

NORMAL MIXTURE GARCH(1,1): APPLICATIONS TO EXCHANGE RATE MODELLING

CAROL ALEXANDER AND EMESE LAZAR*

ISMA Centre, University of Reading, Reading RG6 6BA, UK

SUMMARY

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 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. Copyright © 2006 John Wiley & Sons, Ltd.

1. INTRODUCTION

The classic econometric approach to volatility modelling is the generalized autoregressive conditional heteroscedasticity (GARCH) framework that was pioneered by Engle (1982) and Bollerslev (1986). Whilst some degree of leptokurtosis in the unconditional returns distribution can be captured by the normal GARCH(1,1) model, Bollerslev (1987), Baillie and Bollerslev (1989), Hsieh (1989a,b), Nelson (1996), Johnston and Scott (2000) and many others have all concluded that, in daily or higher frequency data, the observed non-normalities in both conditional and unconditional returns is higher than can be predicted by normal GARCH(1,1) models. Consequently, a number of non-normal conditional densities have been considered in the GARCH framework. Notably, Bollerslev (1987) introduced the Student's *t*-GARCH and, more recently, non-normalities have also been captured by GARCH models with a flexible parametric error distribution such as those based on the exponential generalized beta (Wang *et al.*, 2001).¹ Alternatively, the inclusion of a trend for the long-term volatility can increase the unconditional leptokurtosis, as in the component model of Engle and Lee (1999). Non-distributional models, such as the semi-parametric ARCH model of Engle and Gonzalez-Rivera (1991), have also been considered.

These developments in GARCH models are clearly important for modelling exchange rate returns where non-normalities are highly significant (Boothe and Glassman, 1987; de Vries, 1994; Huisman *et al.*, 2002). Amongst the earliest applications of GARCH models to exchange rates were Engle and Bollerslev (1986) and Hsieh (1988), who rejected the hypothesis that these data have a heavy-tailed distribution with fixed parameters over time. The normal GARCH model

* Correspondence to: Emese Lazar, ISMA Centre, University of Reading, Whiteknights Park, PO Box 242, Reading RG6 6BA, UK. E-mail: e.lazar@ismacentre.rdg.ac.uk

¹ A review of the earlier literature on non-normal GARCH models can be found in Bollerslev *et al.* (1992).

cannot account for the entire leptokurtosis in exchange rate data (Hsieh, 1989b) and a better fit is obtained using non-normal distributions such as Student's t , GED, normal–Poisson mixture and normal–lognormal distributions, or the exponential GARCH model (Hsieh, 1989a; Baillie and Bollerslev, 1989, 1991; Engle *et al.*, 1990; Engle and Gau, 1997; Johnston and Scott, 2000).

Another disadvantage of the normal GARCH(1,1) model is that the conditional excess kurtosis is zero, and both unconditional and conditional skewness are zero. Some of the more recent GARCH models, such as the Student's t , long memory and component GARCH models, cannot account either for the time variability of the conditional higher moments unless this is explicitly modelled, as for instance in Hansen (1994), Harvey and Siddique (1999), Brooks *et al.* (2002).² But, 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, 2001b, pp. 30–31). The assumption that these conditional moments 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), Alexander (2001a), and many others. Bakshi *et al.* (2003) also emphasize 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 purpose of this paper is to examine a general class of GARCH 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 and Palm (1993) considered another restricted form of normal mixture 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 these models is appealing and the improvement in fit over the basic normal GARCH(1,1) model can be dramatic. Other restricted normal mixture GARCH models include those of Bauwens *et al.* (1999) and Bai *et al.* (2001, 2003), in which there are again two conditional variances but now their ratio is assumed constant, forcing the conditional kurtosis to be time-invariant. Also, one of the GARCH models proposed by Ding and Granger (1996, pp. 200–203) to capture volatility components of differing persistence can be viewed as a restricted version of the normal mixture GARCH(1,1) model—here the individual unconditional variances are the same and no skewness is modelled. So there has been a considerable amount of research on various restricted versions of normal mixture GARCH models. This motivated Haas *et al.* (2004a) to introduce a general class of normal mixture GARCH(p, q) models. These models have very flexible individual variance processes and have the advantage that they capture time variation in both conditional skewness and kurtosis.

This paper derives the general properties of normal mixture GARCH(1,1) models and examines their suitability for capturing the dynamics of exchange rate processes. It is outlined as follows.

² More flexible models have been built that can account for higher-order moments, such as the approach of Lye and Martin (1994) which has recently been used to model options by Lim *et al.* (2005).

Section 2 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. Here we also clarify its relationship with other well-known GARCH models and discuss the advantage of using normal mixture conditional densities rather than Student's t parameterizations. 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. We show that these models are estimated imprecisely because they are over-parameterized. At least one component will have very low weight in the mixture and this induces a substantial bias in the parameter estimates. Section 4 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 15 different models are based on eight types of model selection criteria, including likelihood tests, moment tests, 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 and standard errors are here based on the use of the analytic derivatives of the likelihood function that are stated in Appendix A. Our main results and conclusions are summarized in Section 5.

2. PROPERTIES OF NORMAL MIXTURE GARCH(1,1) MODELS

We use the notation

$$X \sim \text{NM}(p_1, \dots, p_K; \mu_1, \dots, \mu_K; \sigma_1^2, \dots, \sigma_K^2)$$

for a random variable whose distribution is characterized by a normal mixture (NM) density of the form

$$\eta(x) = \sum_{i=1}^K p_i \varphi_i(x) \quad \sum_{i=1}^K p_i = 1 \quad \varphi_i(x) = \varphi(x; \mu_i, \sigma_i^2)$$

where $[p_1, p_2, \dots, p_K]$ is the positive *mixing law* and φ denotes the normal density function. Normal mixture densities accommodate skewness and can be mesokurtic/leptokurtic when the constituent means are different, but are always symmetric and leptokurtic when the constituent means are equal.

The general model with K symmetric normal GARCH(1,1) components may have an interdependent autoregressive evolution for the variance series:³

$$y_t = \mathbf{X}_t' \gamma + \varepsilon_t \quad \varepsilon_t | I_{t-1} \sim \text{NM}(p_1, \dots, p_K; \mu_1, \dots, \mu_K; \sigma_{1t}^2, \dots, \sigma_{Kt}^2) \quad \sum_{i=1}^K p_i = 1$$

$$\sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \sum_{k=1}^K \beta_{ik} \sigma_{k,t-1}^2 \quad \text{for } i = 1, \dots, K$$

where I_{t-1} represents the information set available at time $t - 1$. Thus the individual time-varying variance components are related through their common dependence on ε_t and also through cross-equation effects where lagged values of all individual variances can affect the current value of

³ This model extends to higher-order GARCH structures in a straightforward manner, see Haas *et al.* (2004a).

each variance component. However Haas *et al.* (2004a) 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. For this reason we shall here examine in detail only the following form of the model, which we label the NM(K)–GARCH(1,1) model:⁴

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

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

so that the conditional density $\eta(\varepsilon_t)$ is the mixture density of K (≥ 2) normal density functions with each individual variance at time t given by a GARCH(1,1) process.

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 general NM(K)–GARCH(1,1) model has $5K - 2$ parameters: $\theta = (p_1, \dots, p_{K-1}; \mu_1, \dots, \mu_{K-1}; \omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2, \dots, \omega_K, \alpha_K, \beta_K)'$. Models characterized by zero-mean component densities in the mixture will be called *symmetric* NM(K)–GARCH(1,1) models and these will exhibit, conditionally and unconditionally, zero skewness but non-zero unconditional excess kurtosis. Allowing non-zero means in the components gives the *asymmetric* or ‘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.

In the general formulation of NM(K)–GARCH(1,1) the conditional variance, skewness and kurtosis are changing in time due to the time variability of the individual variances. The conditional and unconditional moments of the NM(K)–GARCH(1,1) model are given in Appendix A. Clearly parameter conditions are required to ensure these exist and the even moments are positive. The following parameter restrictions are necessary and sufficient for all the component unconditional variances and the overall unconditional variance to be finite and positive:

$$0 < p_i < 1 \quad \text{for } i = 1, \dots, K-1$$

$$\sum_{i=1}^{K-1} p_i < 1 \quad 0 \leq \alpha_i, \quad 0 \leq \beta_i < 1, \quad \text{for } i = 1, \dots, K \quad (1)$$

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

$$\omega_i + \alpha_i \frac{m}{n} > 0 \quad \text{for } i = 1, \dots, K \quad (2)$$

Note that it is not necessary to set all ω ’s positive; some of them may be negative but, given (2), they must have small absolute value. Also, it is too strict a condition to force $\alpha_i + \beta_i < 1$ for all individual variance components.⁵ Since $E(\varepsilon_t^2)$ is the weighted average of the individual

⁴ 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 $y_t = \varepsilon_t$.

⁵ This condition allowing $\alpha_i + \beta_i > 1$ for some of the variance components parallels the stationarity conditions for the periodic GARCH model of Bollerslev and Ghysels (1996).

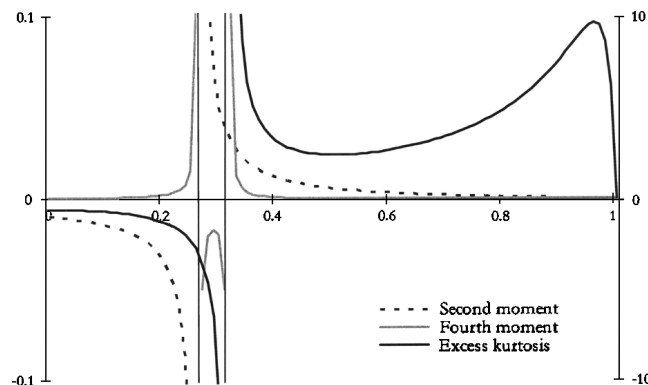
long-term variances, there exists at least one i , $1 \leq i \leq K$, such that $E(\sigma_t^2) < E(\sigma_{it}^2)$, which can be expressed as

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

and it may be that the left-hand side (and may be the right one as well) is negative.

Estimated parameters must also satisfy conditions for the third and fourth moments to be finite and for the fourth moment to be positive. Note that the skewness is finite if each component variance is finite, so (1) and (2) are already sufficient for this. The expression for the fourth moment (given in Appendix A) is complex, and it is not possible to relate existence conditions to simple conditions on estimated parameters. However, (1) and (2) allow some of the components to have $\alpha_i + \beta_i > 1$ and in this case the fourth moment might not exist 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 other parameters constant) is shown in Figure 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 last condition that the parameters must satisfy is that the estimated fourth moment is positive.

