Beta-t-(E)GARCH

Andrew Harvey and Tirthankar Chakravarty

September 2008

CWPE 08340

Beta-t-(E)GARCH

Andrew Harvey and Tirthankar Chakravarty,
Faculty of Economics, Cambridge University.
ACH34@ECON.CAM.AC.UK

September 18, 2008

Abstract

The GARCH-t model is widely used to predict volatilty. However, modeling the conditional variance as a linear combination of past squared observations may not be the best approach if the standardized observations are non-Gaussian. A simple modification lets the conditional variance, or its logarithm, depend on past values of the score of a t-distribution. The fact that the transformed variable has a beta distribution makes it possible to derive the properties of the resulting models. A practical consequence is that the conditional variance is more resistant to extreme observations. Extensions to deal with leverage and more than one component are discussed, as are the implications of distributions other than Student's t.

KEYWORDS: Conditional heteroskedasticity; leverage; robustness; score; Student's t; volatility.

JEL classification; C22, G10

1 Introduction

An established feature of stock returns is that they exhibit volatility clustering. The generalized autoregressive conditional heteroskedastic (GARCH) model, introduced by Bollerslev (1986) and Taylor (1986), is probably the most widely used model for capturing changing variance. The basic GARCH(1,1) specification is

$$y_t = \sigma_{t|t-1} z_t, \qquad t = 1, ..., T, \tag{1}$$

where y_t is the $t-th$ observation, z_t is standard normal serially independent random variable and the conditional variance is

$$\sigma^2_{t|t-1} = \delta + \beta\sigma^2_{t-1|t-2} + \alpha y^2_{t-1}, \quad \gamma > 0, \beta \geq 0, \alpha \geq 0 \tag{2}$$

The notational convention follows that in Andersen et al (2006); the use of $\sigma^2_{t|t-1}$, rather than simply σ^2_t, serves as a reminder that it is given by a filter and depends on information at time $t-1$. We can also write

$$\sigma^2_{t|t-1} = \delta + (\alpha + \beta)\sigma^2_{t-1|t-2} + \alpha\sigma^2_{t-1|t-2}u_{t-1}, \tag{3}$$

where $u_{t-1} = y^2_{t-1}/\sigma^2_{t-1|t-2} - 1$ is a martingale difference (MD). The conditional variance and the observations themselves are weakly stationary if $\alpha + \beta < 1$. The integrated GARCH (IGARCH) model is obtained when $\alpha + \beta = 1$. In this case, repeated substitution shows that $\sigma^2_{t|t-1}$ can be written as an exponentially weighted moving average (EWMA) of past squared observations.

Another stylized fact about returns is that their distributions typically have heavy tails. Although the GARCH structure induces excess kurtosis in the returns, it is not usually enough to match the data. As a result, it is now customary to assume that z_t has a Student t_ν-distribution, where ν denotes degrees of freedom. The GARCH-t model, which was originally proposed by Bollerslev (1987), is widely used in empirical work and as a benchmark for other models; see, for example, Kim, Shephard and Chib (1998) and Zhang and King (2008 p559).

The t-distribution is employed in the predictive distribution of returns and used as the basis for maximum likelihood (ML) estimation of the parameters, but it is not acknowledged in the design of the equation for the conditional variance. The specification of $\sigma^2_{t|t-1}$ as a linear combination of squared observations is taken for granted, but the consequences are that it responds too much to extreme observations and the effect is slow to dissipate. These features of GARCH are well-known and have prompted the development of a number of nonparametric procedures for robustification; see Sakata and White (1998) and Muler and Yohai (2008). Here we approach the problem in a different way by asking what the assumption of a t_ν-distribution for z_t implies about the specification of an equation for the conditional variance. The possible inappropriateness of letting $\sigma^2_{t|t-1}$ be a linear function of past squared observations when ν is finite becomes apparent on noting that, if the variance were constant, the sample variance would be an inefficient estimator of it[1]. The proposed modification replaces u_t in the conditional variance equation, (3), by another MD

$$u_t = \frac{(\nu+1)y_t^2}{(\nu-2)\sigma_{t|t-1}^2 + y_t^2} - 1, \quad -1 \leq u_t \leq \nu, \quad \nu > 2. \tag{4}$$

This variable is proportional to the score of the conditional distribution of y_t. The recognition that $(u_t+1)/(\nu+1)$ has a beta distribution facilitates the derivation of the model's properties. The definition can be modified to deal with $\nu \leq 2$ by measuring volatility in terms of a scale parameter.

A general model, which we call $Beta-t-GARCH(p,q)$, may be defined. The conditional variance is a linear function of q lagged conditional variances and p lagged u_t's. A similar equation may be constructed for the logarithm of the conditional variance. This model, denoted $Beta-t-EGARCH(p,q)$, belongs to the exponential GARCH (EGARCH) class of models introduced by Nelson (1991). In the formulation proposed by Nelson, the unconditional moments of the observations typically do not exist for a t_ν-distribution with finite degrees of freedom. This is not the case with the $Beta-t-EGARCH$ model.

Section 2 of the article sets out the statistical reasoning underlying the choice of u_t as the variable driving the conditional variance. The motivation is based on the idea of local likelihood, where the more recent observations receive a bigger weight. This idea is first applied to the estimation of variance, or rather scale, for a generalized error distribution (GED). The normal distribution is a special case of the GED and when exponential discounting is used to define the local likelihood function, IGARCH is obtained. Applying the same principle to a t_ν-distribution leads to the proposed class of models.

The properties of $Beta-t-EGARCH(p,q)$ models are obtained in section 3 and it is shown that they have the advantages of EGARCH without some of the disadvantages. Exploiting properties of the beta distribution enables expressions for the moments of the observations and the autocorrelations of powers of absolute values to be derived. Section 4 analyses the $Beta-t-GARCH$ model. In section 5, the models are fitted to Dow-Jones and FTSE returns and the resulting conditional variances are contrasted with those produced by GARCHt in the vicinity of the great crash of October 1987. Extensions to handle leverage effects, asymmetry and models with

[1]Efficiency is $(\nu+3)(\nu-4)/\{\nu(\nu-1)\}$.

long and short-run components are described in sections 6, 7 and 8. Section 9 develops a filter for handling a $Beta-t-EGARCH$ model in which the observation level changes over time and illustrates how it may be applied to the rate of inflation. Section 10 concludes.

2 Filters and distributions

A filter is a scheme for weighting current and past observations in order to estimate an unobserved component or future values of the series. In a linear Gaussian state space model, the minimum mean square error estimator (MMSE) of any linear combination of state vector elements is linear and given by the Kalman filter. Once the normality assumption is dropped, linear estimators are no longer optimal. Computer intensive solutions to problems of this kind are an active research area, but such procedures are rarely transparent and are not always guaranteed to converge to a sensible result.

As a specific example consider an unobserved components model made up of a random walk level plus white noise, that is

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim IID\left(0, \sigma_\varepsilon^2\right), \qquad t = 1, ..., T \\ \mu_t &= \mu_{t-1} + \eta_t, \quad \eta_t \sim IID(0, \sigma_\eta^2), \end{aligned} \tag{5}$$

where the irregular and level disturbances, ε_t and η_t respectively, are mutually independent and the notation $IID\left(0, \sigma^2\right)$ denotes independently and identically distributed with mean zero and variance σ^2. When σ_η^2 is zero, the level is constant. When both disturbances are Gaussian, the MMSE of the current level, in the steady-state, is an EWMA of past observations. Updating is carried out by the recursion

$$\widetilde{\mu}_t = \omega \widetilde{\mu}_{t-1} + (1-\omega) y_t, \qquad 0 < \omega < 1,$$

where the smoothing constant, ω, depends on the signal-noise ratio, $q = \sigma_\eta^2 / \sigma_\varepsilon^2$. Now suppose that ε_t has a Student t_ν-distribution. In this case the EWMA is no longer the MMSE although it is the MMSLE. An efficient estimator can be computed by Monte Carlo methods as described in Durbin and Koopman (2001, p 233-5). We might also consider whether it is possible to develop a simple approximate filter. Let the filter for the Gaussian model be written

$$\widetilde{\mu}_t = \sum_{i=1}^{t} w_{i,t} y_i \tag{6}$$

where $w_{i,t}$, $i = 1, ..., t$, denotes a set of weights which in the random walk plus noise case are proportional to ω^{t-i}. These weights are then used to construct a log-likelihood function for a time-invariant t-distribution which is then maximized with respect to the location parameter μ. The idea is similar to that of local likelihood where a kernel is applied to a likelihood function; see, for example, Fan and Gijbels (1996). For the normal distribution this device simply gives $\widetilde{\mu}_t$. For the t-distribution it leads to filter that depends on the score. This filter gives the mean of a conditional t-distribution which now defines the model. If the weights are taken to depend on a parameter, such as ω, the likelihood function from the new model is maximized with respect to this parameters and ν; see section 9.

The main objective here is to apply the idea of local likelihood to the estimation of variance or, more generally, a scale parameter. Again a local likelihood is constructed with weights as in (6). For a Gaussian distribution with zero mean, maximizing this local likelihood with respect to the variance gives a linear function of squared observations, that is $\widetilde{\sigma}_t^2 = \sum_{i=1}^{t} w_{i,t} y_i^2$. When the weights are proportional to ω^{t-i}, IGARCH is obtained. The first sub-section shows this result as a special case of the generalized error distribution and in doing so gives an interesting insight into a number of models in the ARCH family. The second sub-section provides the rationale for the proposed $Beta-t-EGARCH$ model.

2.1 Changing variance and the generalized error distribution

The local log-likelihood kernel at time t for the generalized error distribution is

$$L = -\sum_{i=1}^{t} w_{i,t} \ln \phi_t - \frac{1}{2\phi_t^p} \sum_{i=1}^{t} w_{i,t} \left| y_i \right|^p, \qquad t = 1, ..., T$$

where the $w_{i,t}' s$ are weights. The scale parameter, ϕ_t, is related to the variance by the formula $\sigma_t^p = 2\Gamma\left(3/p\right)\Gamma\left(1/p\right)^{-p/2}\phi_t^p$.

The score is

$$\frac{\partial L}{\partial \phi_t} = -\frac{1}{\phi_t}\sum_{i=1}^{t} w_{i,t} + \frac{p}{2\phi_t}\sum_{i=1}^{t} w_{i,t}\left|y_i/\phi_t\right|^p$$

and setting it to zero yields the estimator

$$\tilde{\phi}_t^p = \frac{p}{2}\frac{\sum w_{i,t}\left|y_i\right|^p}{\sum w_{i,t}}, \qquad t = 1, ..., T$$

If $w_{i,t} = \omega^{t-i}$, then $\tilde{\phi}_t^p$ is an EWMA of absolute values raised to the $p-th$ power. This estimator may be regarded as a predictor of ϕ_{t+1}^p, and as such we will write it as $\phi_{t+1|t}^p$. The EWMA recursion may then be written as

$$\phi_{t+1|t}^p = \omega\phi_{t|t-1}^p + (p/2)\left(1-\omega\right)\left|y_t\right|^p, \quad 0 < \omega \leq 1, \tag{7}$$

or

$$\begin{aligned} \phi_{t+1|t}^p &= \phi_{t|t-1}^p + (1-\omega)\left((p/2)\left|y_t\right|^p - \phi_{t|t-1}^p\right) \\ &= \phi_{t|t-1}^p + (1-\omega)\,\phi_{t|t-1}^p v_t \end{aligned} \tag{8}$$

where $v_t = (p/2)\left|y_t/\phi_{t|t-1}\right|^p - 1$ is a MD. Setting $p = 2$ gives IGARCH.

Equation (7) may be generalized by adding lags of $\phi_{t|t-1}^p$ and $\left|y_t\right|^p$. Rewriting $\phi_{t|t-1}^p$ in terms of $\sigma_{t+1|t}^p$ yields the power ARCH or APARCH class of models of Ding, Granger and Engle (1993), but without leverage effects. Setting $p = 1$ gives the model proposed by Taylor (1986).

