 *(4.00)*	0	0.3903 *(2.78)*
β3										0.9368 *(62.16)*	0	0.9300 *(28.39)*	0.9365 *(62.03)*	0	0.9321 *(38.36)*
Unconditional σ	11.93%	11.82%	11.80%	11.66%	11.85%	12.01%	11.64%	11.83%	11.98%	11.76%	12.07%	12.05%	11.74%	12.45%	12.36%
Unconditional σ1				9.39%	9.63%	9.37%	9.36%	9.58%	9.35%	8.44%	9.32%	8.74%	8.44%	9.72%	9.48%
Unconditional σ2				23.08%	25.63%	23.11%	22.59%	24.85%	22.83%	13.66%	28.47%	11.96%	13.58%	29.64%	11.25%
Unconditional σ3										34.97%	11.43%	29.31%	32.28%	11.34%	28.37%
Unconditional *τ*	0	0	0.2518	0	0	0	0.4145	0.4768	0.4023	0	0	0	0.5708	0.7874	0.6156
Unconditional *k*	0.3005	13.3912	4.0507	3.2773	3.9239	5.3516	2.8279	3.3082	4.8922	5.3013	14.2563	14.1908	3.8622	34.4349	25.9532

(continuation of Table 6.)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Loglikelihood	2632.5	2810.5	2816.1	2796.3	2791.4	2806.6	2804.7	2799.6	2815.2	2808.4	2816.2	2817.1	2816.0	2825.7	2826.3
AIC	-1.520	-1.622	-1.624	-1.613	-1.610	-1.618	-1.617	-1.614	-1.622	-1.619	-1.622	-1.622	-1.622	-1.627	-1.626
BIC	-1.514	-1.615	-1.616	-1.604	-1.601	-1.605	-1.607	-1.604	-1.608	-1.606	-1.606	-1.602	-1.606	-1.607	-1.603
Likelihood Ratio				20.41**	30.34**	17.37**	21.10**	31.34**		17.34**	1.75	18.43**	20.52**	1.07	
Moment specification tests:															
1st moment	6.0E-5	0.56	0.06	0.58	0.61	0.58	3.6E-3	0.02	6.9E-4	7.1E-1	6.8E-1	6.4E-1	4.5E-3	2.3E-2	3.6E-4
2nd moment	3.54	0.03	0.16	12.03**	22.50**	7.83**	4.84*	10.27**	3.20	6.13*	2.16	1.64	1.08	0.18	2.06
3rd moment	13.95**	4.52*	0.33	6.98**	6.24*	7.10**	0.39	0.39	0.15	5.53*	5.70*	5.68*	0.12	0.11	0.01
4th moment	30.57**	0.34	0.01	34.39**	46.06**	11.52**	28.54**	33.36**	8.31**	10.30**	1.57	0.53	1.36	0.13	0.07
1st moment AC(1)	0.20	0.36	0.50	0.24	0.21	0.36	0.36	0.32	0.51	0.24	0.32	0.32	0.37	0.45	0.41
1st moment AC(2)	0.07	0.70	0.93	0.63	0.64	0.61	0.94	0.96	0.91	0.79	0.68	0.72	1.09	1.05	1.02
1st moment AC(3)	0.14	0.01	0.00	0.03	0.05	0.00	0.03	0.05	0.00	0.03	0.01	0.01	0.03	0.01	0.00
1st moment AC(4)	0.15	0.01	0.01	0.04	0.05	0.02	0.04	0.05	0.02	0.03	0.02	0.02	0.03	0.03	0.02
2ndmoment AC(1)	1.73	0.21	0.10	3.20	4.92*	0.24	2.03	3.43	0.05	1.39	0.02	0.00	0.71	0.09	0.79
2ndmoment AC(2)	0.42	0.33	0.28	0.92	1.17	0.39	0.88	1.19	0.45	0.83	0.35	0.31	0.72	0.26	0.06
2ndmoment AC(3)	1.26	2.30	2.30	2.34	2.64	2.39	2.43	2.75	2.49	2.89	2.62	2.41	3.05	2.52	2.09
2ndmoment AC(4)	0.03	2.10	2.32	2.45	3.20	1.76	2.73	3.53	2.04	2.93	2.17	2.13	3.04	2.42	1.97
3rd moment AC(1)	0.87	0.29	0.11	6.41*	16.37**	0.16	3.95*	9.19**	0.02	4.57*	0.13	0.19	2.24	0.00	0.03
3rd moment AC(2)	1.89	0.63	0.37	0.56	0.40	1.28	0.18	0.10	0.71	0.25	0.98	0.84	0.04	0.27	0.38
3rd moment AC(3)	2.08	1.36	1.36	0.58	0.24	3.38	0.41	0.13	3.26	0.31	2.22	2.21	0.16	1.42	1.97
3rd moment AC(4)	2.14	2.23	2.21	2.97	2.98	2.58	2.59	2.53	2.45	2.91	2.63	2.64	2.52	2.41	2.39
4th moment AC(1)	1.30	0.67	0.44	17.16**	298.2**	1.45	12.90**	225.3**	0.61	28.99**	0.50	0.41	15.13**	0.00	0.12
4th moment AC(2)	0.00	0.27	0.19	0.83	1.18	0.44	0.49	0.86	0.38	1.44	0.47	0.44	0.88	0.05	0.01
4th moment AC(3)	1.25	0.92	1.02	0.70	0.56	1.79	0.70	0.63	1.84	1.84	2.04	1.95	2.39	2.30	2.06
4th moment AC(4)	1.11	0.06	0.00	0.05	0.01	0.43	0.00	0.02	0.10	0.04	0.09	0.08	0.10	0.00	0.06
Cumulative test	72.73**	50.92**	35.96*	196.8**	668.6**	188.2**	75.81**	462.2**	46.58**	342.8**	217.7**	250.2**	78.46**	34.30*	31.53*
Rejections at 5%	2	1	0	5	6	3	4	4	1	5	1	1	1	0	0
Rejections at 1%	2	0	0	4	4	3	2	4	1	2	0	0	1	0	0
Density	3.0715	1.2032	0.9963	1.2338	1.3992	1.1256	0.8520	1.0280	0.7245	1.0843	1.1126	1.1632	0.8727	0.9598	0.9080
ACF	0.0972	1.5871	0.9413	0.0915	0.0998	0.1118	0.0920	0.1005	0.1069	0.0922	0.1178	0.1117	0.0918	0.1471	0.1593
VaR ranking	15	13	11	10	8	7	2	1	3	11	5	14	9	6	4