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 squared residuals follow an ARMA(K,K) process and 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 A). The representation as a GARCH(K,K) process does not imply equivalence,



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

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

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

Figure 1. Non-existence of the fourth moment

but that the GARCH(K, K) model (having only $2K + 1$ parameters) is a restricted form of the NM(K)–GARCH(1,1) model.

The NM(K)–GARCH(1,1) model is also related to another important GARCH model with non-normal error distribution, the Markov switching (MS) GARCH model introduced by Hamilton and Susmel (1994) and Cai (1994). Both models assume more than one volatility regime and both have K individual conditional variance equations. The difference between the two models is that, whilst the MS-GARCH model is characterized by a time-varying probability that each observation belongs to a given volatility regime, for the normal mixture GARCH model what matters is only the overall (time-invariant) probability that a given regime occurs over the entire sample. A different MS-GARCH model was presented by Gray (1996) and a modification of this was suggested by Klaassen (1998), but the most tractable model is that of Haas *et al.* (2004b). The only difference between this last model and the normal mixture GARCH model is that the MS model gives a more general framework with the introduction of transition matrices. In the NM-GARCH model the forecasted state probabilities do not depend on the ruling state.

We conclude this section with a discussion of some theoretical advantages of normal mixture GARCH models, particularly when compared with Student's t -GARCH models for conditional returns distributions. One clear advantage is the ability to model time-variation in conditional skewness and kurtosis. These are time-invariant in the standardized symmetric and skewed Student's t -GARCH models (see Appendix B). Another advantage is that intuitive interpretations can be placed on the normal mixture framework. For example, Ball and Torous (1983) interpreted individual distributions in the mixture to represent different market states with the mixing law determining 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. There are also behavioural models to support the use of normal mixtures on market data. For instance, Epps and Epps (1976) argue that the normal densities in the mixture arise from the existence of different types of traders in the market, having different expectations regarding returns and volatilities according to which they form their own prices and trade. The heterogeneous ARCH model of Müller *et al.* (1997), based on high-frequency intervals of different lengths, also considers the presence of different groups of investors with different strategies. In this context the mixing law determines the proportions of different types of traders in a heterogeneous population.⁶

Student's t -GARCH models are less tractable than normal mixture GARCH models. Analytic derivatives are too complex to derive for the model parameter estimates and their standard errors, necessitating the use of numerical methods.⁷ By contrast with normal GARCH, the connection with stochastic volatility option pricing models is opaque, due to the complexity of the diffusion limit of the t -GARCH process. 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.⁸

On the other hand, the analytic results on option prices for normal GARCH models may easily be translated into the normal mixture setting. Alexander and Lazar (2004a) show that,

⁶ One of the referees kindly pointed out that 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.

⁷ 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.

⁸ In fact Duan (1995) highlighted the fact that unique 'locally' risk-neutral option prices do not exist when the underlying returns variance process follows a Student's t -GARCH.

instead of the accepted ‘GARCH diffusion limit’ proposed by Nelson (1990), the continuous limit of the normal GARCH(1,1) process is naturally expressed as a deterministic volatility process. Moreover they show that the normal mixture GARCH (and the MS-GARCH model of Haas *et al.*, 2004b) has a continuous limit that is a stochastic volatility model with uncertainty over the states that govern the volatility and this implies that the log price densities will be normal mixtures.⁹ As a result normal mixture prices of European options are weighted sums of Black–Scholes prices.¹⁰ However, the continuous limit of normal mixture GARCH is not equivalent to the normal mixture local volatility model of Brigo and Mercurio (2001). This will only exhibit a persistent volatility term structure when volatility uncertainty is added to the model, as shown by Alexander (2004). Alexander and Lazar (2004b) compare the physical skews in equity index markets that are generated by several asymmetric GARCH models including those with normal, Student’s *t*, and normal mixture distributions for the errors. Only the normal mixture GARCH models are able to capture the empirical characteristics of the skew surfaces. Even without a risk premium for the state uncertainty, the normal mixture GARCH skew surfaces exhibit realistic maturity effects due to the time-variation in the conditional skewness and kurtosis.

3. MONTE CARLO SIMULATIONS

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 and the Hessian matrices that are required to estimate the parameters and their standard errors. In Appendix A we state the first- and second-order derivatives of the likelihood function with respect to the parameters for the NM(*K*) model. Consequently, these analytic forms have been used in a more efficient implementation of the optimization algorithm to obtain the simulation and historical estimation results reported in this paper.

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). Thus the base parameter values chosen for these simulations were:

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

⁹ In fact, normal mixture log price densities can result from other stochastic volatility models: see Andersen *et al.* (2002), Barndorff-Nielsen and Shephard (2002).

¹⁰ Melick and Thomas (1997) also assumed lognormal mixture stock prices and applied European option prices as weighted averages of Black–Scholes prices.

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. Second, 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. So to investigate the sensitivity of the bias to p we performed nine 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.¹¹ For non-extreme values of the mixing parameters a high percentage of the simulated time series led to realistic estimates;¹² 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.

To investigate the effect of sample size on the bias these nine simulations were performed three times, with different sample sizes of 1000, 2000 and 4000. Naturally the size of bias decreased as the sample size increased, but even with 4000 simulations we found that the ω and α parameters have a positive bias and the β parameters are biased downwards. The left-hand graphs in Figure 2 plot the bias as a function of p for the three parameters of the first variance component. Clearly the bias in the estimation of these parameters is inversely proportional to the mixing parameter associated with that variance component. In fact, though not shown here, even the mixing parameter has a small upward bias when it takes very low values. Interestingly, the overall long-term volatility and the kurtosis are both estimated *without* bias for any value of p , even for sample sizes of only 1000, and the individual component long-term volatilities also have no bias, except for extremely low values of p .¹³

If the estimation is efficient, the standard errors of the estimated parameters should decrease as the sample size increases and they should be around the Cramér–Rao lower bound.¹⁴ On comparing the estimated standard errors on the parameter estimates with their Cramér–Rao bound for different sample sizes, as in the right-hand graphs in Figure 2, we find that these are closely matched except for small values of p , indicating the efficiency of the estimation. Thus the estimation of the variance parameters for a component of the mixture becomes more exact as the mixing parameter associated with it increases and the sample size increases. It is important to note that in all cases the bias is less than the standard error of the estimation, as shown in Figure 2.¹⁵

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 the sample size increases and as the weight

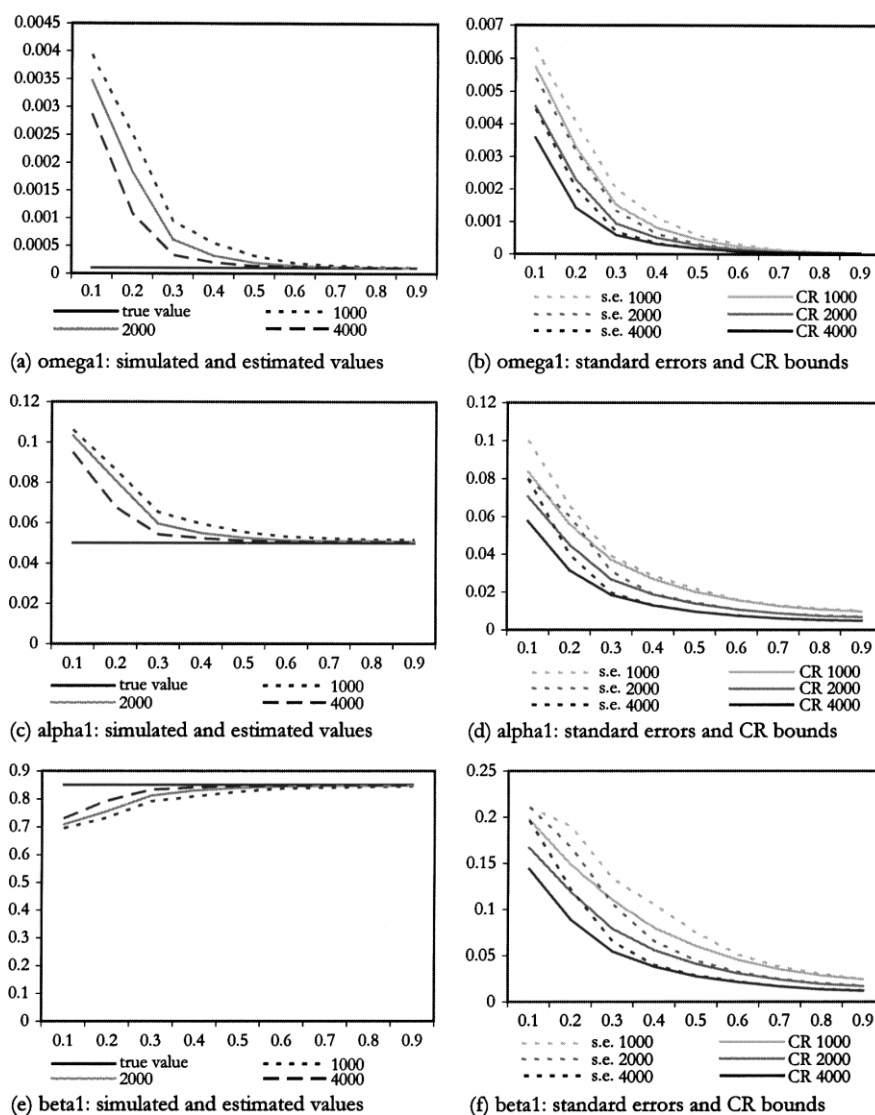
¹¹ The sensitivity of the estimation to changes in the other parameters has been studied as well. The estimation is stable and, similarly, in the vicinity of the fourth-moment non-existence boundary biased parameter estimates have been noticed. See Figure S2 of the Supplemental Appendix.

¹² An estimation is considered *unrealistic* in the following three cases: (1) when convergence is not achieved; (2) when boundary values are obtained for one of the α 's, β 's or the moments; 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. The simulation results are presented in Figure S1 of the Supplemental Appendix.

¹³ The bias in the estimation of all the parameters is presented in Figure S3 of the Supplemental Appendix.

¹⁴ 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.

¹⁵ The efficiency in the estimation of all the parameters is presented in Figure S4 of the Supplemental Appendix.