2.2 Scale and variance in the t-distribution

The t_ν-distribution with a location (median) of μ and scale of $\exp(\lambda/2)$ has *pdf*

$$f\left(y_t;\lambda,\nu\right) = \frac{\Gamma\left(\left(\nu+1\right)/2\right)}{\Gamma\left(1/2\right)\Gamma\left(\nu/2\right)\left(\nu e^{\lambda}\right)^{1/2}}\left(1+\frac{\left(y_t-\mu\right)^2}{\nu e^{\lambda}}\right)^{-(\nu+1)/2}, \quad t = 1, ..., T, \tag{9}$$

where ν is a positive parameter indicating the degrees of freedom. When $\nu > 2$,

$$Var\left(y_t\right) = \sigma^2 = \left\{\nu/\left(\nu-2\right)\right\}\exp\left(\lambda\right). \tag{10}$$

For known degrees of freedom and $\mu = 0$, the kernel of the discounted log-likelihood function at time t is

$$L = -\frac{1}{2}\sum_{i=1}^{t} w_{i,t}\lambda_t - \frac{(\nu+1)}{2}\sum_{i=1}^{t} w_{i,t}\ln\left(1+\frac{y_i^2}{\nu e^{\lambda_t}}\right).$$

Differentiating with respect to λ_t gives

$$\frac{\partial L}{\partial \lambda_t} = \frac{1}{2}\sum_{i=1}^{t} w_{i,t}\left[\frac{(\nu+1)\,y_i^2}{\nu e^{\lambda_t}}\left(1+\frac{y_i^2}{\nu e^{\lambda_t}}\right)^{-1} - 1\right] \tag{11}$$

The expectation of the term in square brackets is zero; see Taylor and Verbyla (2004, p. 96). The information for λ_t is

$$I(\lambda) = \frac{\nu}{2(\nu+3)}\sum w_{i,t} = \frac{\nu}{2(\nu+3)(1-\omega)}$$

and so the scoring algorithm is

$$\tilde{\lambda}_{j,t} = \tilde{\lambda}_{j-1,t} + \kappa\sum_{i=1}^{t} w_{i,t}\xi_i(\tilde{\lambda}_{j-1,t}), \quad j = 1, 2, ...$$

where $\kappa = (\nu+3)(1-\omega)/\nu$ and $\xi_i(\tilde{\lambda}_{j-1,t})$ is the term in square brackets in (11). The notation $\xi_i(\tilde{\lambda}_{j-1,t})$ indicates that the score has been evaluated at $\tilde{\lambda}_{j-1,t}$, the estimate at the $j-1$ iteration. The scoring algorithm is iterated to convergence, at $j = J$, whereupon $\tilde{\lambda}_{J,t}$ becomes the initial value for scoring at time $t+1$, that is $\tilde{\lambda}_{0,t+1} = \tilde{\lambda}_{J,t}$.

If $w_{i,t} = \omega^{t-i}$, the first iteration in the scoring algorithm at time t is

$$\tilde{\lambda}_{1,t} = \tilde{\lambda}_{J,t-1} + \ \kappa\sum_{i=1}^{t}\omega^{t-i}\xi_i(\tilde{\lambda}_{J,t-1}) \tag{12}$$

We then write

$$\sum_{i=1}^{t}\omega^{t-i}\xi_i = \omega\sum_{i=1}^{t-1}\omega^{t-1-i}\xi_i + \xi_t$$

and if the iteration at time $t-1$ has converged, the first term on the right-hand side is zero. Then (12) can be written

$$\tilde{\lambda}_{1,t} = \tilde{\lambda}_{J,t-1} + \kappa\xi_t\left(\tilde{\lambda}_{J,t-1}\right).$$

Setting J to unity gives a recursion for estimating λ_t. As in the previous subsection, this estimator may be regarded as a predictor of λ_{t+1} and written as $\lambda_{t+1|t}$. The recursion is then

$$\lambda_{t+1|t} = \lambda_{t|t-1} + \kappa \left[\frac{(\nu+1)y_t^2}{\nu \exp(\lambda_{t|t-1}) + y_t^2} - 1 \right], \qquad t = 1, ..., T. \tag{13}$$

The recursion in (13) may be expressed in terms of the variance as

$$\ln \sigma^2_{t+1|t} = \ln \sigma^2_{t|t-1} + \kappa \left[\frac{(\nu+1)y_t^2}{(\nu-2)\sigma^2_{t|t-1} + y_t^2} - 1 \right], \qquad \nu > 2 \tag{14}$$

and so $\sigma^2_{t+1|t}$ is positive by construction. The term in square brackets is u_t, the MD defined in (4).

It is important to appreciate that the recursion in (13) is not intended to be an exact solution to the maximization of the local likelihood function. The local likelihood is simply a device which gives a suitable filter for a scale parameter that changes over time. The filter suggests that we define a new model in which the conditional distribution of y_t, that is $y_t \mid Y_{t-1}$, is t_ν with scale parameter $\exp(\widetilde{\lambda}_{t|t-1}/2)$. The link between κ and ω now loses its significance and it is better to replace κ by a new parameter θ in (13). The parameters ν and θ may be estimated by maximizing the likelihood function formed from the conditional distributions.

The local likelihood can be maximized with respect to the variance rather than with respect to λ_t. Proceeding in this way leads to the equation

$$\sigma^2_{t|t-1} = \sigma^2_{t-1|t-2} + \theta \sigma^2_{t-1|t-2} u_{t-1}, \tag{15}$$

where θ is a parameter. The recursion in (14) is close to that in (15) as can be seen by writing it as

$$\begin{aligned} \sigma^2_{t|t-1} &= \sigma^2_{t-1|t-2} \exp \left[\frac{\theta(\nu+1)y_{t-1}^2}{(\nu-2)\sigma^2_{t-1|t-2} + y_{t-1}^2} - \kappa \right] \\ &\simeq \sigma^2_{t-1|t-2} \left[1 + \frac{\theta(\nu+1)y_{t-1}^2}{(\nu-2)\sigma^2_{t-1|t-2} + y_{t-1}^2} - \theta \right]. \end{aligned}$$

3 Beta-t-EGARCH models

Let the observations be written as

$$y_t = \exp(\lambda_{t|t-1}/2)\varepsilon_t, \quad t = 1, ..., T, \tag{16}$$

where ε_t has a $t_\nu-$distribution and is serially independent. Alternatively

$$y_t = \sigma_{t|t-1} z_t, \quad \nu > 2, \tag{17}$$

where $z_t = ((\nu - 2)/\nu)^{1/2}\varepsilon_t$ still has a $t_\nu-$distribution, but standardized so as to have unit variance.

The principal feature of the Beta-t-EGARCH class is that $\lambda_{t|t-1}$ is a linear combination of past values of the variable

$$u_t = \frac{(\nu+1)y_t^2}{\nu\exp(\lambda_{t|t-1}) + y_t^2} - 1, \quad -1 \leq u_t \leq \nu, \quad \nu > 0. \tag{18}$$

This variable may be expressed as

$$u_t = (\nu+1)b_t - 1, \tag{19}$$

where

$$b_t = \frac{y_t^2/\nu\exp(\lambda_{t|t-1})}{1 + y_t^2/\nu\exp(\lambda_{t|t-1})}, \quad 0 < \nu < \infty, \tag{20}$$

is distributed as $Beta(1/2, \nu/2)$, a beta distribution of the first kind; see Stuart and Ord (1987, ch 2). Since $E(b_t) = 1/(\nu+1)$ and $Var(b_t) = 2\nu/\{(\nu+3)(\nu+1)^2\}$, it follows that u_t is a martingale difference with variance $2\nu/(\nu+3)$.

Figure 1 plots u_t against y_t for $\nu = 3$ and 10 with $\lambda_{t|t-1}$ set to zero. For $\nu = 3$, an extreme observation has only a moderate impact as it is regarded as coming from a $t_\nu-$ distribution rather than from a normal distribution with an abnormally high variance. As $|y_t| \to \infty$, $u_t \to \nu$. When $\nu = \infty$, the relationship between u_t and y_t is quadratic. The distribution of $u_t + 1$ is then χ_1^2.

In the $Beta - t - EGARCH(p, q)$ model, $\lambda_{t|t-1}$ in (16) is given by

$$\widetilde{\lambda}_{t|t-1} = \delta + \phi_1\widetilde{\lambda}_{t-1|t-2} + ... + \phi_q\widetilde{\lambda}_{t-q|t-q-1} + \theta_1 u_{t-1} + ... + \theta_p u_{t-p}. \tag{21}$$

The $Beta - t - EGARCH$ model belongs to the EGARCH family introduced by Nelson (1991). Stationarity depends on the roots of the autoregressive

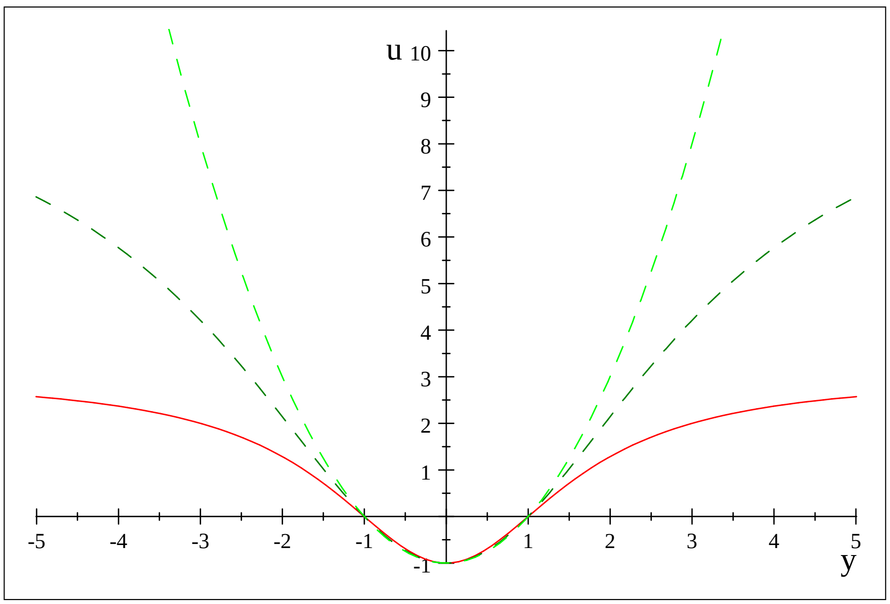

Figure 1: Impact of u_t on scale (variance) for $\nu = 3$ (solid line), $\nu = 10$ (dash) and $\nu = \infty$ (top line)

polynomial lying outside the unit circle, as in an ARMA model. Thus the first-order model

$$\widetilde{\lambda}_{t|t-1} = \delta + \phi\widetilde{\lambda}_{t-1|t-2} + \theta u_{t-1}, \tag{22}$$

is stationary if $|\phi| < 1$, with

$$E\left(\lambda_{t|t-1}\right) = \frac{\delta}{1-\phi} \quad \text{and} \quad Var\left(\lambda_{t|t-1}\right) = \frac{\theta^2}{1-\phi^2}\frac{2\nu}{\nu+3}.$$

The implications for stationarity of the observations themselves are determined in sub-section 3.2.

Long memory may be introduced, simply by pre-multiplying $\widetilde{\lambda}_{t|t-1}$ and its lags by $(1-L)^d$, so providing an alternative to the FIEGARCH model of Bollerslev and Mikkelsen (1996).

3.1 EGARCH

In the EGARCH model the conditional standard deviations in (17) are given by

$$\ln \sigma^2_{t|t-1} = \gamma_t + \sum_{k=1}^{\infty} \psi_k g\left(z_{t-k}\right), \quad \psi_1 = 1 \tag{23}$$

where γ_t and $\psi_k, k = 1, .., \infty$ are real and nonstochastic. The analysis in Nelson (1991) focusses on the specification

$$g\left(z_t\right) = \alpha z_t + \beta\left[\left|z_t\right| - E\left|z_t\right|\right], \tag{24}$$

where α and β are parameters. By construction $g\left(z_t\right)$ is a MD which is able to respond asymmetrically to rises and falls in stock price. Such leverage effects will be incorporated into (21) later.

Theorem 2.1 in Nelson (1991, p. 351) states that for model (17) and (23), with $g(\cdot)$ as in (24), $\sigma^2_{t|t-1}\exp\left(-\gamma_t\right), y_t \exp\left(-\gamma_t/2\right)$ and $\ln \sigma^2_{t|t-1} - \gamma_t$ are strictly stationary and ergodic, and $\ln \sigma^2_{t|t-1} - \gamma_t$ is covariance stationary[2] if and only if $\sum_{k=1}^{\infty} \psi_k^2 < \infty$. His theorem 2.2 demonstrates the existence of

[2] Nelson (1991,p 352) notes that his theorem 2.1 still applies to long memory models for $\ln \sigma^2_{t|t-1}$.

moments of $\sigma^2_{t|t-1}\exp(-\gamma_t)$ and $y_t\exp(-\gamma_t/2)$ for a generalized error distribution provided the index, p, is greater than one.[3] However, Nelson's application suggests that the GED does not adequately deal with outliers. Furthermore, Andersen et al (2006, p804) observe that a practical drawback to using absolute values, as in (24), is that their nondifferentiability makes such EGARCH models '.. more difficult to estimate and analyse numerically'.

