IMF Working Paper

WP/98/29

INTERNATIONAL MONETARY FUND

Policy Development and Review Department

Parametric Distributional Flexibility and Conditional Variance Models with an Application to Hourly Exchange Rates

Prepared by Jenny N. Lye, Vance L. Martin, and Leslie E. Teo[1]

Authorized for distribution by Robert Sharer

March 1998

Abstract

This paper builds on the ARCH approach for modeling distributions with time-varying conditional variance by using the generalized Student t distribution. The distribution offers flexibility in modeling both leptokurtosis and asymmetry (characteristics seen in high-frequency financial time series data), nests the standard normal and Student t distributions, and is related to the Gram Charlier and mixture distributions. An empirical ARCH model based on this distribution is formulated and estimated using hourly exchange rate returns for four currencies. The generalized Student t is found to better model the empirical conditional and unconditional distributions than other distributional specifications.

JEL Classification Numbers: C10, C50, F31

Keywords: ARCH, Generalized Student t Distributions, Modeling Variance, Exchange Rates

Author's E-Mail Address: lteo@imf.org

---

[1] Jenny Lye and Vance Martin are at the University of Melbourne, Australia. All correspondence should be directed to Teo at the International Monetary Fund, Washington, DC 20431, USA. We would like to thank three referees, C. Shiells, and P. Christoffersen for providing many invaluable comments and insights on a previous version of this paper. Special thanks to Tim Bollerslev for providing the data.

Contents

Figures

SUMMARY

The ARCH framework has been found to be very successful in capturing many of the empirical characteristics of exchange rates and other financial variables. However, there are situations when the approach has not been able to explain all of the empirical characteristics of the data. These include when data are highly skewed and leptokurtic, such as intraday financial data or when large movements in asset prices occur.

The paper builds on the ARCH framework for modeling distributions with time-varying conditional variances by using a subordinate class of distributions from the generalized exponential family known as the generalized Student *t* distribution. The distribution offers flexibility in modeling both leptokurtosis and asymmetry, nests the standard normal and Student *t* distributions, and is related to the Gram Charlier and mixture distributions. An empirical model is formulated and estimated using hourly exchange rate returns for four currencies that combine ARCH-type conditional variances and nonnormal conditional distributions. Overall, the generalized Student *t* distribution is found to perform better overall in terms of modeling both the empirical conditional and unconditional distributions than other conditional variance models based on alternative distributional specifications.

I. INTRODUCTION

It is now common practice in the modelling of foreign exchange rates to use models that allow for time variation in the second moment based on the ARCH framework introduced by Engle (1982) and extended by Bollerslev (1986) and Nelson (1991).[1] Key features of the earliest versions of the ARCH class of models are that the conditional distribution is assumed to be normal, the unconditional distribution is leptokurtic as it exhibits fatter tails and a sharper peak than the normal distribution, and because of the normality assumption the conditional distribution is not skewed.

For a range of data sets on exchange rates, the ARCH framework has been found to be very successful in capturing many of the empirical characteristics in the data. However there are situations when some empirical characteristics are left unexplained. For example, Baillie and Bollerslev (1989) replaced the conditional normality assumption by a Student t distribution and a power exponential distribution to model the excess kurtosis in the conditional distribution of exchange rate returns, whereas Hsieh (1989) experimented with a range of conditional nonnormal distributions. In studying the British Pound/US$ exchange rate, Gallant, Hsieh and Tauchen (1991) and Engle and González-Rivera (1991), found evidence of skewness in the conditional distribution.[2] These authors also found evidence of multimodality in the conditional distribution with side lobes occurring in the tails of the distribution. Gallant, Hsieh and Tauchen used a seminonparametric approach by defining the conditional distribution as the product of a normal distribution and a polynomial based on a truncated Hermite expansion. A similar approach was adopted by Lee and Tse (1991) in which they referred to the distribution as a Gram Charlier distribution. In contrast, Engle and González-Rivera used a semiparametric approach with the conditional distribution approximated by a kernel density.[3]

This paper introduces a class of parametric ARCH models that is based on a density that provides flexibility in modelling not only symmetric fat-tailed distributions, but also distributions that are skewed. Such characteritics are seen in high frequency data and when there are large movements in the data (for example, during a asset market crash). This distribution is called the generalized Student t, which is a subordinate of the generalized exponential distribution (Lye and Martin (1993)). This distribution is shown to contain the standard Student t and normal distributions as special cases. It is also shown to be related to the seminonparametric class of Gallant, Hsieh and Tauchen (1991) and Lee and Tse (1991), as well as to mixture distributions which have been investigated by Friedman and Vandersteel (1982) and studied recently by Phillips (1994) in the robust estimation literature.

[1]For a review of this literature and a discussion of further generalizations, see Bollerslev, Engle and Nelson (1994). For a review of much of the applied work see Bollerslev, Chou and Kroner (1992).

[2]It has been argued by de Vries (1994) that the distribution of exchange rate returns can exhibit skewness when countries operate different monetary policies.

[3]An alternative approach to specifying a parameteric density, which is related to the semiparametric approach of Engle and Gonzalez-Rivera (1991), is the adaptive estimator suggested by Linton (1993).

The use of the generalized Student t in modelling the distribution of financial asset prices is shown to offer many advantages. First, this distribution can display a range of distributional properties such as skewness and kurtosis, with a parsimonious set of parameters. Second, as the generalized Student t includes the normal and Student t distributions, this facilitates the use of standard testing procedures based on Lagrange multiplier tests to discriminate between alternative models. Third, as the parameters of the model can be estimated easily by conventional maximum likelihood algorithms, the use of the generalized Student t distribution is potentially more simple, and possibly more convenient, than other distributions based on either mixture distributions or kernel densities. Fourth, the parameter estimates may, in general, be more efficient than those based on quasi-maximum likelihood methods where the conditional distribution is assumed to be normal. Even though Weiss (1986) and Bollerslev and Wooldridge (1992) have shown that the quasi-maximum likelihood estimator is consistent assuming the mean and the variance of the process are specified correctly, in general the estimates are inefficient with the degree of inefficiency increasing the more that the true conditional distribution departs from normality.[4] Finally, the density is always well defined which contrasts with the Gram Charlier distribution of Lee and Tse (1991) where it is possible for the density to be negative over certain regions.[5]

The paper proceeds as follows. The generalized Student t distribution is discussed in Section II. Special attention is given to discussing the relationships between this distribution and distributions of the standard Pearson family, and both the Gram Charlier and mixture distributions. A class of parametric ARCH models is motivated in Section III by showing that the generalized Student t distribution arises as the solution of a continuous time approximation to a discrete ARCH model. Also discussed in this section are estimation and computational issues. In Section IV, the model is used to characterize the distribution of hourly returns of four exchange rates. The results obtained are also contrasted for a number of alternative distributional models, including the normal, Student t, Gram Charlier, mixture and a skewed Student t density which is also known as Pearson's Type IV density. Concluding comments are given in Section V.

II. DISTRIBUTIONAL FLEXIBILITY AND GENERALIZED Student t DISTRIBUTIONS

The generalized Student t distribution represents an extension of the Pearson exponential family. The properties of this family of distributions have previously been studied by O'Toole (1933a, 1933b), Cobb (1978), Cobb, Koppstein and Chen (1983), Cobb and Zacks (1985), Martin (1990), and Lye and Martin (1993). The generalized Student t distribution can be derived from a generalization of the Pearson differential equation as follows

$$\frac{df}{dz} = \frac{-g(z)f(z)}{h(z)}, \tag{1}$$

[4] See Engle and González-Rivera (1991).

[5] See, for example, Kendall and Stuart (1963).

where $g(z)$ and $h(z)$ are polynomials in the random variable z, and $f(z)$ is the density function of z. In the standard Pearson system, $g(z)$ is a polynomial in z of degree less than or equal to one, whereas $h(z)$ is a polynomial in z of degree less than or equal to two. The general solution of (1) is

$$f(z) = \exp\left[-\int^{z}\left(\frac{g(s)}{h(s)}\right)ds - \eta\right], \qquad z \in D, \tag{2}$$

where the normalizing constant is given by

$$\eta = \ln\int \exp\left[-\int^{z}\left(\frac{g(s)}{h(s)}\right)ds\right]dz. \tag{3}$$

The domain D, of $f(z)$ in (2) is the open interval where $h(z)$ is positive.

The choice of $g(.)$ and $h(.)$ for the generalized Student t distribution are[6]

$$g(z) = \sum_{i=0}^{M-1}\alpha_i z^i, \tag{4}$$

$$h(z) = \gamma^2 + z^2. \tag{5}$$

By substituting (4) and (5) in (2), the generalized Student t distribution is given by

$$f(z) = \exp\left[\theta_1\tan^{-1}(z/\gamma) + \theta_2\ln(\gamma^2+z^2) + \sum_{i=3}^{M}\theta_i z^{i-2} - \eta\right], \quad -\infty < z < \infty, \tag{6}$$

where

$$\eta = \log\int \exp\left[\theta_1\tan^{-1}(z/\gamma) + \theta_2\ln(\gamma^2+z^2) + \sum_{i=3}^{M}\theta_i z^{i-2}\right]dz, \tag{7}$$

and the distribution parameters θ_j, are functions of the parameters $\{\alpha_j, j = 0, 1, ..., M-1; \gamma\}$ given in (4) and (5). Provided that in (6) $\theta_M < 0$, all moments of the distribution exist. This distribution can exhibit a range of shapes, including fat tails, sharp peaks, and even multimodality.[7]

A. Relationship With Pearson Distributions

As the generalized Student t distribution given by (6) is derived from an extension of the Pearson family, it directly contains many of the Pearson subordinate distributions as special cases. In particular, from the point of view of existing ARCH models, these special cases include the normal and Student t distributions. The standard normal distribution occurs when $\theta_3 = -0.5$, and all remaining parameters are zero. The Student t distribution occurs when $\theta_2 = -0.5(1+\gamma^2)$, and all remaining parameters are zero.

---

[6]See Lye and Martin (1993).

[7]Lye and Martin (1993) provide a discussion of the properties of the generalized Student t distribution.

