The University of Reading

THE BUSINESS SCHOOL
FOR FINANCIAL MARKETS

Symmetric Normal Mixture GARCH

ISMA Centre Discussion Papers in Finance 2003-09
May 2003

Carol Alexander
ISMA Centre, University of Reading, UK

Emese Lazar
ISMA Centre, University of Reading, UK

Copyright 2003 Carol Alexander and Emese Lazar. All rights reserved.

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

Normal mixture (NM) GARCH models are better able to account for leptokurtosis in financial data and offer a more intuitive and tractable framework for risk analysis and option pricing than student's *t*-GARCH models. We present a general, symmetric parameterisation for NM-GARCH(1,1) models, derive the analytic derivatives for the maximum likelihood estimation of the model parameters and their standard errors and compute the moments of the error term. Also, we formulate specific conditions on the model parameters to ensure positive, finite conditional and unconditional second and fourth moments. Simulations quantify the potential bias and inefficiency of parameter estimates as a function of the mixing law. We show that there is a serious bias on parameter estimates for volatility components having very low weight in the mixing law. An empirical application uses moment specification tests and information criteria to determine the optimal number of normal densities in the mixture. For daily returns on three US Dollar foreign exchange rates (British pound, euro and Japanese yen) we find that, whilst normal GARCH(1,1) models fail the moment tests, a simple mixture of two normal densities is sufficient to capture the conditional excess kurtosis in the data. According to our chosen criteria, and given our simulation results, we conclude that a two regime symmetric NM-GARCH model, which quantifies volatility corresponding to 'normal' and 'exceptional' market circumstances, is optimal for these exchange rate data.

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 Prof. Chris Brooks, a valued colleague at the ISMA Centre, for many enlightening discussions throughout the course of this research.

This discussion paper is a preliminary version designed to generate ideas and constructive comment. The contents of the paper are presented to the reader in good faith, and neither the author, the ISMA Centre, nor the University, will be held responsible for any losses, financial or otherwise, resulting from actions taken on the basis of its content. Any persons reading the paper are deemed to have accepted this.

1. Introduction

One of the main lines of research in finance is focused on finding an appropriate quantitative description for market returns. Already forty years ago, Mandelbrot (1963) and Fama (1965), followed by many others, showed that time invariant normal distributions do not offer an appropriate framework, given the excess kurtosis and volatility clustering that characterize returns in financial markets. Consequently there has been a keen interest in developing tractable non-normal models 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 of the form:

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

where $[p_1, p_2, \ldots, p_K]$ is the positive *mixing law*, with $\sum_{i=1}^{K} p_i = 1$ and $\phi_i(x) = \phi(x; \mu_i, \sigma_i^2)$ are normal density functions. We use the notation $X \sim NM(p_1,\ldots,p_K;\mu_1,\ldots,\mu_K;\sigma_1^2,\ldots,\sigma_K^2)$ for a random variable whose distribution is characterised by a density function of this form.

In stock and exchange rate (FX) markets, leptokurtic densities are known to offer a better description of the unconditional returns densities than the normal density, when returns are measured at daily or intra-day frequency. 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.

One advantage of the normal mixture over the student's *t* model for unconditional 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 circumstances and the mixing law gives the probabilities of these states. In the case of a mixture of two normals we can 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, Petit and Sung (1982). There are also behavioural models to support the

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

Copyright © 2003 Carol Alexander and Emese Lazar.

use of normal mixtures on market data. For example, the normal densities in the mixture may arise from the different types of 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).

Turning now to the literature on conditional densities of market returns, a key issue for modelling returns in all financial markets is the time-variation in volatility. The option theoretic approach treats volatility as a continuous-time process.[2] In this paper we focus on the classic econometric approach to volatility modelling – the generalised autoregressive conditional heteroscedasticity (GARCH) models that were 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:

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

and a symmetric deterministic model for the conditional variance of these errors is given by:[3]

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

In such a model (with daily data) today's conditional variance (σ_t^2) is affected by yesterday's squared unexpected return (ε_{t-1}^2) and yesterday's conditional variance. Heavy tails in unconditional returns distributions can be captured by even the simplest GARCH(1,1) models, where the conditional returns distributions are assumed to be normal.

Westerfield (1977), McFarland, Petit and Sung (1982), Boothe and Glassman (1987), Hsieh (1989) and Johnston and Scott (2000) have concluded that, in daily or higher frequency data, the observed leptokurtosis in both conditional or 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) and developed by Harvey and Siddique (1999) and Brooks, Burke and Persand (2002), amongst others. Non-distributional models, like the semi-parametric ARCH model of Engle and Gonzalez-Rivera (1991) have also been considered. 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.

[2] 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).
[3] Parameter conditions are needed in order to ensure positivity for the second moment.

Copyright © 2003 Carol Alexander and Emese Lazar.

Numerical methods would also have to be applied if diffusion limits of the *t*-GARCH (or semi-parametric GARCH) process were used for stochastic volatility option pricing models. The price density would have no simple analytic properties, and the relationship between these option prices and Black-Scholes prices would be very complex indeed.

These observations lead to another advantage of the normal mixture over the student's *t* framework for modelling conditional returns distributions. Analytic results for normal models may be easily translated into the normal mixture setting. Closed-forms for normal GARCH option prices have a straightforward extension to analytic forms for normal mixture GARCH option prices.[4,5] Also, analytic derivatives for GARCH model parameter estimation may be derived, as in this paper.

Recently, several authors have examined the class of 'NM-GARCH models', i.e. GARCH models where errors have a normal mixture conditional distribution. 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 components has constant variance. Earlier, Vlaar and Palm (1993) considered another restricted form of NM-GARCH, assuming a mixture of two normal distributions where the difference between the instantaneous variances of the components was constant, this way incorporating only constant jumps in the level of the variance. 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 instantaneous variances is constant, so the instantaneous kurtosis is constrained to be constant. 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:

$$
\begin{aligned}
&y_t = \boldsymbol{X_t}'\boldsymbol{\gamma} + \varepsilon_t \\
&\varepsilon_t \big| I_{t-1} \sim NM(p_1,\ldots,p_K;\mu_1,\ldots,\mu_K;\sigma_{1t}^2,\ldots,\sigma_{Kt}^2), \qquad \sum_{i=1}^{K} p_i = 1, \\
&\sigma_{it}^2 = \omega_i + \sum_{j=1}^{q} \alpha_{ij}\varepsilon_{t-j}^2 + \sum_{k=1}^{K}\sum_{j=1}^{p} \beta_{ikj}\sigma_{k,t-j}^2 \quad \text{for } i = 1,\ldots,K
\end{aligned}
\tag{3}
$$

[4] The tractability of normal mixture densities has led to their extensive use for option pricing. Recently Brigo and Mercurio (2000, 2001) proved that if the risk neutral density of the log price is a normal mixture then, under certain conditions, the local volatility function has a simple analytic form. A consequence is that option prices and hedge ratios are simply averages of Black-Scholes prices and hedge ratios weighted by the mixing law. Consequently, closed-form normal mixture GARCH option prices will be weighted sums of the closed-from normal GARCH option prices derived by Heston and Nandi (2000), where the weights are those given in the mixing law.

[5] Also, the short-term implied volatility smile effect of the normal mixture diffusion local volatility model matches the term structure of excess kurtosis that is implied by statistical volatility models, such as GARCH (Alexander et. al., 2003). Although the diffusion limit of the normal mixture GARCH will be a stochastic volatility model, recent research on the normal mixture diffusion encompasses this interpretation, where the mixing law gives the probabilities that the volatility takes a specific value. Normal mixture price (and discrete time returns) densities also result from other stochastic volatility models – see Andersen, Bollerslev and Diebold (2002) and Barndorff-Nielsen and Shephard (2002).

Copyright © 2003 Carol Alexander and Emese Lazar.

The individual variances are related through their common dependence on ε_t and also through cross-equation effects (where lagged values of the kth variance component affect the current value of each variance component). However, as Haas, Mittnik and Paolella have observed, the cross dependence of individual variances does not appear to lead to significant improvements of the model.

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. The difference between the two models is that, whilst a MS-GARCH model estimates the probability (time-varying) that each observation belongs to a given volatility regime, for the NM-GARCH model what is important is the overall probability (time-invariant) that a given regime occurs over the entire sample. Hamilton and 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 instantaneous 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 is the ruling regime. 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 compute the moments of the error term; (3) to derive explicit parameter conditions for the positivity of the second and fourth moments of the error term; (4) to assess the potential bias and inefficiency in parameter estimates, as a function of the mixing law parameters; and (5) to determine the optimal number of normal densities in the mixture when the model is applied to historical returns. To achieve this last aim we apply symmetric 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 describes the symmetric NM(K)-GARCH(1,1) model: a symmetric GARCH(1,1) with errors having conditional densities that are a mixture of K zero mean normal densities. Section three presents the simulation results for the estimation accuracy in the NM(2)-GARCH(1,1) model. Section four surveys the existing literature on

Copyright © 2003 Carol Alexander and Emese Lazar.

exchange rate modelling, describes the historical data used in this study and discusses our empirical results. Section five concludes.

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

The general model (3) as formulated by Haas, Mittnik and Paolella (2002) has a very large number of parameters. Since there is 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 $y_t = \varepsilon_t$.[6] Finally, and only this assumption has an obvious limit on our results, we shall assume the error term follows a conditional normal mixture distribution with *zero mean*:

$$\varepsilon_t \big| I_{t-1} \sim NM(p_1,\ldots,p_K;0_1,\ldots,0_K;\sigma_{1t}^2,\ldots,\sigma_{Kt}^2), \qquad \sum_{i=1}^{K} p_i = 1, \tag{4}$$

where the density $\eta(\varepsilon_t) = \sum_{i=1}^{K} p_i \phi_i(\varepsilon_t)$ is the normal mixture density with ϕ_i, i = 1,…,*K* representing *K* zero-mean normal density functions with variance at time t given by σ_{it}^2. A mixture density of normal densities with the same means has leptokurtosis – that is, it has heavier tails than the normal density of the same variance – but no skewness. Thus our results can only be applied to markets where the underlying price densities are expected to be symmetric and fat-tailed: that is, the foreign exchange markets.

The NM(*K*)-GARCH(1,1) model requires *K* equations to specify the conditional variance, and in the symmetric case the variance of each normal in the mixture is 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 $p_1 + \ldots + p_K = 1$, the last mixing parameter can be expressed as a function of the others, so the symmetric NM(*K*)-GARCH(1, 1) model that we study here has 4*K*-1 parameters:

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

To obtain the optimal values for the parameter estimates, we maximize $\sum_{t=1}^{T} \ln[\eta(\varepsilon_t)]$ using a gradient method.[7,8] In the earlier studies on normal mixture GARCH models, numerical approximations were

[6] after demeaning the series

[7] 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

Copyright © 2003 Carol Alexander and Emese Lazar.

implemented for the gradient vectors used to derive the optimal values for the parameter estimates and the Hessian matrices that are required to estimate the standard errors of the parameters. Here, in Appendix A, we derive analytic expressions for the first and second order derivatives of the likelihood function with respect to the parameters for the symmetric 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. Also, in order to avoid estimation problems and to ensure the conditional (instantaneous) and unconditional (long-term) variance and kurtosis of the error term exist and are positive and finite, we have imposed restrictions on the parameters (in the case that $K \geq 2$) as follows:

R1. To avoid negative values for the conditional variances all parameters need to be positive. Also, to ensure that the variance processes are not explosive, each β_i must be less than one. Thus we have the following first set of restrictions:

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

R2. All individual long-term variances must exist and be finite and positive. In fact, a finite and positive overall long-term variance is sufficient to ensure that all individual long-term variances are finite and positive. This observation follows from the restrictions R1 and the following relationship between the individual and overall long-term variances:

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

Appendix B derives the expressions for the long-term variances and kurtosis, giving the following expression for the overall long-term variance in the case of a symmetric NM(K)-GARCH(1,1) model:

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

The numerator of $E(\varepsilon_t^2)$ is always finite and positive, given R1, so a (necessary and sufficient) additional condition for a finite positive second moment is:

$$\sum_{i=1}^{K} \frac{p_i(1-\alpha_i-\beta_i)}{1-\beta_i} > 0 \tag{R2}$$

important one of these is the Broyden-Fletcher-Goldfarb-Shanno algorithm, based on an approximating and updating method for the information matrix.