Nelson notes that if z_t is t_ν distributed, the conditions needed for the existence of the moments of $\sigma^2_{t|t-1}\exp(-\gamma_t)$ and $y_t\exp(-\gamma_t/2)$ are rarely satisfied in practice. *However, if, as in* (21),

$$g(z_t)/\theta_1 = \frac{(\nu+1)\,z_t^2/(\nu-2)}{1+z_t^2/(\nu-2)} - 1,$$

the moments of $\sigma^2_{t|t-1}\exp(-\gamma_t)$ *always exist and the moments of* $y_t\exp(-\gamma_t/2)$ *exist provided the corresponding moments of* z_t *exist.* This result follows because u_t has bounded support for finite ν and so all its moments exist; see Stuart and Ord (1987 p215). Similarly its exponent has bounded support for $0<\nu<\infty$ and so $E[\exp(ab_t)]<\infty$ for $|a|<\infty$. These results on the existence of moments for y_t can be adapted to so as to cover any t_ν-distribution with $\nu>0$ by working with $\exp\left(\lambda_{t|t-1}\right)$ rather than $\sigma^2_{t|t-1}$. The first sub-section below derives expressions for the moments of the observations, while the second sub-section obtains autocorrelations functions for powers of absolute values of the observations.

3.2 Moments of observations

If $\lambda_{t|t-1}$ is covariance stationary, the $m^{th}-$moment of y_t exists for $m>\nu$ and may be written as

$$E(y_t^m) = E(\varepsilon_t^m)\,E\left(e^{\lambda_{t|t-1}m/2}\right), \quad m=1,2,3,... \tag{25}$$

Write (21) as

$$\lambda_{t|t-1} = \gamma + \sum_{k=1}^{\infty}\psi_k u_{t-k}$$

where $\gamma=\delta/(1-\Sigma\phi_j)$ and $\Sigma\psi_k^2<\infty$. The model could be generalized by letting γ be a deterministic function of time, as in (23), but to do so would complicate the exposition unnecessarily. In (22), $\psi_k=\theta\phi^{k-1}$.

[3] Thus the Laplace distribution is excluded, though the argument in sub-section 2.1 suggests that absolute values, as in (24), might be appropriate for this distribution.

Since the mean of y_t is assumed to be zero, the variance is

$$E\left(y_t^2\right) = E\left(\varepsilon_t^2\right) E\left(e^{\lambda_{t|t-1}}\right), \qquad \nu > 2, \tag{26}$$

while the fourth moment is

$$E\left(y_t^4\right) = E\left(\varepsilon_t^4\right) E\left(e^{2\lambda_{t|t-1}}\right), \qquad \nu > 4$$

Hence the kurtosis of y_t exceeds that of ε_t, denoted κ_ν, since

$$K_\nu = \frac{E\left(y_t^4\right)}{\left(Ey_t^2\right)^2} = \kappa_\nu \frac{E\left(e^{2\lambda_{t|t-1}}\right)}{\left[E\left(e^{\lambda_{t|t-1}}\right)\right]^2} \geq \kappa_\nu, \tag{27}$$

where

$$\kappa_\nu = \frac{E\left(\varepsilon_t^4\right)}{\left(E\left(\varepsilon_t^2\right)\right)^2} = \frac{3\left(\nu-2\right)}{\left(\nu-4\right)}$$

The last term in (25) is the moment generating function (MGF) of $\lambda_{t|t-1}/2$ and, by the law of iterated expectations (LIE), this term may be written as a convolution of MGFs for lagged $u_t's$, that is

$$E\left(e^{\lambda_{t|t-1}m/2}\right) = e^{\gamma m/2} \prod_{j=1}^{\infty} E_{t-j-1}\left(e^{\psi_j u_{t-j} m/2}\right) \tag{28}$$

where E_{t-j-1} denotes the expectation conditional on information at time $t-j-1$. This expression may be evaluated numerically. The calculation requires computing terms of the form $E\left(e^{au_t}\right)$, where $a = \psi_j m/2$. Substituting for u_t from (19) gives

$$E\left(e^{au_t}\right) = e^{-a} E\left(e^{a(\nu+1)b_t}\right) \tag{29}$$

and from the formula for the MGF of a $Beta(1/2, \nu/2)$ distribution,

$$E[\exp((\nu+1)ab_t)] = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1} \frac{1+2r}{\nu+1+2r}\right)\frac{a^k(\nu+1)^k}{k!}, \quad 0 < \nu < \infty. \tag{30}$$

When $\nu = \infty$, $u_t = \varepsilon_t^2 - 1$ and it follows from the formula for the MGF of a chi-square that $E\left(e^{au_t}\right) = e^{-a}(1-2a)^{-1/2}$ for $a < 1/2$. Thus the factor by which the kurtosis of the t-distribution increases in (27) is

$$K_\infty = \frac{E\left(e^{2\lambda_{t|t-1}}\right)}{\left[E\left(e^{\lambda_{t|t-1}}\right)\right]^2} = \frac{\prod_{j=1}^{\infty}(1-2\psi_j m)^{-1/2}}{\left(\prod_{j=1}^{\infty}(1-\psi_j m)^{-1/2}\right)^2},$$

provided that $2m\psi_j < 1,\ j = 1, 2, ...$

In the first-order model, (22), setting θ and ϕ to 0.06 and 0.98 respectively gives $K_\infty = 1.24$ when the products are truncated at $j = 1000$. For $\nu = 5$, we find $K_5 = 1.13$ using (30). The strength of volatility can be measured by $K_\nu - 1$, which is the ratio of the variance of $\sigma^2_{t|t-1}$ to the square of its expected value.

3.3 Autocorrelation functions of squares and powers of absolute values

Assuming that $\lambda_{t|t-1}$ is covariance stationary, the variance of y_t^2 is

$$Var\left(y_t^2\right) = E\left(\varepsilon_t^4\right) E\left(e^{2\lambda_{t|t-1}}\right) - \left(E\left(\varepsilon_t^2\right) E\left(e^{\lambda_{t|t-1}}\right)\right)^2, \quad \nu > 4.$$

To obtain the covariances it is first necessary to find

$$E\left(y_t^2 y_{t-\tau}^2\right) = E\left(\varepsilon_t^2 e^{\lambda_{t|t-1}} \varepsilon_{t-\tau}^2 e^{\lambda_{t-\tau|t-\tau-1}}\right), \quad \tau = 1, 2, ...$$

This is not straightforward because of the dependence between $e^{\lambda_{t|t-1}}$ and $\varepsilon_{t-\tau}^2$. However, applying the LIE we eventually obtain[4]

$$E\left(y_t^2 y_{t-\tau}^2\right) = E\left(\varepsilon_t^2\right).e^{2\gamma} \prod_{j=1}^{\tau-1} E_{t-j-1}\left(e^{\psi_j u_{t-j}}\right).B_\tau. \prod_{i=1}^{\infty} E_{t-\tau-i-1}\left(e^{\left(\psi_{\tau+i}+\psi_i\right)u_{t-\tau-i}}\right), \tag{31}$$

where

$$B_\tau = E_{t-\tau-1}\left(\varepsilon_{t-\tau}^2 e^{\psi_\tau u_{t-\tau}}\right), \quad \tau = 1, 2, ...$$

Re-arranging (19) gives $\varepsilon_t^2 = \nu b_t/(1-b_t)$. Therefore, dropping subscripts,

$$B = E\left(\frac{\nu b}{1-b} e^{\psi(\nu+1)b-\psi}\right) = \nu e^{-\psi} E\left(\frac{b}{1-b} e^{\psi(\nu+1)b}\right).$$

[4] For $\tau = 1$, the product over j is set to one.

Since b has a $Beta(1/2, \nu/2)$ distribution

$$\begin{aligned}
B &= \nu e^{-\psi} \frac{1}{B(1/2,\nu/2)} \int \frac{b}{1-b} b^{-1/2} (1-b)^{\nu/2-1} e^{\psi(\nu+1)b} db \\
&= \nu e^{-\psi} \frac{B(3/2,\nu/2-1)}{B(1/2,\nu/2)} . \frac{1}{B(3/2,\nu/2-1)} \int b^{1/2} (1-b)^{\nu/2-2} e^{\psi(\nu+1)b} db \\
&= \nu e^{-\psi} \frac{B(3/2,\nu/2-1)}{B(1/2,\nu/2)} E(e^{\psi(\nu+1)b})
\end{aligned}$$

where $B(.,.)$ is the beta function. Hence

$$B_\tau = \frac{\nu}{\nu-2} E(e^{\psi_\tau(\nu+1)b_{t-\tau} - \psi_\tau}), \quad \tau = 1, 2, ...$$

where it is understood that the expectation is taken with respect to a $Beta(3/2, \nu/2-1)$ distribution. Since $E\left(\varepsilon_t^2\right) = \nu/(\nu-2)$, substituting in (31) gives

$$E\left(y_t^2 y_{t-\tau}^2\right) = (E\left(\varepsilon_t^2\right))^2 e^{2\gamma} E(e^{\psi_\tau u_\tau}) \prod_{j=1}^{\tau-1} E\left(e^{\psi_j u_{t-j}}\right) . \prod_{i=1}^{\infty} E\left(e^{\left(\psi_{\tau+i}+\psi_i\right) u_{t-\tau-i}}\right)$$

The autocorrelations are then

$$\rho(\tau; y_t^2) = \frac{C_\tau - \left[E\left(e^{\lambda_{t|t-1}}\right)\right]^2}{\kappa_\nu E\left(e^{2\lambda_{t|t-1}}\right) - \left[E\left(e^{\lambda_{t|t-1}}\right)\right]^2}, \quad \tau = 1, 2, ... \tag{32}$$

where

$$C_\tau = e^{2\gamma} E(e^{\psi_\tau(\nu+1)b_{t-\tau} - \psi_\tau}) \prod_{j=1}^{\tau-1} E\left(e^{\psi_j u_{t-j}}\right) . \prod_{i=1}^{\infty} E\left(e^{\left(\psi_{\tau+i}+\psi_i\right) u_{t-\tau-i}}\right),$$

but the factor $\exp(2\gamma)$ cancels. The ACF may be evaluated numerically. The calculation requires that we compute $E\left(e^{au_t}\right)$ from (29), where $a = \psi_i, 2\psi_i$ or $\psi_{\tau+i} + \psi_i$. The first expectation in C_τ is evaluated with $1 + 2r$ replaced by $3 + 2r$ in the product term of the expansion, (30).

The same approach may be used to find an expression for the ACF of $|y_t|^c$ for $0 < c < \nu/2$. Formula (32) becomes

$$\rho(\tau; |y_t|^c) = \frac{C_\tau(c) - \left[E\left(e^{\lambda_{t|t-1} c/2}\right)\right]^2}{\kappa_\nu(c) E\left(e^{\lambda_{t|t-1} c}\right) - \left[E\left(e^{\lambda_{t|t-1} c/2}\right)\right]^2}, \quad \tau = 1, 2, ...$$

where

$$\kappa_\nu(c) = \frac{\Gamma(c+1/2)\Gamma(-c+\nu/2)\Gamma(1/2)\Gamma(\nu/2)}{\{\Gamma(c/2+1/2)\Gamma(-c/2+\nu/2)\}^2}, \qquad 0 < c < \nu/2,$$

and

$$C_\tau(c) = e^{c\gamma} E(e^{\psi_\tau u_\tau c/2}) \prod_{j=1}^{\tau-1} E\left(e^{\psi_j u_{t-j} c/2}\right) . \prod_{i=1}^{\infty} E\left(e^{\left(\psi_{\tau+i}+\psi_i\right) u_{t-\tau-i} c/2}\right).$$

The result follows as $C_\tau(c)$ is obtained from $e^{-\gamma c} B_\tau(c) F_\tau / D$, where[5]

$$D = E(|\varepsilon_t|^c) = \nu^{c/2}\Gamma(c/2+1/2)\Gamma(-c/2+\nu/2)/\{\Gamma(1/2)\Gamma(\nu/2)\},$$

$$B_\tau(c) = E\left(|\varepsilon_{t-\tau}|^c \, e^{\psi_\tau u_{t-\tau} c/2}\right), \quad \tau = 1, 2, ...$$

and

$$F_\tau = e^{c\gamma} \prod_{j=1}^{\tau-1} E\left(e^{\psi_j u_{t-j} c/2}\right) . \prod_{i=1}^{\infty} E\left(e^{\left(\psi_{\tau+i}+\psi_i\right) u_{t-\tau-i} c/2}\right).$$

Noting that $|\varepsilon_t|^c = \nu^{c/2} b_t^{c/2}/(1-b_t)^{c/2}$, and proceeding as before we find

$$\begin{aligned} B_\tau(c) &= \nu^{c/2} \frac{B(c/2+1/2, -c/2+\nu/2)}{B(1/2, \nu/2)} E(e^{\psi_\tau((\nu+1)b_{t-\tau}-1)c/2}) = \\ &= E(|\varepsilon_{t-\tau}|^c) E(e^{\psi_\tau((\nu+1)b_{t-\tau}-1)c/2}) \end{aligned}$$

with the expectation taken with respect to a $Beta((c+1)/2, \nu/2 - c/2)$ distribution. Thus $1 + 2r$ is replaced by $c + 1 + 2r$ in the product term of the expansion, (30). Since $B_\tau(c)/D = E(e^{\psi_\tau((\nu+1)b_{t-\tau}-1)c/2})$, the expression for $C_\tau(c)$ is immediately obtained.