A special case which terms out to be important in the empirical application is given by (6), with $\theta_i = 0$, $i > 2$, and $\theta_2 = -0.5(1+\gamma^2)$

$$f(z) = \exp\left[\theta_1 \tan^{-1}(z/\gamma) - 0.5(1+\gamma^2)\ln(\gamma^2+z^2) - \tilde{\eta}\right], \quad -\infty < z < \infty, \tag{8}$$

where $\tilde{\eta}$, is the normalizing constant. This distribution is referred to as the skewed Student t distribution where skewness is controlled by the parameter θ_1.[8] When $\theta_1 = 0$, there is no skewness and the distribution becomes the Student t distribution.

B. Relationship With Gram Charlier Distributions

In formulating an ARCH model of the term structure, Lee and Tse (1991) used a conditional nonnormal distribution based on a Gram Charlier distribution. One form of this distribution is

$$f^{GC}(z) = \phi(z)\left[1 + \frac{\lambda_3}{6}H_3(z) + \frac{\lambda_4}{24}H_4(z)\right] = \phi(z)\varphi(z), \tag{9}$$

where $\phi(z)$ is the standard normal density function, and $H_3(z)$ and $H_4(z)$ are the Hermite polynomials defined by

$$H_3(z) = z^3 - 3z; \; H_4(z) = z^4 - 6z^2 + 3. \tag{10}$$

The parameters λ_3 and λ_4, represent the standardized measures of skewness and kurtosis respectively. The function $f^{GC}(z)$ represents the first three terms of the Gram Charlier series (Kendall and Stuart (1963, Vol.1, pp.156-157)).

A simple relationship between the Gram Charlier and generalized Student t distributions can be identified as follows. Consider the generalized Student t distribution with $M = 4$ and $\theta_1 = \theta_2 = 0$, which is also known as the generalized normal distribution

$$f^{GN}(z) = \exp\left[\theta_1 z + \theta_2 z^2 + \theta_3 z^3 + \theta_4 z^4 - \overline{\eta}\right], \tag{11}$$

where $\overline{\eta}$ is the normalizing constant. This distribution can be decomposed as:

$$f^{GN}(z) = \exp\left[-\eta'\right]\exp\left[-0.5z^2 - 0.5\ln(2\pi)\right]\exp\left[\theta_1 z + (\theta_2+0.5)z^2 + \theta_3 z^3 + \theta_4 z^4\right],$$

where $\exp[-\overline{\eta}] = \exp\left[-\eta'\right]\exp[-0.5\ln(2\pi)]$. Notice that the second term is just $\phi(z)$, the standard normal distribution given in (9). Now using the properties of a Taylor series expansion of order one for the exponential function $exp[w]$, around $w = 0$, then

$$\exp\left[\theta_1 z + (\theta_2+0.5)z^2 + \theta_3 z^3 + \theta_4 z^4\right] \cong 1 + \theta_1 z + (\theta_2+0.5)z^2 + \theta_3 z^3 + \theta_4 z^4,$$

[8] The skewed Student t distribution is also known as Pearson's Type IV distribution; see Kendall and Stuart (1963, pp.151-152). The former name has been adopted here to emphasize its relationship with the Student t distribution as well as to highlight its ability to generate Student t densities which are skewed.

so that $f^{GN}(z)$ is approximately equal to

$$f^{GN}(z) \simeq \exp\left[-\eta'\right] \exp\left[\theta_2' z^2 - 0.5\ln(2\pi)\right]\left[1 + \theta_1 z + (\theta_2 + 0.5)z^2 + \theta_3 z^3 + \theta_4 z^4\right]. \quad (12)$$

If the parameters in (12) are constrained as

$$\theta_1 = -\frac{\lambda_3}{2};\ \theta_2 = -\frac{(2+\lambda_4)}{4};\ \theta_3 = -\frac{\theta_1}{3};\ \theta_4 = -\frac{1+2\theta_2}{12},$$

then (12) corresponds to the Gram Charlier distribution given by (9) and (10). Hence the Gram Charlier type distribution and the generalized Student t distribution model the higher order skewness and kurtosis moments in a similar way.

A potential problem with the Gram Charlier formulation adopted by Lee and Tse (1991) is that $\varphi(z)$ may take negative values, and consequently the density may not always be well defined (Kendall and Stuart (1963, Vol.1., p.160)). In contrast, this is not a problem with the generalized Student t distribution as it is constrained to be positive and hence it is always well defined. An alternative approach that overcomes the negativity problem, but still is in the spirit of the Gram Charlier distribution is to follow Gallant, Hsieh and Tauchen (1991) by writing the distribution as

$$f^{GHT}(z) = \xi\phi(z)\varsigma(z)^2, \quad (13)$$

where $\phi(z)$ is as before, the standard normal distribution, $\varsigma(z)$ is a polynomial in z, and the normalizing constant is $\xi^{-1} = \int \phi(z)\varsigma(z)^2 dz$. If $\varsigma(z)$ is chosen as a quadratic, then

$$\begin{aligned}\varsigma(z)^2 &= \left(1 + \varsigma_1 z + \varsigma_2 z^2\right)^2 \\ &= 1 + 2\varsigma_1 z + \left(\varsigma_1^2 + 2\varsigma_2\right) z^2 + 2\varsigma_1\varsigma_2 z^3 + \varsigma_2^2 z^4, \end{aligned} \quad (14)$$

and the relationship with the parameters of the generalized Student t distribution, as approximated by (12), is

$$\theta_1 = 2\varsigma_1;\ \theta_2 = \varsigma_1^2 + 2\varsigma_2 - 0.5;\ \theta_3 = 2\varsigma_1\varsigma_2;\ \theta_4 = \varsigma_2^2. \quad (15)$$

The relationship between (13) and a Gram Charlier distribution is preserved by recognizing that the polynomial in (14) can be re-expressed in terms of Hermite polynomials.

C. Relationship With Mixture Distributions

The use of a variance mixture of normal distributions has been a popular method for modelling nonnormalities given the wide range of shapes that can be obtained by varying the parameters of this class of distributions; for a discussion of the properties of this class of distributions.[9]

[9] See for example, Eisenberger (1964); Robertson and Fryer (1969); and Behboodian (1970); while for an application to financial markets, see Jorion (1988).

A mixture of normal distributions is defined as

$$f^{MN}(z) = (1-\kappa) f(z|0,\sigma_1^2) + \kappa f(z|0,\sigma_2^2), \tag{16}$$

where $f(w|\mu_w,\sigma_w^2) = \left(2\pi\sigma_w^2\right)^{-1/2} \exp\left[-\left(w-\mu_w\right)^2/2\sigma_w^2\right]$, and κ is a weighting parameter with the property $0<\kappa<1$. It is apparent that the likelihood function becomes unbounded when one of the variances in (16) approaches zero. A natural way to impose this constraint in the context of ARCH models, which allows for further simplification in estimation, is to let $\sigma_2^2 = 1$, represent the benchmark for the standard normal case, and constrain $\sigma_1^2 = g > 1$, to be the high variance case.[10] Equation (16) is now rewritten as

$$f^{MN}(z) = (1-\kappa) f(z|0,g) + \kappa f(z|0,1), \tag{20}$$

where $g>1$ and $0<\kappa<1$, are parameters to be estimated. This reparameterization has the added benefit that the standard normal distribution is directly given as a special case when $\kappa = 1$, and that values of $\kappa < 1$, provides evidence of nonnormalities.

To highlight the relationship between the generalized normal and mixture distributions, define the exponential terms in (20) as $exp[w]$ and expand around $w=0$ in a Taylor series expansion of order two. This approximation implies that the mixture of normal distributions can be written as

$$f^{MN}(z) \simeq \left(\frac{1-\kappa+\kappa g^{1/2}}{(2\pi g)^{1/2}}\right) - \left(\frac{1-\kappa+\kappa g^{3/2}}{(2\pi g)^{1/2} 2g}\right) z^2 + \left(\frac{1-\kappa+\kappa g^{5/2}}{(2\pi g)^{1/2} 4g^2}\right) z^4 \tag{21}$$

Equation (21) shows that the mixture distribution in (20) can be approximated by a distribution containing a constant and the terms z^2 and z^4. To show the relationship between the mixture and generalized Student t distribution, consider the following special case of the generalized Student t distribution, known as the symmetric generalized normal distribution, which is obtained by setting $\theta_4, \theta_6 \neq 0$, while all remaining parameters in (6) are set to zero

$$f^{SGN}(z) = \exp\left[\theta_4 z^2 + \theta_6 z^4 - \widehat{\eta}\right], \tag{22}$$

---

[10] Following Phillips (1991, 1994), one way to bound the likelihood function associated with this density is to impose the constraints

$$c \leq \frac{\sigma_2}{\sigma_1} \leq 1, \tag{17}$$

where $c \in (0,1)$, is known. Equation (17) forces σ_2^2, to be the low variance and σ_1^2 to be the high variance, with the added condition that the two variances cannot deviate from each other too much. These constraints can be imposed conveniently by re-expressing (17) as

$$\sigma_1 = \delta\sigma_2; \qquad \sigma_2 = c\sigma_1 + \gamma. \tag{18}$$

where $\delta \geq 1$, $\gamma \geq 0$. By solving for σ_1 and σ_2

$$\sigma_1 = \frac{\delta\gamma}{1-c\delta}; \qquad \sigma_2 = \frac{\gamma}{1-c\delta}, \tag{19}$$

the standard deviations are now expressed in terms of the parameters $\{\delta,\gamma,c\}$, which can be estimated conveniently in a standard maximum likelihood algorithm.

where $\widehat{\eta}$ is the normalizing constant. This distribution can be written to have the same form as (21) by expanding $exp[w]$ in a Taylor series expansion of order one around $w = 0$

$$f^{SGN}(z) \cong \left(1 + \theta_4 z^2 + \theta_6 z^4\right) \exp(-\widehat{\eta}). \tag{23}$$

Matching terms with (21) shows a relationship between the generalized Student t parameters $\{\theta_4, \theta_6\}$, and the mixture parameters $\{\kappa, g\}$.