[8] 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 (2000), an initial grid search is performed. However, the difficulty of optimisation increases with the number of parameters, thus with the number of components in the mixture.

Copyright © 2003 Carol Alexander and Emese Lazar.

It can be noticed that if the sum $\alpha_i + \beta_i < 1$ for all i, then this 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 probability-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 is negative. In fact, our own empirical results in section four and the results of Haas, Mittnik and Paolella (2002) have shown that the α parameter of the highest variance component may have to take values higher than 1 in order to capture a high level for the individual 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. Our third and final condition refers to the existence (positivity) of the fourth moment. If one of the components has $\alpha_i + \beta_i > 1$, the fourth moment might not exist for certain values of the parameters. Such a region for the mixing parameter, keeping the other parameters constant, is shown in Fig. 1. The graph also shows that the positivity of the second moment does not ensure the existence of the fourth moment. The final condition is therefore:

$$E(\varepsilon^4) > 0 \tag{R3}$$

[Fig. 1 about here]

3. Simulation Results

One of the main problems with the NM(3) and higher order GARCH models is that often at least one of the mixing parameters takes very low values. In this section we show that biased parameter estimates result for the variance component(s) having low weight in the mixture and the entire model will be adversely affected.[9] We used Monte Carlo simulations to check the accuracy of the symmetric NM(2)-GARCH(1,1) model parameter estimates. The reason for choosing a NM(2) model is that it has only seven parameters (so the likelihood surface is better conditioned than it is for higher order normal mixture GARCH models) but it can still account for the excess kurtosis in the data. A

[9] For an empirical illustration of this phenomenon, see the NM(3) results for the Japanese yen in tables 4 and 5.

Copyright © 2003 Carol Alexander and Emese Lazar.

symmetric NM(2)-GARCH(1,1) model is sufficient to illustrate the dependency of the estimation of the parameters on the mixing law.

The simulation has the following steps: first, the individual variance processes of the model are simulated together with the time series for the error term. Secondly, the parameters of this simulated process and their standard errors are estimated. We generated 5000 time series of length 2000 and hence estimated the symmetric NM(2)-GARCH(1,1) model 5000 times. We expect the histograms of the estimated parameters to be centred on their true values, and the standard error of the estimated parameters should be around the Cramér-Rao lower bound.[10]

The estimated means and standard errors of the parameter estimates are sensitive to the values chosen for the model parameters, and to the value of the mixing parameter p in particular. In order to investigate the sensitivity to p further, the simulations and model estimations were performed 17 times, for mixing parameter values of 0.1 up to 0.95 (with a step of 0.05) but with a fixed set for the other parameters.[11] The individual GARCH(1,1) parameters were chosen to be close to their empirical estimates when the NM(2)-GARCXH(1,1) model was implemented on historical daily exchange rates (see section 4). The parameter values chosen for these simulations were: $\omega_1 = 0.00001$, $\alpha_1 = 0.03$, $\beta_1 = 0.9$, $\omega_2 = 0.0001$, $\alpha_2 = 0.041$, $\beta_2 = 0.96$, $p = 0.1, 0.15, \ldots, 0.95$.

Fig. 2 summarizes our preliminary classification of estimation results into 'good' and 'unrealistic' estimates. It shows that, for non-extreme values of the mixing parameter, a higher percentage of the simulated time series lead to realistic estimates. However, the more extreme the value of p, the worse the estimation is – in the sense that marginal values of the parameters are obtained such as a β higher than 0.99 which is likely to result from a local, and not a global, optimum.

[Fig. 2 about here]

Collection 1 presents the estimation results for the parameters, illustrating the bias of the estimation as a function of the mixing parameter. In particular, ω and α have a positive bias whilst β is biased downwards; and the size of the bias is an inverse function of the mixing parameter associated with the individual density whose variance is being modelled. That is, the bias on the parameters of the first individual variance is higher for low values of p and the bias on the parameters of the second individual variance is higher for high values of p. Even the mixing parameter has a small upward bias,

[10] 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.
[11] p=0.05 is excluded because it does not satisfy the parameter conditions.

Copyright © 2003 Carol Alexander and Emese Lazar.

but only when it takes values close to zero. However, the overall and individual long-term volatilities are only very slightly biased downwards for very low values of p. The last graph shows that the excess kurtosis has an upward bias for extreme values of the mixing parameter.

Collection 2 shows the efficiency of the estimation comparing the estimated standard errors with the Cramér-Rao bound. Our results show that the estimation of the volatility parameters for a component of the mixture becomes more exact as the mixing parameter associated with this component increases. Intuitively, when the mixing parameter associated with a component is high, we have many observations drawn from this density leading to a better precision of the estimation of the associated parameters. Also, we see how nicely the standard errors of the estimation match the Cramér-Rao bound, indicating the efficiency of the estimation.[12] One good thing to conclude, comparing the two sets of graphs, is that in all cases the average bias is less than the standard error of the estimation.

[Collections 1 and 2 about here]

4. Empirical Application of the Model

The use of a GARCH framework for exchange rate modelling is an established approach in the literature. Applications arguably began with Hsieh (1988), who examined the statistical properties of daily exchange rates and concluded that exchange rate returns have a distribution which 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 (1989) proved 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. Similar results were found by Johnston and Scott (2000). Other studies using a GARCH framework to model the statistical properties of exchange rate rates include Baillie and Bollerslev (1989, 1991), Engle, Ito and Lin (1990) and Engle and Gau (1997).

The empirical evidence on the leptokurtosis 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. 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

[12] The reason for getting standard error estimates in some cases lower than the Cramér-Rao bound is that we use restrictions for the parameters and also that with the initial grid search we have a better chance of finding the global and not only local maximum.

Copyright © 2003 Carol Alexander and Emese Lazar.

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 rate rates, although they found that models based on leptokurtic distributions do better than those based on Gaussian distributions. They conclude that the best models do not provide a significantly better forecast than the 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 fully explain the excess kurtosis in the data. A possible explanation for their results is that symmetric student's 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 mixture of just two normal densities already has five parameters), so they are more flexible in capturing leptokurtosis.

Normal mixture distributions have been applied to unconditional 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.

Data

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. Daily returns are computed as the (annualised) difference in the logarithm of daily closing prices.[13] The time evolution of the returns is presented in Fig. 3, and Table 1 reports some statistical properties of the data.[14] From the first four moments of the unconditional distributions, we see that the mean returns are not significantly different from zero for

[13] Zero returns have been removed as they most often indicate missing data and distort the likelihood surface.

[14] 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.

Copyright © 2003 Carol Alexander and Emese Lazar.

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 currencies is high, especially for the Japanese yen. Also, the Ljung-Box statistic shows that the data provide no evidence of autocorrelation.

[Fig. 3 and Table 1 about here]

Since the focus of this section is to estimate the parameters of symmetric NM(K)- GARCH(1,1) models, for different values of K, and to test the moment specifications of these models, and since these models are designed to capture, specifically, the time-variation in the second and fourth moments of conditional distributions of returns, we need to test for the existence of these moments in the sample data used. We therefore compute the maximal moment exponent $a = \sup\{b > 0 : E|\varepsilon_t|^b < \infty\}$, as in Hill (1975). Let

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

where s is a positive integer and $\varepsilon_1 \leq \ldots \leq \varepsilon_n$ are the ordered returns. Hall (1982) showed that if n is large enough, $s=s(n)$ is a function of n of a pre-specified order, s/n small enough and the tails of the distribution have the asymptotic Pareto-Lévy form then $\sqrt{s}(\hat{a}_s - a)$ has an approximate normal distribution with standard deviation a. Table 2 presents the estimates for the maximal moment exponent and the test results for the existence of the fourth moment. According to the results, we cannot reject the existence of the fourth moment.

[Table 2 about here]

Results

Three symmetric NM(K)-GARCH(1,1) 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.[15] Of course, the NM(1)-GARCH(1,1) model is equivalent to the normal GARCH(1,1). In the second and third model, the error term follows a mixture of two and three normal densities with GARCH(1,1) conditional variance equations, respectively. Though useful for estimating models with many parameters – and the NM(3)-GARCH(1,1) has eleven – many 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 each model is estimated on three sub-periods: 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).

[15] Asymmetric and Exponential normal GARCH(1,1) models have also been fitted to the data but they provided no significant improvement related to the basic GARCH(1,1) model.

Copyright © 2003 Carol Alexander and Emese Lazar.

This split of the sample could also be useful for checking the robustness of our parameter estimates. However, from both Fig. 3 and Table 6 we see that, although the three sub-periods were simply chosen to be of roughly equal length, they do 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 1.58, the second sub-period is more stable, with a lower average volatility of 8.4%, but has a higher excess kurtosis, of 2.86, and the last sub-period has a very low average volatility, only 7.2%, and the excess kurtosis decreases to 1.12. Similarly, in the case of the euro, the first period is the most volatile, and the excess kurtosis, probably due to the outliers, 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 1.23. For the Japanese yen, the last period is the most volatile and also has an exceptionally high excess kurtosis, of 8.91. Given the different characteristics of each of the sub-periods, we do not expect the model parameter estimates to be nearly identical in all three periods.

Tables 3, 4 and 5 present the estimation results for the three exchange rates for the entire period and the three sub-periods. Twelve samples are considered (three exchange rates, each having four different sample periods) and three models are estimated for each sample. Standard errors are computed as the square root of the diagonal elements of the information matrix (see Appendix A for its derivation).

[Tables 3, 4 and 5 about here]

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

Note that our simulation results indicate that the parameter estimates are likely to be biased, or indeed convergence problems may be encountered in their estimation, if one of the regimes occurs with very low frequency. However, for the EUR and JPY rates, all estimates of the mixing law parameters are far from the boundary, so the parameters in the variance equations are estimated without bias. Also for the GBP rate, the mixing parameters are almost always higher than 0.3, so again no significant biases are expected in the estimation. The only exception is the 1993-1997 period, when the estimated mixing law is [0.88, 0.12] so that our estimate of 0.732 for α_2 is upwards biased and our estimate of

Copyright © 2003 Carol Alexander and Emese Lazar.

zero for β_2 is downwards biased. Obviously, the downwards bias on the estimate of β_2 is very substantial in this case.

As expected a high value for the unconditional kurtosis is associated with a high value for the extreme long-term volatility and/or a low probability for this extreme. For instance, the JPY rate has the highest long-term individual volatilities and these extreme volatilities are associated with very low probabilities, leading to a high model excess kurtosis, especially during the last two sub-periods.

When fitting a NM(3) model for the GBP and EUR rates, 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.[16] 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. In the case of the GBP rate, the 1993-1997 and 1998-2002 periods have a very low value for the third mixing parameter (0.05 and 0.02, respectively) and this could be the reason why, for these periods, our estimates of α_3 are greater than one and our estimate of β_3 is very low. For the EUR rate, all parameter estimates seem to be reasonable although the moment specification tests for the 1989-1992 and 1998-2002 periods – discussed in detail below – reveal that probably a local and not the global maximum has been reached.

The JPY rate has the peculiarity that the highest probability is associated with the component with the lowest long-term volatility and the component with highest long-term volatility has the smallest mixing parameter. However, from Table 4 we see that the estimation of the third volatility component for the Japanese yen is highly problematic and unrealistic parameter estimates were obtained.

To decide which model has the best in-sample fit, we have first applied the usual model selection criteria, namely the Akaike Information Criterion (AIC) and Schwartz's Bayesian Information Criterion (BIC) and the results are reported in the lower rows of Tables 3, 4 and 5. Schwarz's criteria always prefers the NM(2) specification, whilst the AIC is oscillating between the NM(2) and NM(3) models. In no case, for any exchange rate, and over any sub-period, is the normal GARCH(1,1) model the preferred specification, according to any of these criteria.