Further simplification leads to

$$\rho(\tau; |y_t|^c) = \frac{G_\tau(c) - 1}{\kappa_\nu(c) K_\nu(c) - 1}, \quad \nu > 2c, \quad \tau = 1, 2, ..., \tag{33}$$

where

$$K_\nu(c) = \frac{E\left(e^{\lambda_{t|t-1} c}\right)}{\left[E\left(e^{\lambda_{t|t-1} c/2}\right)\right]^2},$$

[5]If ε_t is t_ν, then $|\varepsilon_t|^c$ can be expressed in terms of chi-square variables raised to the power $c/2$. The expected value of a chi-square with ν df raised to the power $c/2$ is $2^{c/2}\Gamma(c/2+1/2)/\Gamma(\nu/2)$, $c > -\nu$.

with $E\left(e^{\lambda_{t|t-1}c/2}\right)$ as in (28) and $G_\tau(c) = e^{-\gamma c}C_\tau(\tau)/\left[E\left(e^{\lambda_{t|t-1}c/2}\right)\right]^2$ so

$$G_\tau(c) = \frac{E(e^{\psi_\tau(\nu+1)b_{t-\tau}c/2})\prod_{j=1}^{\tau-1}E\left(e^{\psi_j(\nu+1)b_{t-j}c/2}\right)\prod_{i=1}^{\infty}E\left(e^{\left(\psi_{\tau+i}+\psi_i\right)(\nu+1)b_{t-\tau-i}c/2}\right)}{\left(\prod_{j=1}^{\infty}E\left(e^{\psi_j(\nu+1)b_{t-j}c/2}\right)\right)^2}.$$

For a normal distribution, the algebraic manipulations in the appendix lead to the following modifications in the formula for $\rho(\tau;|y_t|^c)$ in (33):

$$G_\tau(c) = \frac{(1-c\psi_\tau)^{-(1+c)/2}\left(\prod_{j=1}^{\tau-1}(1-c\psi_j)^{-1/2}\right)\left(\prod_{i=1}^{\infty}(1-c(\psi_{\tau+i}+\psi_i))^{-1/2}\right)}{\left(\prod_{j=1}^{\infty}(1-c\psi_j)^{-1/2}\right)^2}$$

and

$$K_\nu(c) = \left(\prod_{j=1}^{\infty}(1-2c\psi_j)^{-1/2}\right)\left(\prod_{j=1}^{\infty}(1-c\psi_j)^{-1/2}\right)^{-2}.$$

For the first-order model, where $\psi_k = \theta\phi^{k-1}$, typical values of θ and ϕ are 0.06 and 0.98 respectively; see Taylor (2005, p203). The above expressions for a normal distribution are easily computed. For squared observations, that is $c = 2$, the autocorrelations at $\tau = 1, 2$ and 10 are 0.148, 0.145, and 0.118 respectively. With $c = 1$, that is absolute values, the corresponding figures are 0.127, 0.124, and 0.104. These figures are very similar to those obtained with GARCH(1,1) and it is the case that $\rho(\tau;|y_t|^c) \simeq \phi^{\tau-1}\rho(1;|y_t|^c)$. This relationship is exact for the autocorrelations of squared observations from a GARCH(1,1) model. The similarity in the ACFs is not surprising in view of the approximation given at the end of section 2.

For the Student-t model, the implications of choosing different values of c are likely to be similar to those for the stochastic volatility model, analysed in Harvey and Streibel (1998, pp 180-3). The kurtosis of $|\varepsilon_t|^{c/2}$, $\kappa_\nu(c)$, plays an important role and as ν becomes smaller, $\kappa_\nu(c)$ increases, but is smaller, the smaller is c. For $\nu = 5$, and θ and ϕ as before, evaluation of $\rho(1;|y_t|^c)$, as in (33), gives 0.071 for $c = 1$ as opposed to 0.031 for $c = 2$.

4 Beta-t-GARCH

For $\nu > 2$, a $Beta-t-GARCH(p,q)$ model, having a form analogous to that of the standard $GARCH(p,q)$ model, may be set up as in (17) with

$$\sigma^2_{t|t-1} = \delta + \phi_1\sigma^2_{t-1|t-2} + ... + \phi_q\sigma^2_{t-q|t-q-1} + \theta_1\sigma^2_{t-1|t-2}u_{t-1} + ... + \theta_p\sigma^2_{t-p|t-p-1}u_{t-p} \tag{34}$$

where u_t is as in (4). If $\nu \leq 2$ we can work with an equation for $\exp(\lambda_{t|t-1})$ instead of $\sigma^2_{t|t-1}$ so the condition $\nu > 2$ is not restrictive.

It is convenient to set $q = \max(p,q)$ on the understanding that some coefficients may be set to zero. The model can be re-written as

$$\begin{aligned} \sigma^2_{t|t-1} &= \delta + \beta_1\sigma^2_{t-1|t-2} + ... + \beta_q\sigma^2_{t-q|t-q-1} \\ &\quad + \alpha_1\sigma^2_{t-1|t-2}(\nu+1)b_{t-1} + ... + \alpha_q\sigma^2_{t-q|t-q-1}(\nu+1)b_{t-1-q} \end{aligned} \tag{35}$$

where $\alpha_i = \theta_i$ and $\beta_i = \phi_i - \alpha_i, i = 1, ..., q$. In the limit as $\nu \to \infty$, $(\nu+1)b_t = y_t^2$ leading to the standard GARCH specification. As in GARCH, a sufficient condition for the conditional variance to remain positive is $\delta > 0$, $\beta_i \geq 0$, and $\alpha_i \geq 0$, $i = 1, ..., q$.

The $Beta-t-GARCH(1,1)$ model is

$$\sigma^2_{t|t-1} = \delta + \phi\sigma^2_{t-1|t-2} + \theta\sigma^2_{t-1|t-2}u_{t-1}. \tag{36}$$

This model is a member of the class of models defined by He and Terasvirta (1999) in which $\sigma_{t|t-1}$ in (17) is given by

$$\widetilde{\sigma}^d_{t|t-1} = a(z_{t-1}) + c(z_{t-1})\widetilde{\sigma}^d_{t-1|t-2}$$

In (36), $d = 2$, $a(z_{t-1}) = \delta$ and $c(z_{t-1}) = \phi + \theta u_{t-1}$. From He and Terasvirta (1999), y_t is second-order stationary if $\nu > 2$ and $E(\phi + \theta u_{t-1}) = \phi < 1$. Furthermore the series is strictly stationary and ergodic; see also theorem 1 in Ling and McAleer (2002). If $Ez_t^j < \infty$, a necessary and sufficient condition for the existence of the $j-th$ moment of y_t is

$$E[c(z_t)]^{j/2} < 1, \quad j = 2, 4, ...$$

For $j = 4$, the condition is

$$\phi^2 + \theta^2 E(u_t^2) = \phi^2 + \theta^2\frac{2\nu}{\nu+3} < 1, \quad \nu > 4,$$

or, if we write $c(z_t) = \beta + \alpha(\nu + 1)b_t$,

$$\beta^2 + 2\alpha\beta + \frac{3\alpha^2(\nu+1)}{\nu+3} < 1, \quad \nu > 4.$$

In the limit as $\nu \to \infty$ the above expression tends to the standard GARCH(1,1) condition for the existence of fourth moments.

Given finite fourth moments, squaring (36), taking (unconditional) expectations and carrying out some algebraic manipulations leads to the following expression for the kurtosis of the observations:

$$\kappa_y = E\left(y_t^4\right) / \left(Ey_t^2\right)^2 = \kappa_\nu K \geq \kappa_\nu$$

where κ_ν was defined in (27) and

$$K = \frac{E\left(\sigma^4_{t|t-1}\right)}{\left(E\sigma^2_{t|t-1}\right)^2} = \frac{1-\phi^2}{1-\phi^2-\theta^2 2\nu/(\nu+3)}.$$

For a normal distribution, subtracting three from κ_y yields the formula given for excess kurtosis by Bollerslev (1986).

As in the standard GARCH(1,1) model, the autocorrelation function of the $Beta-t-GARCH(1,1)$ model is of the form $\rho(\tau; y_t^2) = \phi^{\tau-1}\rho(1; y_t^2)$ for $\tau \geq 1$, but $\rho(1; y_t^2)$ now depends on ν as well as ϕ and θ.

The derivation of $\rho(\tau; y_t^2)$ follows from first taking conditional expectations at time $t-1$ of

$$y_t^2 y_{t-\tau}^2 = z_t^2 \sigma^2_{t|t-1} z^2_{t-\tau} \sigma^2_{t-\tau|t-\tau-1}, \qquad \tau = 1, 2, ...$$

to give

$$E_{t-1}(y_t^2 y_{t-\tau}^2) = \sigma^2_{t-1|t-2}(\phi + \theta u_{t-1}) z^2_{t-\tau}\sigma^2_{t-\tau|t-\tau-1} + \delta z^2_{t-\tau}\sigma^2_{t-\tau|t-\tau-1}.$$

Applying the LIE we obtain

$$\begin{aligned} E_{t-\tau-1}\left(y_t^2 y_{t-\tau}^2\right) &= \phi^{\tau-1}\sigma^2_{t-\tau|t-\tau-1}(\phi + \theta E_{t-\tau-1}(z^2_{t-\tau}u_{t-\tau})) \\ &\quad + \delta(1 + \phi + .. + \phi^{\tau-1})\sigma^2_{t-\tau|t-\tau-1}. \end{aligned}$$

Now $z_t^2 = (\nu - 2)b_t/(1 - b_t)$. Proceeding as in sub-section 3.3,

$$\begin{aligned} E_{t-\tau-1}\left(z^2_{t-\tau}u_{t-\tau}\right) &= \frac{(\nu-2)(\nu+1)}{B(1/2,\nu/2)}\int \frac{b^2}{1-b} b^{-1/2}(1-b)^{\nu/2-1}db - 1 \\ &= \frac{B(5/2,\nu/2-1)}{B(1/2,\nu/2)} - 1 = 2. \end{aligned}$$

Noting that $Ey_t^2 = E(z_t^2\sigma_{t|t-1}^2) = E\sigma_{t|t-1}^2 = \delta/(1-\phi)$, and that $Ey_t^2 = E(z_t^4\sigma_{t|t-1}^4) = \kappa_\nu E\sigma_{t|t-1}^4$, the autocorrelations are found to be

$$\rho(\tau; y_t^2) = \frac{\phi^{\tau-1}(\phi+2\theta)K - \phi^\tau}{\kappa_y - 1}, \quad \tau = 1, 2, ... \tag{37}$$

The formula reduces to that of the standard GARCH(1,1) model in the limit as $\nu \to \infty$.

5 Example

The first-order Beta-t-GARCH and Beta-t-EGARCH models were fitted to 6235 de-meaned daily returns for Dow-Jones and FTSE from 3/1/84 to 27/11/07. Table 1 gives the ML estimates of the parameters, together with estimates for IGARCHt. Stationary models were also estimated but the results are not reported here as the parameter estimates were close to the IGARCH boundary and the conditional variances are almost the same[6]. Our models were coded in the Ox language of Doornik (2007), with optimization carried out by feasible sequential quadratic programming or simulated annealing as described in Goffe *et al* (1994).

The estimated volatilites for the Beta-t-EGARCH and Beta-t-GARCH are very similar. The only marked differences between their conditional standard deviations (SDs) and those obtained from the GARCHt model are immediately after extreme values. Figure 2 shows the estimated SDs for Dow-Jones returns around the great crash of October 1987 - observation 987 on the graph. (The value is 22.5 but the y axis has been truncated). As might be expected, the GARCHt filter reacts more strongly to the crash and takes some time to return to a stable level. Figure 3 shows the standardized residuals, $e_t = y_t/\widetilde{\sigma}_{t|t-1}, t = 1, ...T$. The GARCHt residuals appear abnormally small for about 20 observations after 995.

	FTSE			DOW-JONES		
Parameter	B-t-G	B-t-EG	IGARCHt	B-t-G	B-t-EG	IGARCHt
θ	0.077	0.084	0.063	0.059	0.067	0.043
ν	10.611	11.003	11.47	5.74	5.79	6.12
LogL	24256	24270	24268	24603	24612	24602

[6] Very similar estimates were obtained using the G@RCH Oxmetrics program of Laurent (2007).