Table 7.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
p1				0.7730	0.9063	0.6396	0.7657	0.9056	0.6368	0.5093	0.6056	0.6039	0.5146	0.6009	0.6023
				(18.05)	*(43.96)*	*(12.46)*	*(17.72)*	*(43.84)*	*(12.43)*	*(3.80)*	*(9.07)*	*(7.42)*	*(3.85)*	*(8.87)*	*(7.43)*
μ1				0	0	0	0.0012	0.0009	0.0012	0	0	0	0.0021	-0.0007	-0.0011
							(0.83)	*(0.83)*	*(0.70)*				*(0.86)*	*(-0.33)*	*(-0.40)*
ω1	6.7E-5	4.7E-5	4.7E-5	-9.0E-6	9.0E-6	1.0E-5	-1.0E-5	8.0E-6	9.0E-6	-3.4E-5	2.6E-5	1.4E-5	-3.4E-5	2.6E-5	1.5E-5
	(5.28)	*(2.59)*	*(2.59)*	*(-0.82)*	*(0.85)*	*(0.67)*	*(-0.92)*	*(0.77)*	*(0.61)*	*(-1.99)*	*(1.21)*	*(0.78)*	*(-2.04)*	*(1.19)*	*(0.82)*
α1	0.0427	0.0435	0.0434	0.0368	0.0342	0.0308	0.0367	0.0341	0.0305	0.0372	0.0393	0.0273	0.0370	0.0394	0.0276
	(10.02)	*(6.66)*	*(6.64)*	*(5.94)*	*(7.69)*	*(4.47)*	*(5.88)*	*(7.68)*	*(4.45)*	*(3.15)*	*(5.31)*	*(3.73)*	*(3.18)*	*(5.26)*	*(3.76)*
β1	0.9491	0.9520	0.9521	0.9488	0.9557	0.9392	0.9488	0.9560	0.9396	0.9466	0.9661	0.9716	0.9470	0.9661	0.9716
	(192.78)	*(135.27)*	*(135.25)*	*(117.41)*	*(168.54)*	*(75.20)*	*(116.52)*	*(169.92)*	*(75.44)*	*(63.53)*	*(152.70)*	*(145.02)*	*(64.23)*	*(151.80)*	*(145.61)*
p2 (d.f. for (2),(3))		6.1214	6.1114	0.2270	0.0937	0.3604	0.2343	0.0944	0.3632	0.4340	0.3657	0.2962	0.4286	0.3704	0.3028
		(9.19)	*(9.18)*							*(3.73)*	*(5.45)*	*(3.74)*	*(3.67)*	*(5.44)*	*(3.84)*
μ2 (Asym. for (3))			-0.0100	0	0	0	-0.0040	-0.0088	-0.0020	0	0	0	-0.0017	0.0014	0.0013
			(-0.45)										*(-0.40)*	*(0.59)*	*(0.44)*
ω2				0.0005	0.0236	0.0002	0.0005	0.0236	0.0002	1.7E-4	1.2E-5	3.0E-6	1.7E-4	1.4E-5	4.0E-6
				(1.62)	*(0.02)*	*(2.50)*	*(1.66)*	*(0.02)*	*(2.52)*	*(1.11)*	*(0.40)*	*(0.09)*	*(1.10)*	*(0.46)*	*(0.11)*
α2				0.0368	0	0.0697	0.0367	0	0.0695	0.0372	0.0428	0.0481	0.0370	0.0425	0.0480
				(5.94)		*(3.75)*	*(5.88)*		*(3.73)*	*(3.15)*	*(3.13)*	*(2.68)*	*(3.18)*	*(3.14)*	*(2.71)*
β2				0.9488	0	0.9515	0.9488	0	0.9512	0.9466	0.8769	0.8588	0.9470	0.8781	0.8594
				(117.41)		*(74.83)*	*(116.52)*		*(74.27)*	*(63.53)*	*(23.10)*	*(16.52)*	*(64.23)*	*(23.28)*	*(16.75)*
p3										0.0567	0.0287	0.0999	0.0568	0.0287	0.0949
μ3										0	0	0	-0.0059	-0.0028	0.0030
ω 3										0.0012	0.0300	0.0008	0.0012	0.0301	0.0008
										(0.36)	*(0.01)*	*(0.97)*	*(0.36)*	*(0.01)*	*(0.95)*
α3										0.0372	0	0.2042	0.0370	0	0.2154
										(3.15)		*(1.09)*	*(3.18)*		*(1.06)*
β3										0.9466	0	0.8946	0.9470	0	0.8916
										(63.53)		*(12.15)*	*(64.23)*		*(11.55)*
Unconditional σ	9.03%	10.14%	10.14%	8.66%	8.93%	9.09%	8.67%	8.96%	9.09%	8.71%	8.92%	9.04%	8.72%	8.93%	9.04%
Unconditional σ1				7.22%	7.97%	6.59%	7.20%	8.00%	6.58%	6.81%	10.00%	9.13%	6.82%	10.02%	9.19%
Unconditional σ2				12.36%	15.37%	12.33%	12.29%	15.36%	12.32%	9.18%	5.36%	5.29%	9.22%	5.38%	5.32%
Unconditional σ3										16.58%	17.33%	15.20%	16.57%	17.34%	15.34%
Unconditional *τ*	0	0	-0.0242	0	0	0	-0.0427	-0.0600	-0.0321	0	0	0	-0.0473	-0.0226	0.0086
Unconditional *k*	0.8581	-586.14	1.7216	1.4090	1.6151	3.3547	1.3897	1.5869	3.2713	1.9042	2.4176	4.0986	1.8854	2.4151	4.2291