To check the adequacy of each model to capture the higher moments of the conditional returns densities, Tables 3, 4 and 5 also report the results of moment specification tests. In order to test whether the moments of the error densities match the ones specified by the estimated normal mixture distribution, the errors must be transformed into a series that has a standard normal distribution under

[16] For the first and last sub-period in the case of the EUR rate and for the third sub-period for the JPY rate, for the NM(3) model the results suggest that only local optimum was achieved. We consider that this is due to the irregularities of the likelihood surface and to the low number of observations, less than 1000 when estimating a high number of parameters (11).

Copyright © 2003 Carol Alexander and Emese Lazar.

the null hypothesis that the NM(K) 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 kth 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 i.i.d. standard normally distributed. Following 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 squares or fourth 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 even powers of the error term mentioned above are jointly true. Following Newey (1985), Engle, Lilien and Robins (1987), Nelson (1991), Greene (2000) and Brooks, Burke and Persand (2002), a Wald test approach is used and the derivation of the test statistic is given in Appendix C.

The above conditions test whether the estimated model's first four moments match the empirical ones – note that we are not modelling skewness – and whether there is any first order autocorrelation in these moments.[17] Our results show that the GARCH(1,1) model has severe rejections of model appropriateness, especially when testing the fourth moment. In a few isolated cases, the NM(2) model also rejects the 4th moment tests, showing that this model cannot always account for the excess kurtosis present in the data. Still, it gives a major improvement on the normal GARCH model. The NM(3) model can account for almost all excess kurtosis in the data, since the residuals show no clear-cut evidence of non-normality, but this is at the expense of the estimation problems discussed above. The NM(3) model has eleven parameters to be estimated by maximizing a highly non-linear likelihood function, and we have seen, both from our simulations and from our empirical results, that

[17] Moment tests for higher order of autocorrelation were also performed generally leading to non-rejection of normality.

Copyright © 2003 Carol Alexander and Emese Lazar.

this can lead to a substantial bias in the estimated parameters. It should also be noted that too good a fit is not always desirable because it might indicate noise-fitting, and poor out-of-sample performance.

Finally, we consider the relative properties of the three models when estimated over the entire fourteen year sample period, comparing the conditional volatility and excess kurtosis amongst the three models, and comparing the long-term excess kurtosis from each of the three models with that computed directly from the data. In making this last comparison, note that differences can arise from three sources: First, the mixing parameters could be close to boundary values, leading to biased parameter estimates, as shown by the simulation results; Secondly, just a few outliers can artificially influence the estimates (and this is the case in all GARCH models); Thirdly, when the unconditional parameters are estimated directly from the data we compute 'average' volatility and excess kurtosis within the sample, but when estimated via a normal mixture GARCH model, we are estimating long-term parameter values, being steady state limits of the conditional volatility and excess kurtosis series.

Collection 3 presents the conditional volatility and conditional excess kurtosis series derived from the estimated NM(2)- and NM(3)-GARCH(1,1) for the three exchange rates. These have a strong time-varying pattern, but the two volatility series shown on each graph are similar, for all three exchange rates.[18] The excess kurtosis graphs derived from the NM(2) and NM(3) models are more different: they do display a similar shape for the GBP and EUR rates (as expected, the excess kurtosis often increases at times when the conditional volatility decreases) but the NM(3) excess kurtosis is both higher and more volatile than the NM(2) excess kurtosis, particularly for the EUR rate. In the case of the JPY rate, the conditional excess kurtosis series from the NM(3) model is totally useless, probably due to the bias generated by the very low third mixing parameter.[19]

Table 6 presents the unconditional 'realised' excess kurtosis and the long-term excess kurtosis derived from each of the three models, estimated over the whole sample period and sub-periods.[20] Clearly the GARCH(1,1) model is not able to capture the full extent of the leptokurtosis in the data. The NM(2)-GARCH(1,1) model is closest to the realized excess kurtosis in 11 of the 12 cases and the NM(3)-GARCH(1,1) model greatly overestimates the excess kurtosis in the data.

[Collection 3 and Table 6 about here]

---

[18] They are also similar to the NM(1)-GARCH(1,1) conditional volatility series, although this series has not been shown.
[19] Note that the conditional kurtosis for the NM(1)-GARCH(1,1) model is zero.

[20] The unconditional realized excess kurtosis is computed as $K = \frac{\sum_{t=1}^{T}(\varepsilon_t - \mu)^4}{\sum_{t=1}^{T}(\varepsilon_t - \mu)^2} - 3$

Copyright © 2003 Carol Alexander and Emese Lazar.

6. Summary and Conclusions

This paper has examined the properties of the symmetric GARCH(1,1) model where the error term follows a normal mixture distribution, thus integrating the two major approaches in volatility modelling in a unified framework. Our contributions may be summarized as follows:

1. The analytic framework for the estimation of symmetric NM(K)-GARCH(1,1) models has been developed: we have derived expressions for the derivatives of the likelihood function, derived the moments of the error term and set specific parameter conditions for finite positive second and fourth moments;
2. Simulations have been used to examine the efficiency of parameter estimates, and to indicate the direction and size of the potential bias in the estimation, as a function of the mixing law;
3. Symmetric NM(K)-GARCH(1,1) models have been estimated using daily data for major US dollar exchange rates, over several different time periods. Information criteria and moment based specification tests on these models, long-term kurtosis estimates and observations on our simulation results lead us to conclude that the mixture of two normlas model is the preferred specification for all these series.

The analytic framework (detailed Appendix A) improves the efficiency of estimation algorithms and the specific conditions we set allow $\alpha + \beta > 1$ for some components. Simulations indicate potential convergence problems and a possible bias on parameter estimates for the variance components having low weights in the mixture. Our empirical results have a natural and appealing interpretation: for each of the data series, the estimated parameters of the conditional variance equations can easily be distinguished. In particular, in almost all cases the component with the lower volatility in the NM(2) model has the higher value for the mixing parameter. Thus the model is able to quantify two components in exchange rate volatility, a 'normal market circumstances' volatility process which occurs most of the time, and an 'extreme market circumstance' volatility process which occurs only rarely. When fitting a NM(3) model to the exchange rates, the component with the highest mixing parameter has an average long-term volatility. In this case the model quantifies two 'exceptional' market circumstances, i.e. relatively tranquil, and relatively volatile markets. However, the NM(3) model has 11 parameters, and this leads to estimation problems, the most severe of which is the bias on parameters of the 'high' and 'low' volatility regimes that occur only rarely. Based on all our results, we conclude that the NM(2)-GARCH(1,1) model is preferred.

Among the possible extensions of the model, the different types of asymmetric parameterisations would be of interest, as well as the multivariate case. Potential uses of the model include the use of the analytic term structure forecasts for the excess kurtosis and its application to option pricing and hedging models hypothesizing that the risk neutral density is a normal mixture.

Copyright © 2003 Carol Alexander and Emese Lazar.

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., Brintalos, G. and Nogueira, L. (2003): 'Short and Long Term Smile Effects: The Binomial Normal Mixture Diffusion Model', *ISMA Centre Discussion Papers in Finance* 2003-06

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 on Realized Exchange Rate Volatility', *Journal of the American Statistical Association*, Vol. 96, 43-76

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): 'The Kurtosis of GARCH and Stochastic Volatility Models' *Journal of Econometrics*, forthcoming

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

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

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

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

Copyright © 2003 Carol Alexander and Emese Lazar.

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

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 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., 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

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

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

Copyright © 2003 Carol Alexander and Emese Lazar.

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

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

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

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

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

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

McFarland, J.W., Petit, 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

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

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

Copyright © 2003 Carol Alexander and Emese Lazar.

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

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

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 Heteroscedasaticity, and Jumps', *Journal of Business & Economic Statistics*, Vol. 11, No. 3, 351-60

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

Copyright © 2003 Carol Alexander and Emese Lazar.

Appendix A. 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, which is a probability weighted average of K zero-mean normal density functions:

$$\varepsilon_t \big| I_{t-1} \sim NM(p_1,\ldots, p_K; 0_1,\ldots,0_K; \sigma_{1t}^2,\ldots,\sigma_{Kt}^2), \qquad \sum_{i=1}^{K} p_i = 1,$$

where the density of the error term is represented by the following mixing law (ϕ_i, $i = 1,\ldots, K$ represent zero-mean normal density functions):

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

Since the sum of the mixing law weights is one, it is only necessary to use (K-1) parameters for the probabilities. We use the following notation for the parameters:

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

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

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 $(y_1,\ldots, y_T)$.

To ease the analysis, in the following let g_i denote $p_i \frac{1}{\sigma_{it}} e^{-\frac{1}{2}\frac{\varepsilon_t^2}{\sigma_{it}^2}}$, $i = 1,\ldots, K$. It can be easily seen that g_i is a function of p_i and $\boldsymbol{\gamma}_i$ only, for $i = 1,\ldots, K - 1$ and g_K is a function of $\boldsymbol{p}$ and $\boldsymbol{\gamma}_K$. Also, we denote $\ln\left(\sum_{i=1}^{K} g_i\right)$ by $m_t(\boldsymbol{\theta})$. Using the new notation, our objective is to maximize $M(\boldsymbol{\theta}) = \sum_{t=1}^{T} m_t(\boldsymbol{\theta})$. The Newton-Raphson procedure 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 R1, R2, and R3 derived in Section 2.

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 (2000), an initial grid search is performed. However, the difficulty of optimisation increases with the number of parameters, thus with the number of components in the mixture.

Copyright © 2003 Carol Alexander and Emese Lazar.

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:

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

To compute the Hessian matrix and the gradient, 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:

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

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

Also, we have that

$$\frac{\partial \ln g_{\boldsymbol{i}}}{\partial \boldsymbol{\gamma_i}} = \left(\frac{1}{g_i}\right)\left(\frac{\partial g_{\boldsymbol{i}}}{\partial \boldsymbol{\gamma_i}}\right) \tag{2}$$

This way we get:

$$\frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i}} = \left(\frac{g_i}{\sum_{k=1}^{K} g_k}\right)\left(\frac{\partial \ln g_i}{\partial \boldsymbol{\gamma_i}}\right) \tag{3}$$

where

$$\frac{\partial \ln g_{\boldsymbol{i}}}{\partial \boldsymbol{\gamma_i}} = \left(\frac{1}{2\sigma_{it}^2}\right)\left(\frac{\varepsilon_t^2}{\sigma_{it}^2} - 1\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right) \tag{4}$$

From (3) and (4) we obtain the following result:

$$\frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i}} = \left(\frac{1}{2\sigma_{it}^2}\right)\left(\frac{g_i}{\sum_{k=1}^{K} g_k}\right)\left(\frac{\varepsilon_t^2}{\sigma_{it}^2} - 1\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right) \tag{5}$$

The second order derivatives are:

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial p_i \partial p_j} = -\frac{\left(\frac{g_i}{p_i} - \frac{g_K}{p_K}\right)\left(\frac{g_j}{p_j} - \frac{g_K}{p_K}\right)}{\left(\sum_{k=1}^{K} g_k\right)^2} \tag{6}$$

Copyright © 2003 Carol Alexander and Emese Lazar.