The estimated values of the NM(2)–GARCH(1,1) model parameters for the first volatility component, averaged over all the simulations for which realistic estimates were obtained. 5000 simulations were based on N realizations of the following model:

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

$$\sigma_{1t}^2 = 0.0001 + 0.05\varepsilon_{t-1}^2 + 0.85\sigma_{1,t-1}^2; \sigma_{2t}^2 = 0.01 + 0.1\varepsilon_{t-1}^2 + 0.8\sigma_{2,t-1}^2$$

for $N = 1000, 2000$ and 4000 . The estimates of the mixing parameter, the two-component unconditional volatilities (together with the overall unconditional volatility) and the unconditional kurtosis have little or no large sample bias. The efficiency of the estimators is given by a comparison of the Cramér–Rao bounds (denoted by CR) 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.

Figure 2. Bias and efficiency in parameter estimates

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 likelihood surface.

4. APPLICATION TO EXCHANGE RATES

In this section we 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. In currency markets an exchange rate (i.e. the value of one currency in units of a second currency) and its inverse play symmetric roles. However, in the derivative markets demand can exceed supply for out-of-the-money put options, resulting in a 'risk reversal' element that skews the currency option volatility smile. Even in the physical measure the foreign exchange rate returns density is not necessarily symmetric, and this is indeed the case in our data (see Table I). So here we test the specifications of both symmetric and asymmetric normal mixture GARCH models for USD exchange rates and use many model selection criteria to compare them with the fitted normal, and standardized symmetric and standardized skewed t -GARCH(1,1) models (see Fernandez and Steel, 1998; Lambert and Laurent, 2001a,b).

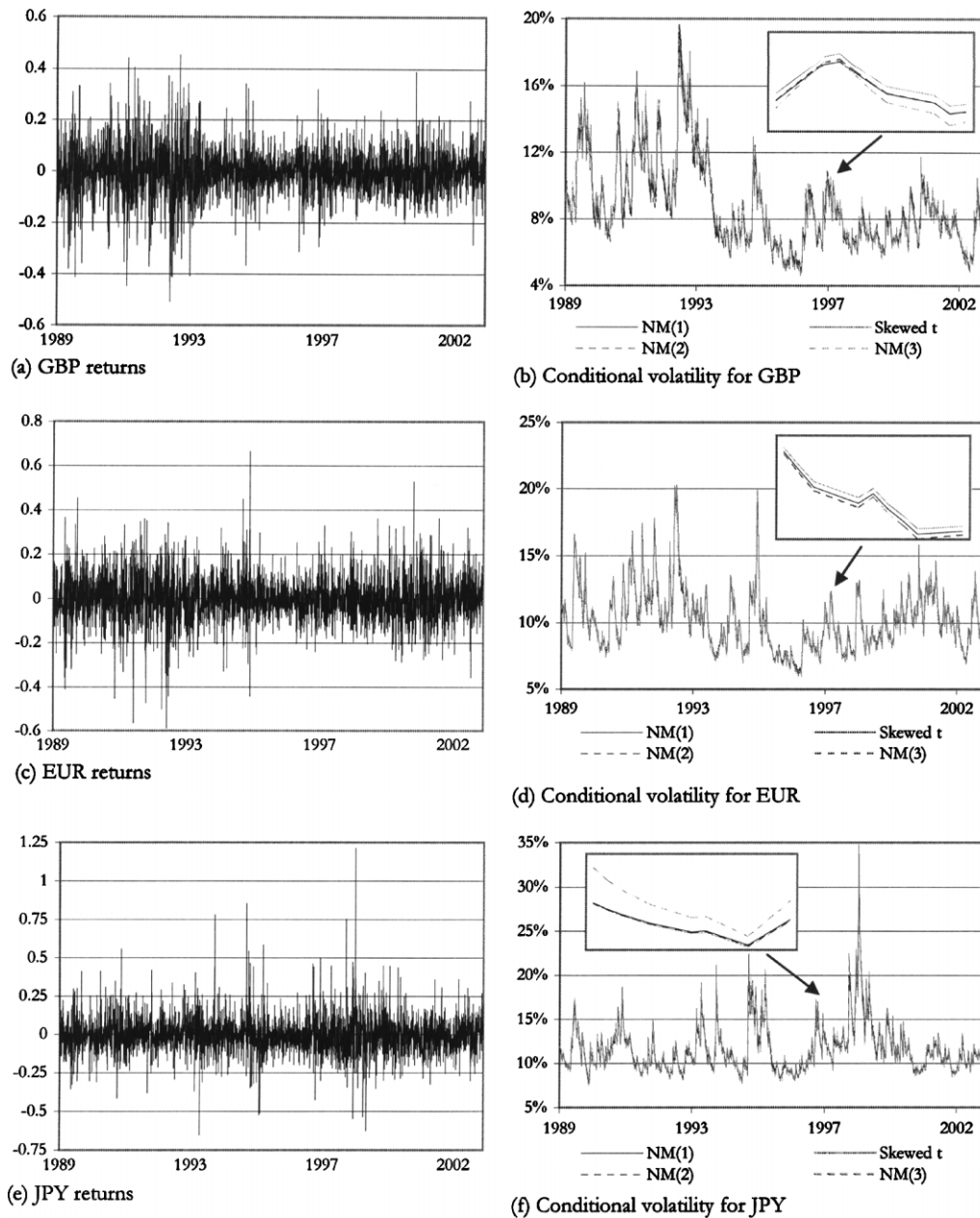
4.1. Data

Our data consists of daily prices of three foreign currencies (British pound, euro and Japanese yen) in terms of US dollar, covering a 14-year period from 2 January 1989 to 31 December 2002 (a total of 3652 observations), provided by Datastream. We denote these exchange rates GBP, EUR

Table I. Statistical description of the data

Returns on exchange rate	GBP	EUR	JPY
Mean	-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—levels (four lags)	2.5708	3.5032	1.1121
AC(1) of squared returns	0.1223***	0.0551***	0.1076***
AC(5) of squared returns	0.1389***	0.0748***	0.0448**
AC(10) of squared returns	0.0927***	0.0777***	0.0056
AC(50) of squared returns	0.0482**	0.0728***	0.0042
AC(200) of squared returns	0.0295*	0.0257	-0.0199

Notes: 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 $k = E[(X - \mu)^4]/\sigma^4 - 3$. In the table *, ** and *** represent results significantly different from zero at the 5%, 1% and 0.1% levels, respectively. The standard errors (s.e.) of the sample estimates of these parameters (assuming i.i.d. returns) are: s.e. sample mean = σ/\sqrt{T} , s.e. sample variance = $\sqrt{2\sigma^2/T}$, s.e. sample skewness = $\sqrt{6/T}$ and (further assuming normality for returns) s.e. sample excess kurtosis = $\sqrt{24/T}$, where T represents the total number of observations.



Daily returns are computed as the first difference of the logarithm of the exchange rates over the period 3 January 1989 to 31 December 2002 (a total of 3651 observations). 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. The small boxes on each conditional volatility graph magnify just a few of the data points so that small differences between the volatilities can be noticed.

Figure 3. The returns on the three exchange rates and their conditional volatilities

and JPY, respectively. Daily returns are computed as the (annualized) difference in the logarithm of daily closing prices.¹⁶ The time evolution of the returns is shown in the left-hand graphs of Figure 3, and Table I reports some statistical properties of the data. From the first four moments of the unconditional distributions, we see that 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.¹⁷ Studying the autocorrelations of the squared returns, we see that these are significant for around 200 lags for the GBP, 50 lags for the EUR and only 5 lags for the JPY rate. Since the normal mixture models are specifically designed to capture the time variation in the higher moments of conditional distributions of returns, we have tested for the existence of these moments in the sample data used. Huisman *et al.* (2001, 2002) support the existence of the fourth moment for all major US dollar exchange rates, based on daily data from 1979 to 1996 and using their methodology, and that of Hill (1975), we have reached the same conclusion for our data set.¹⁸

4.2. Model Estimation

The normal, and the symmetric and asymmetric t - and NM(2)–GARCH(1,1) models are estimated for each of the three exchange rate series, using the whole 14 years of data. Table II presents the estimation results.¹⁹ When fitting the NM(2) models 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.

To decide which model has the best fit, we have applied many model selection criteria. These results are summarized in Table III. First, the value of the Schwartz’s Bayesian information criterion (BIC) is examined—this favouring the t -GARCH models. To check the adequacy of each model to capture the higher moments of the conditional returns densities, the tables also report results of the moment specification tests described by Newey (1985). That is, we test for normality checking the first four moments of the standardized residuals and for zero autocorrelations in the powers, using a Wald test. Our results show that the normal GARCH(1,1) model has severe rejections in many moment tests, the t -GARCH models pass them and the NM(2) models have rejections for the fourth moment test.

One of the most important criteria for GARCH models is related to the unconditional distribution, based on the comparison of the empirical returns density with a simulated returns density generated by the estimated GARCH model. For this criterion we simulate returns, based on the estimated

¹⁶ Zero returns have been removed as they most often indicate missing data and distort the likelihood surface, namely 100, 160 and 190 data points have been removed from the GBP, EUR and JPY series, respectively.

¹⁷ Based on the Bayesian information criterion, the following AR(1) model is selected for the returns on the British pound: $r_t = y_t + 0.06061 * r_{t-1}$. Similarly, the following AR(1) model is chosen for the EUR/USD rate: $r_t = y_t - 0.03115 * r_{t-1}$. Subsequently the terms GBP and EUR will signify the residuals from these regressions (y_t). No autoregressive effects were found necessary in the conditional mean dollar returns for the Japanese yen. As a last step, we de-mean these series.

¹⁸ These are available in the Supplemental Appendix—see Table S1.

¹⁹ The results were generated using C++, Ox version 3.30 (Doornik, 2002) and the G@rch package version 3.0 (Laurent and Peters, 2002).