(continuation of Table 7.)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Loglikelihood	3744.5	3826.0	3826.1	3816.1	3806.0	3827.3	3816.4	3806.4	3827.6	3821.2	3834.6	3838.3	3821.7	3834.8	3838.4
AIC	-2.108	-2.153	-2.153	-2.147	-2.141	-2.152	-2.147	-2.141	-2.152	-2.149	-2.155	-2.156	-2.148	-2.154	-2.155
BIC	-2.103	-2.146	-2.144	-2.138	-2.133	-2.140	-2.136	-2.131	-2.138	-2.137	-2.140	-2.137	-2.132	-2.135	-2.133
Likelihood Ratio				22.53**	42.65**	0.50	22.29**	42.44**		34.22**	7.45*	0.21	33.46**	7.29*	
Moment specification tests:															
1st moment	0.42	0.43	0.37	0.39	0.34	0.47	0.27	0.20	0.37	0.39	0.39	0.46	0.26	0.31	0.46
2nd moment	0.33	1.35	1.46	0.10	12.47**	0.04	0.10	16.11**	0.23	1.09	0.80	1.51	0.60	0.63	1.91
3rd moment	0.70	0.28	0.74	0.04	0.00	0.50	0.41	0.20	1.12	0.03	0.03	0.40	0.40	0.17	0.39
4th moment	42.15**	1.27	1.17	126.8**	25.88**	8.97**	111.98**	22.34**	8.38**	137.49**	3.19	0.03	136.34**	2.53	0.05
1st moment AC(1)	0.00	0.24	0.25	0.17	0.15	0.26	0.19	0.17	0.27	0.20	0.22	0.27	0.23	0.24	0.27
1st moment AC(2)	0.29	0.22	0.20	0.22	0.21	0.23	0.19	0.18	0.20	0.22	0.21	0.23	0.19	0.19	0.23
1st moment AC(3)	1.70	2.56	2.59	2.69	2.78	2.49	2.73	2.81	2.52	2.77	2.64	2.64	2.81	2.67	2.66
1st moment AC(4)	0.81	0.34	0.34	0.45	0.50	0.34	0.42	0.47	0.34	0.43	0.43	0.40	0.42	0.44	0.41
2ndmoment AC(1)	0.60	1.57	1.54	1.44	4.25*	0.94	1.40	4.17*	0.92	1.38	1.46	0.52	1.36	1.43	0.49
2ndmoment AC(2)	1.08	0.17	0.17	0.16	0.00	0.59	0.16	0.00	0.59	0.16	0.43	0.76	0.15	0.43	0.79
2ndmoment AC(3)	0.09	0.00	0.00	0.02	0.12	0.04	0.01	0.07	0.06	0.01	0.01	0.25	0.00	0.01	0.25
2ndmoment AC(4)	0.42	0.24	0.27	0.00	0.12	0.41	0.00	0.08	0.45	0.01	0.04	0.69	0.00	0.06	0.68
3rd moment AC(1)	5.49*	4.06*	4.01*	4.67*	5.72*	4.38*	4.59*	5.36*	4.37*	4.38*	5.72*	3.98*	4.32*	5.66*	4.07*
3rd moment AC(2)	0.19	0.90	0.88	0.19	0.14	0.52	0.16	0.12	0.49	0.21	0.43	0.55	0.20	0.43	0.57
3rd moment AC(3)	0.76	0.26	0.26	0.00	0.02	0.19	0.01	0.03	0.19	0.01	0.07	0.04	0.02	0.06	0.04
3rd moment AC(4)	4.30*	3.95*	3.93*	4.15*	4.05*	4.16*	4.01*	3.85*	4.07*	4.17*	4.30*	4.38*	4.06*	4.27*	4.50*
4th moment AC(1)	0.84	0.00	0.00	0.07	0.85	0.20	0.03	0.55	0.26	0.13	0.13	0.18	0.08	0.10	0.18
4th moment AC(2)	5.15*	2.74	2.81	0.44	0.02	3.63	0.50	0.04	3.76	0.29	1.14	3.13	0.32	1.19	3.18
4th moment AC(3)	0.36	0.31	0.29	0.48	0.79	0.06	0.40	0.67	0.04	0.59	0.84	0.01	0.52	0.81	0.01
4th moment AC(4)	1.36	0.93	1.12	0.11	0.42	1.16	0.05	0.30	1.41	0.14	0.03	1.53	0.08	0.01	1.44
Cumulative test	73.55**	68.77**	72.41**	176.55**	82.02**	59.15**	162.10**	79.99**	59.02**	188.25**	47.82**	52.12**	187.39**	46.73**	53.14**
Rejections at 5%	1	2	2	3	5	3	3	5	3	3	2	2	3	2	2
Rejections at 1%	4	0	0	1	2	1	1	2	1	1	0	0	1	0	0
Density	2.5931	1.0962	1.0315	1.2805	1.8524	0.9339	1.1665	1.7491	0.7814	1.2070	1.1368	1.1284	1.1057	0.9831	1.0677
ACF	0.0814	2.6095	0.7775	0.1407	0.1872	0.1011	0.1428	0.1869	0.0927	0.0927	0.0389	0.1869	0.1713	0.0390	0.1425
VaR ranking	15	14	13	8	9	11	4	10	12	7	3	5	6	1	2

Table 8.