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i'} = \left(\frac{\partial \ln g_i}{\partial \boldsymbol{\gamma}_i}\right) \frac{\partial \left(\dfrac{g_i}{\sum_{k=1}^{K} g_k}\right)}{\partial \boldsymbol{\gamma}_i'} + \left(\frac{g_i}{\sum_{k=1}^{K} g_k}\right)\left(\frac{\partial^2 \ln g_i}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i'}\right) \tag{7}$$

To compute the partial derivatives on the right hand side of (7), we use (2) to derive the following expression:

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

which, using (4), simplifies to:

$$\frac{\partial \left(\dfrac{g_i}{\sum_{k=1}^{K} g_k}\right)}{\partial \boldsymbol{\gamma}_i'} = \left(\frac{1}{2\sigma_{it}^2}\right)\left(\frac{g_i\left(\sum_{k=1}^{K} g_k - g_i\right)}{\left(\sum_{k=1}^{K} g_k\right)^2}\right)\left(\frac{\varepsilon_t^2}{\sigma_{it}^2} - 1\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma}_i}\right)' \tag{8}$$

Using equation (4) again we obtain that:

$$\frac{\partial^2 \ln g_i}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i'} = \left(\frac{1}{2\sigma_{it}^2}\right)\left[\left(\frac{1}{\sigma_{it}^2}\right)\left(1 - \frac{2\varepsilon_t^2}{\sigma_{it}^2}\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma}_i}\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma}_i}\right)' + \left(\frac{\varepsilon_t^2}{\sigma_{it}^2} - 1\right)\left(\frac{\partial^2 \sigma_{it}^2}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i'}\right)\right] \tag{9}$$

Combining and grouping (4), (7), (8) and (9) we obtain:

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

$$+ \left(\frac{1}{2\sigma_{it}^2}\right)\left(\frac{g_i}{\sum_{k=1}^{K} g_k}\right)\left(\frac{\varepsilon_t^2}{\sigma_{it}^2} - 1\right)\left(\frac{\partial^2 \sigma_{it}^2}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i'}\right) \tag{10}$$

Using (5), we compute the cross-derivatives as:

Copyright © 2003 Carol Alexander and Emese Lazar.

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i} \partial p_j} = \left(\frac{1}{2\sigma_{it}^2}\right)\left(\frac{\varepsilon_t^2}{\sigma_{it}^2} - 1\right)\frac{\partial\left(\dfrac{g_i}{\sum_{k=1}^{K} g_k}\right)}{\partial p_j}\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right) \tag{11}$$

where

$$\frac{\partial\left(\dfrac{g_i}{\sum_{k=1}^{K} g_k}\right)}{\partial p_i} = \frac{\left(\dfrac{g_i}{p_i}\right)\left(\sum_{k=1}^{K} g_k\right) - \left(\dfrac{g_i}{p_i} - \dfrac{g_K}{p_K}\right) g_i}{\left(\sum_{k=1}^{K} g_k\right)^2} \qquad i = 1,\ldots, K-1 \tag{12A}$$

$$\frac{\partial\left(\dfrac{g_i}{\sum_{k=1}^{K} g_k}\right)}{\partial p_j} = \frac{-\left(\dfrac{g_j}{p_j} - \dfrac{g_K}{p_K}\right) g_i}{\left(\sum_{k=1}^{K} g_k\right)^2} \qquad i = 1,\ldots, K-1;\ \ j = 1,\ldots, K-1;\ \ i \neq j \tag{12B}$$

$$\frac{\partial\left(\dfrac{g_K}{\sum_{k=1}^{K} g_k}\right)}{\partial p_i} = \frac{-\left(\dfrac{g_K}{p_K}\right)\left(\sum_{k=1}^{K} g_k\right) - \left(\dfrac{g_i}{p_i} - \dfrac{g_K}{p_K}\right) g_K}{\left(\sum_{k=1}^{K} g_k\right)^2} \qquad i = 1,\ldots, K-1 \tag{12C}$$

Combining (11) and (12) we obtain:

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i} \partial p_i} = \left(\frac{-1}{2\sigma_{it}^2}\right)\left(\frac{\left(\dfrac{g_i}{p_i}\right)\left(\sum_{k=1}^{K} g_k\right) - \left(\dfrac{g_i}{p_i} - \dfrac{g_K}{p_K}\right) g_i}{\left(\sum_{k=1}^{K} g_k\right)^2}\right)\left(1 - \frac{\varepsilon_t^2}{\sigma_{it}^2}\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right) \qquad i = 1,\ldots, K-1 \tag{13A}$$

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i} \partial p_j} = \left(\frac{1}{2\sigma_{it}^2}\right)\left(\frac{\left(\dfrac{g_j}{p_j} - \dfrac{g_K}{p_K}\right) g_i}{\left(\sum_{k=1}^{K} g_k\right)^2}\right)\left(1 - \frac{\varepsilon_t^2}{\sigma_{it}^2}\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right) \qquad i = 1,\ldots, K-1\ j = 1,\ldots, K-1;\ \ i \neq j \tag{13B}$$

Copyright © 2003 Carol Alexander and Emese Lazar.

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_K} \partial p_i} = \left(\frac{1}{2\sigma_{3t}^2}\right)\left(\frac{\left(\frac{g_K}{p_K}\right)\left(\sum_{k=1}^{K} g_k\right) + \left(\frac{g_i}{p_i} - \frac{g_K}{p_K}\right) g_K}{\left(\sum_{k=1}^{K} g_k\right)^2}\right)\left(1 - \frac{\varepsilon_t^2}{\sigma_{Kt}^2}\right)\left(\frac{\partial \sigma_{3t}^2}{\partial \boldsymbol{\gamma_K}}\right) \qquad i = 1,\ldots, K-1 \tag{13C}$$

Again using (5), we can write the cross-derivative with respect to $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_j$ as:

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

Using (2) and (4) the above implies:

$$\frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_j}'} = \left(\frac{-1}{4\sigma_{it}^2 \sigma_{jt}^2}\right)\left(\frac{g_i g_j}{\left(\sum_{k=1}^{K} g_k\right)^2}\right)\left(1 - \frac{\varepsilon_t^2}{\sigma_{it}^2}\right)\left(1 - \frac{\varepsilon_t^2}{\sigma_{jt}^2}\right)\left(\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}}\right)\left(\frac{\partial \sigma_{jt}^2}{\partial \boldsymbol{\gamma_j}}\right)' \tag{14}$$

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

$$\frac{\partial \sigma_{it}^2}{\partial \boldsymbol{\gamma_i}} = \boldsymbol{z_{it}} + \beta_i \frac{\partial \sigma_{it-1}^2}{\partial \boldsymbol{\gamma_i}} \tag{15}$$

where $\boldsymbol{z_{it}} = (1, \varepsilon_{t-1}^2, \sigma_{it-1}^2)'$. 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 derivative is:

$$\frac{\partial^2 \sigma_{it}^2}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} = \boldsymbol{w_{it}} + \beta_i \frac{\partial^2 \sigma_{it-1}^2}{\partial \boldsymbol{\gamma_i} \partial \boldsymbol{\gamma_i}'} \tag{16}$$

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

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

Copyright © 2003 Carol Alexander and Emese Lazar.

Appendix B. Derivation of the Moments of the Error Term

The model is specified as:

$$y_t = \varepsilon_t \qquad \varepsilon_t \sim NM(p_1,...,p_K,0_1,...,0_K,\sigma_{1t}^2,...,\sigma_{Kt}^2) \qquad \sum_{i=1}^{K} p_i = 1$$

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

where the conditional variance is:

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

We use the following notations: $x = E(\varepsilon_t^2) = E(\sigma_t^2)$ and $y_i = E(\sigma_{it}^2)$ $i = 1,\ldots, K$. Taking expectations of (1) and (2) gives:

$$x = \sum_{i=1}^{K} p_i y_i \tag{3}$$

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

Equation (4) is equivalent to:

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

After solving the system formed by (3) and (4) we obtain:

$$x = E(\varepsilon_t^2) = E(\sigma_t^2) = \frac{\sum_{j=1}^{K} \frac{p_j \omega_j}{(1-\beta_j)}}{\left(1 - \sum_{j=1}^{K} \frac{p_j \alpha_j}{(1-\beta_j)}\right)} \tag{6}$$

$$y_i = E(\sigma_{it}^2) = \frac{\frac{\omega_i}{(1-\beta_i)} + \sum_{j=1}^{K} p_j \frac{(\alpha_i \omega_j - \alpha_j \omega_i)}{(1-\beta_i)(1-\beta_j)}}{\left(1 - \sum_{j=1}^{K} \frac{p_j \alpha_j}{(1-\beta_j)}\right)}$$

Now let:

$$z = E(\varepsilon_t^4) = 3\sum_{i=1}^{K} p_i E(\sigma_{it}^4)$$

Using the notations $v_i = E(\sigma_{it}^4)$, $\boldsymbol{v} = (v_1,\ldots,v_K)'$ and $\boldsymbol{p} = (p_1,\ldots,p_K)'$ we obtain

$$z = 3\boldsymbol{p}'\boldsymbol{v} \tag{7}$$

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

$$v_i(1-\beta_i^2) = \omega_i^2 + 2\omega_i\alpha_i x + 2\omega_i\beta_i y_i + \alpha_i^2 z + 2\alpha_i\beta_i E(\varepsilon_t^2\sigma_{it}^2) \tag{8}$$

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 $E(\varepsilon_t^2\sigma_{it}^2)$ so (8) becomes:

$$v_i(1-\beta_i^2) = w_i + \alpha_i^2 z + 2\alpha_i\beta_i m_i \tag{9}$$

Copyright © 2003 Carol Alexander and Emese Lazar.

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

Now by (1):

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

Using the notation $r_{ik} = \omega_i \omega_k + (\omega_i \alpha_k + \omega_k \alpha_i)x + \omega_i \beta_k y_k + \omega_k \beta_i y_i$ 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 (9) becomes:

$$m_i \left(1 - \sum_{\substack{k=1 \\ k\neq i}}^{K} \frac{p_k \alpha_k \beta_i}{1 - \beta_i \beta_k} \right) = p_i v_i + \left(\sum_{\substack{k=1 \\ k\neq i}}^{K} \frac{p_k r_{ik}}{1 - \beta_i \beta_k} \right) + \left(\sum_{\substack{k=1 \\ k\neq i}}^{K} \frac{p_k \alpha_i \alpha_k}{1 - \beta_i \beta_k} \right) z + \left(\sum_{\substack{k=1 \\ k\neq i}}^{K} \frac{p_k \alpha_i \beta_k}{1 - \beta_i \beta_k} m_k \right) \tag{10}$$

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

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

So (10) may be written in matrix form as:

$$\boldsymbol{m} = \boldsymbol{A}^{-1} \begin{pmatrix} p_1 v_1 + \left(\sum_{\substack{k=1 \\ k\neq 1}}^{K} \frac{p_k r_{1k}}{1 - \beta_1 \beta_k} \right) + \left(\sum_{\substack{k=1 \\ k\neq 1}}^{K} \frac{p_k \alpha_1 \alpha_k}{1 - \beta_1 \beta_k} \right) z \\ \vdots \\ p_K v_K + \left(\sum_{\substack{k=1 \\ k\neq K}}^{K} \frac{p_k r_{Kk}}{1 - \beta_K \beta_k} \right) + \left(\sum_{\substack{k=1 \\ k\neq K}}^{K} \frac{p_k \alpha_K \alpha_k}{1 - \beta_K \beta_k} \right) z \end{pmatrix}$$

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

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

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

Copyright © 2003 Carol Alexander and Emese Lazar.