Table II. Parameter estimates of selected GARCH(1,1) models

Model	ω_1	α_1	β_1	ω_2	α_2	β_2	p_1	μ_1	μ_2	v	γ	In L
GBP												
Normal	7.0E-5	0.0439	0.9476									3740.6
	(5.36)	(10.33)	(191.6)									
Student-symmetric	4.8E-5	0.0440	0.9513							6.0897		3823.2
	(2.61)	(6.71)	(134.5)							(9.23)		
Student-skewed	4.8E-5	0.0439	0.9514							6.0784	-0.0113	3823.3
	(2.61)	(6.69)	(134.4)							(9.23)	(-0.51)	
NM(2)-symmetric	1.0E-5	0.0292	0.9434	0.0002	0.0786	0.9465	0.6595	0	0			3824.3
	(0.71)	(4.55)	(82.65)	(2.46)	(3.77)	(68.06)	(12.92)					
NM(2)-skewed	9.0E-6	0.0295	0.9424	0.0002	0.0766	0.9471	0.6509	0.0012	-0.0023			3824.6
	(0.64)	(4.51)	(80.62)	(2.52)	(3.81)	(69.59)	(12.77)	(0.74)				
EUR												
Normal	1.5E-4	0.0473	0.9393									3098.1
	(4.84)	(9.69)	(145.2)									
Student-symmetric	1.3E-4	0.0460	0.9427							6.5384		3175.3
	(3.03)	(6.41)	(105.4)							(9.14)		
Student-skewed	1.3E-4	0.0463	0.9429							6.5239	0.0641	3179.5
	(2.99)	(6.47)	(106.5)							(9.07)	(2.82)	
NM(2)-symmetric	4.7E-5	0.0256	0.9495	0.0005	0.1124	0.9241	0.7299	0	0			3172.8
	(1.69)	(4.43)	(87.24)	(1.75)	(3.04)	(35.81)	(15.07)					
NM(2)-skewed	4.6E-5	0.0248	0.9481	0.0004	0.1066	0.9272	0.6927	-0.0055	0.0124			3176.6
	(1.61)	(4.20)	(79.80)	(1.88)	(3.46)	(42.64)	(14.36)	(-2.66)				
JPY												
Normal	2.4E-4	0.0387	0.9448									2632.5
	(7.55)	(13.07)	(220.5)									
Student-symmetric	2.6E-4	0.0390	0.9426							4.8394		2810.5
	(3.34)	(5.17)	(86.88)							(12.52)		
Student-skewed	2.5E-4	0.0386	0.9431							4.9109	0.0737	2816.1
	(3.34)	(5.19)	(88.07)							(12.10)	(3.41)	
NM(2)-symmetric	1.7E-4	0.0271	0.9360	0.0005	0.1141	0.9599	0.8734	0	0			2806.6
	(3.32)	(5.36)	(79.96)	(1.32)	(2.55)	(56.05)	(42.16)					
NM(2)-skewed	1.7E-4	0.0270	0.9358	0.0004	0.1186	0.9587	0.8766	-0.0060	0.0425			2815.2
	(3.35)	(5.34)	(79.52)	(1.23)	(2.49)	(55.46)	(44.59)	(-3.95)				

Notes: Parameters are estimated by MLE. Numbers in parentheses represent t -values.

parameters, and to ensure that the simulated density is not affected by small sample size we use 50 000 replications. Also, to avoid any influence of the starting values, each simulation has 1000 steps ahead in time but we only use the last simulated return. Subsequently we estimate the returns histogram of the model using a nonparametric kernel approach. Several alternatives are available for the kernel, our chosen function being that of Epanechnikov (1969). Then the model selection criterion is based on the modified Kolmogorov–Smirnov (KS) statistic (Kolmogoroff, 1933; Smirnov, 1939; Massey, 1951; Khamis, 2000). For all three currencies it is always the NM(2) model which minimizes the modified KS statistic.²⁰

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*

²⁰ We only report the KS statistic and do not apply the KS statistical test. This is because the KS test is designed for comparison with an arbitrary distribution and to compare, as we do, the data with a simulated density based on estimated parameters we would need to modify the distribution of the test statistic (see Massey, 1951, p. 73 for more details).

Table III. Diagnostics of selected GARCH(1,1) models

Model	BIC	M1	M2	M3	M4	M1-ac1	M2-ac1	M3-ac1	M4-ac1	KS stat	ACF
GBP											
Normal	-2.100	0.51	0.80	0.67	43.07**	0.01	0.55	5.75*	0.98	2.6186	0.0833
Student-symmetric	-2.145	0.54	1.55	0.30	1.18	0.20	1.60	4.28*	0.00	1.1047	2.5289
Student-skewed	-2.142	0.47	1.12	0.82	1.34	0.21	1.57	4.23*	0.01	1.0312	0.7558
NM(2)-symmetric	-2.138	0.59	0.20	0.51	7.49**	0.21	1.01	4.62*	0.23	0.9672	0.1081
NM(2)-skewed	-2.136	0.47	0.20	1.21	7.64**	0.22	0.96	4.61*	0.31	0.8075	0.0999
EUR											
Normal	-1.768	0.08	0.46	3.25	27.05**	0.86	0.53	1.31	2.29	2.1097	0.1741
Student-symmetric	-1.810	3.5E-5	1.82	2.70	0.01	0.47	0.36	2.07	0.42	1.0569	1.5414
Student-skewed	-1.810	0.09	2.27	0.05	2.4E-3	0.58	0.29	2.18	0.31	0.6976	0.0514
NM(2)-symmetric	-1.802	7.2E-4	0.38	4.19*	9.05**	0.48	0.72	3.09	1.85	1.0435	0.1455
NM(2)-skewed	-1.802	0.10	0.00	0.26	9.60**	0.58	0.73	3.04	1.98	0.6759	0.1813
JPY											
Normal	-1.514	6.0E-5	3.54	13.95**	30.57**	0.20	1.73	0.87	1.30	3.0715	0.0972
Student-symmetric	-1.615	0.56	0.03	4.52*	0.34	0.36	0.21	0.29	0.67	1.2032	1.5871
Student-skewed	-1.616	0.06	0.16	0.33	0.01	0.50	0.10	0.11	0.44	0.9963	0.9413
NM(2)-symmetric	-1.605	0.58	7.83**	7.10**	11.52**	0.36	0.24	0.16	1.45	1.1256	0.1118
NM(2)-skewed	-1.608	6.9E-4	3.20	0.15	8.31**	0.51	0.05	0.02	0.61	0.7245	0.1069

Notes: Test statistics for the moment tests have a $\chi^2(1)$ distribution. * and ** signify significance at 5% and 1% significance levels respectively. The KS statistic shows the fit of simulated data with the original data. The ACF statistic shows the fit of the autocorrelation function of the squared residuals of the model to the autocorrelation of the squared returns.

properties of the returns—namely, the empirical autocorrelations of the squared returns. We therefore use the theoretical autocorrelation functions (ACF) of the different models²¹ 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.²² Again the NM(2) model performs best overall. 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. At long lags the over-performance of the NM(2) models over the *t*-GARCH models is even clearer. It is worth mentioning that the GARCH(1,1) model and the restricted normal mixture models also perform quite well according to this criterion.

The right-hand graphs of Figure 3 compare the time series for conditional volatility obtained from the normal GARCH(1,1), skewed *t*-GARCH and the unrestricted normal mixture GARCH(1, 1) models.²³ From this perspective there is no significant difference between the more advanced models and the simple normal GARCH(1,1) model.²⁴ Comparing how closely the unconditional realized skewness and excess kurtosis of the data are matched by the unconditional values derived from the GARCH(1,1), the skewed *t*-GARCH and the unrestricted NM(2) models, we conclude

²¹ These are derived in Appendices A and B.

²² Our criterion is based on the first 250 autocorrelations, the reason being twofold: first, lower-order autocorrelations bear higher significance and second, the length of our data period does not allow us to take arbitrarily high orders for the ACF.

²³ The conditional and unconditional moments of the GARCH(1,1) and the Student's *t*-GARCH(1,1) models are given in Appendix B.

²⁴ However, multi-day volatility forecasts will be easier to differentiate across the models as differences between parameter estimates are compounded.

that the unrestricted NM(2) model gave unconditional moment estimates that were closest to the realized moments in almost every case.²⁵ Figure 4, which is discussed in more detail in the next section, presents the time-varying conditional skewness and kurtosis estimates from the unrestricted NM(2) model (these are constant in the other models) and it is clear that they evolve within reasonable borders. Thus, based on all these results, we have strong reasons to favour the symmetric and asymmetric unrestricted NM(2) specifications.

4.3. Additional Models and Diagnostics

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 (symmetric models); (ii) similar GARCH processes for the individual variances, inspired by the Vlaar and Palm (1993) model, i.e. assuming that $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$; and (iii) a constant variance for one of the components of the mixture density (as in Roberts, 2001) based on the restriction that $\alpha_2 = \beta_2 = 0$.²⁶ These models all have time-varying conditional kurtosis and, except for the symmetric models based on restriction (i), their conditional skewness also displays time variation.

To investigate the effect of parameter restrictions we compare the restricted NM(2) models with their simplest generalizations. The likelihood ratio statistic for this test is chi-squared distributed under the null

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

where L_R and L_U represent 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 model (with no other restrictions) is preferred, whilst for the other two exchange rates the criterion selects the general unrestricted asymmetric model.