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**
p1				0.7797	0.8736	0.7072	0.7535	0.8664	0.6628	0.5544	0.6640	0.5408	0.5551	0.5555	0.5669
				(17.18)	*(30.97)*	*(13.41)*	*(15.99)*	*(29.58)*	*(12.62)*	*(3.82)*	*(8.62)*	*(3.87)*	*(5.15)*	*(7.13)*	*(5.09)*
μ1				0	0	0	-0.0035	-0.0019	-0.0054	0	0	0	-0.0070	-0.0061	0.0055
							(-1.83)	*(-1.28)*	*(-2.53)*				*(-2.24)*	*(-2.46)*	*(1.09)*
ω1	1.4E-4	1.2E-4	1.1E-4	3.3E-5	5.6E-5	3.6E-5	2.6E-5	5.5E-5	3.4E-5	-1.2E-5	3.2E-5	1.1E-4	-1.1E-5	2.9E-5	1.1E-4
	(4.69)	*(2.89)*	*(2.86)*	*(1.23)*	*(2.06)*	*(1.35)*	*(0.98)*	*(2.03)*	*(1.24)*	*(-0.38)*	*(1.02)*	*(1.39)*	*(-0.41)*	*(0.97)*	*(1.54)*
α1	0.0454	0.0449	0.0452	0.0355	0.0355	0.0260	0.0357	0.0359	0.0251	0.0349	0.0292	0.0359	0.0342	0.0229	0.0376
	(9.42)	*(6.34)*	*(6.36)*	*(5.40)*	*(6.44)*	*(4.36)*	*(5.26)*	*(6.39)*	*(4.08)*	*(3.36)*	*(3.97)*	*(2.62)*	*(3.50)*	*(3.19)*	*(2.91)*
β1	0.9415	0.9449	0.9450	0.9427	0.9451	0.9498	0.9417	0.9444	0.9481	0.9423	0.9443	0.9602	0.9419	0.9461	0.9583
	(146.32)	*(108.13)*	*(108.79)*	*(90.23)*	*(109.85)*	*(85.99)*	*(86.04)*	*(107.57)*	*(77.35)*	*(59.57)*	*(72.85)*	*(80.06)*	*(61.26)*	*(60.88)*	*(83.80)*
p2 (d.f. for (2),(3))		6.7003	6.6652	0.2203	0.1264	0.2928	0.2465	0.1336	0.3372	0.4136	0.2008	0.4016	0.4174	0.4371	0.3765
		(8.78)	*(8.79)*							*(3.27)*	*(2.57)*	*(2.37)*	*(4.17)*	*(4.40)*	*(2.85)*
μ2 (Asym. for (3))			0.0608	0	0	0	0.0106	0.0125	0.0106	0	0	0	0.0105	0.0049	-0.0102
			(2.67)										*(1.73)*	*(0.63)*	*(-2.54)*
ω2				0.0009	0.0292	0.0005	0.0008	0.0281	0.0004	4.0E-4	5.4E-5	2.2E-5	4.2E-4	2.5E-4	1.9E-5
				(0.96)	*(0.03)*	*(1.81)*	*(1.08)*	*(0.02)*	*(1.97)*	*(1.19)*	*(0.15)*	*(0.53)*	*(1.32)*	*(1.26)*	*(0.51)*
α2				0.0355	0	0.0980	0.0357	0	0.0918	0.0349	0.1223	0.0265	0.0342	0.0775	0.0242
				(5.40)		*(3.14)*	*(5.26)*		*(3.58)*	*(3.36)*	*(2.25)*	*(2.40)*	*(3.50)*	*(3.98)*	*(2.34)*
β2				0.9427	0	0.9273	0.9417	0	0.9313	0.9423	0.9406	0.9327	0.9419	0.9370	0.9335
				(90.23)		*(37.69)*	*(86.04)*		*(45.96)*	*(59.57)*	*(45.48)*	*(33.78)*	*(61.26)*	*(63.17)*	*(34.21)*
p3										0.0320	0.1353	0.0576	0.0275	0.0074	0.0566
μ3										0	0	0	-0.0180	0.1716	0.0130
ω 3										0.0024	0.0151	0.0015	0.0026	0.0243	0.0014
										(0.11)	*(0.01)*	*(0.56)*	*(0.09)*	*(0.00)*	*(0.58)*
α3										0.0349	0	0.2943	0.0342	0	0.3152
										(3.36)		*(0.65)*	*(3.50)*		*(0.74)*
β3										0.9423	0	0.8769	0.9419	0	0.8763
										(59.57)		*(6.00)*	*(61.26)*		*(6.63)*
Unconditional σ	10.48%	10.76%	10.82%	10.14%	10.28%	10.42%	10.11%	10.26%	10.43%	10.15%	10.45%	10.39%	10.13%	10.40%	10.42%
Unconditional σ1				8.33%	8.87%	7.97%	8.19%	8.82%	7.70%	7.77%	7.93%	11.19%	7.65%	7.17%	11.11%
Unconditional σ2				14.88%	17.08%	14.74%	14.42%	16.78%	14.30%	11.48%	15.30%	6.77%	11.51%	13.12%	6.50%
Unconditional σ3										21.73%	12.29%	19.52%	22.49%	15.59%	19.68%
Unconditional *τ*	0	0	0.1289	0	0	0	0.1073	0.0946	0.1369	0	0	0	0.0302	0.1669	0.1344
Unconditional *k*	0.5635	5.9667	-0.1398	1.4131	1.5967	2.4260	1.3156	1.5252	2.4500	1.9241	2.9678	3.1484	1.9249	1.8567	3.4350

