Two EGARCH models and one fat tail

Michele Caivano and Andrew Harvey

July 2013

CWPE 1326

Two EGARCH models and one fat tail

Michele Caivano and Andrew Harvey*
Bank of Italy and Faculty of Economics, Cambridge University

July 23, 2013

Abstract

We compare two EGARCH models which belong to a new class of models in which the dynamics are driven by the score of the conditional distribution of the observations. Models of this kind are called dynamic conditional score (DCS) models and their form facilitates the development of a comprehensive and relatively straightforward theory for the asymptotic distribution of the maximum likelihood estimator. The EGB2 distribution is light-tailed, but with higher kurtosis than the normal. Hence it is complementary to the fat-tailed t. The EGB2-EGARCH model gives a good fit to many exchange rate return series, prompting an investigation into the misleading conclusions liable to be drawn from tail index estimates.

KEYWORDS: Exchange rates; heavy tails; Hill's estimator, score; robustness; Student's t; tail index

JEL classification; C22, G17

*michele.caivano@bancaditalia.it; ach34@cam.ac.uk

1 Introduction

The exponential generalized autoregressive heteroskedasticity (EGARCH) class of models was introduced by Nelson (1991) as a means of modeling changing volatility. By letting the dynamic equation for the logarithm of variance be driven by the score of the conditional distribution, many of the theoretical problems inherent in EGARCH models are resolved; see Harvey

(2013, ch 4). Furthermore there is already a body of evidence showing that these dynamic conditional score (DCS) EGARCH models perform better on real data than do standard GARCH formulations; see, for example, Creal et al (2011) and Harvey and Sucarrat (2012).

The DCS-EGARCH model based on a conditional Student-t distribution, called Beta-t-EGARCH, is resistant to observations that would be outliers if a Gaussian distribution were used. The reason for this robustness is that the score depends on a beta variable which, of course, is bounded; compare the discussion of robust GARCH models in Muler and Yohai (2008). The t-distribution has fat tails (for finite degrees of freedom) and this property is reflected in the shape of the score function. However, not all variables which are subject to changing volatility have fat tails and the question therefore arises as to what other distributions might be entertained and what is the behaviour of their score functions. One possibility is to assume a general error distribution (GED) leading to the Gamma-GED-EGARCH model. This model has a gamma distributed score and hence its properties may be obtained in much the same way as may the properties of Beta-t-EGARCH. However, although the GED family provides a compromise between the normal and t-distributions, the behaviour of its score function is not ideal from the point of view of robustness. We argue here that a better choice is a new model based on the family of exponential generalized beta distributions of the second kind (EGB2). The EGB2 distribution is light-tailed, but with higher kurtosis than the normal. It has featured in GARCH models before; see Wang et al (2001). But it has not been used in a DCS-EGARCH model and its score function has a form which makes it the ideal complement to Beta-t-EGARCH.

The first contribution of this paper is on what we shall refer to as the EGB2-EGARCH model. We derive its properties and contrast them with those of Beta-t-EGARCH and Gamma-GED-EGARCH. The asymptotic distribution of the maximum likelihood estimator of the dynamic parameters can be derived for EGB2-EGARCH just as it can for the other two models. An analytic expression for the asymptotic covariance matrix can be obtained and the conditions for the asymptotic theory to be valid are easily checked. The theory is much more straightforward than it is for the corresponding GARCH model.

The second aspect of the paper concerns tail indices in financial time series. Tail indices, which are a key feature of fat-tailed distributions, are often computed, and low values are cited as evidence of fat tails and the associated non-existence of higher moments[1]. However, although excess kurtosis is a well-established stylized fact for both unconditional and conditional distributions of financial returns, the issue of whether such series have fat tails is more problematic. While it is undeniable that low tail estimates are a feature of financial returns, we argue that this does not, in itself, provide conclusive evidence of fat-tailed distributions. A subsidiary theme concerns the use of tail index estimates as starting values for the shape parameters of EGB2 and t-distributions, and in fact this provides a convenient lead-in to the discussion.

The article is organized as follows. Section 2 discusses classifications of tail behaviour in distributions and makes an important connection with the conditional score function. Section 3 describes the Beta-t-EGARCH and Gamma-GED-EGARCH models, and the associated asymptotic theory for the maximum likelihood estimator. The DCS scale model with an EGB2 distribution is analysed in Section 4 and the asymptotic theory is shown to extend to this case. Fitting Beta-t-EGARCH and EGB2-EGARCH to various returns series in Section 5 indicates that in a significant number of cases the EGB2 model gives a better fit, indicating that the conditional distribution does not have heavy tails. Since a DCS-EGARCH model cannot induce fat-tails, there is a paradox to be resolved. Section 6 analyses the problem and offers an explanation. Section 7 concludes.

2 Tails and tail indices

The Gaussian distribution has kurtosis of three and a distribution is said to exhibit *excess kurtosis* if its kurtosis is greater than three. Although many researchers take excess kurtosis as defining heavy tails, it is not, in itself, an ideal measure, particularly for asymmetric distributions. Most classifications in the insurance and finance literature begin with the behaviour of the upper tail for a non-negative variable, or one that is only defined above a minimum value; see Asmussen (2003) or Embrechts, Kluppelberg and Mikosch (1997). The two which are relevant here are as follows.

[1] For example, Loretan and Phillips (1994) report (modified) Hill's estimates of between 3 and 4 for the unconditional distributions of many daily and monthly stock and exchange rate returns series. They conclude that fourth moments do not exist for such series.

A distribution is said to be *heavy-tailed* if

$$\lim_{y\to\infty} \exp(y/\alpha)\overline{F}(y) = \infty \qquad \text{for all } \beta > 0, \tag{1}$$

where $\overline{F}(y) = \Pr(Y > y) = 1 - F(y)$ is the survival function. When y has an exponential distribution, $\overline{F}(y) = \exp(-y/\alpha)$, so $\exp(y/\alpha)\overline{F}(y) = 1$ for all y. Thus the exponential distribution is not heavy-tailed.

A distribution is said to be *fat-tailed* if, for a fixed positive value of η,

$$\overline{F}(y) = cL(y)y^{-\eta}, \qquad \eta > 0, \tag{2}$$

where c is a non-negative constant and $L(y)$ is slowly varying, that is $\lim_{y\to\infty}(L(ky)/L(y)) = 1,\ k > 0$.

The parameter η is the (right) tail index. The implied PDF is a *power law* PDF

$$f(y) \sim cL(y)\eta y^{-\eta-1}, \qquad as \ \ y \to \infty, \quad \eta > 0, \tag{3}$$

where $\sim$ is defined such that $a(x) \sim b(x)$ as $x \to x_0$ if $\lim_{x\to x_0}(a/b) \to 1$. The $m-$th moment exists if $m < \eta$. The Pareto distribution is a simple case in which $\overline{F}(y) = y^{-\eta}$ for $y > 1$. If a distribution is fat-tailed then it must be heavy-tailed; see Embrechts, Kluppelberg and Mikosch (1997, p 41-2). On the other hand, not all heavy–tailed distributions are fat-tailed; the lognormal is an example.