III. GENERALIZED PARAMETRIC ARCH MODELS

In this section, a class of ARCH models based on the generalized Student t distribution is presented. A maximum likelihood algorithm is also given for estimating the parameters of the model. To help motivate the use of the generalized Student t distribution in formulating ARCH models, it is shown that this distribution represents the solution of a continuous time model which is used to approximate a discrete time ARCH(1) model. An important outcome of this result is that as it is possible to derive explicit relationships between the parameters of the model and the parameters governing the distribution, it is possible to give the shape of the distribution an economic interpretation.

A. A Continuous Time Approximation

Consider the following simple integrated ARCH(1) model for y_t

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \sqrt{h_t} z_t, \tag{24}$$

where the conditional variance is specified as[11]

$$h_t = \gamma^2 + y_{t-1}^2, \tag{25}$$

and the error term is assumed to be distributed as $z_t \sim N(0,1)$. Combining (24) and (25), and rearranging gives

$$y_t - y_{t-1} = \alpha_0 + (\alpha_1 - 1)\, y_{t-1} + \sqrt{\gamma^2 + y_{t-1}^2}\, z_t. \tag{26}$$

The discrete time ARCH model in (26) does not admit an analytical expression for the unconditional distribution of y. Adopting the lead of Nelson (1990), there is however, much to be obtained

[11]For the typical nonintegated ARCH specification where the conditional mean is $\alpha_0 + \alpha_1 y_{t-1}$, the conditional variance would usually be specified as

$$h_t = \gamma^2 + \beta_1 \left(y_{t-1} - \alpha_0 - \alpha_1 y_{t-2}\right)^2.$$

This only serves to complicate slightly the derivation of the model set out below without significantly changing the thrust of the argument.

from investigating the properties of the continuous time analogue of (26). In particular,equation (26) can be thought of as a disretization based on an Euler scheme of the following continuous time process

$$dy = \mu(y)dt + \sigma(y)dW, \tag{27}$$

where the instantaneous mean and variance are respectively $\mu(y) = \alpha_0 + (\alpha_1 - 1)\,y, \sigma^2(y) = \gamma^2 + y^2$, and $dW \sim N(0, dt)$ is a Wiener process.[12] The unconditional distribution $f(y)$, is derived from the Kolmogorov forward equation

$$\frac{\partial f}{\partial t} = -\frac{\partial \mu(y)}{dy} + \frac{1}{2}\frac{\partial^2 \sigma^2(y)}{\partial y^2}, \tag{28}$$

and setting $\partial f/\partial t = 0$ (Creedy and Martin (1994)). The solution is

$$f(y) = \exp\left[\phi_1 \tan^{-1}(y/\gamma) + \phi_2 \ln\left(\gamma^2 + y^2\right) - \eta^S\right], \tag{29}$$

where the parameters governing the distribution are

$$\begin{aligned} \phi_1 &= 2\alpha_0/\gamma, \\ \phi_2 &= 2(\alpha_1 - 2), \end{aligned} \tag{30}$$

and the normalizing constant is

$$\eta^S = \ln \int_{-\infty}^{\infty} \exp\left[\phi_1 \tan^{-1}(s/\gamma) + \phi_2 \ln\left(\gamma^2 + s^2\right)\right] ds, \tag{31}$$

which is determined from the boundary conditions of (28) which ensure that the density is proper; namely $\int_{-\infty}^{\infty} f(y)dy = 1$.

The unconditional density in (29) highlights in an explicit way, the relationship between the structural parameters $\{\alpha_0, \alpha_1, \gamma^2\}$ and the distributional parameters $\{\phi_1, \phi_2, \gamma^2\}$. More importantly, it shows how the choice of the economic model can be used to identify the form of the probability distribution. To highlight this point, inspection of (29) shows that the standard Student t distribution is a special case, occurring when $\phi_1 = 0$ and $\phi_2 = -(1 + \gamma^2)/2$. From the definition of ϕ_2 in (30), this implies that

$$\alpha_1 = \frac{7 - \gamma^2}{4}. \tag{32}$$

As all of the parameters of (31) are not identified, (32) serves as a natural identifying restriction. Further, the parameter γ^2, which is obtained from the conditional variance specification of an ARCH(1) model, corresponds to the "degrees of freedom" parameter of the distribution.

The distribution given by (29) is skewed when $\phi_1 \neq 0$. For this reason (29) is referred to as the skewed Student t distribution. From the definition of ϕ_1 in (30), this implies that $\alpha_0 \neq 0$.

[12] To reduce the amount of notation, the parameters of the continuous time model are not distinguished from the parameters of the discrete model.

Inspection of (24) shows that this result makes sense as the occurrence of drift causes the process y_t to drift continually in one direction resulting in the distribution being skewed. When $\alpha_0 = 0$, there is no tendency for y_t to drift in either direction thereby resulting in the distribution being symmetrical.

An important feature of the stochastic differential equation in (27) is that the instantaneous mean $\mu(y) = \alpha_0 + (\alpha_1 - 1)\, y$, is linear in the random variable y. Following recent work by Conley, Hansen, Luttmer and Scheinkman (1995), Ait-Sahalia (1995), and Creedy, Lye and Martin (1996), on nonlinear stochastic differential equations, the mean is now generalized to a polynomial of order greater than unity

$$dy_t = \left(\alpha_0 + (\alpha_1 - 1)\, y_t + \alpha_2 y_t^2 + \alpha_3 y_t^3 + \alpha_4 y_t^4 + \alpha_5 y_t^5\right) dt + \sqrt{\gamma^2 + y_t^2} dW_t. \tag{33}$$

The Kolmogorov forward equation is now

$$\frac{\partial f}{\partial t} = -\frac{\partial}{dy}\left(\alpha_0 + (\alpha_1 - 1)\, y + \alpha_2 y^2 + \alpha_3 y^3 + \alpha_4 y_t^4 + \alpha_5 y_t^5\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\gamma^2 + y^2\right), \tag{34}$$

and the unconditional distribution, which is found by setting $\partial f / \partial t = 0$, is

$$f(y) = \exp\left[\theta_1 \tan^{-1}(y/\gamma) + \theta_2 \ln\left(\gamma^2 + y^2\right) + \theta_3 y + \theta_4 y^2 + \theta_5 y^3 + \theta_6 y^4 - \eta\right], \tag{35}$$

where η is the normalizing constant which is determined from the boundary conditions of (34) which ensure that $\int_{-\infty}^{\infty} f(y) dy = 1$. The distribution parameters are related to the structural parameters in (33) as follows

$$\begin{aligned}
\theta_1 &= 2\left(\alpha_0 - \gamma\alpha_2 + \gamma^2\alpha_4\right), \\
\theta_2 &= \alpha_1 - \gamma^2\alpha_3 + \gamma^4\alpha_5 - 3, \\
\theta_3 &= 2\left(\alpha_2 - \gamma^2\alpha_4\right), \\
\theta_4 &= \alpha_3 - \gamma^2\alpha_5, \\
\theta_5 &= 2\alpha_4/3, \\
\theta_6 &= \alpha_5/2.
\end{aligned} \tag{36}$$

Note that it is the ARCH specification in (25), and hence the variance specification, which is governing the form of the unconditional distribution. This is most readily seen by replacing (25) by the unconditional variance $h_t = \gamma^2$. Solving for the unconditional distribution of y_t, results in a normal distribution. Further, the results derived in this section emphasizes the earlier result of Engle (1982) that the combination of conditional normality with an ARCH conditional variance model yields a nonnormal unconditional distribution with fatter tails than the normal distribution, and also reinforces Nelson's (1990) earlier result that an explicit expression for the unconditional distribution can be obtained by using the continuous time framework to approximate the discrete time ARCH model.

B. Basic Framework

Given the result that the unconditional distribution of exchange rate returns is generalized Student t when the underlying continuous time process given by (33) is assumed to approximate the discrete time process, this suggests that a natural starting model is given by (24), except that z_t is no longer assumed to be normal, but generalized Student t (Lim, Lye, Martin and Martin (1996)). A particular form of this model which is used in the empirical investigation is

$$y_t = \sqrt{h_t} z_t, \tag{37}$$

where the conditional variance is given by either a GARCH or EGARCH specification

$$\begin{array}{llcl} GARCH: & h_t & = & \alpha_0^G + \alpha_1^G y_{t-1}^2 + \alpha_2^G h_{t-1}, \\ EGARCH & \ln(h_t) & = & \alpha_0^E + \alpha_1^E \ln(h_{t-1}) + \alpha_2^E \left| y_{t-1}/\sqrt{h_{t-1}} \right| + \alpha_3^E \left(y_{t-1}/\sqrt{h_{t-1}} \right), \end{array} \tag{38}$$

and the distribution of z_t is *iid* with zero mean, unit variance and is distributed as generalized Student t.