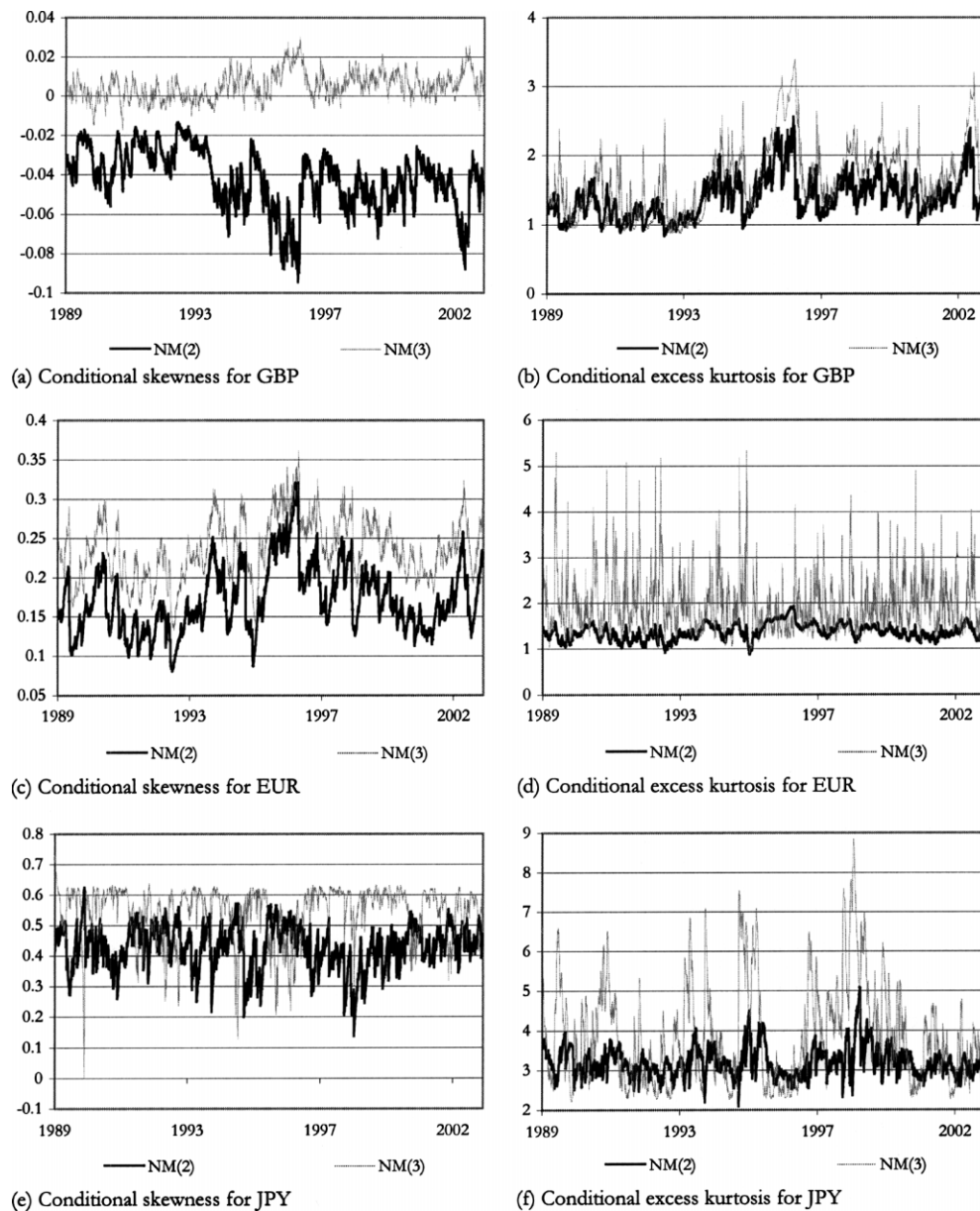
Table IV presents the diagnostics for the restricted models. The first thing to notice is that the moment test results are worse than for the unrestricted models—here we have rejections even for the second and third moments. The KS statistic reveals that the restricted models offer a worse density fit than the unrestricted ones. The ACF results are inconclusive and the evolution of the conditional moments is unrealistic.²⁷ It is therefore clear that unrestricted NM(2) models perform better than restricted ones for these exchange rates.

We also fitted restricted and unrestricted NM(3) models to the data. In this case, 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. All parameter estimates are in line with the stylized 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, all β parameter estimates lie in the range [0.93, 0.95]

²⁵ That is, for each of the three exchange rates and over four samples for each rate (1989–2002; 1989–1992; 1993–1997; 1998–2002) both the unconditional skewness and the unconditional excess kurtosis estimated from the unrestricted NM(2) model were closest to the realized sample statistics in 20 out of 24 cases. The results are available in Table S3 of the Supplemental Appendix.

²⁶ The complete set of results is available in Tables S4 to S6 in the Supplemental Appendix.

²⁷ See Figure S5 in the Supplemental Appendix.



Conditional skewness and conditional excess kurtosis for the three exchange rates estimated via the NM(2)– and NM(3)–GARCH(1,1) model respectively. The conditional skewness of the two models have a similar shape, but their sensitivities to changes in the returns are different. Note that the conditional excess kurtosis estimated via the NM(3)–GARCH(1,1) model is more volatile than that of the NM(2) models, as the third conditional volatility component is highly sensitive to unusually large squared returns.

Figure 4. The unrestricted NM(2) and NM(3) conditional skewness and excess kurtosis

Table IV. Diagnostics of restricted NM(2)–GARCH(1,1) models

Model	BIC	M1	M2	M3	M4	M1-ac1	M2-ac1	M3-ac1	M4-ac1	KS stat	ACF
GBP											
NM(2)-symmetric: $\alpha_1 = \alpha_2; \beta_1 = \beta_2$	−2.136	0.48	0.00	0.02	140.5**	0.13	1.49	5.05*	0.04	1.3112	0.1545
NM(2)-symmetric: $\alpha_2 = \beta_2 = 0$	−2.130	0.42	14.94**	0.00	22.46**	0.10	4.16*	6.35*	0.63	1.9210	0.2035
NM(2)-skewed: $\alpha_1 = \alpha_2; \beta_1 = \beta_2$	−2.134	0.33	0.18	0.40	121.8**	0.14	1.45	4.97*	0.01	1.1873	0.1568
NM(2)-skewed: $\alpha_2 = \beta_2 = 0$	−2.128	0.25	17.11**	0.18	21.47**	0.12	4.08*	5.90*	0.38	1.7971	0.2042
EUR											
NM(2)-symmetric: $\alpha_1 = \alpha_2; \beta_1 = \beta_2$	−1.803	1.4E-3	2.05	1.98	27.93**	0.41	0.22	1.11	0.11	0.9966	0.0428
NM(2)-symmetric: $\alpha_2 = \beta_2 = 0$	−1.799	0.00	16.28**	1.45	19.44**	0.40	0.11	0.89	0.02	0.9624	0.0525
NM(2)-skewed: $\alpha_1 = \alpha_2; \beta_1 = \beta_2$	−1.801	0.05	0.62	0.07	33.08**	0.48	0.19	1.11	0.08	0.8685	0.0436
NM(2)-skewed: $\alpha_2 = \beta_2 = 0$	−1.798	0.04	13.37**	0.06	24.05**	0.45	0.07	0.88	0.10	1.0883	0.0535
JPY											
NM(2)-symmetric: $\alpha_1 = \alpha_2; \beta_1 = \beta_2$	1.604	0.58	12.03**	6.98**	34.39**	0.24	3.20	6.41*	17.16**	1.2338	0.0915
NM(2)-symmetric: $\alpha_2 = \beta_2 = 0$	−1.601	0.61	22.50**	6.24*	46.06**	0.21	4.92*	16.37**	298.2**	1.3992	0.0998
NM(2)-skewed: $\alpha_1 = \alpha_2; \beta_1 = \beta_2$	−1.607	3.6E-3	4.84*	0.39	28.54**	0.36	2.03	3.95*	12.90**	0.8520	0.0920
NM(2)-skewed: $\alpha_2 = \beta_2 = 0$	−1.604	0.02	10.27**	0.39	33.36**	0.32	3.43	9.19**	225.3**	1.0280	0.1005

Notes: Test statistics for the moment tests have a $\chi^2(1)$ distribution. * and ** signify significance at 5% and 1% significance levels, respectively. The KS statistic shows the fit of simulated data with the original data. The ACF statistic shows the fit of the autocorrelation function of the squared residuals of the model to the autocorrelation of the squared returns.

and all α parameter estimates lie in the range [0.02, 0.12]. However, the unrestricted symmetric and asymmetric models have a larger α and/or a smaller β for all three exchange rates.²⁸ It may be that the third (lowest weight) component, which is most sensitive to large absolute returns, is attempting to approximate jumps in the price process, as in the GARCH-jump model of Maheu and McCurdy (2004). Also, as shown by our simulation results, there can be a large upwards bias on α and a large downwards bias on the β parameter estimate of the third component.

The diagnostic results for NM(3) are broadly in line with those for NM(2): some of the restricted models have rejections in the moment tests, and in terms of density fitting and ACF criteria, the NM(3) never over-performs the unrestricted NM(2). However, on examining the evolution of the conditional moments in the unrestricted NM(3) model we find important pitfalls, of which users should be aware.²⁹ Specifically, the highest volatility component in the NM(3) model is very unstable and inspection of the higher conditional moments (Figure 4) shows that the NM(3) model conditional kurtosis estimates are excessively jumpy. We also found that all models with restrictions other than the zero-mean restriction produce conditional skewness and kurtosis estimates that

²⁸ For the GBP and JPY rates this is also observed in the restricted models.

²⁹ These are available in the Supplemental Appendix, see Figure S5.

fluctuated rapidly, between unrealistically high and low values.³⁰ Based on these results, we again have strong reasons to favour the symmetric and asymmetric unrestricted NM(2) specifications.

We have also estimated the models over subperiods of length 4–5 years.³¹ It was interesting to observe how the parameter estimates for all 15 models changed according to the market conditions at the time. Problems were again experienced with the NM(3) models; this time a lack of convergence often occurred as 4–5 years of data were insufficient to identify a third component.

To study the effect of exceptional events on the estimation we repeated all the estimations after excluding the most extreme returns.³² Namely, a (daily) return of -3.3% on 16 September 1992 for the GBP rate, one of 4.2% on 26 May 1995 for the EUR, and one of 7.7% on 7 October 1998 for the JPY rate. Although these returns were each very extreme, no significant changes in the parameter estimates were found.³³ Also, although NM(3) parameter estimates are more sensitive to extreme observations than NM(2) models, the problems with their estimation outlined above remain. Thus the large bias on the third component parameter estimates in these two models is not caused by extreme values in the data. The diagnostic results in this case are in line with the results applied to the entire data set.

4.4. VaR Comparisons

When a statistical volatility model is applied to Value-at-Risk (VaR) estimation the accuracy of quantile predictions is a very important property. 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 $\alpha\%$ forecasted VaR, and then compares this with its theoretical value, α . We shall do this for $\alpha = 1\%, 5\%$ and 10% . That is, variances are forecasted for each model for each day and the forecasted VaR estimates 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 β_α . To avoid underestimating errors in the 1% VaR, we translate each exceptions proportion into a percentage, relative to the significance level:

$$\gamma_\alpha = \frac{\beta_\alpha - \alpha}{\alpha}$$

Then, to consolidate all this information we average the γ ’s across the different significance levels to obtain two exceptions percentages:

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

This γ gives our criterion for the VaR ranking of the models, reported in Table V. 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 most cases, the general NM(2) model performs better than the skewed t -GARCH specification; furthermore, the restricted NM(2) models also perform well.

³⁰ Results are presented in Figures S6 and S7 of the Supplemental Appendix.

³¹ Results are presented in Tables S7–S9 of the Supplemental Appendix.

³² 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.

³³ Results are available in the Supplemental Appendix, see Tables S10–S12.