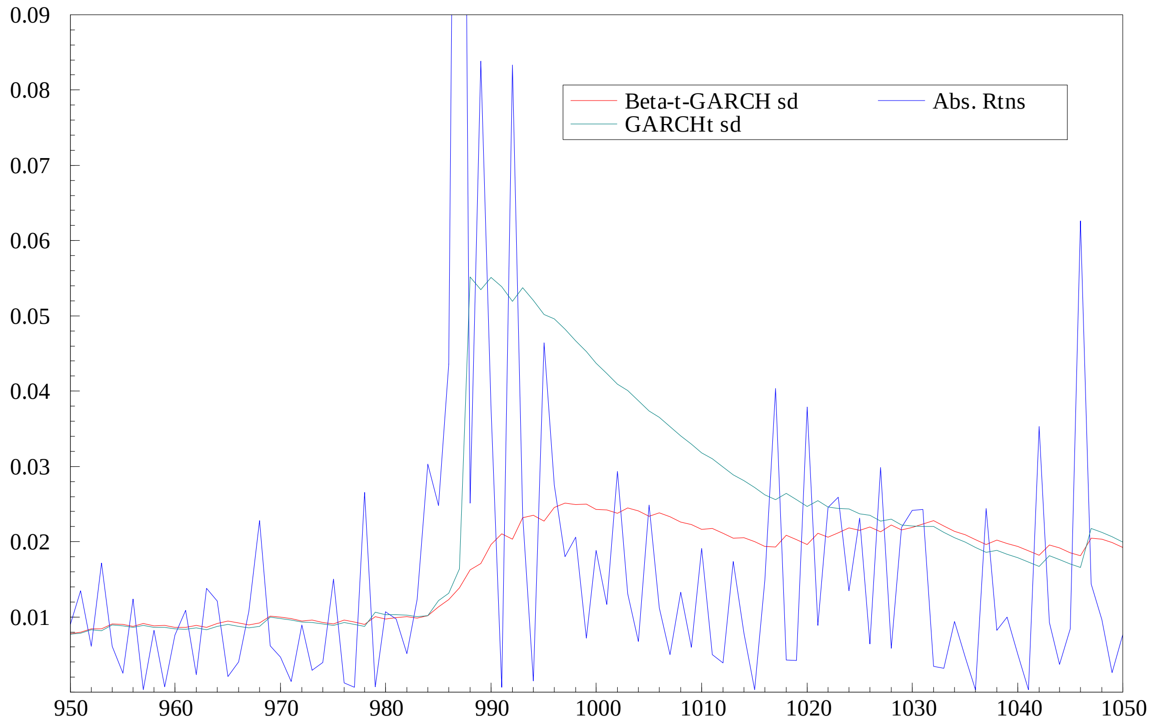

Figure 2: Dow-Jones absolute returns (minus mean) around the great crash of October 1987, together with estimated standard deviations for Beta-t-GARCH and GARCHt.

Table 1 Parameter estimated for Integrated Beta-t-GARCH and Beta-t-EGARCH models together with IGARCH

Figure 4 shows the Beta-t-EGARCH standard deviations for the FTSE around the October 1987 crash. As with the Dow-Jones there is little perceptible difference between these SDs and those from Beta-t-GARCH.

Finally, figure 5 shows the filtered estimates of the standard deviation from a stochastic volatility model estimated by QML as described in Harvey, Ruiz and Shephard (1994). As can be seen they respond very slowly to the increased volatility after the crash, only reaching a level similar to that shown in the figures of the previous section (for all models) approximately 20 observations after the crash. At this point the SV estimates are rising while

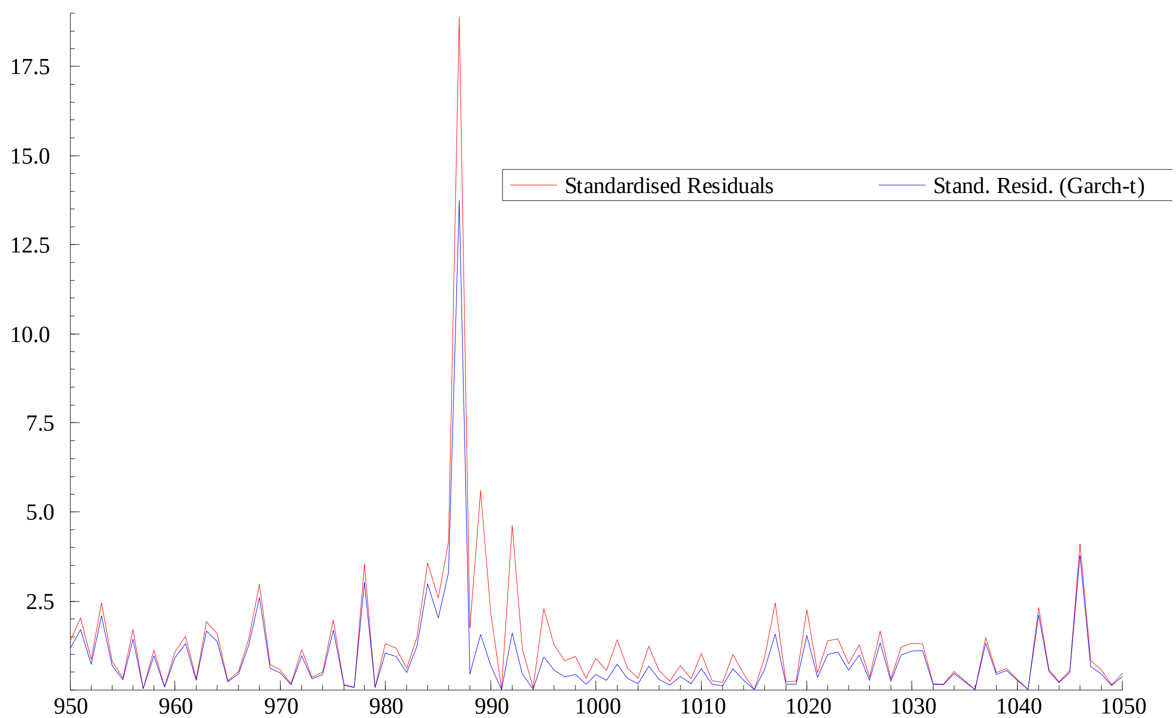

Figure 3: Dow-Jones - residuals from Beta-t-GARCH and GARCHt models around the great crash of October 1987

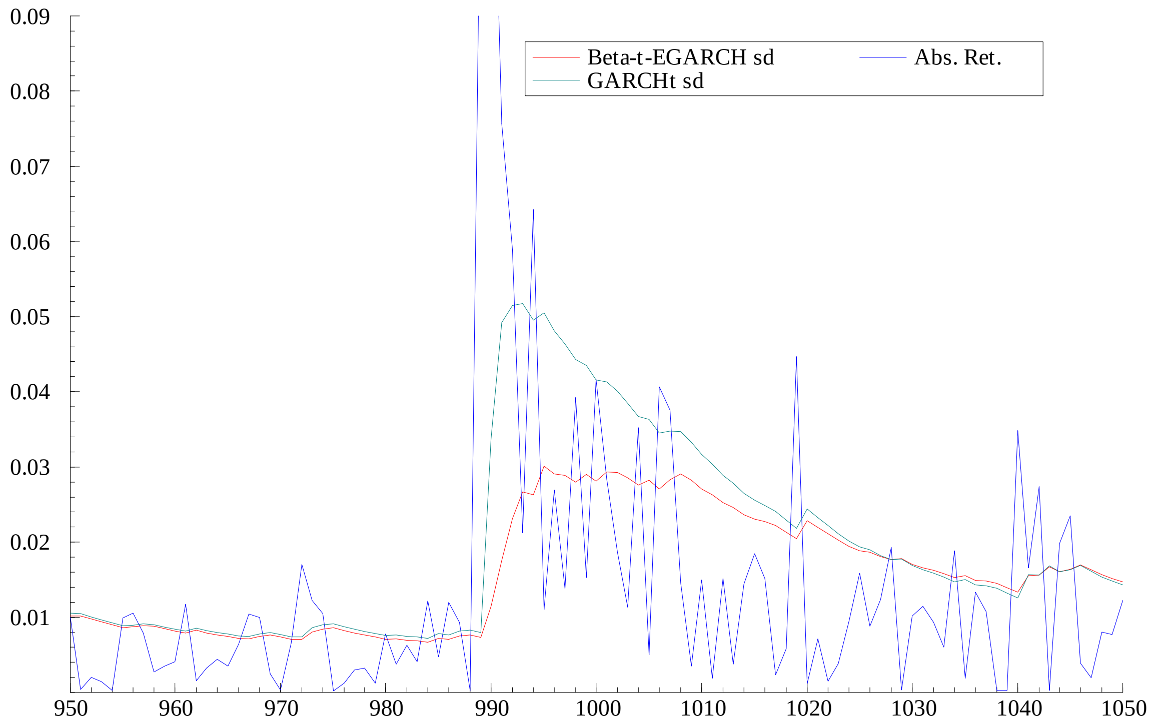

Figure 4: FTSE absolute returns (minus mean) around the great crash of October 1987, together with estimated standard deviations for Beta-t-EGARCH and GARCHt.

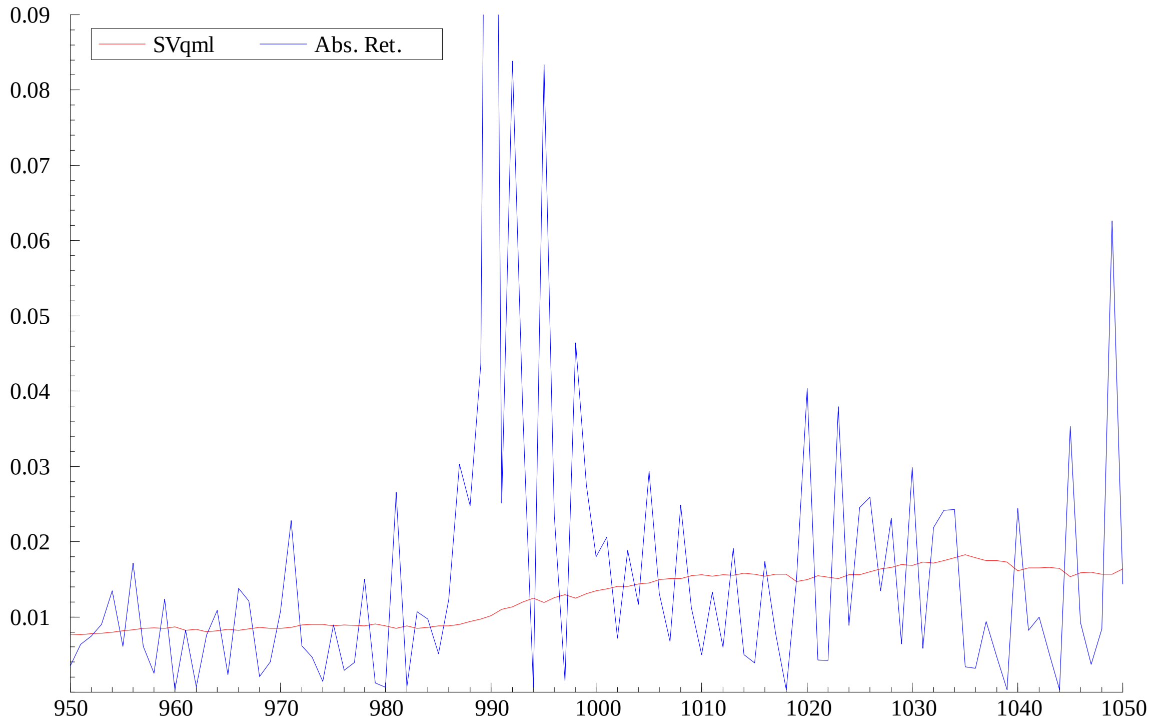

Figure 5: Filtered estimated for Dow-Jones from a stochastic volatility model estimated by QML

those from the Beta-t-GARCH model are falling. The filtered estimates from a simulation based method of estimation may be different; see the discussions of Markov chain Monte Carlo and importance sampling in Andersen *et al.*(2006) and Durbin and Koopman (2001). The relationship between SV and GARCH filters is worth exploring but will not be pursued here.

6 Leverage effects

The standard way of incorporating leverage effects in GARCH models is by including the variable $I(y_t < 0)u_t$, where $I(y_t < 0)$ is the indicator taking the value one for $y_t < 0$ and zero otherwise. This additional variable is an

MD and putting it in the first-order exponential model, (22), gives

$$\widetilde{\lambda}_{t|t-1} = \delta + \phi\widetilde{\lambda}_{t-1|t-2} + \theta u_{t-1} + \theta^* I(y_{t-1} < 0)u_{t-1},$$

where θ^* will usually be less than or equal to zero. In the EGARCH equation, (23), $g(z_t) = [1 + (\theta^*/\theta)I(y_t < 0)]u_t$. The existence of moments of the observations is not affected.