The standardized generalized Student t distribution is derived as follows. Define the standardized random variable as

$$z = \frac{w - \mu_w}{\sigma_w}, \tag{39}$$

where w has the following generalized Student t distribution

$$f(w) = \exp\left[\theta_1 \tan^{-1}(w/\gamma) - 0.5\left(1+\gamma^2\right)\ln\left(\gamma^2 + w^2\right) + \theta_4 w^2 - \eta^w\right], \tag{40}$$

with respective mean and variance

$$\mu_w = \int w f(w) dw, \tag{41}$$

$$\sigma_w^2 = \int (w - \mu_w)^2 f(w) dw. \tag{42}$$

Using the change of variable technique, the distribution of z is

$$\begin{aligned} f(z) = \sigma_w \exp\Big[\theta_1 \tan^{-1}\left(\left(\mu_w + \sigma_w z\right)/\gamma\right) - 0.5\left(1+\gamma^2\right)\ln\left(\gamma^2 + \left(\mu_w + \sigma_w z\right)^2\right) \\ + \theta_4 \left(\mu_w + \sigma_w z\right)^2 - \eta\Big], \end{aligned} \tag{43}$$

and where the normalizing constant is given by

$$\begin{aligned} \eta = \ln \int \sigma_w \exp\Big[\theta_1 \tan^{-1}\left(\left(\mu_w + \sigma_w z\right)/\gamma\right) - 0.5\left(1+\gamma^2\right)\ln\left(\gamma^2 + \left(\mu_w + \sigma_w z\right)^2\right) \\ + \theta_4 \left(\mu_w + \sigma_w z\right)^2\Big] dz, \end{aligned} \tag{44}$$

C. Estimation Issues

The parametric ARCH class of models can be estimated conveniently using maximum likelihood methods. The logarithm of the likelihood at the t^{th} observation is

$$\begin{aligned} \ln L_t &= -0.5 \ln h_t + 0.5 \ln \sigma_w^2 + \theta_1 \tan^{-1}\left(\left(\mu_w + \sigma_w y_t/\sqrt{h_t}\right)/\gamma\right) \\ &\quad -0.5\left(1+\gamma^2\right)\ln\left(\gamma^2 + \left(\mu_w + \sigma_w y_t/\sqrt{h_t}\right)^2\right) \\ &\quad +\theta_4\left(\mu_w + \sigma_w y_t/\sqrt{h_t}\right)^2 - \eta. \end{aligned} \tag{45}$$

For a sample of $t = 1, 2, ..., T$, observations, the log of the likelihood function is

$$\ln L = \sum_{t=1}^{T} \ln L_t. \tag{46}$$

This expression can be maximized with respect to the conditional variance parameters in (38) and the distribution parameters $\{\gamma, \theta_1, \theta_4\}$, using standard gradient optimization algorithms. As a result of the nonlinearities in the model, it is convenient to use numerical derivatives. The integrals involved in computing μ_w, σ_w, η, in (41), (42) and (44) respectively, can be computed numerically using standard quadrature methods. All programs are written using the GAUSS computer language. The optimization algorithm chosen is MAXLIK, while the integrations are calculated using the procedure INTQUAD1.

In practice it was found that it was best to split the optimizations in terms of the conditional variance and distribution parameters before maximizing (46) with respect to all parameters jointly. Hence the following algorithm is proposed:

1. Choose some initial values for the conditional variance parameters α_i. One possibility is to estimate these parameters assuming conditional normality.
2. Form the standardized residuals $z_t = y_t/\sqrt{h_t}$, and estimate the distribution parameters $\{\gamma, \theta_1, \theta_4\}$, by maximizing the likelihood function in terms of z_t

$$\begin{aligned} \ln L_t(z_t) &= 0.5 \ln \sigma_w^2 + \theta_1 \tan^{-1}\left(\left(\mu_w + \sigma_w z_t\right)/\gamma\right) \\ &- 0.5\left(1+\gamma^2\right)\ln\left(\gamma^2 + \left(\mu_w + \sigma_w z_t\right)^2\right) + \theta_4\left(\mu_w + \sigma_w z_t\right)^2 - \eta. \end{aligned} \tag{47}$$

3. Estimate the conditional variance parameters by maximizing the log likelihood function in (46), with the distribution parameters fixed at the estimated values obtained from the previous step.
4. Repeat steps 2 and 3 until convergence of all the parameters is obtained.
5. Finally, maximize (46) with respect to all parameters and compute standard errors.

The properties of the maximum likelihood estimator when the conditional variance follows an ARCH structure were first studied by Weiss (1986). Bollerslev and Wooldridge (1992) showed that when the conditional distribution is assumed to be normal, but the true distribution is nonnormal, the (quasi) maximum likelihood parameter estimates of the model are still consistent provided that both the mean and the variance are specified correctly. There is generally however, a loss of efficiency with the quasi maximum likelihood estimator. This property was noted by Engle and González-Rivera (1991), who showed that the loss of efficiency of the quasi maximum likelihood estimator could be substantial, especially when the underlying distribution is asymmetric. To help overcome this loss of efficiency, the ARCH model described above where the pertinent conditional distribution is specified as generalized Student t, provides a parametric approach for modelling skewness and achieving improvements in efficiency over the quasi maximum likelihood estimator.

IV. APPLICATION TO HOURLY EXCHANGE RATES

There is widespread agreement that the distribution of exchange rate returns is nonnormal; see, for example, de Vries (1994). There is also general agreement that models that combine the assumption of conditional normality with ARCH conditional variance specifications, explain only part of this nonnormality, albeit a significant part. This has led to the development of more general classes of models which replace the assumption of conditional normality by a particular nonnormal distribution. Hsieh (1989) investigated a broad range of parametric nonnormal distributions for a range of exchange rates; Engle and González-Rivera (1991) used a semiparametric approach to model the distribution of the returns on the British pound/US exchange rate; while Gallant, Hsieh and Tauchen (1991) used a seminonparametric approach. Baillie and Bollerslev (1989) used a quasi maximum likelihood estimator to estimate ARCH models of intra-day exchange rate returns, while for the same data set Harvey, Ruiz and Sentana (1992) used the STAR model.

The data set chosen here consists of hourly exchange rates recorded in 1987 from 0.00 a.m., January 2, to 11.00 a.m., July 15th. The currencies are the British pound (BP), Deutschmark (DM), Japanese Yen (JY) and the Swiss frank (SF), all relative to the US dollar.[13]

A. Parameter Estimates

Tables 1 to 8 contain the parameter estimates of the model given by (37) with conditional variance specification given by (38), for the returns on the four bilateral exchange rates. Returns are computed as

$$y_t = \ln(e_t/e_{t-1}), \tag{48}$$

where e_t is the bilateral exchange rate, which are then standardized to have zero mean and unit variance. Following Harvey, Ruiz and Sentana (1992), hourly dummy variables have not been included in the model. For each exchange rate six conditional distributions are specified: normal, Student t, Gram Charlier, Mixture, Skewed Student t, generalized Student t. The generalized

[13] For further discussion of the data, see Baillie and Bollerslev (1989).

Table 1:

GARCH model estimates for alternative conditional distributions: BP
(Asymptotic robust standard errors in brackets)

Parameter	Normal	Stud. t	G. Charlier	Mixture	Skew. Stud. t	Gen. Stud. t
α_0^G	0.336	0.484	0.349	0.321	0.345	0.314
	(0.040)	(0.096)	(0.074)	(0.051)	(0.050)	(0.046)
α_1^G	0.212	0.559	0.298	0.326	0.403	0.349
	(0.026)	(0.117)	(0.103)	(0.062)	(0.058)	(0.054)
α_2^G	0.471	0.411	0.429	0.423	0.414	0.419
	(0.048)	(0.053)	(0.099)	(0.063)	(0.053)	(0.056)
γ^2		2.554			2.524	1.342
		(0.128)			(0.134)	(0.236)
λ_3			0.005			
			(0.005)			
λ_4			2.389			
			(0.137)			
κ				0.723		
				(0.041)		
g				9.851		
				(0.789)		
θ_1					-0.039	-0.048
					(0.129)	(0.054)
θ_4						-0.018
						(0.005)
Mean $\ln L_t$	-1.381	-1.255	-1.291	-1.260	-1.254	-1.250

Table 2:

GARCH model estimates for alternative conditional distributions: DM
(Asymptotic robust standard errors in brackets)

Parameter	Normal	Stud. t	G.Charlier	Mixture	Skew. Stud. t	Gen. Stud. t
α_0^G	0.293	0.384	0.319	0.287	0.299	0.277
	(0.033)	(0.103)	(0.081)	(0.061)	(0.061)	(0.055)
α_1^G	0.176	0.521	0.230	0.317	0.403	0.365
	(0.025)	(0.136)	(0.048)	(0.055)	(0.071)	(0.064)
α_2^G	0.550	0.475	0.528	0.476	0.454	0.463
	(0.042)	(0.078)	(0.089)	(0.079)	(0.076)	(0.072)
γ^2		2.591			2.558	1.620
		(0.046)			(0.148)	(0.309)
λ_3			-0.090			
			(0.058)			
λ_4			2.417			
			(0.165)			
κ				0.786		
				(0.046)		
g				9.915		
				(1.039)		
θ_1					0.144	0.128
					(0.045)	(0.045)
θ_4						-0.012
						(0.005)
Mean $\ln L_t$	-1.391	-1.252	-1.303	-1.265	-1.251	-1.248

Table 3:

GARCH model estimates for alternative conditional distributions: JY
(Asymptotic robust standard errors in brackets)

Parameter	Normal	Stud. t	G. Charlier	Mixture	Skew. Stud. t	Gen. Stud. t
α_0^G	0.278	0.011	0.030	0.011	0.003	0.003
	(0.041)	(0.008)	(0.018)	(0.006)	(0.002)	(0.002)
α_1^G	0.279	0.190	0.069	0.085	0.074	0.071
	(0.036)	(0.071)	(0.018)	(0.021)	(0.019)	(0.019)
α_2^G	0.494	0.924	0.916	0.921	0.938	0.939
	(0.057)	(0.017)	(0.025)	(0.018)	(0.015)	(0.015)
γ^2		2.281			2.243	1.891
		(0.110)			(0.113)	(0.220)
λ_3			-0.103			
			(0.067)			
λ_4			2.854			
			(0.145)			
κ				0.831		
				(0.031)		
g				14.556		
				(1.991)		
θ_1					0.133	0.131
					(0.041)	(0.039)
θ_4						-0.003
						(0.002)
Mean $\ln L_t$	-1.368	-1.168	-1.250	-1.187	-1.165	-1.164

Table 4:

GARCH model estimates for alternative conditional distributions: SF
(Asymptotic robust standard errors in brackets)

Parameter	Normal	Stud. t	G. Charlier	Mixture	Skew. Stud. t	Gen. Stud. t
α_0^G	0.441	0.718	0.466	0.368	0.444	0.376
	(0.079)	(0.200)	(0.139)	(0.115)	(0.085)	(0.102)
α_1^G	0.173	0.642	0.214	0.274	0.410	0.291
	(0.028)	(0.200)	(0.066)	(0.083)	(0.081)	(0.073)
α_2^G	0.397	0.366	0.352	0.413	0.354	0.395
	(0.092)	(0.096)	(0.160)	(0.149)	(0.092)	(0.129)
γ^2		2.403			2.335	0.334
		(0.131)			(0.137)	(0.250)
λ_3			-0.043			
			(0.030)			
λ_4			2.298			
			(0.124)			
κ				0.655		
				(0.062)		
g				10.262		
				(0.975)		
θ_1					0.116	0.098
					(0.039)	(0.025)
θ_4						-0.036
						(0.018)
Mean $\ln L_t$	-1.394	-1.270	-1.309	-1.277	-1.269	-1.258