The complement to the Pareto distribution is the power function distribution for which $\overline{F}(y) = y^{\overline{\eta}}$, $0 < y < 1$, $\overline{\eta} > 0$. More generally,

$$\overline{F}(y) = cL(y)y^{\overline{\eta}}, \qquad 0 < y < 1, \quad \overline{\eta} > 0,$$

where $\overline{\eta}$ is the left tail index. Hence $f(y) \sim cL(y)\overline{\eta}y^{\overline{\eta}-1}$ as $y \to 0$.

The above criteria are related to the behavior of the conditional score and whether or not it discounts large observations. This, in turn, connects to robustness, as shown in Caivano and Harvey (2013). More specifically, consider a power law PDF, (3), with y divided by a scale parameter, φ, so that $\overline{F}(y/\varphi) = cL(y/\varphi)(y/\varphi)^{-\eta}$ and $f(y) \sim cL(y)\varphi^{-1}\eta(y/\varphi)^{-\eta-1}$. Then

$$\partial \ln f/\partial\varphi \sim \eta/\varphi \quad as \quad y \to \infty \tag{4}$$

and so the score is bounded. With the exponential link function, $\varphi = \exp(\lambda)$,

$\partial \ln f / \partial \lambda \sim \eta$ as $y \to \infty$. Similarly as $y \to 0$, $\partial \ln f / \partial \lambda \sim \overline{\eta}$.

The logarithm of a variable with a fat-tailed distribution has exponential tails. Let x denote a variable with a fat-tailed distribution in which the scale is written as $\varphi = \exp(\mu)$ and let $y = \ln x$. Then for large y

$$f(y) \sim cL(e^y)\eta e^{-\eta(y-\mu)}, \qquad \eta > 0, \quad as \quad y \to \infty,$$

whereas as $y \to -\infty$, $f(y) \sim cL(e^y)\overline{\eta}e^{\overline{\eta}(y-\mu)}$, $\overline{\eta} > 0$. Thus y is not heavy-tailed but it may exhibit excess kurtosis. The score with respect to location, μ, is the same as the original score with respect to the logarithm of scale and so tends to η as $y \to \infty$. If a scale parameter is introduced, its score is bounded when divided by the variable.

The relevance of the above paragraph to this paper is that the (light-tailed) EGB2 variable is obtained by taking the logarithm of a fat-tailed GB2 variable.

3 DCS volatility models

A volatility model is typically of the form

$$y_t = \mu + \varphi_{t|t-1}\varepsilon_t, \qquad t = 1, ..., T, \tag{5}$$

where $\varphi_{t|t-1}$ is a time-varying scale and ε_t is a standardized IID random variable. The scale, $\varphi_{t|t-1}$, is proportional to $\sigma_{t|t-1}$, with the factor of proportionality depending on the shape parameter(s) of the distribution of ε_t. In a DCS model the dynamics are set up by letting the logarithm of a time-varying scale parameter be a linear function of the conditional score. In the case of first-order dynamics,

$$\lambda_{t+1|t} = \omega(1-\phi) + \phi\lambda_{t|t-1} + \kappa u_t, \tag{6}$$

where u_t is the score with respect to $\lambda_{t|t-1} = \ln \varphi_{t|t-1}$. Extensions to higher order models, components, seasonals and explanatory variables are discussed in Harvey (2013, ch 4).

The above model belongs to the EGARCH class introduced by Nelson (1991). The usual formulation has u_t replaced by $|\varepsilon_t|$. Moments of y_t exist for a GED distribution (with the normal being a special case), but Student's t is not viable because y_t has no moments for finite degrees of freedom. The dynamic scale model overcomes this difficulty because the score is a linear function of a variable with a beta distribution.

Like the GED, the EGB2 distribution offers a contimuum of distributions between the normal and Laplace. However, unlike the GED (with shape parameter greater than one), the score with respect to the scale parameter of the EGB2 is bounded when divided by the variable[2]. The associated dynamic scale model is described in Section 4. (The dynamic EGB2 location model is discussed in Caivano and Harvey, 2013).

The GARCH-t model is widely used in empirical finance. The GARCH-EGB2 has been studied by Wang et al (2001) but is far less common. In both models, $y_t = \mu + \sigma_{t|t-1}\epsilon_t$ and the variance is driven by squared observations, that is

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

or, in notation similar to that in (6),

$$\sigma^2_{t+1|t} = \delta + \phi\sigma^2_{t|t-1} + \kappa\sigma^2_{t|t-1}\epsilon_t^2,$$

where $\delta = \omega_\sigma(1-\phi)$, where $\phi = \alpha + \beta$ and $\kappa = \alpha$.

For the DCS-EGARCH models with t and GED conditional distributions, all moments of the score exist and the existence of moments of y_t is not affected by the dynamics. The same is true of the EGB2. On the other hand, the existence of moments for GARCH models is affected by the volatility; see, for example, Mikosch and Starica (2000).

3.1 ML estimation

The ML estimates of the parameters, $\boldsymbol{\psi} = (\kappa,\ \phi,\ \omega)'$, in a DCS model can be obtained by maximizing the log-likelihood function with respect to the unknown parameters. The asymptotic distribution of the ML estimator in

[2]This property features in the robustness literature; see Maronna *et al* (2006, p 34-8).

the first-order case is derived in Harvey (2013). Define

$$
\begin{aligned}
a &= \phi + \kappa E\left(\frac{\partial u_t}{\partial \mu}\right) \\
b &= \phi^2 + 2\phi\kappa E\left(\frac{\partial u_t}{\partial \mu}\right) + \kappa^2 E\left(\frac{\partial u_t}{\partial \mu}\right)^2 \geq 0 \\
c &= \kappa E\left(u_t \frac{\partial u_t}{\partial \mu}\right),
\end{aligned}
\tag{7}
$$

where unconditional and conditional expectations are the same. When scale and shape parameters are known and $b < 1$, the information matrix for a single observation is time-invariant and given by

$$
\mathbf{I}(\boldsymbol{\psi}) = \sigma_u^2 \mathbf{D}(\boldsymbol{\psi}), \tag{8}
$$

where σ_u^2 is the information quantity for a single observation and

$$
\mathbf{D}(\boldsymbol{\psi}) = \mathbf{D}\begin{pmatrix} \widetilde{\kappa} \\ \widetilde{\phi} \\ \widetilde{\omega} \end{pmatrix} = \frac{1}{1-b}\begin{bmatrix} A & D & E \\ D & B & F \\ E & F & C \end{bmatrix}, \tag{9}
$$

with

$$
\begin{aligned}
A &= \sigma_u^2, \qquad B = \frac{\kappa^2\sigma_u^2(1+a\phi)}{(1-\phi^2)(1-a\phi)}, \qquad C = \frac{(1-\phi)^2(1+a)}{1-a}, \\
D &= \frac{a\kappa\sigma_u^2}{1-a\phi}, \qquad E = c(1-\phi)/(1-a) \qquad \textit{and} \qquad F = \frac{ac\kappa(1-\phi)}{(1-a)(1-a\phi)}.
\end{aligned}
$$

The ML estimator is asymptotically normal with covariance matrix given by the inverse of (8).