(continuation of Table 8.)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Loglikelihood	3114.1	3184.9	3188.7	3178.6	3172.8	3184.4	3180.4	3173.7	3187.7	3183.0	3185.0	3189.3	3186.3	3189.9	3194.1
AIC	-1.783	-1.823	-1.824	-1.819	-1.815	-1.821	-1.819	-1.815	-1.822	-1.820	-1.820	-1.821	-1.821	-1.822	-1.823
BIC	-1.778	-1.816	-1.816	-1.810	-1.807	-1.809	-1.809	-1.805	-1.808	-1.808	-1.804	-1.802	-1.805	-1.802	-1.800
Likelihood Ratio				11.66**	23.23**	6.51*	14.65**	28.05**		12.64**	8.62*	9.60**	15.71**	8.46*	
Moment specification tests:															
1st moment	0.08	4.5E-3	0.12	8.4E-5	6.8E-5	3.8E-3	0.05	0.03	0.11	1.1E-4	5.9E-3	3.2E-3	0.03	0.12	0.12
2nd moment	0.47	1.95	2.44	0.52	12.31**	0.08	0.03	10.57	0.07	0.89	0.86	1.70	2.12	1.85	2.38
3rd moment	1.99	2.14	0.01	1.09	0.71	2.71	0.00	0.00	0.04	1.39	3.36	3.06	0.10	0.16	0.06
4th moment	28.45**	0.08	1.4E-2	27.02**	19.17**	6.49*	28.63**	23.14*	7.27**	51.92**	2.82	1.71	28.13**	5.08*	0.63
1st moment AC(1)	0.81	0.46	0.55	0.39	0.38	0.47	0.44	0.41	0.55	0.40	0.43	0.42	0.45	0.53	0.52
1st moment AC(2)	1.87	1.12	1.10	1.11	1.15	1.11	1.10	1.14	1.10	1.08	1.07	1.05	1.06	1.08	1.04
1st moment AC(3)	0.11	0.25	0.31	0.24	0.23	0.26	0.29	0.26	0.32	0.24	0.24	0.21	0.27	0.35	0.28
1st moment AC(4)	0.05	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.01
2ndmoment AC(1)	0.09	0.13	0.09	0.01	0.02	0.30	0.00	0.05	0.28	0.02	0.53	0.13	0.01	0.27	0.14
2ndmoment AC(2)	0.13	0.73	0.57	0.10	0.02	0.89	0.05	0.00	0.71	0.15	0.68	0.79	0.16	0.43	0.81
2ndmoment AC(3)	0.03	0.01	0.00	0.01	0.12	0.11	0.03	0.17	0.06	0.05	0.05	0.02	0.11	0.03	0.00
2ndmoment AC(4)	1.02	1.59	1.45	2.04	2.31	1.36	1.83	2.11	1.18	2.38	2.02	1.97	2.65	0.89	1.75
3rd moment AC(1)	1.03	1.81	1.89	0.85	0.66	2.51	0.81	0.63	2.39	0.85	2.13	2.19	0.84	1.65	2.28
3rd moment AC(2)	3.92*	4.82*	4.63*	3.91*	3.81*	4.23*	3.69	3.60	4.03*	4.88*	4.20*	4.93*	5.01*	3.63	4.85*
3rd moment AC(3)	1.56	0.42	0.29	0.65	0.60	0.55	0.60	0.56	0.47	0.47	0.60	0.65	0.46	0.41	0.52
3rd moment AC(4)	0.02	0.54	0.48	0.12	0.21	0.23	0.07	0.17	0.14	0.32	0.23	0.51	0.24	0.29	0.45
4th moment AC(1)	0.81	0.09	0.06	0.05	1.11	0.63	0.06	1.33	0.72	0.11	0.61	0.19	0.12	0.55	0.25
4th moment AC(2)	0.18	0.04	0.18	0.85	1.15	0.00	1.08	1.33	0.04	0.94	0.02	0.01	0.85	0.36	0.02
4th moment AC(3)	0.09	0.00	0.03	0.01	0.20	0.29	0.04	0.29	0.13	0.14	0.15	0.01	0.26	0.01	0.00
4th moment AC(4)	0.59	0.53	0.45	0.54	0.41	0.33	0.41	0.28	0.28	0.77	0.80	0.64	0.98	0.06	0.52
Cumulative test	50.01**	26.19	21.28	62.82**	50.78**	41.92**	56.88**	48.03**	25.04	100.62**	55.97**	123.60**	57.44**	22.33	24.78
Rejections at 5%	2	1	1	2	3	2	1	1	2	2	1	1	2	1	1
Rejections at 1%	1	0	0	1	2	0	1	0	1	1	0	0	1	0	0
Density	2.0385	1.0095	0.7063	0.9764	0.9264	1.0083	0.7939	1.0263	0.7241	0.9847	1.0178	1.0591	0.7191	0.9173	0.9579
ACF	0.1357	1.5571	0.0446	0.0514	0.0674	0.1072	0.0528	0.0691	0.1323	0.0550	0.1016	0.1173	0.0594	0.1364	0.1462
VaR ranking	13	10	3	9	7	15	14	6	5	8	11	12	4	2	1

Table 9.