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:

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

Substituting this expression into (9) yields:

$$v_i(1-\beta_i^2) = w_i + \alpha_i^2 z + 2\alpha_i \beta_i \left(c_i + d_i z + \sum_{j=1}^{K} e_{ij} v_j \right) \tag{11}$$

Let $\boldsymbol{B}$ denote the matrix:

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

and introduce the vectors: $\boldsymbol{f} = \begin{pmatrix} w_1 + 2\alpha_1\beta_1 c_1 \\ \vdots \\ w_K + 2\alpha_K\beta_K c_K \end{pmatrix}$ and $\boldsymbol{g} = \begin{pmatrix} \alpha_1^2 + 2\alpha_1\beta_1 d_1 \\ \vdots \\ \alpha_K^2 + 2\alpha_K\beta_K d_K \end{pmatrix}$ so that (11) may be written as:

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

Substituting this into (7) yields:

$$z = 3\boldsymbol{p}'\boldsymbol{B}^{-1}(\boldsymbol{f} + \boldsymbol{g}z)$$

which leads to the following result:

$$z = E(\varepsilon_t^4) = \frac{3\boldsymbol{p}'\boldsymbol{B}^{-1}\boldsymbol{f}}{1-3\boldsymbol{p}'\boldsymbol{B}^{-1}\boldsymbol{g}} \tag{12}$$

and the excess kurtosis is:

$$K = \frac{E(\varepsilon_t^4)}{E(\varepsilon_t^2)^2} - 3 = \frac{z}{x^2} - 3 \tag{13}$$

In the case of a mixture of two normal densities, equations (6) and (12) become:

$$E(\varepsilon_t^2) = E(\sigma_t^2) = \frac{\dfrac{p\omega_1}{(1-\beta_1)} + \dfrac{(1-p)\omega_2}{(1-\beta_2)}}{\left(1 - \dfrac{p\alpha_1}{(1-\beta_1)} - \dfrac{(1-p)\alpha_2}{(1-\beta_2)}\right)} \tag{14}$$

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

$$E(\varepsilon_t^4) = 3\frac{pA_1B_1 + (1-p)A_2B_2 + 2p(1-p)A_3B_3}{T_1 + T_2 + T_3 + T_4 + T_5} \tag{15}$$

Copyright © 2003 Carol Alexander and Emese Lazar.

where

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

$$B_1 = \omega_1^2 + 2\omega_1\alpha_1 E(\varepsilon_t^2) + 2\omega_1\beta_1 E(h_{1t})$$

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

$$B_2 = \omega_2^2 + 2\omega_2\alpha_2 E(\varepsilon_t^2) + 2\omega_2\beta_2 E(h_{2t})$$

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

$$B_3 = \omega_1\omega_2 + \left(\omega_1\alpha_2 + \omega_2\alpha_1\right)E(\varepsilon_t^2) + \omega_1\beta_2 E(h_{2t}) + \omega_2\beta_1 E(h_{1t})$$

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

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

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

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

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

Copyright © 2003 Carol Alexander and Emese Lazar.

Appendix C. 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, $\boldsymbol{r}(\hat{\boldsymbol{\theta}})$ can be written as:

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

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 $\hat{\boldsymbol{M}}$ denote the following $T \times J$ matrix:

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

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

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

The null hypothesis is that the restrictions are zero:

H_0: $r(\boldsymbol{\theta}) = 0$

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

$W = \boldsymbol{r}(\hat{\boldsymbol{\theta}})^T \hat{\boldsymbol{\Omega}}^{-1} \boldsymbol{r}(\hat{\boldsymbol{\theta}})$, which has a χ^2_J distribution under the null.

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

$$\hat{\boldsymbol{\Omega}} = \frac{1}{T^2}\left[\hat{\boldsymbol{M}}^T\hat{\boldsymbol{M}} - \hat{\boldsymbol{M}}^T\hat{\boldsymbol{D}}\left(\hat{\boldsymbol{D}}^T\hat{\boldsymbol{D}}\right)^{-1}\boldsymbol{D}^T\hat{\boldsymbol{M}}\right]$$

Copyright © 2003 Carol Alexander and Emese Lazar.

Fig. 1. The long-term volatility, the fourth moment and the excess kurtosis as a function of the first mixing parameter for the NM(2)-GARCH(1,1) model having the parameters $\omega_1 = 0.00001$, $\alpha_1 = 0.03$, $\beta_1 = 0.9$, $\omega_2 = 0.0001$, $\alpha_2 = 0.041$, $\beta_2 = 0.96$.

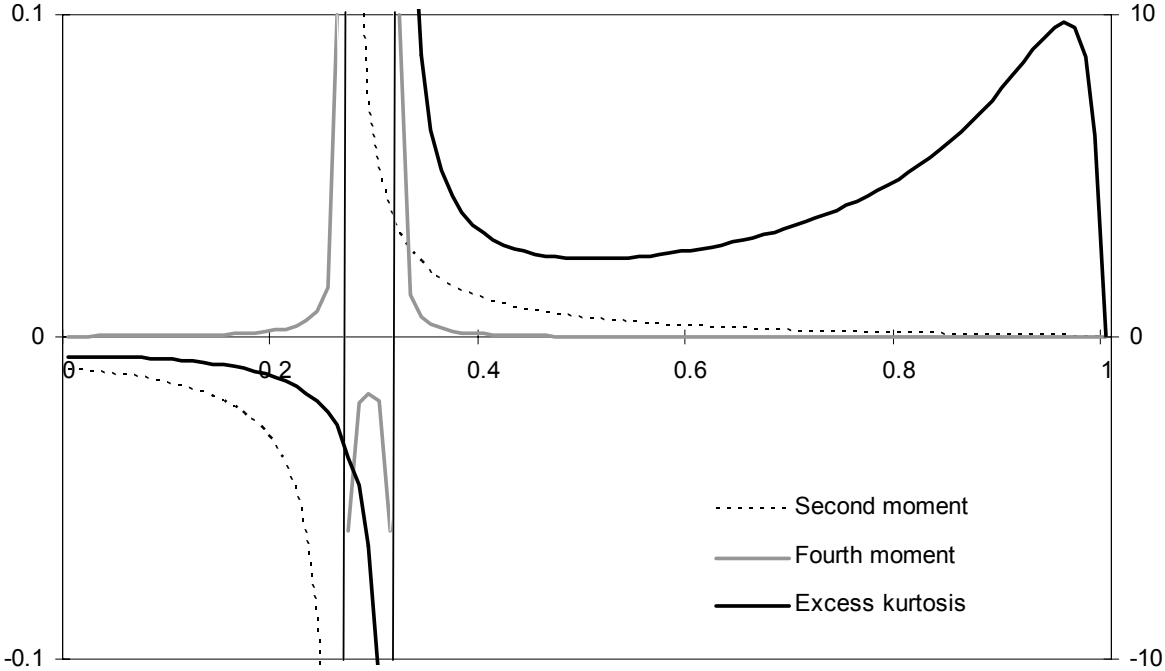
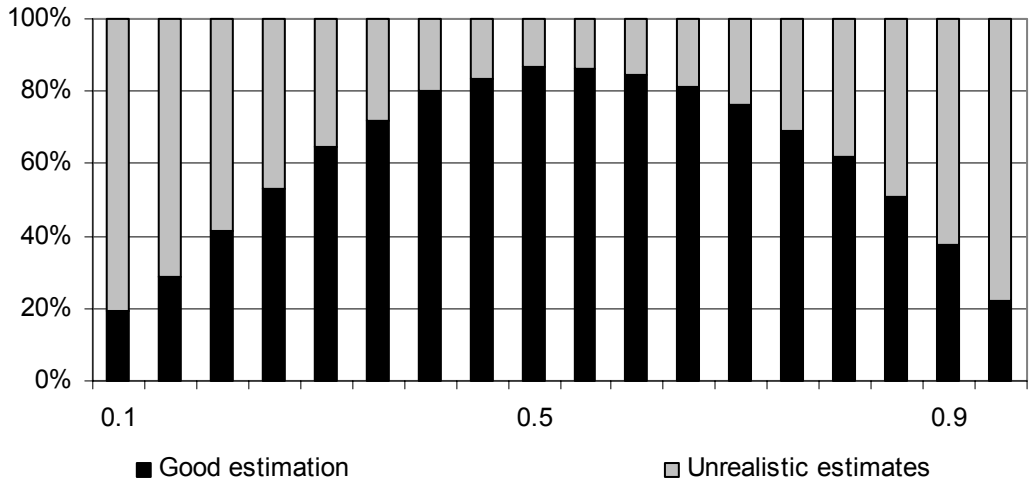
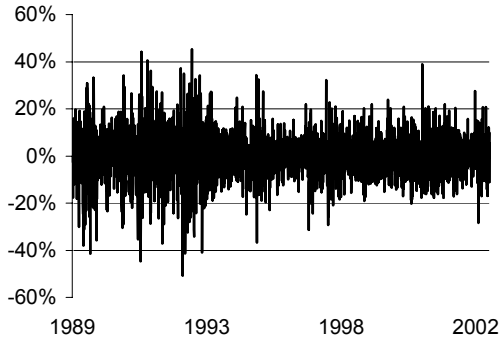


Fig. 2. Classification of Estimation Results

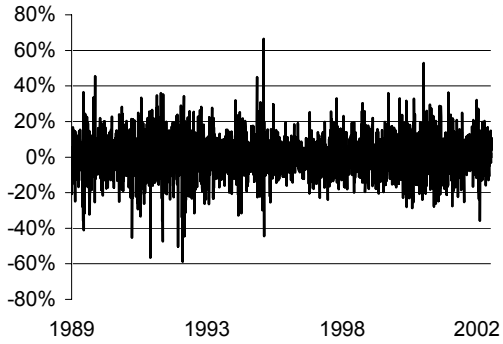


Copyright © 2003 Carol Alexander and Emese Lazar.

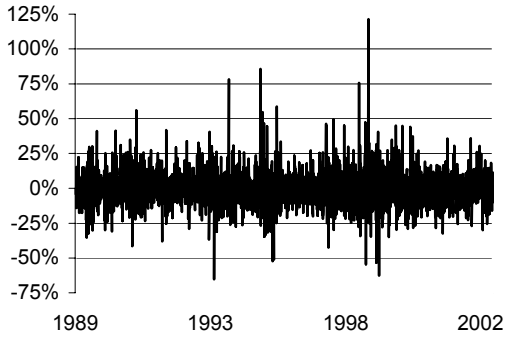
Fig. 3. The returns on the three exchange rates



(a) GBP/USD



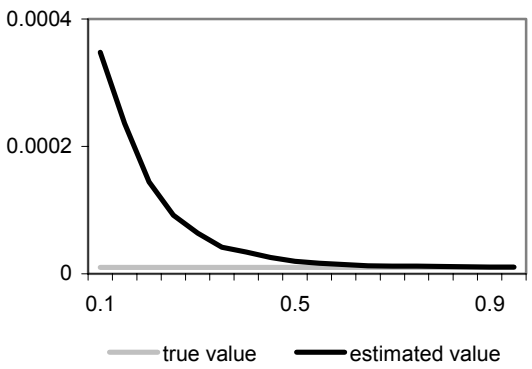
(b) EUR/USD



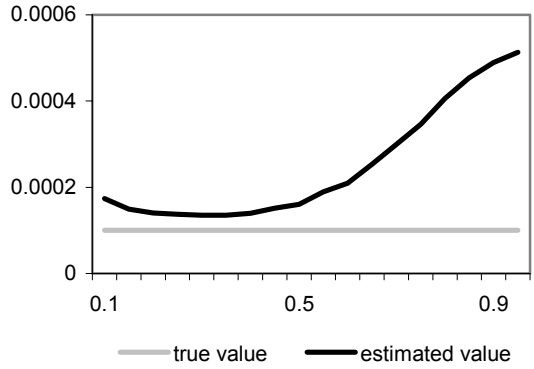
(c) JPY/USD

Copyright © 2003 Carol Alexander and Emese Lazar.