The above result can be easily extended to include the estimation of additional fixed parameters, such as the degrees of freedom in a t-distribution. Let $\boldsymbol{\theta}$ denote a vector of parameters such that $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_2')'$. Suppose that $\boldsymbol{\theta}_2$ consists of $n-1 \geq 1$ fixed parameters, while θ_1 is time-varying and depends on a set of parameters, $\boldsymbol{\psi}$. When the terms in the information matrix of the static model that involve θ_1, including cross-products, do not depend on θ_1,

$$\mathbf{I}\begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\theta}_2 \end{pmatrix} = \begin{bmatrix} E\left(\frac{\partial \ln f_t}{\partial \theta_1}\right)^2 \mathbf{D}(\boldsymbol{\psi}) & \mathbf{d}E\left(\frac{\partial \ln f_t}{\partial \theta_1}\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2'}\right) \\ E\left(\frac{\partial \ln f_t}{\partial \theta_1}\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2}\right)\mathbf{d}' & E\left(\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2}\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}_2'}\right) \end{bmatrix}, \tag{10}$$

where $\mathbf{D}(\boldsymbol{\psi})$ is the matrix in (9) and $\mathbf{d} = (0, 0, (1-\phi)/(1-a))'$. When the asymptotic distributions of the ML estimators of θ_1 and $\boldsymbol{\theta}_2$ are independent, the information matrix is block-diagonal and the top left hand block is as in (8).

3.2 Beta-t-EGARCH

The t_ν-distribution with a location of μ and scale of φ has probability density function

$$f(y;\mu,\varphi,\nu) = \frac{\Gamma\left(\left(\nu+1\right)/2\right)}{\Gamma\left(\nu/2\right)\varphi\sqrt{\pi\nu}}\left(1+\frac{(y-\mu)^2}{\nu\varphi^2}\right)^{-(\nu+1)/2}, \qquad \varphi, \nu > 0,$$

where ν is the degrees of freedom and $\Gamma(.)$ is the gamma function. Moments exist only up to and including $\nu - 1$. For $\nu > 2$, the variance is $\sigma^2 = \{\nu/(\nu-2)\}\varphi^2$. The excess kurtosis is $6/(\nu-4)$, provided that $\nu > 4$. The t_ν distribution has fat tails for finite ν with the tail index given by ν; see McNeil et al (2005, p 293).

When ε_t in (5) is t_ν distributed with $\mu = 0$ and $\varphi = 1$, the conditional score for the time-varying parameter $\lambda_{t|t-1}$ is

$$u_t = \frac{(\nu+1)y_t^2}{\nu\exp(2\lambda_{t|t-1}) + y_t^2} - 1 = \frac{(\nu+1)\varepsilon_t^2}{\nu+\varepsilon_t^2} - 1, \quad -1 \leq u_t \leq \nu, \quad \nu > 0. \tag{11}$$

At the true parameters values, u_t is IID and may be expressed as $u_t = (\nu+1)b_t - 1$, where b_t is distributed as $beta(1/2, \nu/2)$; see Harvey (2013, ch 4). Analytic expressions for the moments and autocorrelations of y_t can be found from the infinite MA representation of $\lambda_{t|t-1}$. The asymptotic distribution for a stationary first-order model, as in (6), can be found from (8).

There are a number of ways in which skewness may be introduced into a t-distribution. One possibility is by the method proposed by Fernandez and Steel (1998); see Harvey and Sucarrat (2012).

3.3 Gamma-GED-EGARCH

The PDF of the general error distribution, denoted $GED(\upsilon)$, is

$$f\left(y;\mu,\varphi,\upsilon\right)=\left[2^{1+1/\upsilon}\varphi\Gamma(1+1/\upsilon)\right]^{-1}\exp(-\left|(y-\mu)/\varphi\right|^{\upsilon}/2),\qquad \varphi,\upsilon>0, \tag{12}$$

where φ is a scale parameter, related to the standard deviation by the formula $\sigma=2^{1/\upsilon}(\Gamma\left(3/\upsilon\right)/\Gamma\left(1/\upsilon\right))^{1/2}\varphi$. The normal distribution is obtained when $\upsilon=2$, in which case $\sigma=\varphi$. Setting $\upsilon=1$ gives the Laplace, or double exponential, distribution, in which case $\sigma=2\sqrt{2}\varphi$. Therefore when $1\leq\upsilon\leq 2$ the GED distribution provides a continuum between the normal and Laplace. The kurtosis is $\Gamma\left(5/\upsilon\right)\Gamma\left(1/\upsilon\right)/\Gamma\left(3/\upsilon\right)$, so for $\upsilon=1$ the excess kurtosis is three.

The conditional score for $\lambda_{t|t-1}=\ln\varphi_{t|t-1}$ is

$$u_t=(\upsilon/2)\left|y_t/\exp(\lambda_{t|t-1})\right|^{\upsilon}-1,\qquad t=1,...,T. \tag{13}$$

The variable u_t is IID and may be expressed as $u_t=(\upsilon/2)g_t-1$, where g_t has a $gamma(2,\ 1/\upsilon)$ distribution. The score gives less weight to outliers than squared observations when $\upsilon<2$, but it is not as robust as a Beta-t-EGARCH model with small degrees of freedom. Unlike the EGB2, the score is not bounded when divided by y_t, unless $\upsilon=1$.

4 EGB2-EGARCH

The exponential generalized beta distribution of the second kind (EGB2) is obtained by taking the logarithm of a variable with a GB2 distribution. The distribution was first analyzed in Prentice (1975) and further explored by McDonald and Xu (1995). The PDF of a $GB2(\alpha,\nu,\xi,\varsigma)$ is[3]

$$f(x)=\frac{\nu(x/\alpha)^{\nu\xi-1}}{\alpha B(\xi,\varsigma)\left[(x/\alpha)^{\nu}+1\right]^{\xi+\varsigma}},\qquad \alpha,\nu,\xi,\varsigma>0, \tag{14}$$

where α is the scale parameter, ν,ξ and ς are shape parameters and $B(\xi,\varsigma)$ is the beta function. GB2 distributions are fat tailed for finite ξ and ς with

[3]The GB2 is described in Kleiber and Kotz (2003, ch6). Note that their convention has the order of α and ν reversed.

upper and lower tail indices of $\eta = \varsigma\nu$ and $\overline{\eta} = \xi\nu$ respectively. The absolute value of a t_f variate is GB2$(\varphi, 2, 1/2, f/2)$ with tail index $\eta = \overline{\eta} = f$.