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**
p1				0.8750	0.8953	0.8595	0.8761	0.8952	0.8632	0.5994	0.5810	0.5376	0.5979	0.7942	0.5528
				(44.74)	*(54.95)*	*(40.38)*	*(46.30)*	*(56.27)*	*(42.09)*	*(4.82)*	*(8.04)*	*(5.61)*	*(4.68)*	*(21.43)*	*(4.83)*
μ1				0	0	0	-0.0056	-0.0053	-0.0059	0	0	0	-0.0065	-0.0065	-0.0041
							(-3.71)	*(-3.74)*	*(-3.77)*				*(-2.29)*	*(-3.16)*	*(-1.22)*
ω1	2.3E-4	2.3E-4	2.3E-4	1.2E-4	1.4E-4	1.4E-4	1.3E-4	1.4E-4	1.5E-4	2.3E-5	-5.5E-5	-7.0E-5	2.2E-5	1.3E-4	-4.4E-5
	(7.15)	*(3.15)*	*(3.17)*	*(2.78)*	*(3.09)*	*(2.90)*	*(2.84)*	*(3.11)*	*(2.95)*	*(0.46)*	*(-1.13)*	*(-1.40)*	*(0.43)*	*(2.53)*	*(-0.84)*
α1	0.0352	0.0389	0.0384	0.0292	0.0291	0.0282	0.0292	0.0291	0.0281	0.0302	0.0442	0.0439	0.0305	0.0284	0.0439
	(11.33)	*(5.18)*	*(5.21)*	*(5.69)*	*(5.80)*	*(5.37)*	*(5.66)*	*(5.76)*	*(5.34)*	*(3.72)*	*(5.12)*	*(4.49)*	*(3.69)*	*(4.95)*	*(4.63)*
β1	0.9481	0.9447	0.9453	0.9405	0.9399	0.9375	0.9400	0.9394	0.9372	0.9382	0.9353	0.9375	0.9378	0.9335	0.9303
	(221.45)	*(88.97)*	*(90.64)*	*(89.01)*	*(89.19)*	*(79.75)*	*(88.13)*	*(88.17)*	*(79.13)*	*(59.60)*	*(82.10)*	*(73.14)*	*(58.62)*	*(68.78)*	*(62.14)*
p2 (d.f. for (2),(3))		4.9616	5.0387	0.1250	0.1047	0.1405	0.1239	0.1048	0.1368	0.3656	0.0700	0.3885	0.3636	0.2049	0.3778
		(12.14)	*(11.78)*							*(3.21)*	*(3.32)*	*(4.24)*	*(3.11)*	*(5.53)*	*(3.45)*
μ2 (Asym. for (3))			0.0718	0	0	0	0.0395	0.0454	0.0371	0	0	0	0.0021	0.0218	-0.0072
			(3.31)										*(0.33)*	*(2.42)*	*(-1.01)*
ω2				0.0024	0.0527	0.0004	0.0023	0.0506	0.0004	6.4E-4	4.1E-4	3.7E-3	6.2E-4	2.0E-4	1.9E-4
				(0.61)	*(0.02)*	*(1.42)*	*(0.61)*	*(0.01)*	*(1.39)*	*(1.15)*	*(0.38)*	*(0.43)*	*(1.14)*	*(1.24)*	*(0.49)*
α2				0.0292	0	0.0660	0.0292	0	0.0717	0.0302	0.1942	0.0186	0.0305	0.0433	0.0033
				(5.69)		*(1.81)*	*(5.66)*		*(1.82)*	*(3.72)*	*(1.69)*	*(0.57)*	*(3.69)*	*(2.36)*	*(0.38)*
β2				0.9405	0	0.9708	0.9400	0	0.9689	0.9382	0.9539	0.6534	0.9378	0.9762	0.9811
				(89.01)		*(64.32)*	*(88.13)*		*(61.74)*	*(59.60)*	*(41.28)*	*(0.85)*	*(58.62)*	*(94.43)*	*(24.85)*
p3										0.4006	0.3490	0.0739	0.0384	0.0008	0.0693
μ3										0	0	0	0.0813	0.7996	0.0722
ω 3										0.0050	0.0119	0.0004	0.0044	0.0018	0.0002
										(0.11)	*(0.01)*	*(0.49)*	*(0.13)*	*(0.00)*	*(0.15)*
α3										0.0302	0	0.1706	0.0305	0	0.2667
										(3.72)		*(1.63)*	*(3.69)*		*(2.07)*
β3										0.9382	0	0.9570	0.9378	0	0.9390
										(59.60)		*(42.47)*	*(58.62)*		*(37.93)*
Unconditional σ	11.73%	11.84%	11.81%	11.52%	11.62%	11.74%	11.50%	11.60%	11.72%	11.58%	11.81%	11.79%	11.55%	11.66%	11.95%
Unconditional σ1				9.25%	9.44%	9.21%	9.23%	9.40%	9.20%	8.32%	9.32%	9.29%	8.30%	8.80%	9.15%
Unconditional σ2				21.49%	22.96%	21.50%	21.15%	22.51%	21.30%	13.00%	26.01%	10.67%	12.84%	18.20%	11.15%
Unconditional σ3										29.67%	10.92%	25.58%	27.70%	4.26%	25.65%
Unconditional τ	0	0	0.2333	0	0	0	0.3550	0.3881	0.3539	0	0	0	0.4525	0.4763	0.5016
Unconditional *k*	0.2438	11.9343	3.2640	2.8452	3.1400	3.6889	2.5180	2.7455	3.4127	4.0099	6.6105	6.0436	3.1186	-0.1725	6.9871

(continuation of Table 9.)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Loglikelihood	2653.4	2818.0	2823.3	2809.5	2806.0	2814.8	2817.0	2813.2	2822.7	2818.5	2822.7	2823.3	2825.4	2833.5	2831.4
AIC	-1.532	-1.626	-1.629	-1.621	-1.619	-1.623	-1.624	-1.622	-1.627	-1.625	-1.626	-1.625	-1.627	-1.631	-1.629
BIC	-1.526	-1.619	-1.620	-1.612	-1.610	-1.610	-1.614	-1.612	-1.612	-1.612	-1.610	-1.606	-1.611	-1.611	-1.606
Likelihood Ratio				10.70**	17.76**	15.75**	11.53**	18.95**		9.74**	1.22	16.07**	12.06**	-4.25	
Moment specification tests:															
1^{st} moment	3.6E-4	0.34	0.01	0.36	0.37	0.35	1.3E-2	0.02	7.3E-3	4.1E-1	4.2E-1	4.0E-1	1.4E-2	1.8E-1	1.0E-2
2^{nd} moment	2.20	0.03	0.07	8.28**	13.88**	6.06*	3.23	6.28*	2.41	3.35	1.64	2.64	0.54	0.21	1.23
3^{rd} moment	6.75**	4.54*	0.39	5.76*	4.73*	7.22**	0.20	0.13	0.30	5.06*	6.07*	6.19*	0.09	0.10	0.10
4^{th} moment	16.09**	0.16	0.00	22.17**	10.53**	10.03**	16.73**	9.56**	8.22**	3.88*	4.43*	5.52*	0.82	0.63	0.75
1^{st} moment AC(1)	0.27	0.47	0.62	0.46	0.45	0.48	0.59	0.57	0.62	0.40	0.43	0.38	0.53	0.59	0.54
1^{st} moment AC(2)	0.36	0.77	1.00	0.75	0.79	0.73	1.06	1.12	1.03	0.87	0.75	0.77	1.15	1.19	1.05
1^{st} moment AC(3)	0.00	0.04	0.02	0.09	0.13	0.03	0.08	0.12	0.02	0.09	0.05	0.04	0.08	0.01	0.03
1^{st} moment AC(4)	0.17	0.02	0.02	0.04	0.05	0.03	0.04	0.05	0.03	0.04	0.03	0.03	0.04	0.03	0.03
2^{nd}moment AC(1)	0.09	0.00	0.02	0.15	0.40	0.03	0.03	0.15	0.00	0.01	0.06	0.62	0.01	0.01	0.24
2^{nd}moment AC(2)	0.31	0.50	0.42	1.36	1.62	0.84	1.23	1.54	0.76	1.17	0.68	0.39	1.00	0.67	0.50
2^{nd}moment AC(3)	0.60	2.36	2.39	2.64	2.80	2.83	2.75	2.89	2.86	2.88	2.78	2.57	3.04	2.94	2.46
2^{nd}moment AC(4)	0.48	2.34	2.53	3.25	4.59*	2.24	3.47	5.05	2.45	3.50	2.44	2.32	3.45	3.34	2.41
3^{rd} moment AC(1)	0.29	0.01	0.09	0.00	0.02	0.09	0.04	0.02	0.26	0.04	0.04	0.00	0.01	0.11	0.16
3^{rd} moment AC(2)	0.24	0.68	0.39	0.51	0.28	1.15	0.14	0.02	0.57	0.35	1.06	0.98	0.07	0.11	0.48
3^{rd} moment AC(3)	1.01	1.28	1.27	0.60	0.17	3.65	0.43	0.08	3.41	0.35	2.38	2.61	0.20	2.39	1.92
3^{rd} moment AC(4)	2.17	2.04	2.06	3.03	3.47	2.83	2.61	2.76	2.63	3.03	2.76	2.84	2.52	3.01	2.46
4^{th} moment AC(1)	0.01	0.00	0.02	0.92	8.22**	0.23	0.42	3.89*	0.03	0.76	0.02	0.05	0.25	0.00	0.15
4^{th} moment AC(2)	0.12	0.47	0.35	1.79	2.23	1.10	1.23	1.67	0.70	2.15	0.94	0.59	1.39	0.66	0.41
4^{th} moment AC(3)	1.26	1.09	1.21	1.59	1.62	2.49	1.74	1.76	2.51	2.48	2.48	2.53	3.06	4.07	2.37
4^{th} moment AC(4)	0.01	0.01	0.00	0.10	0.66	0.16	0.30	1.16	0.01	0.33	0.04	0.07	0.43	0.02	0.00
Cumulative test	59.83**	46.67**	31.85*	143.40**	133.07**	152.17**	47.69**	45.11**	36.96**	251.76**	202.31**	200.47**	34.67*	34.49*	29.36
Rejections at 5%	2	1	0	3	5	3	1	3	1	2	2	2	0	0	0
Rejections at 1%	2	0	0	2	3	2	1	1	1	0	0	0	0	0	0
Density	2.9249	1.1003	1.0334	1.0786	1.2069	1.0386	0.7745	0.9122	0.7910	1.0280	1.0266	1.0379	0.9204	1.2322	0.9522
ACF	0.0469	1.6468	0.8514	0.0553	0.0645	0.0459	0.0559	0.0653	0.0465	0.0555	0.0563	0.0533	0.0554	0.0407	0.0734
VaR ranking	14	12	11	9	5	9	2	1	3	15	5	8	4	13	7