Table 5:

EGARCH model estimates for alternative conditional distributions: BP
(Asymptotic robust standard errors in brackets)

Parameter	Normal	Stud. t	G. Charlier	Mixture	Skew. Stud. t	Gen. Stud. t
α_0^E	-0.299	-0.260	-0.310	-0.360	-0.358	-0.380
	(0.037)	(0.054)	(0.039)	(0.036)	(0.033)	(0.034)
α_1^E	0.678	0.668	0.684	0.671	0.669	0.673
	(0.050)	(0.039)	(0.055)	(0.051)	(0.039)	(0.040)
α_2^E	0.404	0.666	0.442	0.496	0.576	0.528
	(0.046)	(0.071)	(0.059)	(0.048)	(0.046)	(0.045)
α_3^E	0.051	0.011	0.041	0.022	0.010	0.020
	(0.035)	(0.012)	(0.056)	(0.073)	(0.013)	(0.027)
γ^2		2.594			2.559	1.286
		(0.134)			(0.138)	(0.238)
λ_3			0.007			
			(0.017)			
λ_4			2.330			
			(0.127)			
κ				0.706		
				(0.044)		
g				9.646		
				(0.781)		
θ_1					-0.040	-0.049
					(0.083)	(0.029)
θ_4						-0.020
						(0.005)
Mean $\ln L_t$	-1.374	-1.252	-1.287	-1.257	-1.252	-1.247

Table 6:

EGARCH model estimates for alternative conditional distributions: DM
(Asymptotic robust standard errors in brackets)

Parameter	Normal	Stud. t	G. Charlier	Mixture	Skew. Stud. t	Gen. Stud. t
α_0^E	-0.227	-0.248	-0.222	-0.310	-0.319	-0.331
	(0.040)	(0.044)	(0.046)	(0.037)	(0.033)	(0.035)
α_1^E	0.730	0.749	0.734	0.736	0.743	0.742
	(0.051)	(0.047)	(0.059)	(0.050)	(0.047)	(0.046)
α_2^E	0.313	0.564	0.361	0.445	0.501	0.478
	(0.045)	(0.069)	(0.051)	(0.047)	(0.049)	(0.048)
α_3^E	-0.058	-0.035	-0.081	-0.046	-0.028	-0.030
	(0.037)	(0.023)	(0.037)	(0.029)	(0.028)	(0.023)
γ^2		2.637			2.613	1.657
		(0.150)			(0.154)	(0.319)
λ_3			-0.087			
			(0.131)			
λ_4			2.438			
			(0.176)			
κ				0.799		
				(0.043)		
g				9.733		
				(1.076)		
θ_1					0.147	0.131
					(0.052)	(0.049)
θ_4						-0.012
						(0.005)
Mean $\ln L_t$	-1.384	-1.249	-1.297	-1.261	-1.247	-1.245

Table 7:

EGARCH model estimates for alternative conditional distributions: JY
(Asymptotic robust standard errors in brackets)

Parameter	Normal	Stud. t	G. Charlier	Mixture	Skew. Stud. t	Gen. Stud. t
α_0^E	-0.300	-0.135	-0.127	-0.141	-0.142	-0.142
	(0.067)	(0.023)	(0.063)	(0.023)	(0.030)	(0.030)
α_1^E	0.784	0.982	0.952	0.972	0.982	0.979
	(0.096)	(0.008)	(0.043)	(0.012)	(0.009)	(0.010)
α_2^E	0.422	0.287	0.208	0.219	0.223	0.220
	(0.090)	(0.060)	(0.099)	(0.042)	(0.046)	(0.045)
α_3^E	-0.084	-0.037	-0.043	-0.039	-0.027	-0.030
	(0.037)	(0.017)	(0.040)	(0.020)	(0.017)	(0.015)
γ^2		2.397			2.291	1.823
		(0.108)			(0.119)	(0.227)
λ_3			-0.102			
			(0.071)			
λ_4			2.796			
			(0.148)			
κ				0.821		
				(0.034)		
g				13.600		
				(1.665)		
θ_1					0.133	0.132
					(0.042	(0.039)
θ_4						-0.004
						(0.002)
Mean $\ln L_t$	-1.353	-1.163	-1.238	-1.179	-1.161	-1.159

Table 8:

EGARCH model estimates for alternative conditional distributions: SF
(Asymptotic robust standard errors in brackets)

Parameter	Normal	Stud. t	G. Charlier	Mixture	Skew. Stud. t	Gen. Stud. t
α_0^E	-0.231	-0.140	-0.260	-0.304	-0.300	-0.314
	(0.050)	(0.071)	(0.057)	(0.058)	(0.040	(0.048)
α_1^E	0.642	0.662	0.616	0.659	0.654	0.658
	(0.119)	(0.086)	(0.125)	(0.106)	(0.086)	(0.094)
α_2^E	0.315	0.657	0.367	0.432	0.535	0.446
	(0.059)	(0.109)	(0.070)	(0.081)	(0.066)	(0.066)
α_3^E	-0.037	-0.062	-0.007	-0.020	-0.048	-0.057
	(0.029)	(0.044)	(0.016)	(0.018)	(0.031)	(0.036)
γ^2		2.430			2.362	0.347
		(0.129)			(0.138)	(0.237)
λ_3			-0.045			
			(0.050)			
λ_4			2.289			
			(0.124)			
κ				0.658		
				(0.062)		
g				10.133		
				(0.926)		
θ_1					0.115	0.103
					(0.038)	(0.024)
θ_4						-0.036
						(0.016)
Mean $\ln L_t$	-1.389	-1.267	-1.305	-1.274	-1.265	-1.254

Student t distribution is given by (43). The skewed Student t distribution is (43) with $\theta_4 = 0$, whilst the Student t distribution is (43) with $\theta_1 = \theta_4 = 0$. The Gram Charlier and mixture distributions are given by (9) and (20) respectively. The normal distribution serves as a benchmark for comparing the estimates obtained from the nonnormal conditional distributions with the quasi-maximum likelihood estimates. The Student t and mixture distributions allow for kurtosis, but not skewness. The other three distributions, namely Gram Charlier, skewed Student t and generalized Student t distributions, allow for both skewness and kurtosis with varying degrees of flexibility. In total, forty-eight models are estimated.

Parameter estimates assuming a GARCH conditional variance structure for the four currencies as well as for the six conditional distributions, are given in Tables 1 to 4. The corresponding EGARCH results are given in Table 5 to 8. The GARCH parameter estimates are very similar for all currencies and for all conditional distributional models with the exception of the Student t and normal results for the BP and JY currencies respectively. There is also a high degree of consistency in the EGARCH parameter estimates across both currencies and distributional specifications, with the possible exception of the normal results for the JY currency.

A t-test based on $\theta_1 = 0$, using either the skewed Student t or generalized Student t parameter estimates, shows that BP is the only currency where there is no significant skewness present in the conditional distribution. This is true for both GARCH and EGARCH conditional variance models. A similar test can also be conducted based on testing $\lambda_3 = 0$, using the Gram Charlier results. However, as will be demonstrated below, the results based on the Gram Charlier distribution are less reliable than either the skewed or generalized Student t distributions as the Gram Charlier distribution fails a number of diagnostic tests which the other two distributions do not.

The relatively small estimates of the degrees of freedom parameter γ^2, for all models point towards fatness in the tails of the conditional distributions. Further evidence of fatness in the tails is given by the estimates of the mixture distribution. The parameter estimates of g range from 9.646 to 14.556, impling that the variance of the high variance distribution is larger than the variance of the low variance distribution by a factor of more than 10 on average. Furthermore, as the parameter estimates of $1-\kappa$, are around 0.25, this suggests than approximately one quarter of returns come from the high variance distribution.

A test of the generalized Student t distribution can be formulated by constructing a t-test of the hypothesis $\theta_4 = 0$. However, this test statistic does not have the standard asymptotic distribution as the test is being conducted on the boundary of the parameter space. A more convenient approach is to construct a Lagrange multiplier test

$$LM = \left(\frac{\partial \ln L}{\partial \ln \psi}\right)' I(\psi)^{-1} \left(\frac{\partial \ln L}{\partial \ln \psi}\right)\bigg|_{\psi=\widehat{\psi}}, \tag{49}$$

where $I(\psi)$ is the information matrix, $\psi = \{\alpha, \gamma, \theta_1, \theta_4\}$, and $\widehat{\psi} = \{\widehat{\alpha}, \widehat{\gamma}, \widehat{\theta}_1, 0\}$ corresponds to the constrained maximum likelihood estimates associated with the skewed Student t distribution. This

Table 9:

Ljung-Box test of z_t, 20 lags, p-values

Distribution	GARCH				EGARCH			
	BP	DM	JY	SF	BP	DM	JY	SF
Normal	0.006	0.641	0.105	0.615	0.004	0.656	0.106	0.675
Stud. t	0.009	0.771	0.114	0.783	0.008	0.756	0.142	0.827
G. Charlier	0.007	0.668	0.039	0.652	0.004	0.669	0.060	0.698
Mixture	0.008	0.736	0.074	0.743	0.006	0.729	0.102	0.778
Skew. Stud. t	0.009	0.770	0.117	0.788	0.003	0.657	0.088	0.637
Gen. Stud. t	0.009	0.764	0.087	0.751	0.002	0.655	0.062	0.611

Table 10:

Ljung-Box test of z_t^2, 20 lags, p-values

Distribution	GARCH				EGARCH			
	BP	DM	JY	SF	BP	DM	JY	SF
Normal	0.047	0.298	0.259	0.039	0.043	0.205	0.396	0.056
Stud. t	0.033	0.087	0.986	0.006	0.022	0.069	0.807	0.013
G. Charlier	0.041	0.252	0.472	0.026	0.044	0.167	0.307	0.040
Mixture	0.038	0.135	0.908	0.014	0.031	0.096	0.590	0.023
Skew. Stud. t	0.033	0.087	0.975	0.006	0.018	0.225	0.729	0.075
Gen. Stud. t	0.035	0.095	0.984	0.011	0.018	0.225	0.674	0.060

test statistic is distributed asymptotically as χ_1^2; see Rogers (1986). Alternatively, a test of the Student t distribution could be obtained by defining $\widehat{\psi} = \{\widehat{\alpha}, \widehat{\gamma}, \widehat{\theta}_1, 0\}$ where now all parameter estimates are obtained from the Student t distribution. In this case the test statistic (49) is distributed asymptotically as χ_2^2. This latter test was performed for all currencies and for both conditional variance models with the result that there was a very strong rejection of the null hypothesis that the underlying distribution is Student t.