Table V. VaR results of different GARCH(1,1) models

Model	VaR ranking			Model	VaR ranking		
	GBP	EUR	JPY		GBP	EUR	JPY
Normal	9	9	9	NM(2)-symmetric: $\alpha_1 = \alpha_2; \beta_1 = \beta_2$	2	6	6
Student-symmetric	8	7	8	NM(2)-symmetric: $\alpha_2 = \beta_2 = 0$	3	4	5
Student-skewed	7	1	7	NM(2)-skewed: $\alpha_1 = \alpha_2; \beta_1 = \beta_2$	1	2	2
NM(2)-symmetric	5	8	4	NM(2)-skewed: $\alpha_2 = \beta_2 = 0$	4	5	1
NM(2)-skewed	6	3	3				

Notes: VaR rankings are based on a counting method of the losses higher than the estimated VaR.

5. SUMMARY AND CONCLUSIONS

We examine a general class of GARCH models where the error term follows a normal mixture distribution with two or more GARCH(1,1) variance components. Normal mixture GARCH models capture the observed time variation in higher conditional moments. The assumption of different volatility ‘states’ in financial markets is intuitive (for instance, in the NM(2) model the high probability component has lower volatility than the low probability component) and the continuous time limit of these models is a stochastic volatility model with state uncertainty (not a volatility diffusion) where European option prices are just linear combinations of lognormal option prices. For all these reasons normal mixture GARCH should be preferred to Student *t*-GARCH models, but do the data agree?

The aim of this paper is to provide theoretical support and empirical evidence for the use of normal mixture GARCH models. The appendices to this paper summarize only the most important theoretical properties of the models considered: the derivations of these properties and further theoretical and empirical results have been collected in a Supplemental Appendix that is available from the authors (See also: www.ismacentre.rdg.ac.uk/alexander/JAE2005). Our empirical results concern 15 models (normal, standardized symmetric and skewed Student’s *t*-GARCH, six NM(2) and six NM(3) models) fitted to three main US dollar exchange rates, with likelihood, moment specification, density fitting, unconditional and conditional moment analysis, autocorrelation functions and VaR criterion applied to select the best model. Although the NM(3) models appear to perform well according to a few of these criteria, Monte Carlo simulations have indicated potential convergence problems and bias on parameter estimates of such models. Also, these models yield a too volatile kurtosis evolution: consequently we cannot recommend the use of NM(3) and higher-order normal mixture models for any asset class.

Our empirical results do not favour the restricted forms of normal mixture GARCH models that have previously been studied by several authors. Also, whilst 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, they were found to be inferior to NM(2) models for unconditional density and moment fitting and they also performed poorly under the ACF and VaR criteria. Hence our empirical evidence clearly favours NM(2)–GARCH(1,1) as the preferred specification for exchange rates.

A number of extensions of this research are attractive, including: the use of analytic term structure forecasts for excess kurtosis for dynamic delta hedging of options portfolios; empirical applications where equity returns follow a normal mixture process with two asymmetric

GARCH variance components; and applications to term structures of commodity futures within a multivariate normal mixture GARCH framework.

APPENDIX A: PROPERTIES OF THE GENERAL NM(K)–GARCH(1,1) MODEL

(i) GARCH(K, K) Representation³⁴

In the NM(K)–GARCH(1,1) model the overall variance is

$$\sigma_t^2 = \sum_{i=1}^K p_i \sigma_{it}^2 + \sum_{i=1}^K p_i \mu_i^2 \quad \text{where} \quad \sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2 \quad \text{for } i = 1, \dots, K$$

It can be shown, after some calculations, that

$$\begin{aligned} \sigma_t^2 &= A_K + B_K \quad \text{where} \quad A_K = L + \sum_{j=1}^K \varphi_j \varepsilon_{t-j}^2 + c_j + d_j L \quad \text{and} \\ B_K &= \sum_{j=1}^K \theta_j \sigma_{t-j}^2 \quad \text{with} \quad L = \sum_{i=1}^K p_i \mu_i^2 \\ c_j &= (-1)^{j+1} \sum_{\substack{i_1=1 \\ \vdots \\ i_j=1 \\ i_1 \neq \dots \neq i_j}}^K p_{i_1} \omega_{i_1} \prod_{m=2}^j \beta_{i_m}; \quad d_j = (-1)^j \sum_{\substack{i_1=1 \\ \vdots \\ i_j=1 \\ i_1 \neq \dots \neq i_j}}^K \prod_{m=1}^j \beta_{i_m}; \quad \varphi_j = (-1)^{j+1} \\ &\times \sum_{\substack{i_1=1 \\ \vdots \\ i_j=1 \\ i_1 \neq \dots \neq i_j}}^K p_{i_1} \alpha_{i_1} \prod_{m=2}^j \beta_{i_m}; \quad \theta_j = (-1)^{j+1} \sum_{\substack{i_1=1 \\ \vdots \\ i_j=1 \\ i_1 \neq \dots \neq i_j}}^K \prod_{m=1}^j \beta_{i_m} \end{aligned}$$

(ii) Conditional and Unconditional Moments³⁵

The conditional skewness and kurtosis are given by

$$\tau_t = \frac{3 \sum_{i=1}^K p_i \mu_i \sigma_{it}^2 + \sum_{i=1}^K p_i \mu_i^3}{(\sigma_t^2)^{3/2}} \quad k_t = \frac{3 \sum_{i=1}^K p_i (\sigma_{it}^2)^2 + 6 \sum_{i=1}^K p_i \mu_i^2 \sigma_{it}^2 + \sum_{i=1}^K p_i \mu_i^4}{(\sigma_t^2)^2}$$

The overall and individual unconditional variances—denoted x and y_j —are

$$x = E(\varepsilon_t^2) = E(\sigma_t^2) = \frac{\sum_{i=1}^K p_i \mu_i^2 + \sum_{i=1}^K \frac{p_i \omega_i}{(1 - \beta_i)}}{\sum_{i=1}^K \frac{p_i (1 - \alpha_i - \beta_i)}{(1 - \beta_i)}} \quad y_j = E(\sigma_{jt}^2) = \frac{\omega_j + \alpha_j x}{1 - \beta_j}$$

³⁴ Derivation of these results available in Appendix S1 of the Supplemental Appendix.

³⁵ Derivation of these results available in Appendix S2 of the Supplemental Appendix.

Then the unconditional skewness is

$$\sum_{i=1}^K p_i (3y_i \mu_i + \mu_i^3) / x^{3/2}$$

and the unconditional kurtosis may be expressed as:

$$\left(\frac{\mathbf{3p}'\mathbf{B}^{-1}\mathbf{f} + s}{\mathbf{1} - \mathbf{3p}'\mathbf{B}^{-1}\mathbf{g}} \right) / x^2 \quad \text{where} \quad \mathbf{p} = (p_1, \dots, p_{K-1})'$$

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

$$\mathbf{f} = \begin{pmatrix} w_1 + 2\alpha_1\beta_1c_1 \\ \vdots \\ w_K + 2\alpha_K\beta_Kc_K \end{pmatrix} \quad \mathbf{g} = \begin{pmatrix} \alpha_1^2 + 2\alpha_1\beta_1d_1 \\ \vdots \\ \alpha_K^2 + 2\alpha_K\beta_Kd_K \end{pmatrix}$$

$$c_i = \sum_{j=1}^K a_{ij} \left[\left(\sum_{\substack{k=1 \\ k \neq j}}^K \frac{p_k r_{jk}}{1 - \beta_j \beta_k} \right) + y_j q \right]$$

$$d_i = \sum_{j=1}^K a_{ij} \left(\sum_{\substack{k=1 \\ k \neq j}}^K \frac{p_k \alpha_j \alpha_k}{1 - \beta_j \beta_k} \right)$$

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

$$e_{ij} = a_{ij} p_j \quad r_{ik} = \omega_i \omega_k + x(\omega_i \alpha_k + \omega_k \alpha_i) + \beta_i y_i \omega_k + \beta_k y_k \omega_i$$

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

$$s = \sum_{i=1}^K p_i (6\mu_i^2 y_i^2 + \mu_i^4) \quad q = \sum_{i=1}^K p_i \mu_i^2$$

(iii) Autocorrelation Function of Squared Residuals³⁶

$$\rho_k = \frac{c_k - x^2}{z - x^2} \quad \text{where} \quad c_k = x \sum_{i=1}^K p_i \mu_i^2 + \sum_{i=1}^K p_i b_{ik} \quad \text{and} \quad b_{ik} = \omega_i x + \alpha_i c_{k-1} + \beta_i b_{ik-1}$$

$$c_0 = z; \quad b_{i0} = c_i + d_i z + \mathbf{e}_i' \mathbf{B}^{-1} (\mathbf{f} + \mathbf{g}z)$$

(iv) Analytic Derivatives of the Likelihood Function³⁷

We use the following notation for the parameters:

$$\mathbf{P} = (p_1, \dots, p_{K-1})', \quad \boldsymbol{\mu} = (\mu_1, \dots, \mu_{K-1})', \quad \boldsymbol{\gamma}_i = (\omega_i, \alpha_i, \beta_i)' \quad i = 1, \dots, K$$

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

$$= (\mathbf{p}'; \boldsymbol{\mu}; \boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \dots, \boldsymbol{\gamma}_K')'$$

Maximizing $\sum_{t=1}^T \ln[\eta(\varepsilon_t)] + \frac{T}{2} \ln(2\pi)$ gives the optimal parameter values. Let $g_i = \frac{p_i}{\sigma_{it}^2} \exp\left(-\frac{1}{2} \frac{(\varepsilon_t - \mu_i)^2}{\sigma_{it}^2}\right)$. Note that g_i is a function of p_i, μ_i and $\boldsymbol{\gamma}_i$ only, for $i = 1, \dots, K-1$ and g_K is a function of $\mathbf{p}, \boldsymbol{\mu}$ and $\boldsymbol{\gamma}_K$. Let $m_t(\boldsymbol{\theta}) = \ln\left(\sum_{i=1}^K g_i\right)$ so our objective is to maximize $M(\boldsymbol{\theta}) = \sum_{t=1}^T m_t(\boldsymbol{\theta})$. For this we require the first- and second-order derivatives of $m_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$:

$$\frac{\partial m_t(\boldsymbol{\theta})}{\partial p_i} = \left(\frac{1}{\sum_{k=1}^K g_k} \right) \left(\frac{g_i}{p_i} + \frac{\partial g_K}{\partial p_i} \right) \quad \text{where} \quad \frac{\partial g_K}{\partial p_i} = g_K \left(-\frac{1}{p_K} + \left(\frac{\varepsilon_t - \mu_K}{\sigma_{Kt}^2} \right) \left(\frac{\partial \mu_K}{\partial p_i} \right) \right)$$

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

$$\frac{\partial m_t(\boldsymbol{\theta})}{\partial \mu_i} = \left(\frac{1}{\sum_{k=1}^K g_k} \right) \left(\frac{\partial g_i}{\partial \mu_i} + \frac{\partial g_K}{\partial \mu_i} \right) \quad \text{where} \quad \frac{\partial g_i}{\partial \mu_i} = \frac{g_i(\varepsilon_t - \mu_i)}{\sigma_{it}^2}$$

³⁶ Derivation of these results available in Appendix S2 of the Supplemental Appendix.