The Beta-t-GARCH can be similarly handled. In the first order case

$$\begin{aligned} c(z_t) &= \beta + \alpha(\nu+1)b_t + \alpha^* I(y_t < 0)(\nu+1)b_t \\ &= (\nu+1)b_t\{\alpha + \alpha^* I(y_t < 0)\} \end{aligned}$$

The condition for the existence of the second moment, assuming $\nu > 2$, is now $\alpha + \beta + \alpha^*/2 < 1$, while the fourth moment exists if

$$\beta^2 + 2\alpha\beta + \frac{3(\alpha^2 + \alpha^{*2}/4)(\nu+1)}{\nu+3} < 1, \quad \nu > 4.$$

7 Asymmetric *t*

The t_ν-distribution with a location (median) of zero and scale of $\exp(\lambda/2)$ has *pdf* as in Laurent (2007, p31) or Bauwens and Laurent (2005). For known degrees of freedom and skewness, ξ, the kernel of the log-likelihood function at time t is

$$L = -\frac{1}{2}\lambda_t - \frac{(\nu+1)}{2}\ln\left(1 + \frac{s(y_t+m)^2}{\nu e^{\lambda_t}\xi^{2I_t}}\right)$$

where

$$I_t = \begin{cases} 1 \; if & z_t \geq -\frac{m}{s} \\ -1 \; if & z_t < -\frac{m}{s} \end{cases},$$

ξ is the asymmetry parameter, and m and s are given by the following functions of ν and ξ:

$$m = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(\xi - \frac{1}{\xi}\right), \quad and \quad s = \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2}.$$

Differentiating *with respect to* λ_t gives the score as

$$\frac{\partial L}{\partial \lambda_t} = \frac{1}{2}\left[\frac{s\left(\nu+1\right)\left(y_t+m\right)^2}{\nu e^{\lambda_t}\zeta^{2I_t}}\left(1+\frac{s(y_t+m)^2}{\nu e^{\lambda_i}\zeta^{2I_t}}\right)^{-1}-1\right],$$

suggesting that u_t be re-defined as

$$u_t = \frac{s\left(\nu+1\right)\left(y_t+m\right)^2}{\nu e^{\lambda_t}\zeta^{2I_t}+s(y_t+m)^2}-1, \quad -1 \leq u_t \leq \nu, \tag{38}$$

in the Beta-t-EGARCH and Beta-t-GARCH models. ML estimation is carried out with respect to the asymmetry parameter, ξ, as well as ν and the parameter(s) governing the dynamics.

8 Two component models

Engle and Lee (1999) proposed a GARCH model in which the variance is broken into a long-run and a short-run component. The main role of the short-run component is to pick up the temporary increase in variance after a large shock. Another feature of the model is that it can approximate long memory behaviour; see Andersen et al (2006, p 806-7).

Here we adapt the model so that the long-run component is of the Beta-t-(E)GARCH form, while the short-run component responds in the same way as in the conventional GARCH model. The attraction of this formulation is that the long-run component is not sensitive to extreme observations. As a result it may be easier to separate the two components.

8.1 Model formulation

The model in section 4 may be extended to include two factors as follows. In (17),

$$\sigma^2_{t|t-1} = s_{t|t-1} + q_{t|t-1}, \quad t = 1, ..., T, \tag{39}$$

with, if we restrict attention to first-order effects,

$$\begin{aligned} s_{t|t-1} &= (\alpha+\beta)s_{t-1|t-2} + \alpha(y^2_{t-1} - \sigma^2_{t-1|t-2}) \\ &= \beta s_{t-1|t-2} + \alpha(y^2_{t-1} - q_{t-1|t-2}) \end{aligned} \tag{40}$$

and

$$q_{t|t-1} = \delta + \phi q_{t-1|t-2} + \theta\sigma^2_{t-1|t-2}u_{t-1},$$

where u_{t-1} is defined as in (4), that is with respect to the total variance.

The exponential form has $\lambda_{t|t-1}$ in (16) defined by

$$\lambda_{t|t-1} = \lambda^S_{t|t-1} + \lambda^L_{t|t-1}, \quad t = 1, ..., T, \tag{41}$$

where

$$\lambda^S_{t|t-1} = (\alpha + \beta)\lambda^S_{t-1|t-2} + \alpha y^2_{t-1} / \exp(\lambda_{t-1|t-2})$$

and

$$\lambda^L_{t|t-1} = \lambda + \phi\lambda^L_{t-1|t-2} + \theta u_{t-1},$$

with u_{t-1} as in (18). The variances now combine multiplicatively.

The models can be further modified to include leverage effects. In the exponential model, $u^*_{t-1} = I(y_{t-1} < 0)(u_{t-1} + 1)$ can be included in the equation for $\lambda^L_{t|t-1}$, while $I(y_{t-1} < 0)y^2_{t-1}/\exp(\lambda_{t-1|t-2})$ is added to the equation for $\lambda^S_{t|t-1}$. As a variation we might follow Nelson (1991) and set

$$\begin{aligned}\lambda^S_{t|t-1} &= (\alpha + \beta)\lambda^S_{t-1|t-2} + \alpha\{|y_{t-1}| \exp(-\lambda_{t-1|t-2}/2) - E|\varepsilon_{t-1}|)\} \\ &\quad + \alpha^* y_{t-1} \exp(-\lambda_{t-1|t-2}/2);\end{aligned}$$

compare (24). Note that $E|\varepsilon_{t-1}| = (\nu/\pi)^{1/2}\Gamma(\nu/2 - 1/2)/\Gamma(\nu/2)$ for a t_ν.

Similarly in Beta-t-GARCH, $I(y_{t-1} < 0).(y^2_{t-1} - \sigma^2_{t-1|t-2})$ or $I(y_{t-1} < 0).(y^2_{t-1} - q_{t-1|t-2})$ can be added to (40); the second suggestion accords with the specification[7] in Engle and Lee (1999).

8.2 Application

Table 2 shows the results from fitting component models, with and without leverage effects, to Dow-Jones and FTSE returns. As in section 5, we set $\phi = 1$, so the long-run component is non-stationary; the same restriction is made in table 4 of Engle and Lee (1999, p 487).

[7] Actually Engle and Lee (1999, p486) have $I(y_{t-1} < 0)y^2_{t-1} - 0.5q_{t-1|t-2})$, replacing $I(y_{t-1} < 0)$ by its expectation.

	FTSE		DOW-JONES			
Param.	B-t-G	B-t-G + lev	B-t-G	B-t-G + lev	B-t-EG	B-t-EG+lev
α	0.059	0.021	0.033	-0.005	.013	-0.021
β	0.894	0.858	0.918	0.876	.451	0.855
ν	11.14	11.36	5.69	5.90	5.80	5.68
α^*	–	0.060	–	0.054	-	.042
θ	0.030	0.033	0.025	0.029	.058	0.029
$LogL$	24299	24309	24635	24647	24618	24648

Table 2 Two component models with and without leverage in the short-run component

Engle and Lee (1999, p 487) found clear evidence of the leverage effect in the transitory component in their analysis of the S&P 500 index but not in the long-run component. In fact the transitory component is completely dominated by negative shocks. Our findings are similar with a LR test being statistically significant for the leverage variable in the short-run component for both series. The effect is to make the estimate of β smaller so there is a more rapid decline after an extreme observation.

Figures 6 and 7 show the standard deviations of the total and long-run components, $\sigma_{t|t-1}$, and $q_{t|t-1}$, in the Beta-t-GARCH model, without leverage and with leverage. The GARCHt SD shown on the earlier figures is also included. The long-run component is not pulled up as much by the great crash observation of October 19th 1987 as was the single component produced by the model fitted in section 5. Furthermore it dies down more quickly. It can be seen that the effect of including the leverage term is that the total SD is higher immediately after the great crash but thereafter it falls away more rapidly.

Figures 8 and 9 show the conditional scale of the total and long-run components, $\lambda_{t|t-1}$, and $\lambda^L_{t|t-1}$, in the Beta-t-EGARCH model, without leverage and with leverage. The contrast with figures 6 and 7 is interesting. The immediate response of the short-term component in the exponential model is stronger, but it dissappears more rapidly. The response is greater in the leverage model and the long-run component is hardly affected.

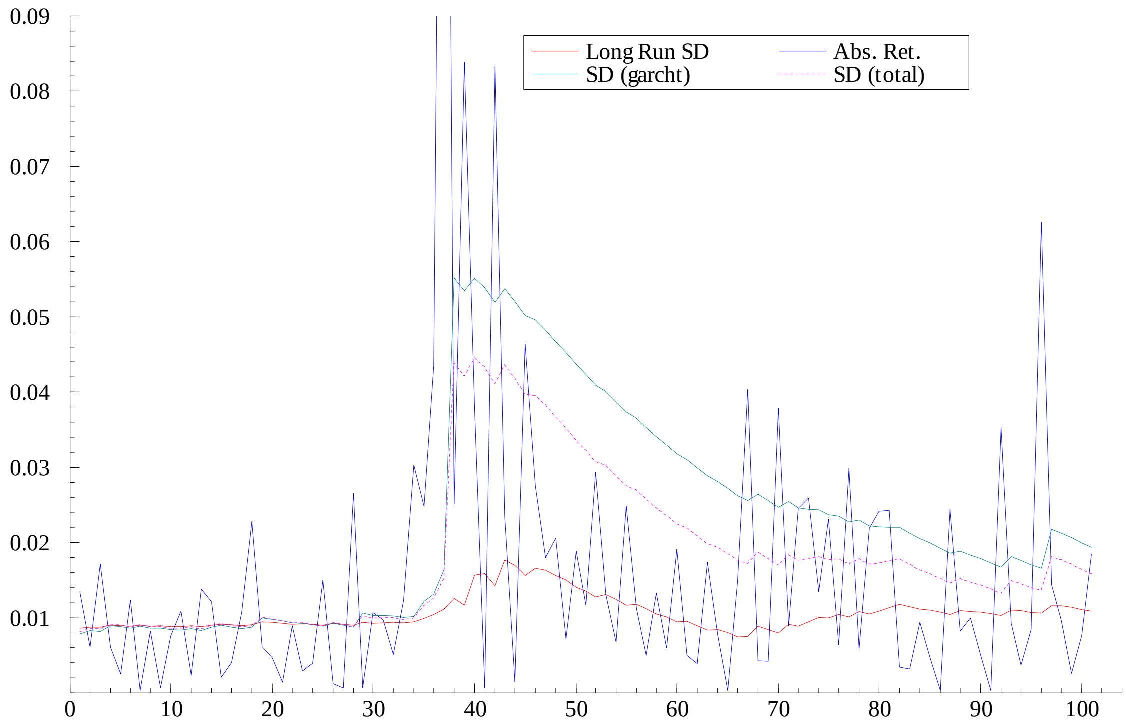

Figure 6: SDs of long-run component and total Sd for Beta-t-GARCH model fitted toDow-Jones returns.

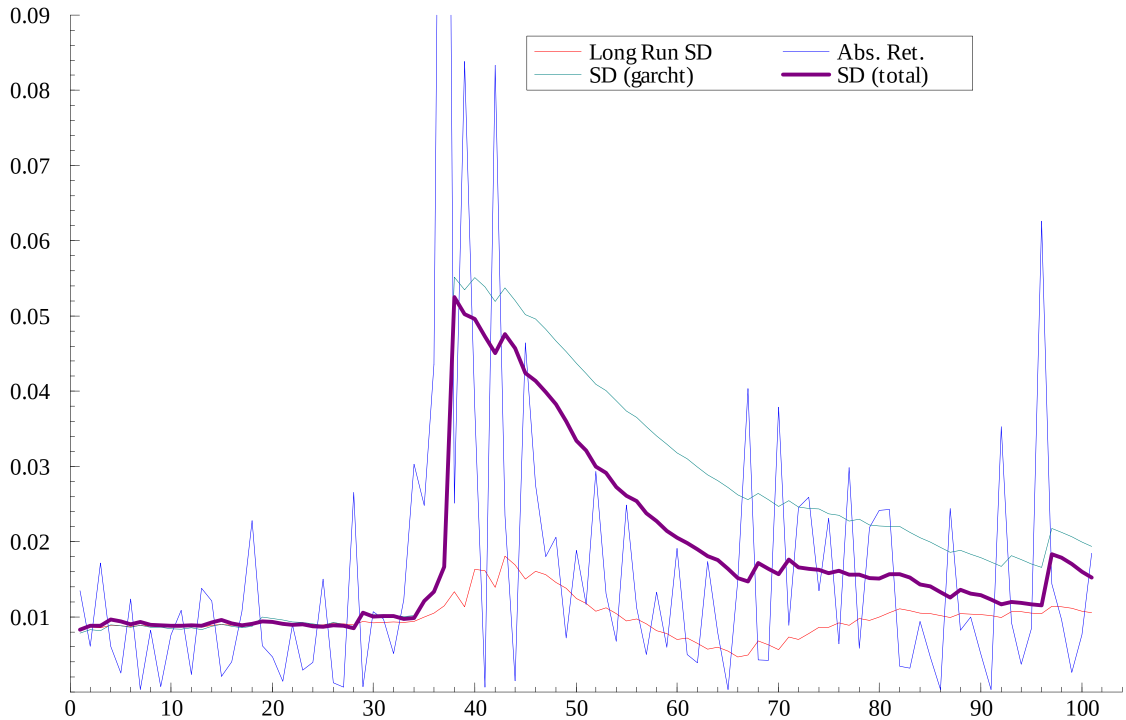

Figure 7: SDs of long-run component and total SD for Beta-t-GARCH with leverage fitted to Dow-Jones returns.