Table 10.

Period	1989-1992			1993-1997			1998-2002		
Model	**(1)**	**(3)**	**(9)**	**(1)**	**(3)**	**(9)**	**(1)**	**(3)**	**(9)**
p1			0.5318			0.8723			0.9186
			(5.73)			*(23.28)*			*(17.75)*
μ1			0.0113			-0.0001			-0.0052
			(2.35)			*(-0.05)*			*(-2.02)*
ω1	4.7E-4	5.3E-4	1.6E-4	6.7E-5	2.7E-5	1.0E-6	1.4E-4	7.2E-5	4.0E-5
	(2.99)	*(2.11)*	*(0.93)*	*(3.45)*	*(1.21)*	*(0.10)*	*(2.11)*	*(1.31)*	*(1.10)*
α1	0.0740	0.0782	0.0433	0.0384	0.0404	0.0243	0.0274	0.0235	0.0167
	(5.20)	*(3.31)*	*(2.06)*	*(6.37)*	*(4.06)*	*(4.16)*	*(2.88)*	*(2.49)*	*(2.42)*
β1	0.8931	0.8879	0.8655	0.9511	0.9573	0.9642	0.9460	0.9629	0.9704
	(42.85)	*(26.78)*	*(13.60)*	*(124.69)*	*(93.26)*	*(116.92)*	*(47.24)*	*(55.93)*	*(69.74)*
p2 (d.f. for (2),(3))		6.0098	0.4682		5.0741	0.1277		8.9053	0.0814
		(4.34)			*(6.76)*			*(4.06)*	
μ2 (Asym. for (3))		-0.1025	-0.0128		0.0073	0.0006		0.0613	0.0586
		(-2.43)			*(0.21)*			*(1.50)*	
ω2			0.0008			0.0028			0.0003
			(1.33)			*(1.21)*			*(0.22)*
α2			0.0986			0.5055			0.4417
			(2.21)			*(2.43)*			*(0.67)*
β2			0.9068			0.7377			0.7885
			(20.74)			*(8.08)*			*(2.82)*
Unconditional σ	12.02%	12.46%	11.97%	7.99%	10.97%	9.17%	7.23%	7.28%	7.33%
Unconditional σ1			7.60%			7.57%			6.62%
Unconditional σ2			15.41%			16.34%			11.32%
Unconditional *τ*	0	-0.2435	-0.1881	0	0.0238	0.0061	0	0.0958	0.3471
Unconditional *k*	0.6128	0.6783	2.1275	0.4939	-14.64	7.1055	0.0880	-1.3815	0.8429
Loglikelihood	753.7	775.7	779.7	1430.2	1481.7	1478.1	1563.5	1580.0	1580.6

Table 11.