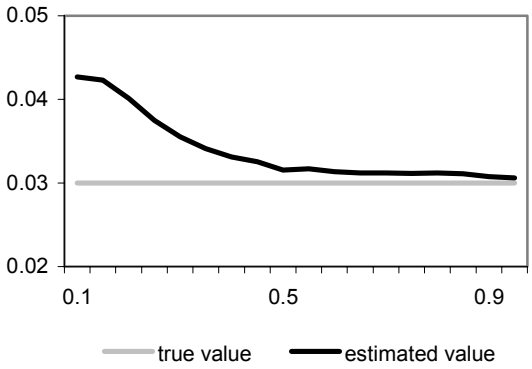
Collection 1. The estimation bias as the difference between the simulated and average estimated values of the parameters, long-term volatilities and excess kurtosis



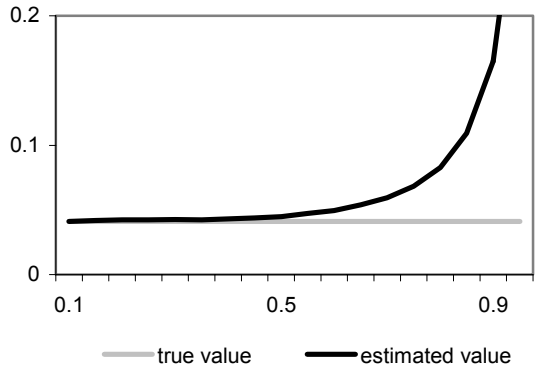
(a) ω_1



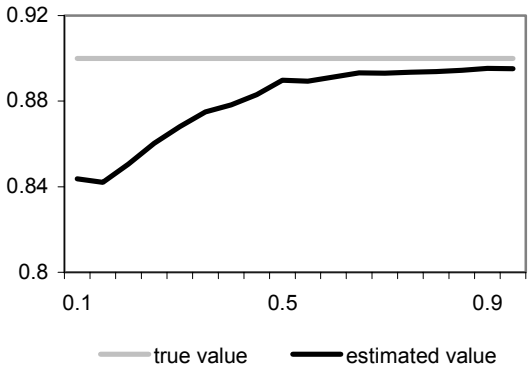
(b) ω_2



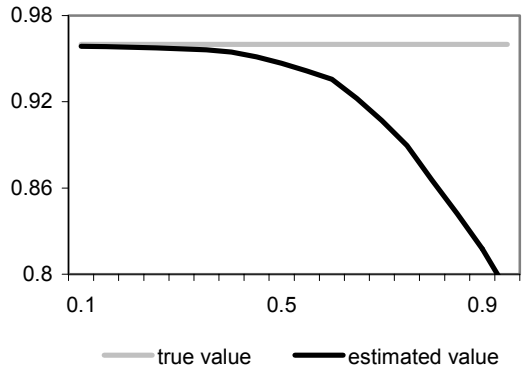
(c) α_1



(d) α_2



(e) β_1



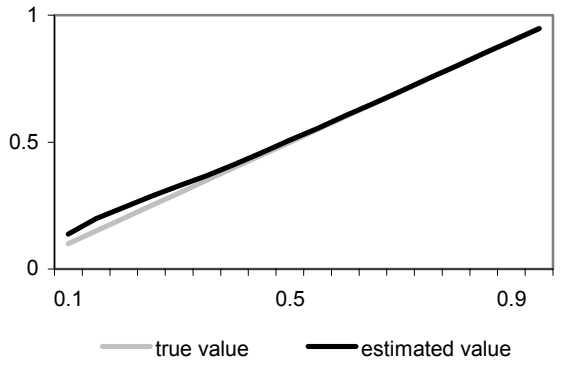
(f) β_2

Note: The data generating process is $y_t = \varepsilon_t, \quad \varepsilon_t \mid I_{t-1} \sim NM(p, 1-p, 0, 0, \sigma_{1t}^2, \sigma_{2t}^2)$

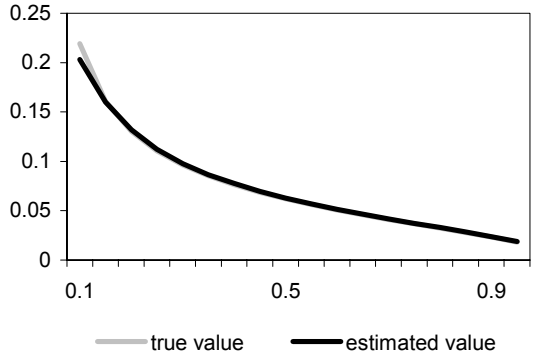
$$\sigma_{1t}^2 = 0.00001 + 0.03\varepsilon_{t-1}^2 + 0.9\sigma_{1,t-1}^2 \quad \sigma_{2t}^2 = 0.0001 + 0.041\varepsilon_{t-1}^2 + 0.96\sigma_{2,t-1}^2$$

Copyright © 2003 Carol Alexander and Emese Lazar.

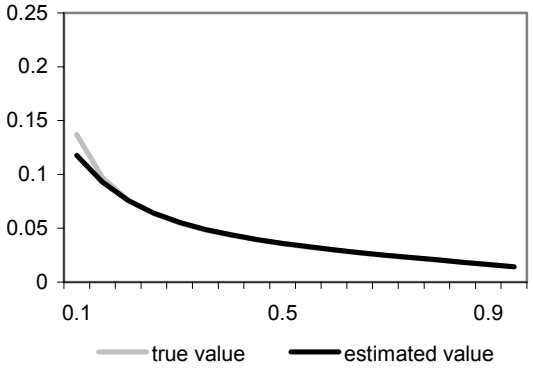
(continuation of Collection 1.)



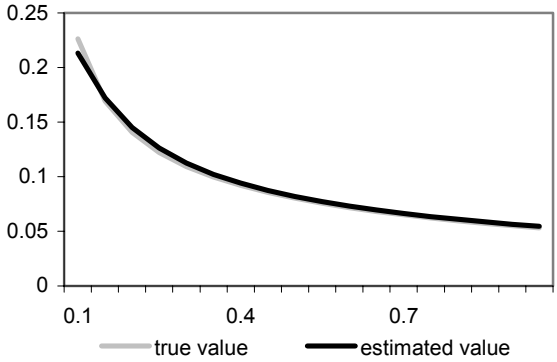
(g) *p*



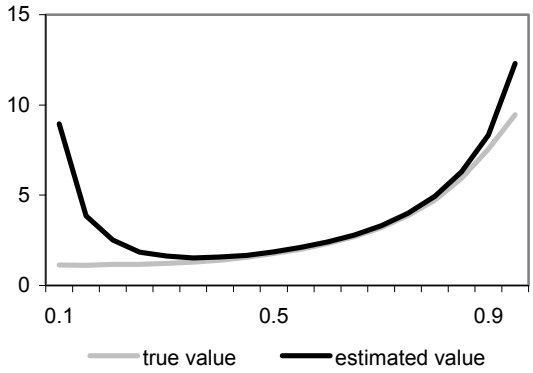
(h) long-term volatility



(i) long-term volatility 1



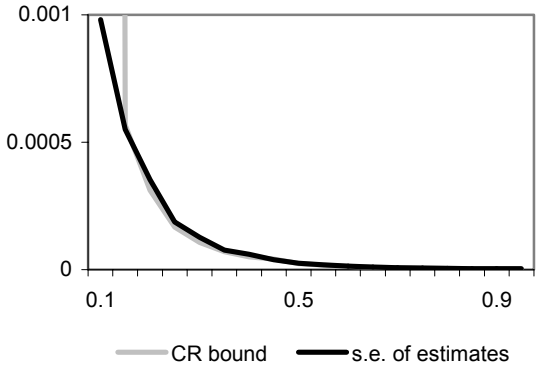
(j) long-term volatility 2



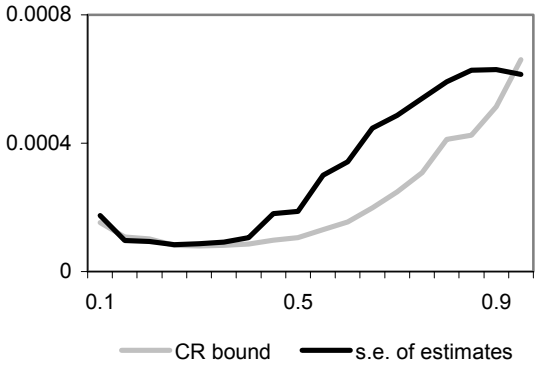
(k) excess kurtosis

Copyright © 2003 Carol Alexander and Emese Lazar.

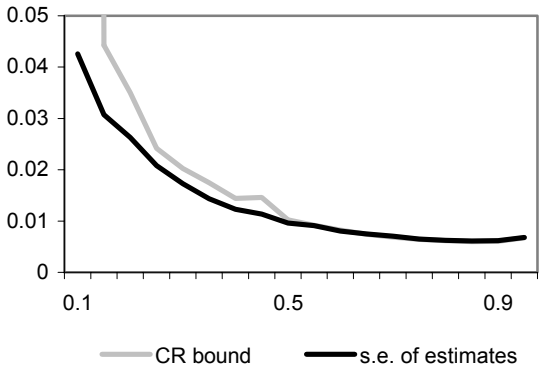
Collection 2. The efficiency of the estimation - a comparison of the Cramér-Rao bounds (where available) and the standard errors of the parameters, long-term volatilities and excess kurtosis



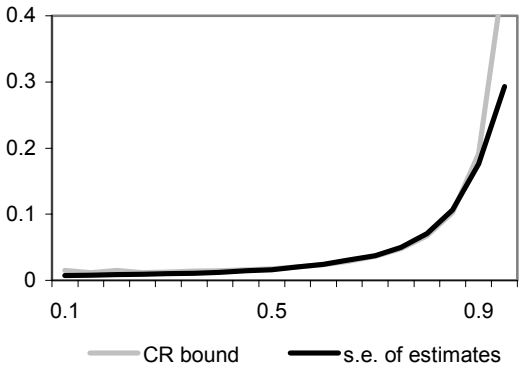
(a) ω_1



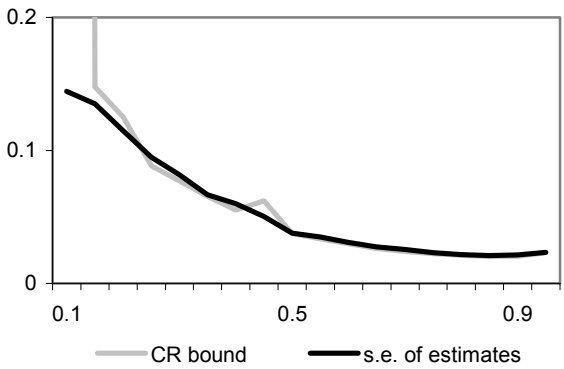
(b) ω_2



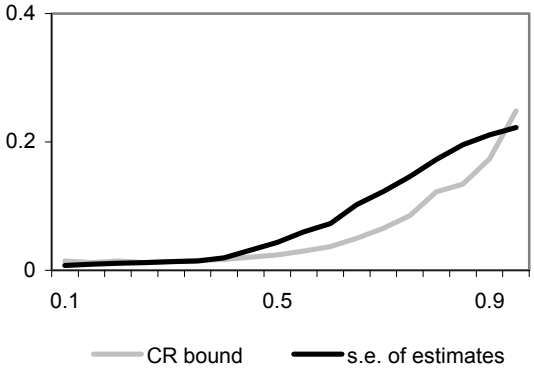
(c) α_1



(d) α_2



(e) β_1



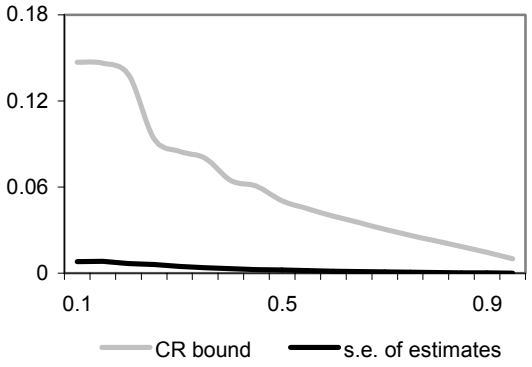
(f) β_2

Note: The data generating process is $y_t = \varepsilon_t, \quad \varepsilon_t \mid I_{t-1} \sim NM(p, 1-p, 0, 0, \sigma_{1t}^2, \sigma_{2t}^2)$

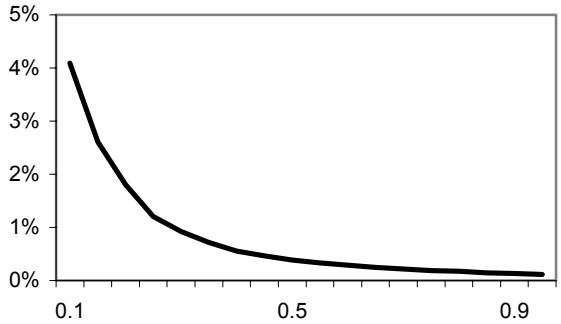
$$\sigma_{1t}^2 = 0.00001 + 0.03\varepsilon_{t-1}^2 + 0.9\sigma_{1,t-1}^2 \quad \sigma_{2t}^2 = 0.0001 + 0.041\varepsilon_{t-1}^2 + 0.96\sigma_{2,t-1}^2$$

Copyright © 2003 Carol Alexander and Emese Lazar.