If x is distributed as GB2$(\alpha, \nu, \xi, \varsigma)$ and $y = \ln x$, the PDF of the EGB2$(\mu, \nu, \xi, \varsigma)$ variate y is

$$f(y; \mu, \nu, \xi, \varsigma) = \frac{\nu \exp\{\xi(y-\mu)\nu\}}{B(\xi, \varsigma)(1 + \exp\{(y-\mu)\nu\})^{\xi+\varsigma}}. \tag{15}$$

The parameter which was the logarithm of scale in GB2 now becomes location in EGB, that is $\ln \alpha$ becomes μ. Furthermore ν is now a scale parameter, but ξ and ς are still shape parameters and they determine skewness and kurtosis.

4.1 Properties of EGB2

All moments of the EGB2 distribution exist. The first four are as follows:

$$\text{Mean: } E(y) = \mu + \nu^{-1}[\psi(\xi) - \psi(\varsigma)] \tag{16}$$

$$\text{Variance: } \sigma^2 = E(y - E(y))^2 = \nu^{-2}[\psi'(\xi) + \psi'(\varsigma)] \tag{17}$$

$$\text{Skewness: } \frac{E(y - E(y))^3}{\sigma^3} = \frac{\psi''(\xi) - \psi''(\varsigma)}{[\psi'(\xi) + \psi'(\varsigma)]^{3/2}} \tag{18}$$

$$\text{Kurtosis: } \frac{E(y - E(y))^4}{\sigma^4} = \frac{\psi'''(\xi) + \psi'''(\varsigma)}{[\psi'(\xi) + \psi'(\varsigma)]^2} + 3, \tag{19}$$

where ψ, ψ', ψ'' and ψ''' are polygamma functions of order 0, 1, 2 and 3 respectively. The EGB2 distribution is positively (negatively) skewed when $\xi > \varsigma$ ($\xi < \varsigma$) and its kurtosis decreases as ξ and ς increase. Skewness ranges between -2 and 2 and kurtosis[4] lies between 3 and 9. There is excess kurtosis for finite ξ and/or ς.

Although ν is a scale parameter, it is the inverse of what would normally be considered a conventional measure of scale. Thus scale is better defined as $1/\nu$ or as the standard deviation

$$\sigma = \sqrt{\psi'(\xi) + \psi'(\varsigma)}/\nu = h(\xi, \varsigma)/\nu = h/\nu. \tag{20}$$

[4]The maximum kurtosis in the symmetric case is 6 and is for $\xi = \varsigma = 0$. The kurtosis of 9 is achieved when ξ(or ς) $= 0$ and ς(or ξ)$=\infty$.

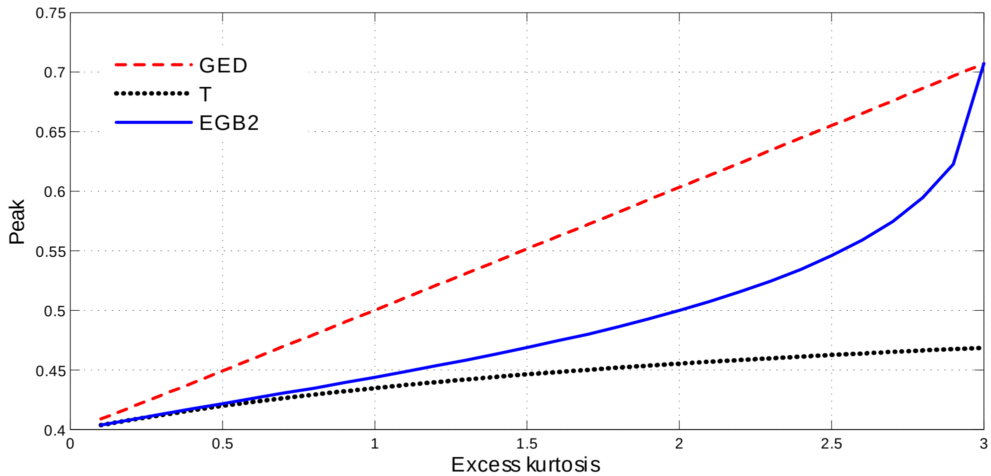

Figure 1: PDF at the mean for GED, EGB2 and t-distributions with the same kurtosis.

Thus

$$f(y;\mu,\sigma,\xi,\varsigma)=\frac{h\exp\{\xi h(y-\mu)/\sigma\}}{\sigma B(\xi,\varsigma)(1+\exp\{h(y-\mu)/\sigma\})^{\xi+\varsigma}}.$$

When $\xi=\varsigma$, the distribution is symmetric; for $\xi=\varsigma=1$ it is a logistic distribution and when $\xi=\varsigma\to\infty$ it tends to a normal distribution. When $\xi=\varsigma=0$ in the EGB2, the distribution is double exponential or Laplace; see Caivano and Harvey (2013). The following results will be used in a number of places when $\xi=\varsigma$: (i) $\xi h^2=2$ as $\xi\to\infty$, and $\xi h\to 2/h\to\infty$,(ii) $\xi h=\sqrt{2}$ for $\xi=0$. Equivalently: (i) $\xi\psi'(\xi)=1$ as $\xi\to\infty$, (ii) $\xi\sqrt{\psi'(\xi)}=1$ for $\xi=0$.

A plot of the (symmetric) EGB2, GED and Student's t with the same excess kurtosis shows them to be very similar. It is difficult to see the heavier tails of the t distribution from the graph, and the only discernible difference among the three distributions is in the peak, which is higher and more pointed for the GED. The EGB2 in turn is more peaked than the t. As the excess kurtosis increases, the differences between the peaks become more marked; see Figure 1.

4.2 Dynamic scale model

The first-order dynamic scale model with EGB2 distributed errors is (5) where ε_t is a standardized ($\mu = 0, \nu = 1$) EGB2, that is $\varepsilon_t \sim EGB2(0, 1, \xi, \varsigma)$. Thus the conditional distribution is

$$f_t(y_t; \mu, \boldsymbol{\psi}, \xi, \varsigma) = \frac{\exp\{\xi(y_t - \mu)e^{-\lambda_{t|t-1}}\}}{e^{\lambda_{t|t-1}} B(\xi, \varsigma)(1 + \exp\{(y - \mu)e^{-\lambda_{t|t-1}}\})^{\xi+\varsigma}},$$

where $\boldsymbol{\psi}$ now denotes the parameters in (6). The conditional score is

$$u_t = \frac{\partial \ln f(y_t)}{\partial \lambda_{t|t-1}} = (\xi + \varsigma)\varepsilon_t b_t - \xi\varepsilon_t - 1, \tag{21}$$

where $\varepsilon_t = (y_t - \mu)e^{-\lambda_{t|t-1}}$ and

$$b_t = \frac{\exp\{(y - \mu)e^{-\lambda_{t|t-1}}\}}{1 + \exp\{(y - \mu)e^{-\lambda_{t|t-1}}\}} = \frac{\exp \varepsilon_t}{1 + \exp \varepsilon_t}.$$

At the true parameters values, $b_t \sim beta(\xi, \varsigma)$.