³⁷ Derivation of these results available in Appendix S3 of the Supplemental Appendix.

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

The second-order derivatives are

$$\begin{aligned} \frac{\partial^2 m_t(\theta)}{\partial p_i \partial p_j} &= - \left(\frac{\partial m_t(\theta)}{\partial p_i} \right) \left(\frac{\partial m_t(\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) \\ &\quad \times \left[\left(\frac{1}{\sigma_{Kt}^2} \right) \left(\frac{\partial \mu_K}{\partial p_i} \right) \left(\frac{\partial \mu_K}{\partial p_j} \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(\theta)}{\partial p_i \partial \mu_i} &= - \left(\frac{\partial m_t(\theta)}{\partial p_i} \right) \left(\frac{\partial m_t(\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] \\ &\quad + \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(\theta)}{\partial p_i \partial \mu_j} &= - \left(\frac{\partial m_t(\theta)}{\partial p_i} \right) \left(\frac{\partial m_t(\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) \\ &\quad + \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], \quad j \neq i \\ \frac{\partial^2 m_t(\theta)}{\partial \mu_i^2} &= - \left(\frac{\partial m_t(\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 \right] \end{aligned}$$

$$\begin{aligned} & - \left(\frac{g_i}{\sigma_{it}^2} - \left(\frac{p_i}{p_K} \right)^2 \left(\frac{g_K}{\sigma_{Kt}^2} \right) \right] \\ \frac{\partial^2 m_t(\theta)}{\partial \mu_i \partial \mu_j} &= - \left(\frac{\partial m_t(\theta)}{\partial \mu_i} \right) \left(\frac{\partial m_t(\theta)}{\partial \mu_j} \right) + \left(\frac{1}{\sum_{k=1}^K g_k} \right) \left[\left(\frac{1}{g_K} \right) \left(\frac{\partial g_K}{\partial \mu_i} \right) \left(\frac{\partial g_K}{\partial \mu_j} \right) \right. \\ & \quad \left. - \left(\frac{p_i}{p_K} \right) \left(\frac{p_j}{p_K} \right) \left(\frac{g_K}{\sigma_{Kt}^2} \right) \right], \quad j \neq i \\ \frac{\partial^2 m_t(\theta)}{\partial p_i \partial \gamma_j} &= \left(\frac{1}{2\sigma_{jt}^2} \right) \left(\left(\frac{\varepsilon_t - \mu_j}{\sigma_{jt}} \right)^2 - 1 \right) \frac{\partial \left(\frac{\frac{g_j}{K}}{\sum_{k=1}^K g_k} \right)}{\partial p_i} \left(\frac{\partial \sigma_{jt}^2}{\partial \gamma_j} \right), \quad j \neq K \end{aligned}$$

where

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

$j \neq K, i$

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

$$\begin{aligned} & \times \left(\frac{\partial \sigma_{Kt}^2}{\partial \gamma_K} \right) \text{ with } \frac{\partial \left(\frac{g_K}{\sum_{k=1}^K g_k} \right)}{\partial p_i} = \frac{\left(\frac{\partial g_K}{\partial p_K} \right) \left(\sum_{k=1}^K g_k \right) - \left(\frac{g_i}{p_i} + \frac{\partial g_K}{\partial p_i} \right) g_K}{\left(\sum_{k=1}^K g_k \right)^2} \\ \\ \frac{\partial^2 m_t(\theta)}{\partial \mu_i \partial \gamma_i} &= \left(\frac{1}{\sum_{k=1}^K g_k} \right) \left\{ \left[\left(\frac{1}{g_i} \right) \left(\frac{\partial g_i}{\partial \mu_i} \right) - \left(\frac{\partial m_t(\theta)}{\partial \mu_i} \right) \right] \left(\frac{\partial g_i}{\partial \gamma_i} \right) - \left(\frac{1}{\sigma_{it}^2} \right) \left(\frac{\partial g_i}{\partial \mu_i} \right) \left(\frac{\partial \sigma_{it}^2}{\partial \gamma_i} \right) \right\}, \\ \\ \frac{\partial^2 m_t(\theta)}{\partial \mu_i \partial \gamma_j} &= \left(\frac{-1}{\sum_{k=1}^K g_k} \right) \left(\frac{\partial m_t(\theta)}{\partial \mu_i} \right) \left(\frac{\partial g_i}{\partial \gamma_j} \right) \quad j \neq K, i \\ \\ \frac{\partial^2 m_t(\theta)}{\partial \mu_i \partial \gamma_K} &= \left(\frac{1}{\sum_{k=1}^K g_k} \right) \left\{ \left[\left(\frac{1}{g_K} \right) \left(\frac{\partial g_K}{\partial \mu_i} \right) - \left(\frac{\partial m_t(\theta)}{\partial \mu_i} \right) \right] \left(\frac{\partial g_K}{\partial \gamma_K} \right) - \left(\frac{1}{\sigma_{Kt}^2} \right) \left(\frac{\partial g_K}{\partial \mu_i} \right) \left(\frac{\partial \sigma_{Kt}^2}{\partial \gamma_K} \right) \right\} \end{aligned}$$

Derivatives with respect to γ_i are more complex. After some calculations:

$$\begin{aligned} \frac{\partial m_t(\theta)}{\partial \gamma_i} &= \left(\frac{1}{2\sigma_{it}^2} \right) \left(\frac{g_i}{\sum_{k=1}^K g_k} \right) \left(\left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 - 1 \right) \left(\frac{\partial \sigma_{it}^2}{\partial \gamma_i} \right) \\ \\ \frac{\partial^2 m_t(\theta)}{\partial \gamma_i \partial \gamma'_i} &= \left(\frac{1}{\sigma_{it}^2} \right)^2 \left[\left(\frac{g_i \left(\sum_{k=1}^K g_k - g_i \right)}{4 \left(\sum_{k=1}^K g_k \right)^2} \right) \left(1 - \left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 \right)^2 + \left(\frac{g_i}{2 \left(\sum_{k=1}^K g_k \right)} \right) \right. \\ &\quad \times \left. \left(1 - 2 \left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 \right) \right] \left(\frac{\partial \sigma_{it}^2}{\partial \gamma_i} \right) \left(\frac{\partial \sigma_{it}^2}{\partial \gamma_i} \right)' + \left(\frac{1}{2\sigma_{it}^2} \right) \left(\frac{g_i}{\sum_{k=1}^K g_k} \right) \left(\left(\frac{\varepsilon_t - \mu_i}{\sigma_{it}} \right)^2 - 1 \right) \left(\frac{\partial^2 \sigma_{it}^2}{\partial \gamma_i \partial \gamma'_i} \right) \end{aligned}$$

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

For first- and second-order derivatives of σ_{it}^2 with respect to γ_i , let $z_{it} = (1, \varepsilon_{t-1}^2, \sigma_{it-1}^2)'$. Then

$$\frac{\partial \sigma_{it}^2}{\partial \gamma_i} = z_{it} + \beta_i \frac{\partial \sigma_{it-1}^2}{\partial \gamma_i} \quad \text{with starting values } \frac{\partial \sigma_{i0}^2}{\partial \gamma_i} = (1, s^2, s^2)', s^2 = \frac{\sum_{t=1}^T \varepsilon_t^2}{T} \\ \frac{\partial^2 \sigma_{it}^2}{\partial \gamma_i \partial \gamma_i'} = w_{it} + \beta_i \frac{\partial^2 \sigma_{it-1}^2}{\partial \gamma_i \partial \gamma_i'} \quad \text{with starting values } \frac{\partial^2 \sigma_{i0}^2}{\partial \gamma_i \partial \gamma_i'} = \frac{1}{(1 - \beta_i)} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & s^2 \\ 1 & s^2 & 2s^2 \end{bmatrix}$$

where $w_{it} = A_{it} + A_{it}^T$ with

$$A_{it} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \left(\frac{\partial \sigma_{it-1}^2}{\partial \gamma_i} \right)' & & \end{bmatrix}$$

APPENDIX B: PROPERTIES OF NORMAL AND STUDENT'S *T*-GARCH(1,1) MODELS

Here $y_t = \varepsilon_t = z_t \sigma_t$ with $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$. The autocorrelation function in all three models is

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

(i) Normal GARCH(1,1)³⁸: $\varepsilon_t | I_{t-1} \sim N(0, \sigma_t^2)$

The conditional skewness and excess kurtosis and unconditional skewness as well are zero. The unconditional kurtosis is $3z/x^2$ where x is the unconditional variance:

$$x = E(\varepsilon_t^2) = E(\sigma_t^2) = \frac{\omega}{1 - \alpha - \beta} \quad \text{and} \quad z = \frac{\omega^2 + 2\omega x(\alpha + \beta)}{1 - \beta^2 - 3\alpha^2 - 2\alpha\beta}$$

(ii) Standardized Student's *t*-GARCH(1,1)³⁹: $z_t | I_{t-1} \sim \text{Standardized Student's } t(v)$

$$\text{The density function is } g(z|v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{(v-2)\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{z^2}{v-2} \right)^{-\frac{v+1}{2}}$$

³⁸ Derivation of these results available in Appendix S4 of the Supplemental Appendix.

³⁹ Derivation of these results available in Appendix S5 of the Supplemental Appendix.

Both variances are the same as for the normal GARCH(1,1) model; the conditional and unconditional skewness are zero. The conditional kurtosis for $v > 4$ is $k = \frac{3v-6}{v-4}$; the unconditional kurtosis is $\left(\frac{k}{x^2}\right) \left(\frac{\omega^2 + 2\omega x(\alpha + \beta)}{1 - \beta^2 - k\alpha^2 - 2\alpha\beta}\right)$ where v is the degrees of freedom parameter.