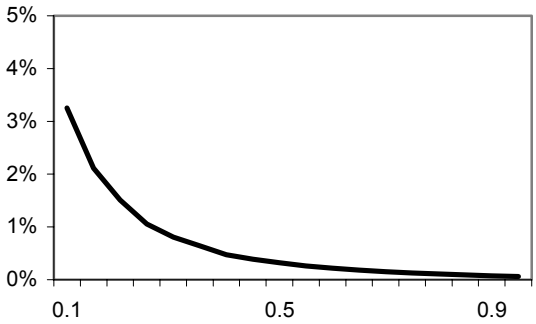
(continuation of Collection 2.)



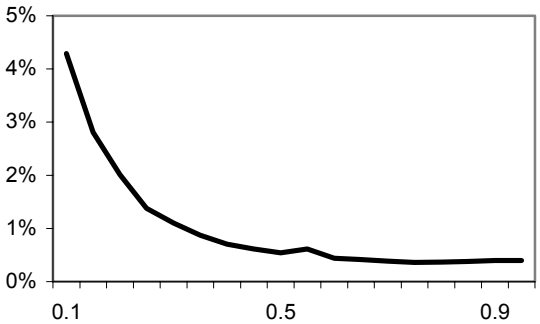
(g) *p*



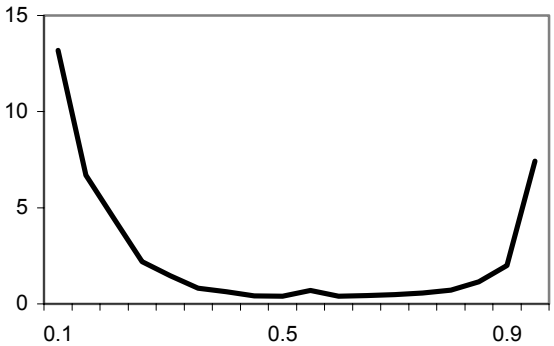
(h) long-term volatility



(i) long-term volatility 1



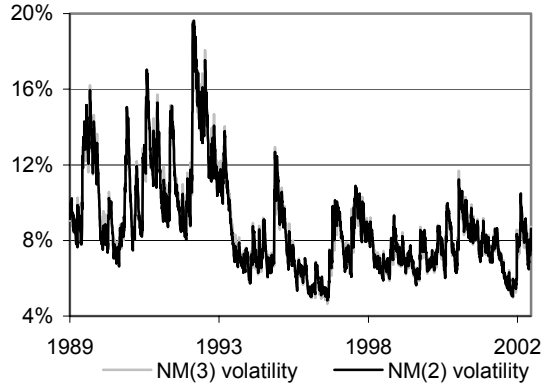
(j) long-term volatility 2



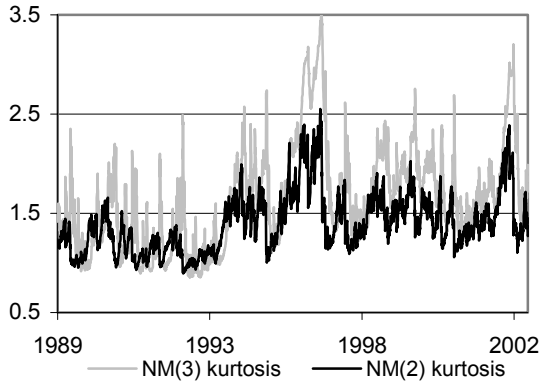
(k) excess kurtosis

Copyright © 2003 Carol Alexander and Emese Lazar.

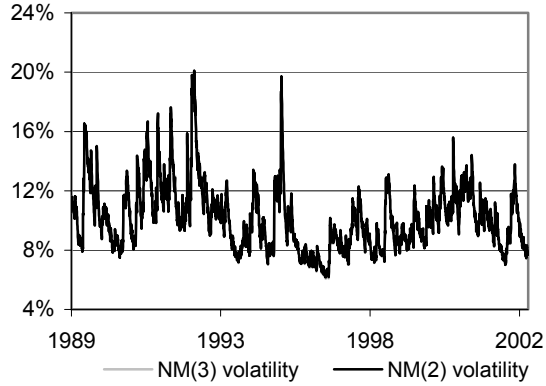
Collection 3. NM(2)- and NM(3)-GARCH(1,1) conditional volatilities and excess kurtosis's for the three exchange rates.



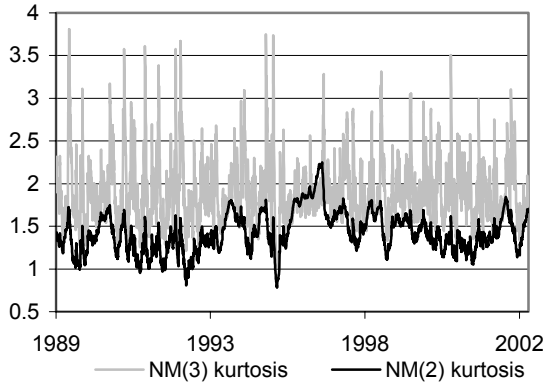
(a) Conditional volatility for GBP



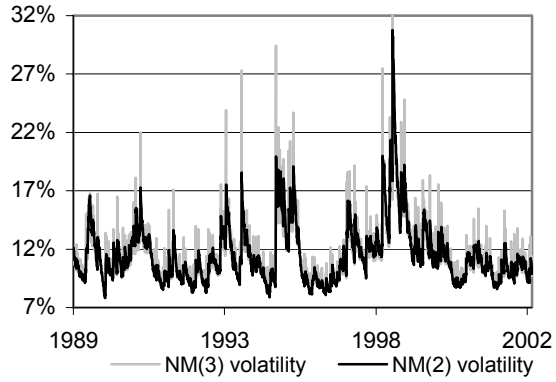
(b) Conditional excess kurtosis for GBP



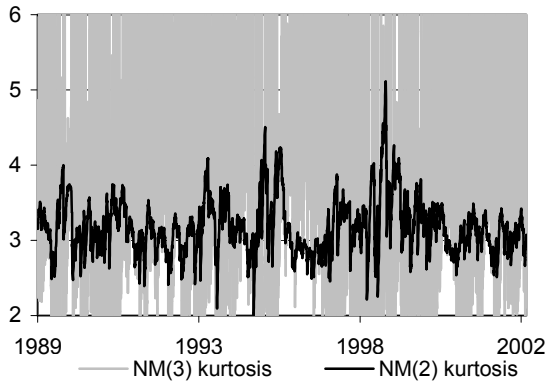
(c) Conditional volatility for EUR



(d) Conditional excess kurtosis for EUR



(e) Conditional volatility for JPY



(f) Conditional excess kurtosis for JPY

Copyright © 2003 Carol Alexander and Emese Lazar.

Table 1. Statistical description of the data

Returns on exchange rate	GBP	EUR	JPY
Mean return	-0.0005	-0.0005	0.0002
Standard deviation	0.0912	0.1035	0.1179
Skewness	-0.1533***	0.0256	0.7711***
Excess kurtosis	2.7285***	2.4232***	7.0853***
Minimum	-0.5089	-0.5885	-0.6536
Maximum	0.4535	0.6652	1.2139
Ljung-Box statistic (4 lags)	2.5708	3.5032	1.1121

Note: The standard errors for the skewness and kurtosis are approximated by $\sqrt{6/T}$ for the skewness and by $\sqrt{24/T}$ for the excess kurtosis, where T represents the total number of observations.
, ** 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

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

Note: s is defined as the percentage of number of the observations in the right (left) tail.

Copyright © 2003 Carol Alexander and Emese Lazar.

Table 3. Estimation results for different NM-GARCH(1,1) models, the GBP/USD exchange rate for different periods of time

Period	**2nd Jan 1989 – 31st Dec 2002**			**2nd Jan 1989 – 31st Dec 1992**			**2nd Jan 1993 – 31st Dec 1997**			**2nd Jan 1998 – 31st Dec 2002**		
Model	**NM(1)**	**NM(2)**	**NM(3)**	**NM(1)**	**NM(2)**	**NM(3)**	**NM(1)**	**NM(2)**	**NM(3)**	**NM(1)**	**NM(2)**	**NM(3)**
p1	*1*	*0.6603****	*0.6031****	*1*	*0.4919****	*0.3990***	*1*	*0.8801****	*0.5910***	*1*	*0.7060****	*0.8460****
ω1	7.0E-5***	9.5E-06	1.36E-5	0.0005**	0.0001	0.0007	0.0001***	8.8E-08	0.00001	0.0001**	8.3E-06	0.00004
α1	0.0442***	0.0290***	0.0261***	0.0756***	0.0509**	0.1170*	0.0384***	0.0266***	0.0286	0.0271**	0.0199**	0.0195**
β1	0.9472***	0.9441***	0.9724***	0.8926***	0.8578***	0.908***	0.9513***	0.9622***	0.9700	0.9466***	0.9677***	0.9735***
p2	-	*0.3397****	*0.2929****	-	*0.5081****	*0.3380***	-	*0.1199****	*0.3460*	-	*0.2940****	*0.1315***
ω2	-	0.0002*	1.26E-06	-	0.0007*	1E-10	-	0.0171	1E-10	-	0.0009	1E-10
α2	-	0.0789**	0.0486***	-	0.0966**	0.0611**	-	0.7320	0.0733**	-	0.0621	0.0221*
β2	-	0.9462***	0.8579***	-	0.9069***	0.841***	-	1E-10	0.8190***	-	0.8675***	0.8380
p3	-	-	*0.1041****	-	-	*0.2630***	-	-	*0.0630*	-	-	*0.0225**
ω 3	-	-	0.0008	-	-	0.0007	-	-	0.0190	-	-	0.0002
α3	-	-	0.2253*	-	-	0.0275	-	-	1.5850	-	-	2.7100
β3	-	-	0.8858***	-	-	0.873***	-	-	3.1E-07	-	-	0.5020
Long-term σ1	9.06%	6.71%	9.07%	12.05%	7.62%	15.98%	8.02%	7.03%	8.72%	7.22%	5.89%	7.72%
Long-term σ2	-	12.57%	5.30%	-	15.13%	7.43%	-	14.90%	5.51%	-	9.68%	2.83%
Long-term σ3	-	-	15.20%	-	-	9.39%	-	-	17.57%	-	-	18.00%
AIC	-2.10557	-2.15061	**-2.15485**	-1.51442	**-1.55244**	-1.54635	-2.24212	-2.31629	**-2.32205**	-2.42921	-2.44790	**-2.45019**
BIC	-2.09862	**-2.13843**	-2.13572	-1.49468	**-1.51789**	-1.49206	-2.22594	**-2.28798**	-2.27756	-2.41315	**-2.41978**	-2.40600
Moment specification tests:												
1st moment	1.803	0.574	0.561	13.073***	0.573	0.555	0.299	0.095	0.108	2.089	0.004	0.005
2nd moment	0.967	0.034	1.726	0.764	0.004	0.713	3.016	0.061	0.840	0.010	0.321	0.355
3rd moment	0.147	0.418	0.369	6.092*	0.803	0.803	1.502	0.090	0.067	6.231*	3.479	2.658
4th moment	44.056***	7.133**	0.077	20.343***	0.008	3.444	22.202***	8.522**	0.347	6.303*	3.069	0.867
1st moment AC(1)	0.002	0.204	0.226	1.269	0.741	0.732	2.395	3.316	3.365	0.001	0.012	0.001
2nd moment AC(1)	0.609	1.040	0.465	0.001	0.041	0.018	2.344	2.388	1.141	1.197	1.739	2.655
3rd moment AC(1)	5.457*	4.627*	4.051*	3.932*	4.133*	4.153*	0.004	0.047	0.026	0.877	0.636	0.447
4th moment AC(1)	0.902	0.217	0.306	0.447	0.564	0.395	1.332	1.645	1.688	5.231*	4.338*	5.908*
Cumulative test	46.240***	12.665*	7.502	20.784***	1.525	5.790	32.052***	25.886***	14.748**	10.904*	10.882*	8.019

*Notes: NM(1), NM(2) and NM(3) represent GARCH(1,1) models with a mixture of 1, 2 and 3 normal densities, respectively. For AIC and BIC, numbers in bold signify chosen models. The cumulative test is a joint test that the moment and AC conditions for the 2nd and 4th moments are met. Test statistics for the moment tests have a $\chi^2(1)$ distribution and for the cumulative test have a $\chi^2(4)$ distribution. *, ** and *** signify significance at 5%, 1% and 0.1% significance level, respectively.*

Copyright © 2003 Carol Alexander and Emese Lazar.