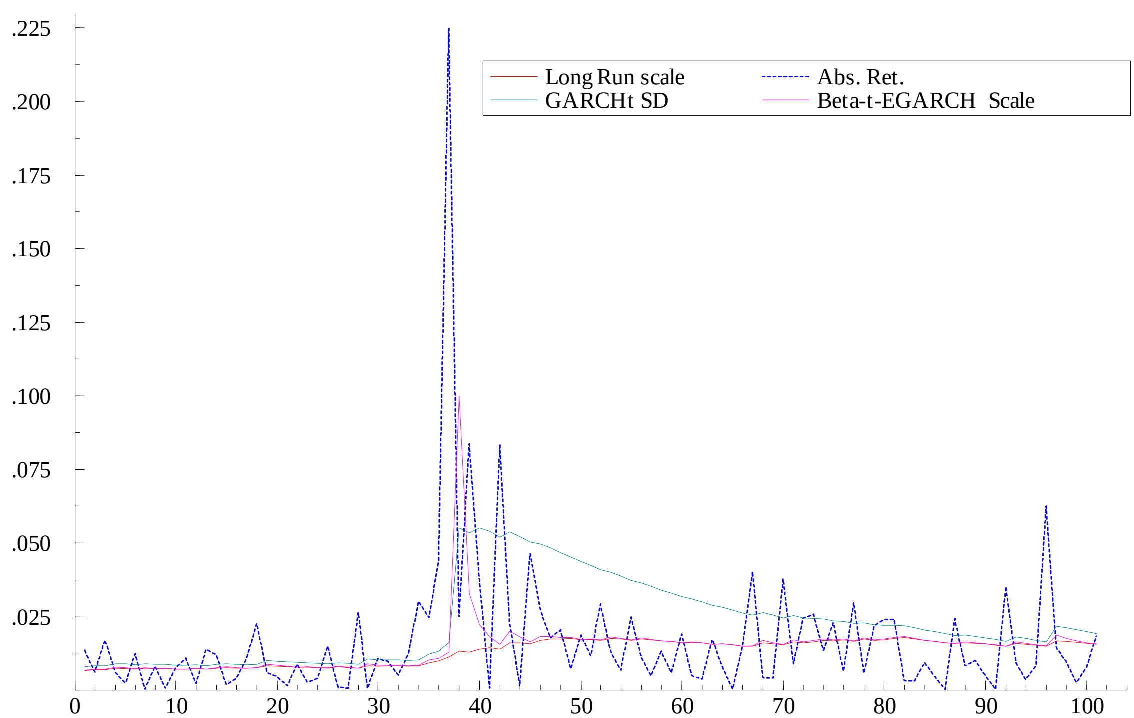

Figure 8: Total scale for Beta-t-EGARCH model fitted to Dow-Jones returns.

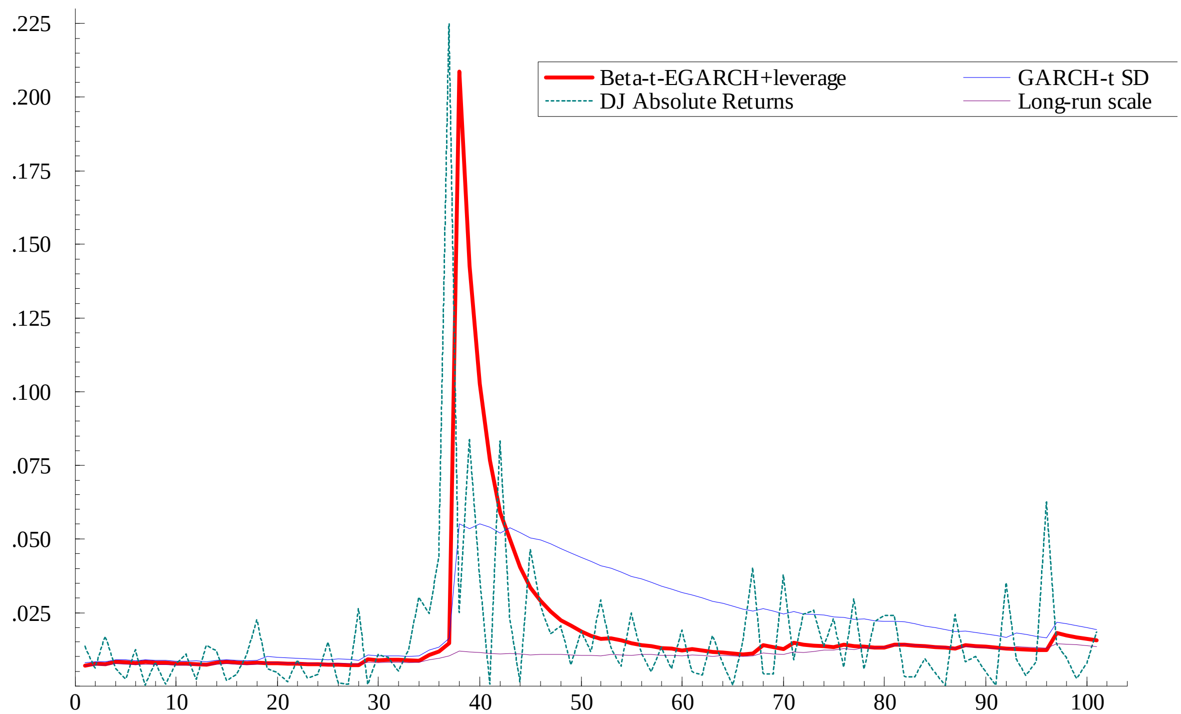

Figure 9: Total scale for Beta-t-EGARCH model with leverage fitted to Dow-Jones returns.

9 Level and scale

When observations are from a t_ν-distribution, as in (9), the sample mean may be a very inefficient[8] estimator of μ for small ν. Similarly the EWMA may be very inefficient when the noise in (5) has a t_ν-distribution[9]. The suggestion in the introduction to section 2 was to set up a local likelihood. For known degrees of freedom and scale, the kernel of the local log-likelihood function at time t is

$$L = -\frac{(\nu+1)}{2}\sum_{i=1}^{t} w_{i,t} \ln\left(1+\frac{(y_i-\mu_t)^2}{\nu e^{\lambda}}\right)$$

where, as before, $w_{i,t}$ denotes a set of weights. Differentiating with respect to μ_t gives the scoring algorithm as

$$\widetilde{\mu}_{t,j} = \widetilde{\mu}_{t,j-1} + \frac{(\nu+3)}{\nu\sum_i w_{i,t}}\sum w_{i,t}\left(1+\frac{(y_i-\mu_t)^2}{\nu e^{\lambda}}\right)^{-1}(y_i-\widetilde{\mu}_{t,j-1}) \qquad (42)$$

and arguing as in sub-section 2.2 suggests the filter

$$\mu_{t+1|t} = \mu_{t|t-1} + \alpha v_t, \qquad t = 1, ..., T, \qquad (43)$$

where α is a parameter and

$$v_t = \left(1+\frac{(y_t-\widetilde{\mu}_{t|t-1})^2}{\nu e^{\lambda}}\right)^{-1}(y_t-\widetilde{\mu}_{t|t-1}) \qquad (44)$$

The model is now defined as one in which $y_t \mid Y_{t-1}$ has a t_ν distribution with mean $\widetilde{\mu}_{t|t-1}$. More generally we might consider an ARMA-type structure for the conditional mean, that is

$$\mu_{t|t-1} = \delta + \phi_1\mu_{t-1|t-2} + ... + \phi_p\mu_{t-p|t-p-1} + \theta_1 v_{t-1} + ... + \theta_q v_{t-q}$$

Another possibility is to take the innovations form of a Kalman filter for an unobserved components model that defines the predicted level, replace

[8]The efficiency is $(\nu+3)(\nu-2)/\{\nu(\nu+1)\}$.

[9]But note that if both the level and irregular noise follow t-distributions constructed as normal variates divided by the square roots of the same chi-square variates then the KF is optimal (though the MSEs of the state will be incorrect).

the innovation, $y_t - \widetilde{\mu}_{t|t-1}$, by v_t and treat the Kalman gain as a vector of parameters to be estimated.

A model for the level may be combined with one for the scale, and the diagonality of the information matrix in the constant parameter case suggests that the recursions can be formulated separately. However, they will run in parallel, so λ in (44) becomes $\lambda_{t|t-1}$ while u_t in (21) is generalized to

$$u_t = \frac{(\nu+1)(y_t - \mu_{t|t-1})^2}{\nu \exp(\lambda_{t|t-1}) + (y_t - \mu_{t|t-1})^2} - 1, \qquad t = 1, ..., T.$$

The parameters are estimated by assuming that $y_t \mid Y_{t-1}$ has a t_ν distribution with mean $\widetilde{\mu}_{t|t-1}$ and scale $\exp(\lambda_{t|t-1}/2)$.

As an example, consider the seasonally adjusted rate of inflation[10] in the US. This variable is often taken to follow a random walk plus noise. Fitting a Gaussian model with the STAMP8 package of Koopman et al (2007), gives an estimate of the parameter corresponding[11] to α of 0.579. Figure 10 shows the filtered level and irregular. The level is clearly sensitive to extreme values and the irregular displays heteroskedasticity. The ACF of the absolute values of the residuals (innovations) also provides strong evidence of serial correlation in variance. Letting the variance of the innovations evolve as a stationary Beta-t-GARCH process gives the following ML estimates: $\widehat{\alpha} = .758, \widehat{\nu} = 4.81$, $\widetilde{\phi} = 0.311$ and $\widehat{\theta} = .187$. Figure 11 shows the filtered level from the Beta-t-GARCH model. The estimates respond less to extreme values than those from the Gaussian model.

In the context of estimating the volatility of stock returns, the main concern is to estimate a constant level. Rather than simply using the mean, the ML estimator of μ can be computed from the standardized returns (which depend on an estimator of μ). Alternatively the estimator can be computed recursively by modifying (43) to

$$\mu_{t+1|t} = \frac{t-1}{t}\mu_{t|t-1} + \frac{\nu+3}{\nu t}v_t, \qquad t = 1, ..., T,$$

with $\mu_{1|0} = 0$. As $\nu \to \infty$, $\mu_{t+1|t} \to \overline{y}$; see Harvey (1989, p. 108).

[10] The (annualized) rate of inflation is measured as the first differences of the quarterly CPI multiplied by four. We have data from 1947(1) to 2007(2), obtained from the U.S. Bureau of Labor Statistics (website: www.bls.gov).

[11] It would be α if ν were infinity.