Period	**1989-1992**			**1993-1997**			**1998-2002**		
Model	**(1)**	**(3)**	**(9)**	**(1)**	**(3)**	**(9)**	**(1)**	**(3)**	**(9)**
p1			0.7831			0.7669			0.5036
			(8.17)			*(14.07)*			*(5.54)*
μ1			0.0044			-0.0039			0.0176
			(0.95)			*(-1.50)*			*(2.72)*
ω1	8.8E-4	6.8E-4	2.2E-4	1.0E-4	1.5E-4	1.3E-4	1.8E-4	1.3E-4	1.6E-4
	(2.61)	*(1.99)*	*(1.33)*	*(3.22)*	*(2.03)*	*(1.71)*	*(2.02)*	*(1.39)*	*(0.96)*
α1	0.0887	0.0730	0.0305	0.0388	0.0507	0.0401	0.0330	0.0324	0.0464
	(4.97)	*(3.39)*	*(2.35)*	*(5.48)*	*(3.62)*	*(2.94)*	*(3.59)*	*(2.96)*	*(1.97)*
β1	0.8545	0.8828	0.9297	0.9491	0.9326	0.9013	0.9491	0.9552	0.9582
	(24.20)	*(23.64)*	*(32.11)*	*(103.12)*	*(52.18)*	*(27.33)*	*(61.45)*	*(57.80)*	*(43.10)*
p2 (d.f. for (2),(3))		7.1122	0.2169		5.6045	0.2331		8.3382	0.4964
		(4.41)			*(6.00)*			*(4.00)*	
μ2 (Asym. for (3))		-0.0278	-0.0159		0.0479	0.0129		0.1450	-0.0178
		(-0.62)			*(1.29)*			*(3.59)*	
ω2			0.0032			0.0002			0.0001
			(0.68)			*(0.91)*			*(0.67)*
α2			0.2770			0.0840			0.0148
			(1.06)			*(1.75)*			*(1.50)*
β2			0.7861			0.9561			0.9506
			(3.64)			*(36.41)*			*(26.86)*
Unconditional σ	12.44%	12.41%	12.29%	9.23%	9.41%	9.32%	9.93%	10.05%	9.92%
Unconditional σ1			9.83%			6.99%			12.14%
Unconditional σ2			18.56%			14.49%			6.51%
Unconditional *τ*	0	-0.0553	-0.1387	0	0.1304	0.1798	0	0.2223	0.2856
Unconditional *k*	0.4988	-0.6506	2.6272	0.4342	1.4344	3.5341	0.1968	-1.2413	1.1051
Loglikelihood	695.4	713.4	715.2	1254.2	1295.6	1295.4	1159.9	1183.2	1181.4

Table 12.

Period	**1989-1992**			**1993-1997**			**1998-2002**		
Model	**(1)**	**(3)**	**(9)**	**(1)**	**(3)**	**(9)**	**(1)**	**(3)**	**(9)**
p1			0.7140			0.8852			0.8921
			(9.97)			*(33.87)*			*(21.79)*
μ1			-0.0071			-0.0088			-0.0047
			(-2.05)			*(-3.38)*			*(-1.61)*
ω1	9.5E-4	2.9E-4	-2.7E-5	2.8E-4	3.5E-4	4.0E-4	9.5E-5	1.9E-4	1.6E-4
	(2.89)	*(1.57)*	*(-0.44)*	*(5.44)*	*(2.10)*	*(2.54)*	*(2.29)*	*(1.96)*	*(2.11)*
α1	0.0716	0.0547	0.0433	0.0383	0.0399	0.0335	0.0302	0.0264	0.0147
	(3.78)	*(2.95)*	*(3.51)*	*(7.69)*	*(2.86)*	*(3.07)*	*(7.29)*	*(2.89)*	*(2.54)*
β1	0.8451	0.9228	0.9273	0.9432	0.9367	0.8876	0.9638	0.9600	0.9608
	(19.22)	*(32.56)*	*(40.52)*	*(157.23)*	*(45.00)*	*(25.97)*	*(163.06)*	*(73.07)*	*(69.92)*
p2 (d.f. for (2),(3))		5.2529	0.2860		3.9845	0.1148		5.7967	0.1079
		(5.49)			*(8.45)*			*(6.55)*	
μ2 (Asym. for (3))		0.1286	0.0178		0.0892	0.0679		0.0051	0.0387
		(3.07)			*(2.61)*			*(0.14)*	
ω2			0.0006			0.0004			-0.0003
			(0.63)			*(0.28)*			*(-0.64)*
α2			0.0257			0.2132			0.1256
			(0.75)			*(1.53)*			*(1.75)*
β2			0.9597			0.9470			0.9719
			(17.05)			*(36.89)*			*(44.73)*
Unconditional σ	10.69%	11.30%	10.62%	12.28%	12.27%	12.79%	12.50%	11.71%	12.23%
Unconditional σ1			7.97%			9.18%			9.85%
Unconditional σ2			15.22%			26.90%			23.79%
Unconditional *τ*	0	0.3728	0.2151	0	0.5276	0.7309	0	0.0133	0.3243
Unconditional *k*	0.2061	2.7534	1.5933	0.2620	-29.36	14.5389	0.5317	0.4543	8.4348
Loglikelihood	816.7	851.7	853.9	925.5	1034.2	1031.9	893.6	937.8	943.4

Table 13.

Exchange rate	*Period*	*Realized excess kurtosis*	*Unconditional excess kurtosis*				*Realized skewness*	*Unconditional skewness*			
			NM(1)	*Skewed t*	*NM(2)*	*NM(3)*		*NM(1)*	*Skewed t*	*NM(2)*	*NM(3)*
GBP	1989-2002	*2.73*	0.88	1.79	**3.46**	4.39	*-0.15*	0	-0.03	**-0.03**	0.00
	1989-1992	*1.60*	0.61	0.68	**2.13**		*-0.35*	0	**-0.24**	-0.19	
	1993-1997	*2.89*	**0.49**	-14.64	7.11		*-0.10*	0	0.02	**0.00**	
	1998-2002	*1.14*	0.09	-1.38	**0.84**		*0.28*	0	0.10	**0.35**	
EUR	1989-2002	*2.43*	0.61	-0.02	**2.89**	6.81	*0.03*	0	**0.14**	0.15	0.23
	1989-1992	*1.92*	0.50	-0.65	**2.63**		*-0.42*	0	-0.06	**-0.14**	
	1993-1997	*3.83*	0.43	1.43	**3.53**		*0.38*	0	0.13	**0.18**	
	1998-2002	*1.24*	0.20	-1.24	**1.11**		*0.39*	0	0.22	**0.29**	
JPY	1989-2002	*7.09*	0.30	4.05	**4.89**	25.95	*0.77*	0	0.25	0.40	**0.62**
	1989-1992	*1.95*	0.21	2.75	**1.59**		*0.37*	0	**0.37**	0.22	
	1993-1997	*6.84*	**0.26**	-29.36	14.54		*0.77*	0	0.53	**0.73**	
	1998-2002	*8.97*	0.53	0.45	**8.43**		*0.95*	0	0.01	**0.32**	