Table 4.Estimation results for different NM-GARCH(1,1) models, the EURO/USD exchange rate for different periods of time

Period	**2nd Jan 1989 – 31st Dec 2002**			**2nd Jan 1989 – 31st Dec 1992**			**2nd Jan 1993 – 31st Dec 1997**			**2nd Jan 1998 – 31st Dec 2002**		
Model	**NM(1)**	**NM(2)**	**NM(3)**	**NM(1)**	**NM(2)**	**NM(3)**	**NM(1)**	**NM(2)**	**NM(3)**	**NM(1)**	**NM(2)**	**NM(3)**
p1	*1*	*0.7299****	*0.5637****	*1*	*0.7465****	*0.4520*	*1*	*0.7567****	*0.5560****	*1*	*0.6022***	*0.7200****
ω1	0.00015***	0.00005	0.00017	0.0009**	0.00019	0.00086	0.0001***	0.00012	1E-10	0.00017**	0.00008	0.00015
α1	0.0475***	0.0254***	0.0441*	0.0908***	0.0300*	0.0639	0.0389***	0.0398***	0.0540**	0.0332***	0.0206*	0.0384**
β1	0.9392***	0.9501***	0.9509***	0.8517***	0.9301***	0.8770	0.9492***	0.9033***	0.9080***	0.9492***	0.9474***	0.9490***
p2	-	*0.2701****	*0.4033****	-	*0.2535****	*0.2690*	-	*0.2433****	*0.2880***	-	*0.3978***	*0.2190*
ω2	-	0.0005	0.00003	-	0.0028	1E-10	-	0.0002	0.0010	-	0.00029	0.00012
α2	-	0.1132**	0.0234**	-	0.2743	0.0174	-	0.0810*	0.0000	-	0.0451	0.0076
β2	-	0.9232***	0.9366***	-	0.7857***	0.9460	-	0.9563***	0.8500	-	0.9555***	0.9300***
p3	-	-	*0.0330***	-	-	*0.0890*	-	-	*0.1560**	-	-	*0.0610*
ω3	-	-	0.00094	-	-	0.0029	-	-	0.00008	-	-	0.00011
α3	-	-	0.3959	-	-	0.5140	-	-	0.1160*	-	-	1E-10
β3	-	-	0.8999***	-	-	0.7830	-	-	0.9570***	-	-	0.9970***
Long-term σ1	10.60%	8.11%	11.59%	12.46%	9.66%	12.21%	9.24%	6.99%	7.17%	9.93%	7.36%	10.47%
Long-term σ2	-	15.29%	6.73%	-	18.12%	7.03%	-	14.47%	8.16%	-	12.82%	5.34%
Long-term σ3	-	-	23.15%	-	-	22.27%	-	-	15.96%	-	-	18.80%
AIC	-1.77315	-1.81420	**-1.81578**	-1.41795	**-1.45048**	-1.44634	-2.01648	**-2.07602**	-2.07244	-1.81669	**-1.83564**	-1.83418
BIC	-1.76610	**-1.80185**	-1.79637	-1.39796	**-1.41548**	-1.39135	-1.99996	**-2.04710**	-2.02700	-1.80051	**-1.80732**	-1.78968
Moment specification tests:												
1st moment	0.135	0.0005	0	3.667	0.197	0.175	0.907	0.011	0.010	1.529	0.023	0.051
2nd moment	0.797	0.007	3.337	0.256	1.455	9.037**	2.217	0.062	0.514	2.354	2.638	113***
3rd moment	3.989*	3.461	3.656	3.290	0.182	0.123	5.764*	2.741	2.336	14.285***	9.577**	6.675**
4th moment	26.992***	9.425***	4.619*	10.913***	1.466	5.334*	10.272**	4.017*	2.009	8.249**	1.640	209***
1st moment AC(1)	0.853	0.482	0.441	0.128	0.116	0.089	0.192	1.168	1.282	2.590	2.781	2.952
2nd moment AC(1)	0.515	0.721	0.339	0.604	0.305	0.322	0.013	0.077	0.037	3.957*	3.215	3.530
3rd moment AC(1)	1.308	3.089	2.271	0.159	0.228	0.125	4.276*	4.354*	4.441*	0.879	1.288	1.533
4th moment AC(1)	2.346	1.895	0.496	2.953	1.298	1.090	0.493	0.174	0.108	11.615***	9.547**	7.401**
Cumulative test	30.129***	13.62**	11.689*	14.902**	6.636	16.532***	14.828**	5.104	3.633	16.739**	14.472**	225.3***

*Notes: NM(1), NM(2) and NM(3) represent GARCH(1,1) models with a mixture of 1, 2 and 3 normal densities, respectively. For AIC and BIC, numbers in bold signify chosen models. The cumulative test is a joint test that the moment and AC conditions for the 2nd and 4th moments are met. Test statistics for the moment tests have a $\chi^2(1)$ distribution and for the cumulative test have a $\chi^2(4)$ distribution. *, ** and *** signify significance at 5%, 1% and 0.1% significance level, respectively.*

Copyright © 2003 Carol Alexander and Emese Lazar.

Table 5.Estimation results for different NM-GARCH(1,1) models, the JPY/USD exchange rate for different periods of time

Period	**2nd Jan 1989 – 31st Dec 2002**			**2nd Jan 1989 – 31st Dec 1992**			**2nd Jan 1993 – 31st Dec 1997**			**2nd Jan 1998 – 31st Dec 2002**		
Model	**NM(1)**	**NM(2)**	**NM(3)**	**NM(1)**	**NM(2)**	**NM(3)**	**NM(1)**	**NM(2)**	**NM(3)**	**NM(1)**	**NM(2)**	**NM(3)**
p1	*1*	*0.8734****	*0.7330****	*1*	*0.7494****	*0.674****	*1*	*0.8838****	*0.5340****	*1*	*0.8700****	*0.7790****
ω1	0.0002***	0.00017**	0.00014*	0.00094**	1E-10	1E-10	0.0003***	0.00045*	0.0011	0.0001**	0.00015*	0.00014*
α1	0.0389***	0.0270***	0.0289***	0.0710***	0.0427**	0.041***	0.0388***	0.0334**	0.0606*	0.0305***	0.0150**	0.0157*
β1	0.9444***	0.9361***	0.9270***	0.8467***	0.9280***	0.926***	0.9427***	0.8815***	0.6080**	0.9634***	0.9600***	0.9570***
p2	-	*0.1266****	*0.2480****	-	*0.2506****	*0.268****	-	*0.1162****	*0.4360****	-	*0.1300****	*0.1610**
ω2	-	0.00049	0.00018*	-	0.00056	0.0044	-	0.0013	0.0002	-	1E-10	1E-10
α2	-	0.1139	0.0277**	-	0.0189	0.1410	-	0.2405	0.0319*	-	0.1080	0.0255*
β2	-	0.9600***	0.9780***	-	0.9693***	0.663***	-	0.9281***	0.9620***	-	0.9687***	0.9880***
p2	-	-	*0.0190***	-	-	*0.058****	-	-	*0.0300****	-	-	*0.0600*
ω3	-	-	0.0530	-	-	0.0003	-	-	0.0909	-	-	0.0173*
α3	-	-	4.0100	-	-	0.0500	-	-	3.2134	-	-	1.7510*
β3	-	-	1E-10	-	-	0.978***	-	-	0.0385	-	-	0.0112
Long-term σ1	11.93%	9.37%	8.54%	10.69%	8.24%	7.95%	12.29%	9.01%	7.11%	12.52%	9.58%	8.90%
Long-term σ2	-	23.11%	16.02%	-	15.90%	13.36%	-	26.45%	13.29%	-	22.33%	16.67%
Long-term σ3	-	-	32.90%	-	-	19.85%	-	-	37.91%	-	-	20.16%
AIC	-1.51958	-1.61830	**-1.62233**	-1.67244	**-1.73789**	-1.73183	-1.49622	-1.65267	**-1.66309**	-1.41679	-1.48825	**-1.49061**
BIC	-1.51247	**-1.60586**	-1.60278	-1.65238	**-1.70278**	-1.67666	-1.47961	**-1.62360**	-1.61741	-1.40044	**-1.45963**	-1.44563
Moment specification tests:												
1st moment	3.486	0.556	0.586	0.040	0.037	0.057	4.910	0.387	0.309	1.097	0.312	0.305
2nd moment	6.669**	5.917*	0.302	0.002	2.029	0.366	2.725	23.057***	0.967	0.196	3.157	7.339**
3rd moment	20.014***	6.343*	4.109*	6.563*	2.769	2.737	14.890***	3.659	2.422	9.631**	0.288	0.055
4th moment	30.900***	11.448***	0.896	18.961***	0.320	4.588*	27.620***	9.588**	3.190	29.545***	1.074	1.790
1st moment AC(1)	0.199	0.358	0.375	0.430	0.282	0.307	1.452	1.623	1.688	0.012	0.060	0.081
2nd moment AC(1)	1.720	0.248	0.107	1.604	0.010	1.572	0.908	0.557	0.131	3.229	0.042	6.548*
3rd moment AC(1)	0.968	0.157	0.183	0.006	1.186	1.020	1.076	0.351	0.017	6.793**	3.311	0.416
4th moment AC(1)	1.304	1.470	0.000	1.703	0.872	3.757	3.551	0.319	1.146	6.834***	10.584***	0.119
Cumulative test	39.392***	17.38**	1.754	22.543***	4.599	7.100	33.788***	26.772***	5.688	38.040***	35.015***	33.700***

*Notes: NM(1), NM(2) and NM(3) represent GARCH(1,1) models with a mixture of 1, 2 and 3 normal densities, respectively. For AIC and BIC, numbers in bold signify chosen models. The cumulative test is a joint test that the moment and AC conditions for the 2nd and 4th moments are met. Test statistics for the moment tests have a $\chi^2(1)$ distribution and for the cumulative test have a $\chi^2(4)$ distribution. *, ** and *** signify significance at 5%, 1% and 0.1% significance level, respectively.*

Copyright © 2003 Carol Alexander and Emese Lazar.

Table 6.The realized and the modelled kurtosis for the three exchange rates for different periods

Exchange rate	***Period***	***Realized excess kurtosis***	***Long-term excess kurtosis***		
			NM(1)	***NM(2)***	***NM(3)***
GBP	1989-2002	*2.73*	0.89	**3.59**	4.33
	1989-1992	*1.58*	0.67	**2.28**	2.78
	1993-1997	*2.86*	0.50	**3.50**	10.14
	1998-2002	*1.12*	0.09	**0.96**	28.76
EUR	1989-2002	*2.42*	0.61	**2.82**	4.75
	1989-1992	*1.89*	0.52	**2.96**	4.54
	1993-1997	*3.80*	0.44	**3.58**	4.74
	1998-2002	*1.23*	0.20	**1.22**	1.59
JPY	1989-2002	*7.08*	0.30	**5.34**	NA
	1989-1992	*1.92*	0.21	**1.77**	2.30
	1993-1997	*6.79*	0.27	**10.46**	NA
	1998-2002	*8.91*	0.54	5.28	**8.47**

Note: Numbers in bold represent values that are closest to the realized excess kurtosis.

Copyright © 2003 Carol Alexander and Emese Lazar.