The model may be parameterized in terms of the standard deviation, $\sigma_{t|t-1}$, by defining $\epsilon_t = \varepsilon_t / h$. Then

$$y_t = \mu + \exp(\lambda_{\sigma,t|t-1})\epsilon_t, \qquad t = 1, ..., T,$$

with the only difference between $\lambda_{\sigma,t|t-1}$ and $\lambda_{t|t-1}$ being in the constant term which in $\lambda_{\sigma,t|t-1}$ is $\omega_\sigma = \omega + \ln h$; see the earlier discussion in sub-section 5.1. Note that the variance of ϵ_t is unity.

Writing the score, (21) as

$$u_t = h(\xi + \varsigma)\epsilon_t b_t - h\xi\epsilon_t - 1, \tag{22}$$

it can be seen[5] that when $\xi = \varsigma = 0, \sqrt{2}\,|\epsilon_t| - 1$ and, when $\xi = \varsigma \to \infty$,

[5] When $\xi = 0$, $\xi h = \sqrt{2}$ and b_t degenerates to a Bernoulli variable such that $b_t = 0$ when $\epsilon_t < 0$ and $b_t = 1$ when $\epsilon_t > 0$. Then $2b_t - 1 = 1$ (-1) for $\epsilon_t > 0$ ($\epsilon_t < 0$) and the score can be written as: $u_t = \sqrt{2}\,|\epsilon_t| - 1$.

As regards $\xi \to \infty$, note that because $\partial b_t / \partial \epsilon_t = h b_t (1 - b_t)$, a first order Taylor expansion of b_t around $\varepsilon_t = 0$ yields $b_t \simeq \frac{1}{2} + \frac{h}{4}\epsilon_t$.Therefore $2b_t - 1 \simeq (h/2)\epsilon_t$ and $u_t \simeq (\xi h^2/2)\epsilon_t^2 - 1$. As $\xi \to \infty$, $\xi h^2 \to 2$.

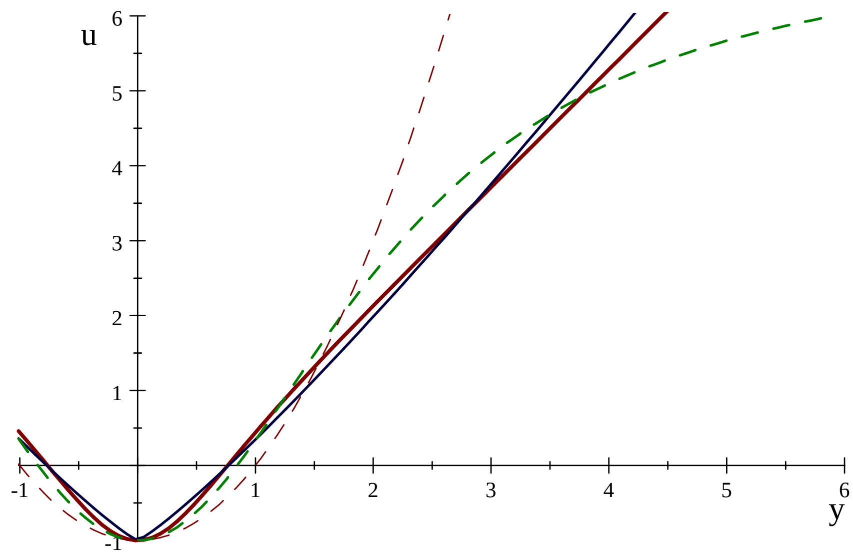

Figure 2: Score functions for EGB2 (thick line), GED (medium line) and t (thick dash), all with unit variance and an excess kurtosis of 2. Thin line shows normal score. (These score functions are even).

$u_t = \epsilon_t^2 - 1$.

Figure 2 compares the way observations are weighted by the score of an EGB2 distribution with $\xi = \varsigma = 0.5$, a Student's t_7 distribution and a $GED(1.148)$. These are the same distributions used in Figure 1; all have excess kurtosis of 2. Dividing (22) by ϵ_t gives a bounded function as $|\epsilon_t| \to \infty$. This is consistent with the 'soft' Winsorizing[6] of the location score; see Caivano and Harvey (2013).

The unconditional mean is given by $E\left(y_t\right) = \mu + E\left(\varepsilon_t\right) E(e^{\lambda_{t|t-1}})$, whereas the $m-th$ unconditional moment about the mean is $E\left(\varepsilon_t^m\right) E(e^{m\lambda_{t|t-1}})$, $m > 1$. In the Beta-t-EGARCH and Gamma-GED-EGARCH models analysed in Harvey (2013, ch4), the expression $E(\exp(m\lambda_{t|t-1}))$ depends on the moment generating functions (MGF) of beta and gamma variates, respectively, which

[6] The M-estimator, which features prominently in the robustness literature, has a Gaussian response until a certain threshold, K, whereupon it is constant; see Maronna *et al* (2006, p 25-31). This is known as Winsorizing as opposed to trimming where observations greater in absolute value than K are set to zero.

have a known form. For EGB2-EGARCH, the unconditional moments depend on the MGF of u_t, ie $E_{EGB2(\xi,\varsigma)}[mu_t]$, where u_t is defined in (21). For the limiting normal and Laplace cases of the EGB2, the score functions and hence the unconditional moments are the same as for $\upsilon = 2$ and $\upsilon = 1$ in Gamma-GED-EGARCH; see Harvey (2013, sub-section 4.2.2). For $\upsilon = 1$ it is necessary to have $m\kappa < 1$ in the first-order model for the $m-th$ moment to exist, whereas for $\upsilon = 2$ the condition is $m\kappa < 1/2$. For $0 < \xi, \varsigma < \infty$ having the last condition hold is therefore sufficient for the existence of the unconditional moments. This being the case, we can at least assert, from Jensen's inequality, that the unconditional moments exceed the conditional moments and that the kurtosis increases; see Harvey (2013, p 102).

The MGF of u_t is also required to find the conditional expectations needed to forecast volatility and volatility of volatility. However, it is the full $\ell - step$ ahead conditional distribution which is often needed in practice and this is easily simulated from standardized beta variates. The quantiles, such as those needed for VaR and the associated expected shortfalls, may be estimated at the same time.

4.3 Maximum likelihood estimation

The asymptotic distribution of the ML estimators of the parameters in a dynamic scale model with a symmetric EGB2 distribution is given in the proposition below. The score and its derivatives are linear combinations of variables of the form $\varepsilon_t^r b_t^h (1 - b_t)^k$, $r, h, k = 0, 1, 2..$ and the properties of these variables are such that the conditions for convergence and asymptotic normality of the maximum likelihood estimator may be verified without too much difficulty. The formulae for the general result on the asymptotic distribution are quite complex; see Caivano and Harvey (2013).