Some diagnostics for each model are given in Tables 9 and 10. The Ljung-Box test of the standardized residuals z_t, shows that there is no significant autocorrelation structure at the one percent level in the DM, JY and SF currencies. This is true for both the GARCH and EGARCH conditional variance versions of the model. In contrast, the BP currency exhibits significant autocorrelation for all six conditional distributions and both conditional variance specifications.[14]

The Ljung-Box test of the squared standardized residuals z_t^2, show that there is no significant

[14]Comparing this result with that obtained by Baillie and Bollerslev (1990), suggests that use of hourly dummies in the conditional mean specification can eradicate the autocorrelelation problem.

ARCH structure at the one percent level in the BP, DM and JY currencies for both GARCH and EGARCH specifications. For the SF currency, there is no presence of ARCH at the one percent level assuming an EGARCH conditional variance. There is significant ARCH at the one percent level assuming a GARCH conditional variance structure for both the Student t and skewed Student t distributions, but not the other conditional distributions.

B. Estimates of the Conditional Distributions

Some selected statistics on the conditional distribution of z_t are given in Table 11. All models generate sample means of z_t close to the theoretical value of 0.0. In terms of the sample variance of z_t, the Student t distribution performs the worst as it results in sample variances between 0.503 and 0.721, which does not compare favourably with the theoretical value of 1.0. The sample skewness statistics also provide evidence against the Student t distribution which has a theoretical value of 0.0. Both the sample skewness and kurtosis statistics provide strong evidence against the normal distribution which has theoretical values of 0.0 and 3.0 respectively.

Further evidence on the suitability of the conditional distributions is given in Table 12 which compares the heights of the empirical (E) and theoretical (T) distributions at $z = 0$. These results show strong rejection of normality with the empirical values consistently around 0.650 for all models, compared with the theoretical value of 0.399. The Gram Charlier distribution also performs badly as it generates heights of the conditional distribution comparable to the normal distribution. For the remaining distributions, namely the Student t, mixture and the skewed and generalized Student t distributions, there is a good match between the heights of the empirical and theoretical densities.

Figures 1 and 2 give further information of the goodness of fit of the theoretical conditional distributions, $f(z_t)$. In Figure 1, the six conditional distributions are compared with the empirical distribution $g(z_t)$, assuming a GARCH conditional variance, for the BP currency. The empirical distribution is computed nonparametrically using a normal kernel. In Figure 2, the conditional variance is EGARCH. A formal test of goodness of fit is given in Table 13, which reports the p-values for the standard chi-square goodness of fit test. These results provide further evidence against the normal as well as the Gram Charlier distributions, while providing strong acceptance of the other four conditional distributions.

C. Estimates of the Unconditional Distributions

It is not possible to obtain closed form, analytical expressions of the unconditional distribution for the discrete ARCH model. The approach adopted here is to simulate this density using the following steps:

1. A sample of $z_1, z_2, ..., z_T$, random numbers are drawn from one of the six conditional densities, where $T = 500$, is chosen.

Table 11:

Selected statistics on the conditional distribuition of z_t

Distribution	GARCH				EGARCH			
	BP	DM	JY	SF	BP	DM	JY	SF
					Mean			
Normal	0.004	-0.003	-0.005	-0.005	0.005	-0.003	-0.004	-0.004
Stud. t	0.004	-0.004	0.001	-0.005	0.003	-0.004	0.002	-0.004
G. Charlier	0.004	-0.003	0.000	-0.006	0.004	-0.003	0.001	-0.005
Mixture	0.005	-0.004	0.001	-0.006	0.004	-0.004	0.002	-0.005
Skew. Stud. t	0.005	-0.005	0.001	-0.007	0.004	-0.002	0.004	-0.002
Gen. Stud. t	0.005	-0.005	0.001	-0.006	0.004	-0.002	0.004	-0.002
					Variance			
Normal	1.000	1.000	1.000	1.000	1.000	1.000	1.002	1.000
Stud. t	0.652	0.697	0.503	0.531	0.674	0.721	0.629	0.555
G. Charlier	0.962	0.918	0.901	0.988	0.971	0.912	0.908	0.988
Mixture	1.000	1.000	1.001	1.000	1.001	0.999	1.007	1.000
Skew. Stud. t	0.905	0.940	1.025	0.871	0.830	0.879	0.944	0.816
Gen. Stud. t	1.002	1.002	1.044	1.001	0.947	0.968	0.992	0.967
					Skewness			
Normal	-0.185	0.180	-0.225	0.102	-0.172	0.189	-0.161	0.135
Stud. t	-0.089	0.106	-0.127	0.026	-0.107	0.105	-0.054	0.047
G. Charlier	-0.165	0.159	-0.088	0.082	-0.170	0.172	-0.046	0.102
Mixture	-0.172	0.177	-0.266	0.096	-0.188	0.175	-0.088	0.115
Skew. Stud. t	-0.145	0.163	-0.490	0.050	-0.127	0.155	-0.050	0.153
Gen. Stud. t	-0.171	0.177	-0.534	0.086	-0.154	0.179	-0.055	0.185
					Kurtosis			
Normal	8.788	10.941	13.474	8.598	8.388	10.741	12.582	8.646
Stud. t	3.833	5.806	6.330	2.587	3.963	6.020	7.801	2.860
G. Charlier	8.222	9.374	12.587	8.485	7.967	9.048	12.033	8.506
Mixture	8.952	11.691	20.387	9.000	8.614	11.355	17.420	9.028
Skew. Stud. t	7.390	10.590	25.458	6.973	5.762	8.267	17.133	5.697
Gen. Stud. t	9.025	11.972	26.529	9.062	7.479	10.015	18.606	7.949

Table 12:

Heights of the empirical (E) and theoretical (T) conditional distributions at $z = 0$

Distribution	Type	GARCH				EGARCH			
		BP	DM	JY	SF	BP	DM	JY	SF
Normal	E	0.630	0.648	0.689	0.685	0.626	0.644	0.678	0.682
	T	0.399	0.399	0.399	0.399	0.399	0.399	0.399	0.399
Stud. t	E	0.782	0.779	0.982	0.944	0.765	0.761	0.861	0.920
	T	0.778	0.760	1.022	0.880	0.759	0.740	0.886	0.857
G. Charlier	E	0.643	0.676	0.722	0.689	0.635	0.675	0.709	0.687
	T	0.399	0.394	0.394	0.397	0.401	0.395	0.397	0.399
Mixture	E	0.631	0.649	0.689	0.686	0.626	0.645	0.675	0.684
	T	0.601	0.581	0.634	0.623	0.601	0.572	0.626	0.620
Skew. Stud. t	E	0.664	0.672	0.679	0.738	0.660	0.667	0.676	0.734
	T	0.663	0.656	0.710	0.693	0.657	0.648	0.701	0.689
Gen. Stud. t	E	0.631	0.650	0.676	0.686	0.626	0.645	0.669	0.683
	T	0.668	0.664	0.720	0.773	0.663	0.656	0.710	0.767

Table 13:

Chi-square goodness of fit test of the conditional distribution: p-values

Distribution	GARCH				EGARCH			
	BP	DM	JY	SF	BP	DM	JY	SF
Normal	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Stud. t	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
G. Charlier	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mixture	0.999	0.894	0.775	0.940	0.997	0.816	0.921	0.852
Skew. Stud. t	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Gen. Stud. t	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

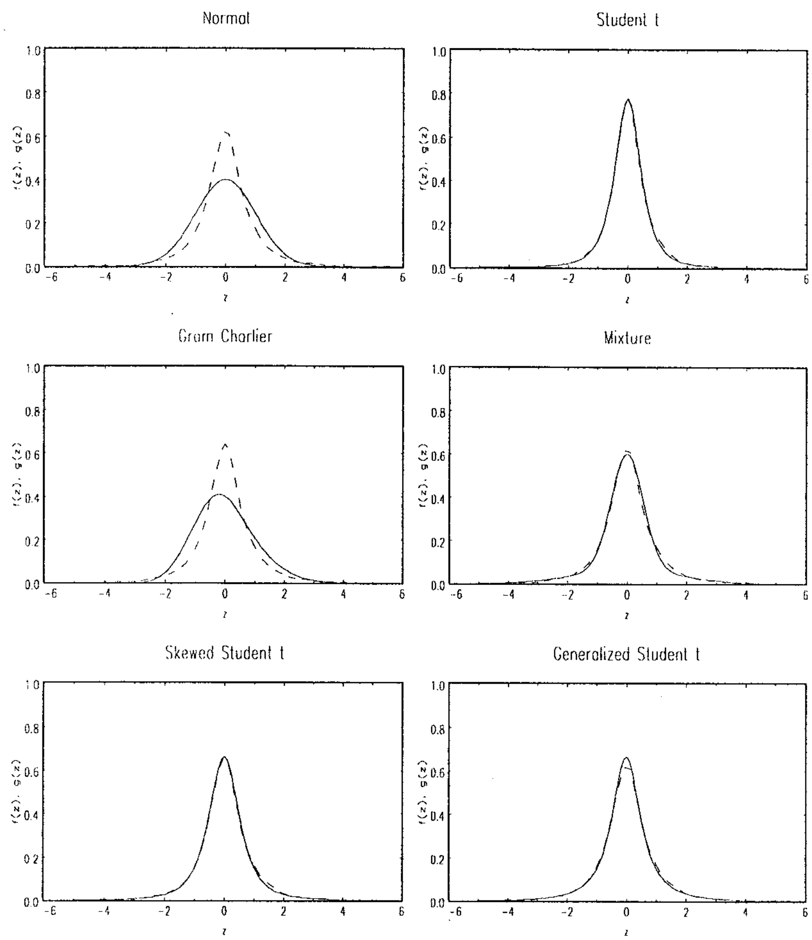

Figure 1: Conditional distributions, BP currency, GARCH: theoretical (continuous line) empirical (broken line).