(iii) Skewed Student's t -GARCH(1,1)⁴⁰: $z_t|I_{t-1} \sim \text{Standardized Skewed Student's } t(v, \gamma)$
The density function is

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

where s is given by $s^2 = e_2 - e_1^2$, I is an indicator function and e_r are the conditional r th moments given by

$$e_r = M_r(v) \left(\frac{\gamma^{r+1} + \frac{(-1)^r}{\gamma^{r+1}}}{\gamma + \frac{1}{\gamma}} \right) \quad M_r(v) = \frac{\Gamma\left(\frac{v-r}{2}\right) \Gamma\left(\frac{1+r}{2}\right) (v-2)^{r/2}}{\sqrt{\pi} \Gamma\left(\frac{v}{2}\right)}$$

$$I_t(z) = \begin{cases} \mathbf{1} & \text{if } z_t \geq -e_1/s \\ -\mathbf{1} & \text{if } z_t < -e_1/s \end{cases}$$

The two variances are the same as above; the conditional and unconditional skewness is $\frac{e_3 - 3e_1e_2 + 2e_1^3}{(e_2 - e_1^2)^{3/2}}$. The conditional kurtosis is $k = \frac{e_4 - 4e_1e_3 + 6e_2e_1^2 - 3e_1^4}{(e_2 - e_1^2)^2}$, the unconditional kurtosis may be expressed in terms of this k as $\left(\frac{k}{x^2}\right) \left(\frac{\omega^2 + 2\omega x(\alpha + \beta)}{1 - \beta^2 - k\alpha^2 - 2\alpha\beta}\right)$.

ACKNOWLEDGEMENTS

This paper has benefited considerably from the extensive and insightful comments and the very helpful suggestions of the editor-in-charge, Professor Tim Bollerslev, and of two anonymous referees. Nevertheless, the responsibility for any errors in the paper is ours.

REFERENCES

- Alexander C. 2001a. Principles of the skew. *Risk* **14**(1): S29–32.
 Alexander C. 2001b. *Market Models: a Guide to Financial Data Analysis*. John Wiley & Sons: Chichester.
 Alexander C. 2004. Normal mixture diffusion with uncertain volatility: modelling short and long term smile effects. *Journal of Banking and Finance*, **28**(12): 2957–2980.
 Alexander C, Lazar E. 2004a. The continuous limit of GARCH models. ISMA Centre Discussion Papers in Finance 2004-10 (available from www.ismacentre.rdg.ac.uk/dp).
 Alexander C, Lazar E. 2004b. The equity index skew, market crashes and asymmetric normal mixture GARCH. ISMA Centre Discussion papers in Finance 2004-14 (available from www.ismacentre.rdg.ac.uk/dp).
 Andersen TG, Bollerslev T, Diebold FX. 2002. Parametric and nonparametric volatility measurement. In *Handbook of Financial Econometrics*, Aït-Sahalia Y, Hansen LP (eds). North-Holland: Amsterdam.

⁴⁰ Derivation of these results available in Appendix S5 of the Supplemental Appendix.

- Bai X, Russell JR, Tiao GC. 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 JR, Tiao GC. 2003. Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics* **114**(2): 349–360.
- Baillie RT, Bollerslev T. 1989. The message in daily exchange rates: a conditional-variance tale. *Journal of Business and Economic Statistics* **7**(3): 297–305.
- Baillie RT, Bollerslev T. 1991. Intra-day and inter-market volatility in foreign exchange rates. *The Review of Economic Studies* **58**(3): 565–585.
- Bakshi G, Kapadia N, Madan D. 2003. Stock return characteristics, skew laws, and the differential pricing for individual equity options. *The Review of Financial Studies* **16**(1): 101–143.
- Ball CA, Torous WN. 1983. A simplified jump process for common stock returns. *Journal of Financial and Quantitative Analysis* **18**(1): 53–65.
- Barndorff-Nielsen OE, Shephard N. 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* **64**(2): 253–280.
- Bates DS. 1991. The crash of '87: was it expected? The evidence from options markets. *Journal of Finance* **46**: 1009–1044.
- Bauwens L, Bos CS, van Dijk HK. 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* **31**: 309–328.
- Bollerslev T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**: 542–547.
- Bollerslev T, Ghysels E. 1996. Periodic autoregressive conditional heteroscedasticity. *Journal of Business and Economic Statistics* **14**: 139–151.
- Bollerslev T, Chou RY, Kroner KF. 1992. ARCH modeling in finance. *Journal of Economics* **52**: 5–59.
- Boothe P, Glassman D. 1987. The statistical distribution of exchange rates: empirical evidence and economic implication. *Journal of International Economics* **22**: 297–319.
- Brigo D, Mercurio F. 2001. *Interest Rate Models: Theory and Practice*. Springer-Verlag: Berlin.
- Brooks C, Burke SP, Persaud 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* **12**: 309–316.
- Derman E. 1999. Regimes of volatility. *Risk Magazine* **12**(4): 55–59.
- de Vries CG. 1994. Stylized facts of nominal exchange rate returns. In *Handbook of International Macroeconomics*, van der Ploeg F (ed.). Basil Blackwell: Oxford; 348–389.
- Ding Z, Granger CWJ. 1996. Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics* **73**: 185–215.
- Doornik JA. 2002. *Object-Oriented Matrix Programming Using Ox*, 3rd edn. Timberlake Consultants Press and Oxford: London (www.nuff.ox.ac.uk/Users/Doornik).
- Duan J-C. 1995. The GARCH option pricing model. *Mathematical Finance* **5**(1): 13–32.
- Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4): 987–1007.
- Engle RF, Bollerslev T. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* **5**: 1–50.
- Engle RF, 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 RF, Lee G. 1999. A permanent and transitory component model of stock return volatility. In *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, Engle RF, White H (eds). Oxford University Press: Oxford; 475–497.
- Engle RF, Gonzalez Rivera G. 1991. Semiparametric GARCH models. *Journal of Business & Economic Statistics* **9**(4): 345–359.
- Engle RF, Ito T, Lin W-L. 1990. Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica* **58**(3): 525–542.

- Epanechnikov V. 1969. Nonparametric estimation of a multidimensional probability density. *Teoriya Veroyatnostej i Ee Primeneniya* **14**: 156–162.
- Epps TW, Epps ML. 1976. The stochastic dependence of security price changes and transaction volumes: implications for the mixture-of-distributions hypothesis. *Econometrica* **44**(2): 305–321.
- Fernandez C, Steel M. 1998. On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association* **93**: 359–371.
- Gray SF. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* **42**: 27–62.
- Haas M, Mittnik S, Paolella MS. 2004a. Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics* **2**: 211–250.
- Haas M, Mittnik S, Paolella MS. 2004b. A new approach to Markov-Switching GARCH models. *Journal of Financial Econometrics* **2**(4): 493–530.
- Hamilton JD, Susmel R. 1994. Autoregressive conditional heteroscedasticity and changes in regime. *Journal of Econometrics* **64**: 307–333.
- Hansen BE. 1994. Autoregressive conditional density estimation. *International Economic Review* **35**: 705–730.
- Harvey CR, Siddique A. 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Studies* **34**(4): 465–487.
- Hill BM. 1975. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**(5): 1163–1174.
- Hsieh DA. 1988. The statistical properties of daily foreign exchange rates: 1974–1983. *Journal of International Economics* **24**: 129–145.
- Hsieh DA. 1989a. Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business and Economic Statistics* **7**(3): 307–317.
- Hsieh DA. 1989b. Testing for nonlinear dependence in daily foreign exchange rates. *The Journal of Business* **62**(3): 339–368.
- Huisman R, Koedijk KG, Kool CJM, Palm F. 2001. Tail-index estimator in small sample. *Journal of Business and Economic Statistics* **19**: 208–216.
- Huisman R, Koedijk KG, Kool CJM, Palm F. 2002. The tail-fatness of FX returns reconsidered. *De Economist* **150**: 299–312.
- Johnston K, Scott E. 2000. GARCH models and the stochastic process underlying exchange rate price changes. *Journal of Financial and Strategic Decisions* **13**(2): 13–24.
- Khamis HJ. 2000. The two-stage δ -corrected Kolmogorov–Smirnov test. *Journal of Applied Statistics* **27**(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* **4**: 83–91.
- Lambert P, 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, 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, Peters J-P. 2002. GARCH 2.2: an Ox package for estimating and forecasting various ARCH models. *Journal of Economic Surveys* **16**(3): 447–485.
- Lim GC, Martin GM, Martin VL. 2005. Parametric pricing of higher order moments in S&P500 options. *Journal of Applied Econometrics*, forthcoming.
- Lye JN, Martin VL. 1994. Non-linear time series modelling and distributional flexibility. *Journal of Time Series Analysis* **15**: 65–84.
- Maheu JM, McCurdy TH. 2004. News arrival, jump dynamics and volatility components for individual stock returns. *Journal of Finance* **59**: 755–793.
- Massey FJ Jr. 1951. The Kolmogorov–Smirnov test for goodness of fit. *Journal of the American Statistical Association* **46**(253): 68–78.
- Melick WR, Thomas CP. 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* **32**(1): 91–115.

- Müller UA, Dacorognal MM, Davé RD, Olsen RB, Pictet OV, von Weizsäcker JE. 1997. Volatilities of different time resolutions—analyzing the dynamics of market components. *Journal of Empirical Finance* **4**: 213–239.
- Nelson DB. 1990. ARCH Models as Diffusion Approximations. *Journal of Econometrics* **45**: 7–38.
- Nelson DB. 1996. Asymptotic filtering theory for multivariate ARCH models. *Journal of Econometrics* **71**: 1–47.
- Newey WK. 1985. Generalized method of moments specification testing. *Journal of Econometrics* **29**: 229–256.
- Roberts MC. 2001. Commodities, options, and volatility: modelling agricultural futures prices. Working Paper, The Ohio State University.
- Smirnov N. 1939. Sur les écarts de la courbe de distribution empirique. *Matematicheskii Sbornik*. **48**(6): 3–26 (in Russian with French summary).
- Vlaar PJG, Palm FC. 1993. The message in weekly exchange rates in the European monetary system: mean reversion, conditional heteroscedasticity, and jumps. *Journal of Business & Economic Statistics* **11**(3): 351–360.
- Wang K-L, Fawson C, Barrett CB, McDonald JB. 2001. A flexible parametric GARCH model with an application to exchange rates. *Journal of Applied Econometrics* **16**: 521–536.