Proposition 1 *Suppose that ε_t in (5) is known to be symmetric with a standardized $EGB2(0, 1, \xi, \xi)$ distribution. Let $\lambda_{t|t-1}$ be generated by (6) with $|\phi| < 1$. Define a, b and c as in (7) with*

$$E(u_t') = \frac{1 - 2\xi^2 \psi'(\xi) - 2\xi}{2\xi + 1} = -\sigma_u^2 \tag{23}$$

$$E(u_t'^2) = \frac{\xi^3 (\xi + 1)}{(2\xi + 3)(2\xi + 1)} (2\psi'''(\xi + 2) + 12\psi'^2(\xi + 2)) + \sigma_u^2 + 1 \tag{24}$$

and

$$E(u_t u_t') = -1. \tag{25}$$

Let $\boldsymbol{\psi} = (\kappa,\ \phi,\ \omega)'$. *Assuming that* $b < 1$ *and* $\kappa \neq 0$, $(\mu, \widetilde{\boldsymbol{\psi}}', \widetilde{\xi},)'$, *the ML estimator of* $(\boldsymbol{\mu}, \boldsymbol{\psi}', \xi)'$, *is consistent and the limiting distribution of* $\sqrt{T}(\widetilde{\mu} - \mu, (\widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi})', \widetilde{\xi} - \xi)'$ *is multivariate normal with mean vector zero and covariance matrix given by* $Var(\widetilde{\mu}, \widetilde{\boldsymbol{\psi}}, \widetilde{\xi},) = \mathbf{I}^{-1}(\mu, \boldsymbol{\psi}, \xi)$, *where the information matrix is*

$$\mathbf{I}\begin{pmatrix} \mu \\ \boldsymbol{\psi} \\ \xi \end{pmatrix} = \begin{bmatrix} \frac{\xi^2}{1+2\xi}E(e^{-2\lambda_{t|t-1}}) & 0 & 0 \\ 0 & \frac{2\xi+2\xi^2\psi'(\xi)-1}{1+2\xi}\mathbf{D}(\boldsymbol{\psi}) & -\frac{1}{\xi}\mathbf{d} \\ 0 & -\frac{1}{\xi}\mathbf{d}' & 2\psi'(\xi) - 4\psi'(2\xi) \end{bmatrix}. \tag{26}$$

The block diagonality of (26) means that the asymptotic variances of ξ and the parameters in $\boldsymbol{\psi}$ can be computed even though an expression for the unconditional expectation of $\exp(\lambda_{t|t-1})$ is difficult to derive for $0 < \xi < \infty$.

Remark 1 *The information matrix is more complicated if* ω_σ *(which is* $\omega + \ln h$*) is used rather than* ω *(although it can still be found). However, standard errors are of little practical importance for the constant term and the standard errors of the other parameters do not depend on its parameterization.*

When $\xi = 0$, so that the distribution is Laplace, $E(u_t') = -1$. Similarly as $\xi \to \infty$, $E(u_t') = -2$, which is the correct result for a Gaussian distribution. In addition, when $\xi = 0$, both $\psi'(\xi+2)$ and $\psi'''(\xi+2)$ are finite, so $E(u_t'^2) = 2$. Hence

$$b = \phi^2 - 2\phi\kappa + 2\kappa^2, \tag{27}$$

which is the same as given by the expression in Harvey (2013, p 120) for b in Gamma-GED-EGARCH when $\upsilon = 1$. (Also $c = -1$.) Similarly for $\xi \to \infty$,

$$b = \phi^2 - 4\phi\kappa + 12\kappa^2.$$

4.4 Tests for serial correlation

Before fitting a model, a test against serial correlation may be carried out. Lagrange multiplier (LM) tests against an MA(P) process are based on Ljung-Box (portmanteau) statistics formed using the score. One possibility is to fit an EGB2 distribution to the raw data and to construct scores using the estimated location, scale and shape parameters. An alternative is simply to assume a Laplace distribution, in which case the scores are just the absolute values of deviations from the median. There is a good deal of evidence to suggest that tests based on absolute values are more powerful than tests based on squares.

Mikosch and Starica (2000) draw attention to the unreliability of sample autocorrelations computed from squared observations when the underlying process is a persistent GARCH(1,1). Fitting a t-distribution under the null hypothesis and carrying out the score based test may mitigate the problem.

5 Exchange rates

Tables 1 and 2 report the full ML estimates of the (symmetric) EGB2-EGARCH and Beta-t-EGARCH models for the returns of exchange rates of developed and emerging countries against the US dollar. Developed countries currencies include the Australian dollar (AUD), the Canadian dollar (CAD), the Swiss franc (CHF), the Denmark krone (DKK), the Euro (EUR), the Pound sterling (GBP), the Japanese yen (JPY), the Norwegian krone (NOK), the New Zealand dollar (NZD) and the Swedish krona (SEK). Emerging countries currencies include the Brazilian Real (BRL), the Chinese renmimbi (CNY), the Hong Kong dollar (HKD), the Indian rupee (INR), the South Korean won (KRW), the Sri Lanka rupee (LKR), the Mexican peso (MXN), the Malaysian ringgit (MYR), the Singapore dollar (SGD) the Thai baht (THB), the Taiwan dollar (TWD) and the South African Rand (ZAR). Exchange rate data are daily and range from 4th January 1999 to 15th March 2013.

As can be seen, the EGB2 gives a better fit for five developed countries, whereas the t is best for four. For the developing countries the situation is very different in that the EGB2 is better than the t in only three cases out of 12. For four currencies the estimated degrees of freedom of the t-distribution is below three and in these cases the ML estimation of the EGB2 model failed to converge[7].

[7] Although it is not the purpose of this exercise to compare DCS EGARCH models with standard GARCH - there is already a good deal of evidence in Creal et al (2011), Harvey and Sucarrat (2012) and elsewhere to suggest that DCS EGARCH tends to be better- we did fit GARCH-t models and found that in only 7 out of 23 cases did they beat

	EGB2					t				
	κ	ϕ	ω	ξ	Log-L	κ	ϕ	ω	ν	Log-L
AUD	0.030	0.991	-5.34	1.29	12523.7	0.030	0.992	-5.04	9.22	**12526.2**
CAD	0.023	0.996	-5.55	1.71	**13823.2**	0.024	0.996	-5.40	12.64	13822.4
CHF	0.018	0.993	-5.56	1.05	12848.2	0.017	0.994	-5.14	8.47	**12849.9**
CHF*	0.017	0.994	-5.47	1.22	**12865.0**	0.016	0.994	-5.13	9.69	12863.1
DKK	0.019	0.995	-5.61	1.12	13086.9	0.018	0.995	-5.22	8.87	13086.9
EUR	0.017	0.995	-5.43	1.46	**13118.4**	0.017	0.995	-5.20	11.13	13117.4
GBP	0.022	0.994	-5.30	2.30	**13575.8**	0.022	0.994	-5.33	16.01	13575.3
JPY	0.024	0.989	-5.87	0.73	13074.7	0.024	0.990	-5.21	6.14	**13078.2**
NOK	0.018	0.997	-5.38	1.54	**12596.8**	0.018	0.997	-5.19	11.03	12596.1
NZD	0.024	0.992	-5.41	1.03	12184.0	0.024	0.992	-4.98	7.77	**12184.7**
SEK	0.018	0.996	-5.15	1.91	**12547.4**	0.018	0.996	-5.07	13.18	12546.9