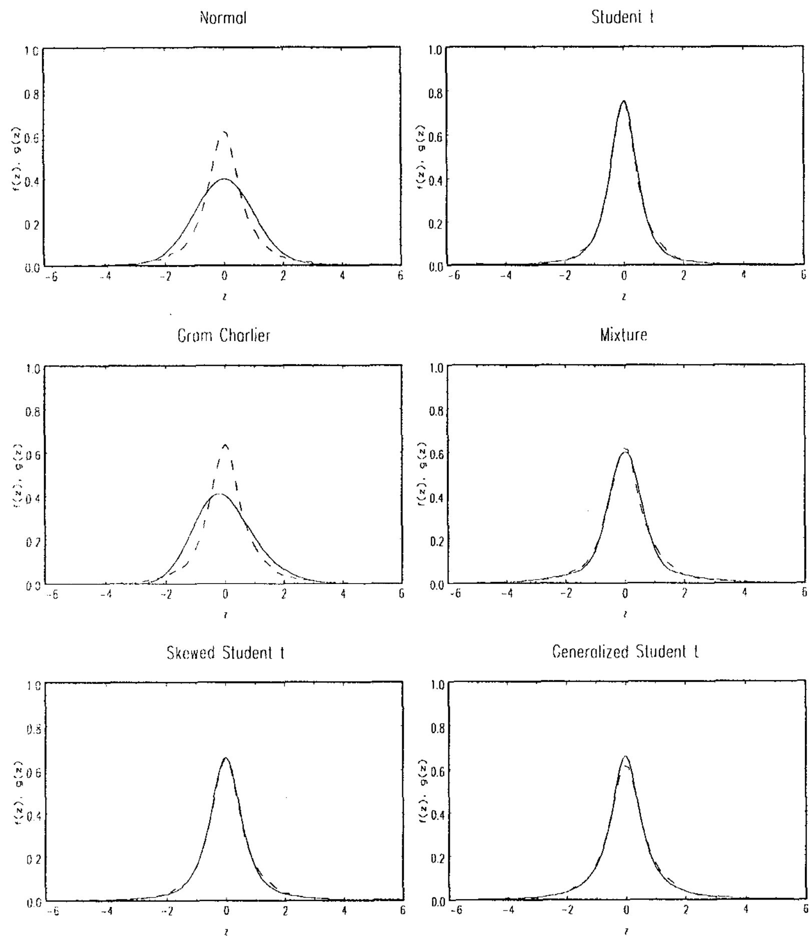

Figure 2: Conditional distributions, BP currency, EGARCH: theoretical (continuous line) empirical (broken line).

2. A sample path of length T, is generated for the simulated data $\widetilde{y}_t$, from the ARCH model

$$\widetilde{y}_t = \sqrt{h_t} z_t$$

where h_t depends on the conditional variance specification

$$\begin{array}{lll} GARCH: & h_t & = \alpha_0^G + \alpha_1^G \widetilde{y}_{t-1}^2 + \alpha_2^G h_{t-1}, \\ EGARCH & \ln(h_t) & = \alpha_0^E + \alpha_1^E \ln(h_{t-1}) + \alpha_2^E \left| \widetilde{y}_{t-1}/\sqrt{h_{t-1}} \right| + \alpha_3^E \left(\widetilde{y}_{t-1}/\sqrt{h_{t-1}} \right), \end{array}$$

and the parameters are chosen as the maximum likelihood estimates given in Tables 1 and 8.

3. Steps 1 and 2 are repeated N times, where $N = 500$, is chosen, thereby generating N sample paths of length T.

4. The distribution of $\widetilde{y}$ is based on the last observation, that is the T^{th} observation, from each of the N sample paths. A kernel density is then used to compute the theoretical density based on the N observations.

Heights of the theoretical distribution, $f(y)$, are compared with the heights of the empirical distribution, $g(y)$, in Table 14. For the empirical distribution, heights are recorded at the mode of the density $[\max g(y)]$, as well as at $y = 0$ $[g(y = 0)]$. The heights of the six theoretical densities are recorded at the mode of the theoretical density $[\max f(y)]$ and at the mode of the empirical density $[f(g_{\text{mode}})]$.

The results in Table 14 provide further evidence against the normal distribution: the peaks of the empirical densities are generally between 0.6 and 0.7, whereas the conditional normal model assuming either GARCH or EGARCH conditional variances achieves much smaller peaks at between 0.4 and 0.45. The Student t distribution model with a GARCH conditional variance specification is the opposite to the normal model as it tends to generate theoretical peaks which are in excess of the empirical peaks. The generalized Student t distribution performs the best overall in terms of all exchange rates and all conditional variance specifications, whereas the Student t distribution with and EGARCH conditional variance performs very well.

Further information of the goodness of fit of the theoretical distributions is given in Figures 3 (GARCH) and 4 (EGARCH), which give plots of the empirical $g(y_t)$, and theoretical $f(y_t)$, unconditional distributions for the BP currency. A more formal test based on the chi-square goodness of fit test is given in Table 15 for each of the six conditional distributions and for both conditional variance specifications. These results show that both the skewed Student t and generalized Student t distributions consistently generate p-values in excess of ten percent for all exchange rates and for both conditional variances specifications.

Table 14:

Heights of the empirical $g(y)$ and theoretical $f(y)$ unconditional distributions

Distribution	Type	GARCH				EGARCH			
		BP	DM	JY	SF	BP	DM	JY	SF
Empirical	$\max g(y)$	0.642	0.652	0.708	0.678	0.642	0.652	0.708	0.678
	$g(y=0)$	0.642	0.652	0.708	0.678	0.642	0.652	0.708	0.678
Normal	$\max f(y)$	0.401	0.420	0.451	0.399	0.418	0.399	0.446	0.426
	$f(g_{\text{mode}})$	0.396	0.404	0.451	0.361	0.399	0.398	0.429	0.400
Stud. t	$\max f(y)$	0.752	0.679	1.710	1.127	0.680	0.635	0.746	0.725
	$f(g_{\text{mode}})$	0.727	0.675	1.710	1.127	0.663	0.633	0.745	0.725
G. Charlier	$\max f(y)$	0.601	0.516	0.597	0.491	0.566	0.513	0.559	0.487
	$f(g_{\text{mode}})$	0.581	0.499	0.592	0.488	0.550	0.499	0.554	0.484
Mixture	$\max f(y)$	0.586	0.530	0.584	0.520	0.577	0.562	0.536	0.538
	$f(g_{\text{mode}})$	0.586	0.525	0.584	0.520	0.577	0.561	0.532	0.538
Skew. Stud. t	$\max f(y)$	0.594	0.557	0.587	0.541	0.583	0.546	0.604	0.528
	$f(g_{\text{mode}})$	0.594	0.554	0.587	0.541	0.583	0.541	0.604	0.528
Gen. Stud. t	$\max f(y)$	0.620	0.571	0.592	0.622	0.611	0.561	0.605	0.611
	$f(g_{\text{mode}})$	0.620	0.571	0.592	0.622	0.611	0.559	0.605	0.611

Table 15:

Chi-square goodness of fit test of the unconditional distribution: p-values

Distribution	GARCH				EGARCH			
	BP	DM	JY	SF	BP	DM	JY	SF
Normal	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Stud. t	0.067	0.994	0.000	0.000	0.792	1.000	0.999	0.157
G. Charlier	0.738	0.658	0.708	0.095	0.378	0.570	0.125	0.037
Mixture	0.992	0.852	0.185	0.197	0.958	0.919	0.024	0.383
Skew. Stud. t	0.983	0.975	0.161	0.652	0.950	0.905	0.481	0.443
Gen. Stud. t	0.999	0.991	0.152	0.993	0.997	0.963	0.446	0.988

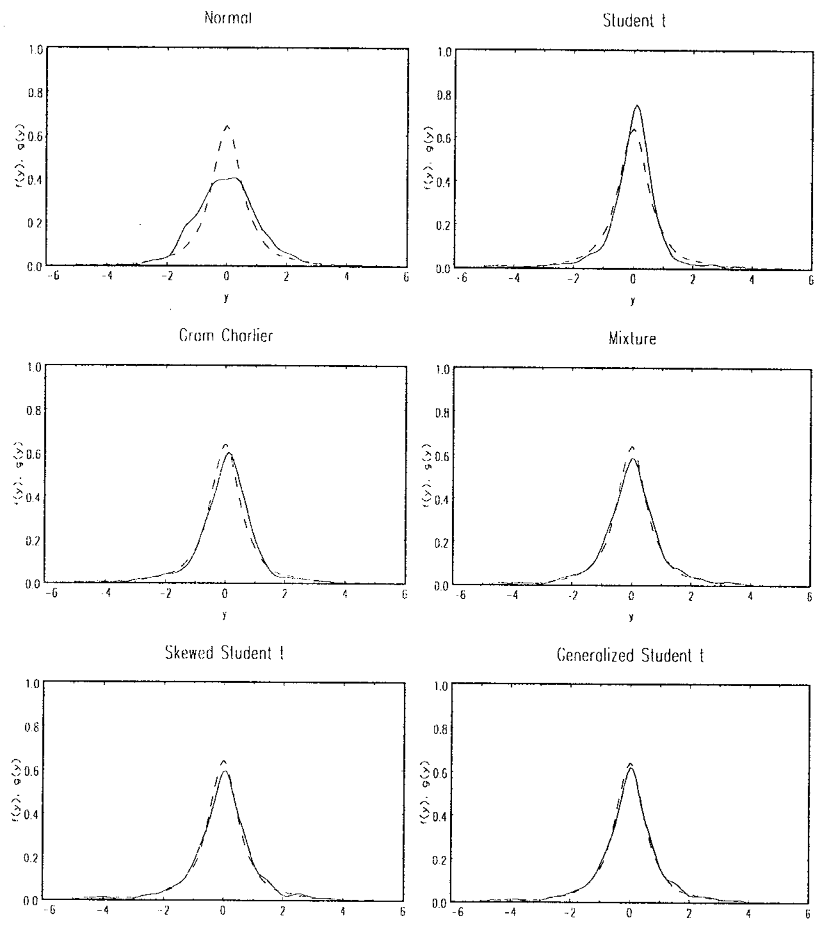

Figure 3: Unconditional distributions, BP currency, GARCH: theoretical (continuous line) empirical (broken line).