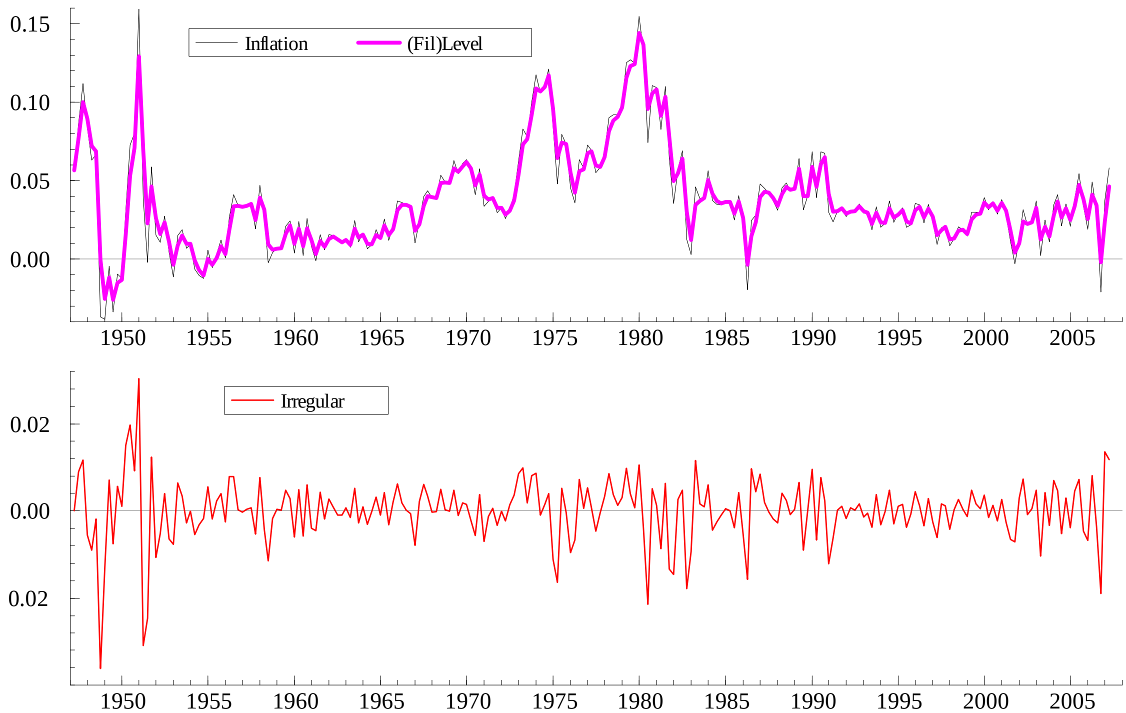

Figure 10: Filtered estimates of level and irregular from a Gaussian random walk plus noise fitted to US inflation.

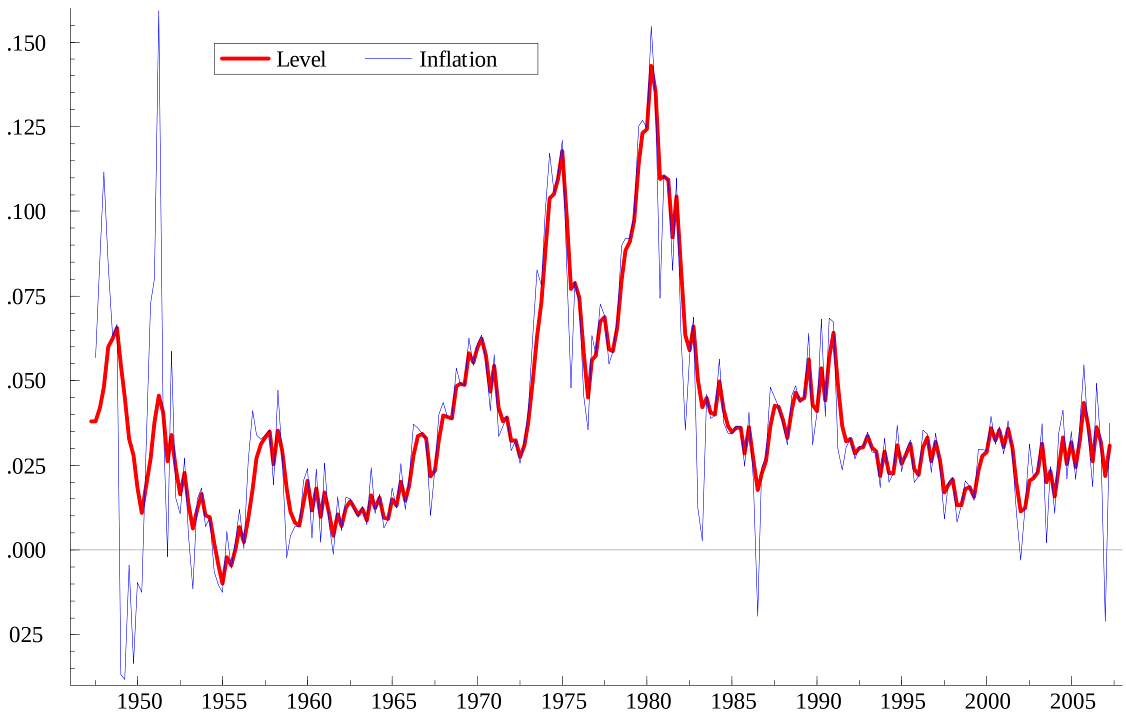

Figure 11: Estimated level from Beta-t-GARCH model

In the ARCH in mean model, ARCH-M, a risk premium term that depends on volatility is added to the expected value of returns. Thus (17) becomes

$$y_t = \mu + \delta\sigma_{t|t-1} + \sigma_{t|t-1}z_t, \quad t = 1, ..., T$$

where δ is a parameter. Adapting the procedure above to deal with this ARCHt-M model is straightforward.

10 Conclusions

We began by questioning whether specifying the conditional variance in a GARCHt model as a linear combination of past squared observations is appropriate. An approach based on local likelihood suggested the use of the score of the t-distribution as an alternative to squared observations. The score transformation can also be used to formulate an equation for the logarithm of the conditional variance, in which case no restrictions are needed to ensure that the conditional variance remains positive. Such a model is a special case of the most general formulation of an EGARCH model. Since the score variables have a beta distribution, we call the model Beta-t-EGARCH. While t-distributed (standardized) variables, with finite degrees of freedom, fail to give moments for the observations when they enter the EGARCH model in the form analysed by Nelson (1991), the transformation to beta variables means that all moments of the conditional variance exist when the equation defining the logarithm of the conditional variance is stationary. Furthermore, it is possible to obtain analytic expressions for the autocorrelations of powers of absolute values of the observations. Some calculations indicate that the autocorrelations tend to be smaller for heavy-tailed distributions, but the reduction is less marked for absolute values than it is for squares. Of course, volatility can be nonstationary, or very close to nonstationarity, but an attraction of the EGARCH model is that, when the equation for the logarithm of the conditional variance has the form of a random walk, it does not lead to the variance collapsing to zero almost surely, as is the case with IGARCH.

When the score variables are used in an equation for the conditional variance, rather than its logarithm, the model has many of the theoretical features of GARCH. However, the main attraction of the Beta-t-GARCH model, which it shares with Beta-t-EGARCH, is that it is resistant to extreme observations. This point is illustrated by fitting the models to daily Dow-Jones and FTSE returns and comparing with GARCH for the period around the great crash of October 1987. When there are no extreme observations the models give similar results.

Beta-t-EGARCH and Beta-t-GARCH may both be modified to include leverage effects. They may also be extended to include long-run and short-run components. Only the long-run component is driven by a beta variable and this makes it easier to separate from the short-run component. Fitting two component models to the Dow-Jones and FTSE returns shows the short-run component to be subject to strong leverage effects. There are marked differences between the Beta-t-EGARCH and Beta-t-GARCH models with the short-run component in the exponential model displaying a bigger response to the great crash but one that dies away faster.

Finally, we show that when the score from the generalized error distribution is used to determine a transformation for the observations in the conditional variance equation, the result is a member of the power ARCH family. The shape of the generalized error distribution may make it less appealing for modeling returns than the t-distribution. For example the heaviest tails are obtained with the double exponential distribution.

APPENDIX

A Autocorrelation function for the Gaussian exponential model

For a normal distribution, $u_t = \varepsilon_t^2 - 1$. Thus terms of the form $E_{t-j-1}\left(e^{\psi_j(\nu+1)b_{t-j}c/2}\right)$ are replaced by

$$E_{t-j-1}\left(e^{\psi_j \varepsilon_{t-j}^2 c/2}\right) = (1 - c\psi_j)^{-1/2}, \qquad c\psi_j < 1, \quad j = 0, 1, 2, ...$$

To evaluate the ACF of $|y_t|^c$ we need to find $E_{t-\tau-1}\left(|\varepsilon_{t-\tau}|^c \, e^{\psi_\tau \varepsilon_{t-\tau}^2 c/2}\right), \tau =$

$1, 2, \ldots$ Since $\varepsilon_t^2 = x$ has a chi-square distribution we have, for $c\psi < 1$,

$$\begin{aligned} E\left(|x|^{c/2}\, e^{\psi xc/2}\right) &= \frac{1}{\Gamma(1/2)}\int x^{c/2} e^{\psi xc/2} (x/2)^{-1/2} e^{-x/2} dx \\ &= \frac{\Gamma((c+1)/2)}{\Gamma(1/2)} . \frac{2^{1/2} 2^{(c+1)/2-1}}{\Gamma((c+1)/2)} \int (x/2)^{(c+1)/2-1} e^{\psi xc/2} e^{-x/2} dx \\ &= 2^{c/2}\Gamma((c+1)/2)\pi^{-1/2}(1-\psi c)^{-(c+1)/2} \\ &= E\left(|\varepsilon_t|^c\right)(1-\psi c)^{-(c+1)/2}. \end{aligned} \tag{45}$$

Acknowledgement *We are grateful to Neil Shephard for helpful comments, but any errors are our responsibility. An earlier version was presented at the 6th Oxmetrics User conference in London and we are grateful to participants for helpful comments. We acknowledge financial support from the ESRC under the grant Time-Varying Quantiles, RES-062-23-0129.*

REFERENCES

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Volatility and correlation forecasting. In: Elliot, G., Granger, C., Timmerman, A. (Eds.), Handbook of Economic Forecasting. 777-878. Amsterdam: North Holland

Bauwens, L., Laurent, S., 2005. A new class of multivariate skew densities, with application to GARCH models. Journal of Business and Economic Statistics 23, 346-354.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327.

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for security prices and rates of return data. Review of Economics and Statistics 59, 542-547.

Bollerslev, T., Mikkelsen, H.O., 1996. Modeling and pricing long memory in stock market volatility. Journal of Econometrics 73, 151–184.

Chib, S., Kim, S., Shephard, N.S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361-393.

Ding, Z., Engle, R.F., Granger, C.W.J., 1993. A long memory model of stock market returns and a new model. Journal of Empirical Finance 1, 107-131.

Doornik, J. A., 2007. Ox: An Object-Oriented Matrix Language. 5th edition. Timberlake Consultants Press: London.

Doornik, J.A., Harvey, A.C., Koopman, S.J., Shephard, N., 2007. STAMP 8.0 structural time series analysis modeller and predictor. Timberlake Consultants Ltd., London.

Durbin, J., Koopman, S.J., 2001. Time series analysis by state space methods. Oxford University Press, Oxford.

Engle, R.F., Lee, G.G.J., 1999. A long-run and short-run component model of stock return volatility. In: Engle, R.F., White, H. (Eds.), Cointegration, causality, and forecasting: a festschrift in honour of Clive W.J. Granger. Oxford University Press, Oxford.

Fan, J., Gijbels, I., 1996. Local polynominal modelling and its applications. Chapman and Hall, London.

Goffe, W.L., Ferrier, G.D., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60, 65-100.

Harvey, A.C., 1989. Forecasting, structural time series models and the Kalman Filter. Cambridge University Press, Cambridge.

Harvey A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies 61, 247-264.

Harvey, A. C., Streibel, M., 1998. Testing for a slowly changing level with special reference to stochastic volatility. Journal of Econometrics 87, 167-189.

He, C., Terasvirta, T., 1999. Properties of moments of a family of GARCH processes. Journal of Econometrics 92, 173-192.

King, M.L., Zhang, X., 2008. Box-Cox stochastic volatility models with heavy-tails and correlated errors. Journal of Empirical Finance 15, 549–566.

Koopman, S-J., Harvey A.C., Doornik J.A., Shephard, N., 2007. STAMP 8.0 Structural Time Series Analysis Modeller and Predictor. Timberlake Consultants Ltd., London.

Laurent, S., 2007. GARCH5. Timberlake Consultants Ltd., London.

Ling, S., McAleer, M., 2002. Stationarity and the existence of moments of a family of GARCH processes. Journal of Econometrics 106, 109-117.

Muler, N., Yohai, V.J., 2008. Robust estimates for GARCH models. Journal of Statistical Planning and Inference 138, 2918-2940.

Nelson, D.B., 1991. Conditional heteroscedasticity in asset returns: a new approach. Econometrica 59, 347-370.

Shinichi, S., White, H., 1998. High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility. Econometrica 66, 529-567.

Stuart, A., Ord, J.K., 1987. Kendall's advanced theory of statistics, originally by Sir Maurice Kendall. Fifth Edition of Volume 1, Distribution theory. Charles Griffin & Company Limited, London.

Taylor, S. J, 1986. Modelling financial time series. Wiley, Chichester.

Taylor, S. J, 2005. Asset price dynamics, volatility, and prediction. Princeton University Press, Princeton.

Taylor, J., Verbyla, A., 2004. Joint modelling of location and scale parameters of the t distribution. Statistical Modelling 4, 91-112.