Table 1 ML estimates for exchange rate data (developed countries)

	EGB2					T				
	κ	ϕ	ω	ξ	Log-L	κ	ϕ	ω	ν	Log-L
BRL	0.082	0.975	-5.25	1.08	**12099.7**	0.087	0.974	-484	8.30	12098.8
CNY	0.011	0.998	-9.73	0.35	23414.1	0.057	0.999	-9.38	5.37	**23919.2**
HKD	-	-	-	-	-	0.213	0.985	-9.15	2.29	**26527.7**
INR	-	-	-	-	-	0.102	0.992	-6.49	2.97	**16136.4**
MYR	0.024	1.000	-10.11	1.16	21792.3	0.078	1.000	-9.81	9.64	**21907.8**
MXN	0.055	0.979	-5.63	1.31	13727.8	0.055	0.980	-5.34	9.32	**13728.9**
ZAR	0.042	0.991	-5.09	1.18	11596.4	0.042	0.991	-4.74	8.75	**11596.5**
SGD	0.032	0.989	-6.48	0.89	15724.7	0.033	0.989	-5.95	7.04	**15726.9**
KRW	0.074	0.985	-6.75	0.33	**14026.1**	0.074	0.984	-5.49	4.83	14020.1
LKR	-	-	-	-	-	0.167	0.974	-7.08	1.89	**18451.6**
TWD	-	-	-	-	-	0.111	0.974	-6.47	2.72	**16727.3**
THB	0.096	0.968	-7.25	0.31	**15410.3**	0.091	0.970	-5.95	4.26	15405.3

Table 2 ML estimates for exchange rate data (emerging countries)

Beta-t-EGARCH.

6 Scale parameters and tail indices

Tail index estimators may be computed prior to fitting DCS-EGARCH volatility models. As such they may be used as starting values for an iterative maximum likelihood estimation procedure. Sub-section 6.1 reviews tail estimators and sub-section 6.2 presents evidence on the accuracy with which they may be expected to estimate the scale parameters of an EGB2 distribution when applied to the residuals from fitting a preliminary model to returns. A similar analysis is conducted on the estimation of the degrees of freedom of a t-distribution from tail indices computed from the logarithms of absolute returns. Sub-section 6.3 returns to the exchange rate data of Section 5 and presents estimates of the tail indices, and implied shape parameters, computed from the residuals from GARCH models. These estimates are much smaller than the corresponding ML estimates.

The use of residuals from a preliminary model can be avoided simply by using the raw data on returns because, in theory, the tail index estimators will still be consistent; see Resnick and Starica (1995). However, it seems that the increased kurtosis induced by dynamic volatility can substantially increase the downward bias. These findings have important implications for the conclusions to be drawn from estimating tail indices by nonparametric methods.

6.1 Tail index estimators

Hill's estimator of the tail index for a fat-tailed distribution is

$$\widehat{\eta} = \left(k^{-1} \sum_{j=1}^{k} \ln x_j - \ln x_k \right)^{-1} = \left(k^{-1} \sum_{j=1}^{k} y_j - y_k \right)^{-1}$$

where x_j and y_j, $j = 1, ..., k$, denote the observations in descending order. Embrechts, Kluppelberg and Mikosch (1997, p 336-7) set out the asymptotic properties for a power law distribution of the form (3). The variance of the limiting (normal) distribution of $\sqrt{k}(\widehat{\eta} - \eta)$ is η^2, so the asymptotic variance of $\ln \widehat{\eta}$ is $1/k$. Note that the asymptotic theory requires not only that T and $k \to \infty$, but that $k/T \to 0$.

A similar estimator, $\widehat{\widehat{\eta}}$, may be constructed for the lower tail index by putting the observations in ascending order and using the smallest observations. When the observations come from a (symmetric) distribution, an estimate of location is subtracted and Hill's estimator is then constructed from the logarithms of absolute values.

It is well-known that Hill's estimator can be quite badly biased; it is usually too low. Various alternatives have been suggested, one of the more recent ones being the OLS estimator from a regression of log rank minus half on log size; see Gabaix and Ibragimov (2011). However, even the improved estimators have bias and this bias turns out to play an important role in influencing the conclusions that one might be tempted to draw.

Because the performance of both Hill's and OLS estimates improves the more observations are excluded from the tail, one might be tempted to exclude as many observations as possible. However, doing so can lead to very imprecise estimates. A careful choice of the truncation point is needed in order to achieve a good bias-variance trade-off; see the plots in Embrechts et al (1997).

6.2 Tail index estimators of shape parameters for EGB2 and Student t distributions

The upper and lower tail indices in the GB2 distribution are $\nu\varsigma$ and $\nu\xi$ respectively. Hence estimators of ς and ξ in the EGB2 model may be obtained from standardized residuals from an initial model by solving the equations $\widehat{\eta} = h(\widehat{\xi}, \widehat{\varsigma})\widehat{\varsigma}$ and $\widehat{\overline{\eta}} = h(\widehat{\xi}, \widehat{\varsigma})\widehat{\xi}$. Note that the lower bound for η $(= \overline{\eta})$ is obtained in the symmetric model when $\xi = \varsigma = 0$ and is $\sqrt{2}$. More generally the lower bound is one for $\widehat{\eta}$ ($\widehat{\overline{\eta}}$) when $\varsigma(\xi) = 0$ and $\xi(\varsigma) > 0$. There is no finite upper bound. In the symmetric case the tail index values implied by various values of $\xi = \varsigma$ - given in brackets - are as follows: 14. 18 (100), 3. 33 (5), 2. 27 (2), 1. 81 (1), 1. 57 (0.5).

When Hill's estimator is constructed from the *logarithms* of absolute values of residuals, it gives an estimator of the degrees of freedom of a t-distribution directly.