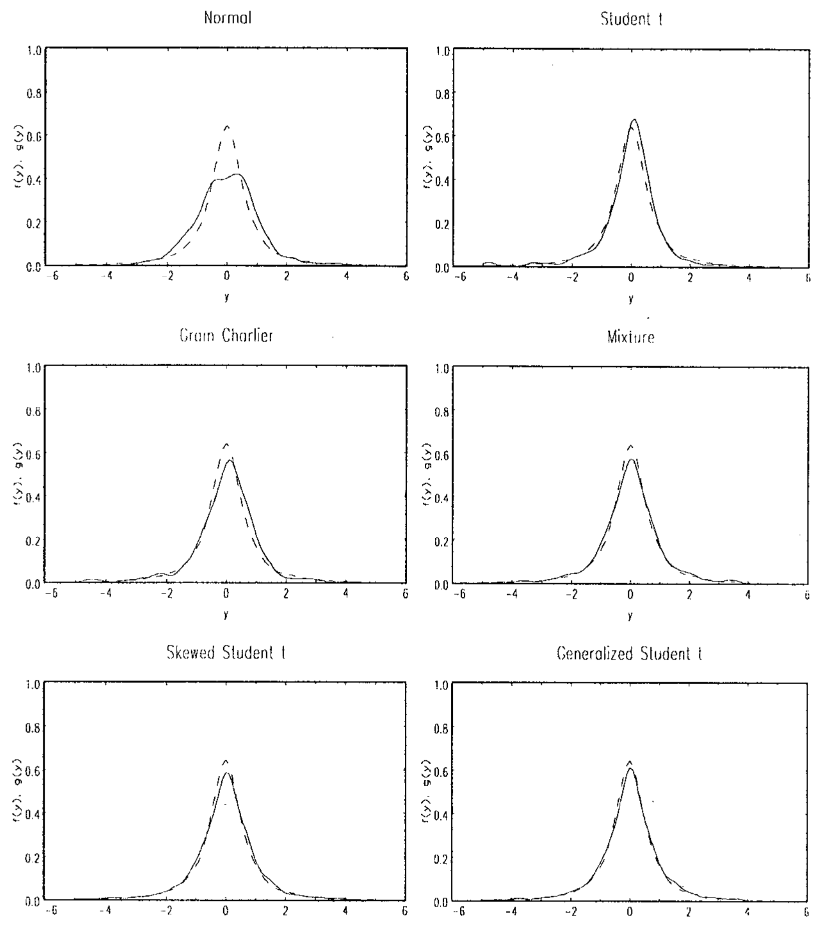

Figure 4: Unconditional distributions, BP currency, EGARCH: theoretical (continuous line) empirical (broken line).

V. CONCLUSIONS

This paper has provided a parametric framework for estimating conditional variance models based on a flexible class of distributions referred to as generalized Student t. An important feature of the generalized Student t distribution is that it is flexible enough to exhibit various distributional characteristics such as leptokurtosis and asymmetry. The generalized Student t distribution was shown to nest both the normal and Student t distributions, and to be related to the Gram Charlier and mixture distributions. This distribution was also shown to correspond to the unconditional distribution of a continuous time model used to approximate a discrete ARCH model.

Empirical models based on GARCH and EGARCH conditional variance specifications (where the conditional distribution was generalized Student t) were formulated and estimated for four currencies using intra-day exchange rate returns. These models were compared with models based on other conditional distributions including normal, Student t, Gram Charlier and mixture of normals, as well as a skewed Student t distribution. The generalized Student t distribution was found to perform uniformly better than these other distributions in modelling both the empirical conditional and unconditional distributions. One reason for the superiority of the generalized Student t distribution was that significant skewness was identified in the DM, JY and SF exchange rate returns that was not captured by the symmetric distributions.

REFERENCES

Ait-Sahalia, Y. (1995), ``Nonparametric Pricing of Interest Rate Derivatives", *Econometrica*, forthcoming.

Baillie, R.T. and T. Bollerslev (1989), ``The Message in Daily Exchange Rates: A Conditional Variance Tale", *Journal of Business and Economic Statistics*, 7, pp. 297-305.

Behboodian, J. (1970), ``On a Mixture of Normal Distributions", *Biometrika*, 57, pp.215-217.

Bollerslev, T. (1986), ``Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, pp. 307-327.

__________ and Wooldridge, J.M. (1992), ``Quasi-Maximum Likelihood Estimation and inference in Dynamic Models with Time-varying Covariances", *Econometric Reviews*, 11, pp.143-172.

__________, Chou, R.Y. and Kroner, K.F. (1992), ``ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics*, 52, pp.5-59.

__________, Engle, R.F. and Nelson, D. (1994), ``ARCH Models", *The Handbook of Econometrics*, Vol.4, North-Holland, forthcoming.

Cobb, L. (1978), ``Stochastic Catastrophe Models and Multimodal Distributions", *Behavioural Science*, 23, pp. 360-374.

__________, P. Koppstein, and N.H. Chen (1983), ``Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family", *Journal of the American Statistical Association*, 78, pp.124-130.

__________ and S. Zachs (1985), ``Applications of Catastrophe Theory for Statistical Modelling in the Biosciences", *Journal of the American Statistical Association*, 80, pp. 793-802.

Conley, T., Hansen, L.P., Luttmer, E. and Scheinkman, J. (1995), ``Estimating Subordinated Diffusions from Discrete data", University of Chicago, mimeo.

Creedy, J. and Martin, V.L. (1994), ``A Model of the Distribution of Prices", *Oxford Bulletin of Economics and Statistics*, 56, pp. 67-76.

__________, Lye, J.N. and Martin, V.L. (1996), ``A Nonlinear Model of the Real US/UK Exchange Rate", *Journal of Applied Econometrics*, forthcoming.

de Vries, C.G. (1994), ``Stylized Facts of Nominal Exchange Rate Returns", *The Handbook of International Economics*, F. van der Ploeg (ed.), Blackwell, pp.348-389.

Eisenberger, I. (1964), ``Genesis of bimodal distributions", *Technometrics*, 6, pp. 357-363.

Engle, R.F. (1982), ``Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica*, 50, pp. 987-1008.

__________ and G. Gonzalez-Rivera (1991), ``Semiparametric ARCH Models", *Journal of Business and Economic Statistics*, 9, pp. 345-360.

Friedman, D. and Vandersteel, S. (1982), ``Short-Run Fluctuations in Foreign Exchange Rates: Evidence from the Data 1973-79", *Journal of International Economics*, 13, pp.171-186.

Gallant, R.A., Hsieh, D.A. and Tauchen, G. (1991), ``On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 974-83", in W.A. Barnett, J. Powell and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics,* Cambridge University Press, Chapter 8.

Harvey, A., Ruiz, E. and Sentana, E. (1992), ``Unobserved Component Time Series Models with ARCH Disturbances", *Journal of Econometrics*, 52, pp.129-157.

Hsieh, D.A. (1989), ``Modelling Heteroskedasticity in Daily Foreign Exchange Rates", *Journal of Business and Economic Statistics*, 7, pp. 307-317.

Jorion, P. (1988), ``On Jump Processes in the Foreign Exchange and Stock Markets", *Review of Financial Studies*, 1, pp.427-445.

Kendall, M.G. and A. Stuart (1963), The Advanced Theory of Statistics, Vol 1, 6th edition, London: Charles Griffen and Co.

Lee, T.K. and Y.K. Tse (1991), ``Term Structure of Interest Rates in the Singapore Asian Dollar Market", *Journal of Applied Econometrics*, 6, pp.143-152.

Lim, G.C., Lye, J.N., Martin, G.M. and Martin, V.L. (1996), ``The Distribution of Exchange Rate returns and the Pricing of Currency Options", mimeo.

Linton. O. (1993), ``Adaptive Estimation in ARCH Models", *Econometric Theory*, 9, pp.539-569.

Lye, J.N. and V.L. Martin (1993), ``Robust Estimation, Non-Normalities and Generalized Exponential Distributions", *Journal of the American Statistical Association*, pp.253-259.

Martin, V.L. (1990), Properties and Applications of Distributions from the Generalized Exponential Family, unpublished Ph.D. Thesis, Monash University.

Nelson, D.A. (1990), ``ARCH Models as Diffusion Approximations", *Journal of Econometrics*, 45, pp.7-38.

__________ (1991), ``Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, 59, pp. 347-370.

O'Toole, A.L. (1933a), ``On the System of Curves for which the Method of Moments is the Best Method of Fitting", *Annals of Mathematical Statistics*, 4, pp. 1-29.

__________ (1933b), ``A Method of Determining the Constants in the Bimodal Fourth Degree Exponential Function", *Annals of Mathematical Statistics*, 4, pp. 80-93.

Phillips, R.F. (1991), ``A Constrained Maximum-Likelihood Approach to Estimating Switching Regressions", *Journal of Econometrics*, 48, pp. 241-262.

__________ (1994), ``Partially Adaptive Estimation via a Normal Mixture", *Journal of Econometrics*, 64, pp.123-144.

Robertson, C.A. and J.G. Fryer (1970), ``The Bias and Accuracy of Moment Estimators", *Biometrika*, 57, pp. 57-65.

Rogers, A.J. (1986), ``Modified Lagrange Multiplier Tests for Problems with One-Sided Alternatives", *Journal of Econometrics*, 31, pp.341-361.

Weiss, A.A. (1986), ``Asymptotic Theory for ARCH Models: Estimation and Testing", *Econometric Theory*, 2, pp.107-131.