In order to assess the accuracy of the Hill's and OLS estimators for the EGB2 and t-distributions, simulations for $T = 10,000$ were carried out using 1,000 replications. The results for an $EGB2(0, 1, \xi, \xi)$ are shown in Figure 3; setting with $\nu = 1$ means that $\eta = \xi$. The OLS estimator dominates Hill's estimator in terms of bias, but it still underestimates the true tail index, with the bias increasing with the shape parameter. The bias also depends on how many observation are included in the tail: when 10% of the observations are

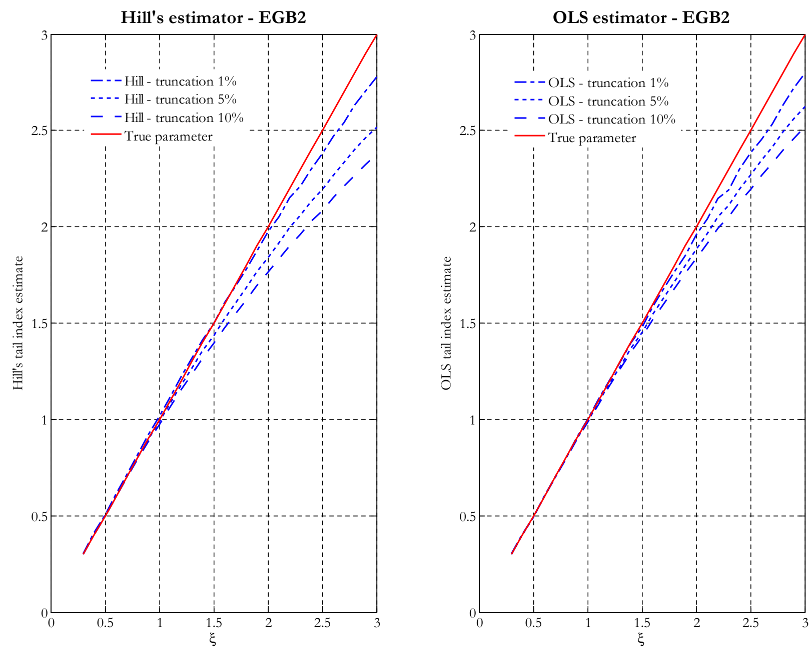

Figure 3: Average estimates of tail index plotted against true tail index for a $GB2$ corresponding to $EGB2(0, 1, \xi, \xi)$.

included, the bias is already non-negligible for $\xi = 1.5$. On the other hand, if we include only 1% of the observations, the estimate is still relatively reliable when $\xi = 2$.

The ML estimates of the EGB2 shape parameters reported in Table 1 are all quite low. Hence the tail index estimates obtained by the Hill and OLS methods will provide good starting values for parameters of this order of magnitude. On the other hand, for a t distribution, the bias in Hill's estimator is large even for a relatively small degrees of freedom and a 1% truncation; see Figure 4. The bias becomes considerably worse as the degrees of freedom increase. The OLS estimator offers some improvement but not a great deal. The same is true of other modifications. Studies of new estimators are often confined to small tail indices; for example, Huisman at al (2001)

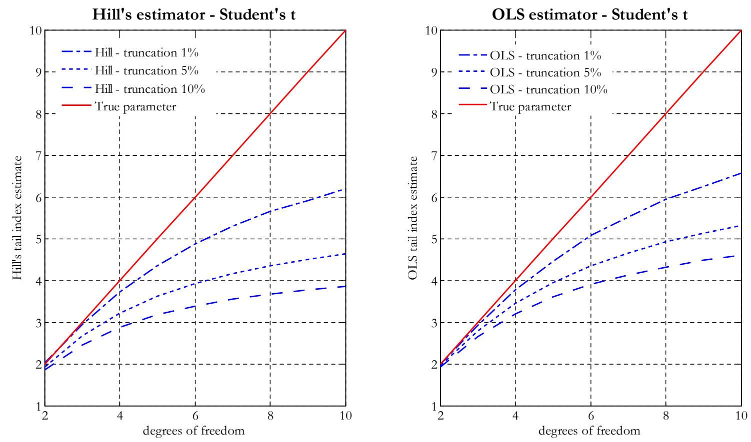

Figure 4: Average estimates of tail index plotted against degrees of freedom for t_ν.

only[8] present results for a t-distribution with $\nu \leq 5$.

The differences in the value of the tail index estimators as starting values for t and EGB2 stems from the fact that the low values of the EGB2 shape parameter, ξ, correspond to tail indices for the GB2 distribution that are much smaller than the the tail indices for the t-distribution. Figure 5 shows the tail index estimators for EGB2 and t-distributions plotted in such a way that the values of ξ on the horizontal axis for EGB2 correspond to a tail index for GB2 that is similar to the degrees of freedom (and hence tail index) for the t.

The above graphs prompt the question as to the behaviour of tail estimators when the distribution does not have fat tails. An analysis of the log-normal distribution provides some insight. The log-normal distribution is sub-exponential[9], but it is not fat-tailed and all its moments exist. Hence the tail index should theoretically be infinite. However, consider Hill's estimator which, as McNeil et al (2005, p 286-7) observe, is motivated by the mean excess function of the logarithm of variable, x, with a fat-tailed distribution,

[8] Nevertheless they conclude on p 214 that '..tail fatness is easily exaggerated in small samples.'

[9] See Embrechts, Kluppelberg and Mikosch (1997, p 34)

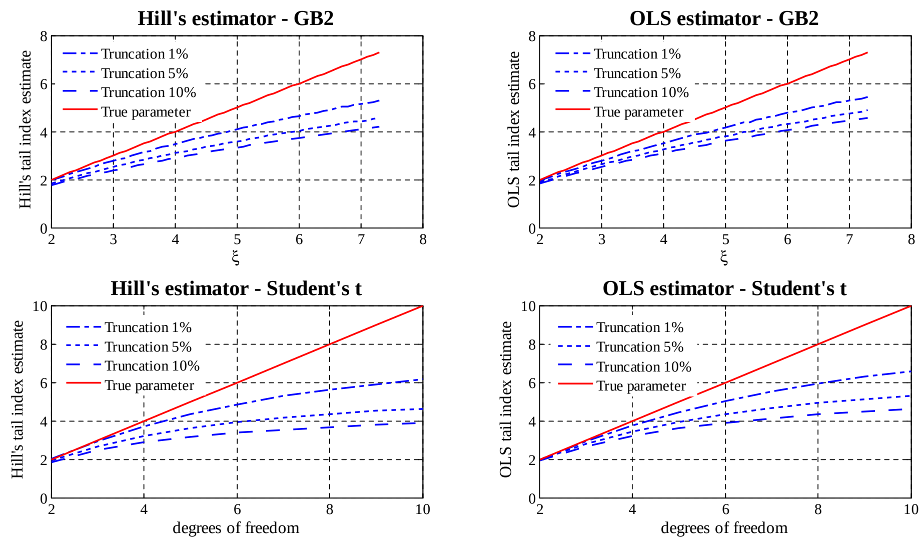

Figure 5: Hill's and OLS estimates compared for GB2 and t distributions

that is

$$e(y^*) = E(y - y^* \mid y > y^*), \tag{28}$$

where $y = \ln x$. Hill's estimator is the inverse of the sample mean excess function. For a Pareto distribution, y is exponentially distributed and $e(0) = 1/\eta$ is just the mean. For the log-normal, we can make use of the relationship between $ES(\alpha)$, the expected shortfall for a Gaussian variable, that is $y \sim N(\mu, \sigma^2)$, beyond the α quantile, and the mean excess function. Specifically,

$$e(y^* = y_\alpha) = ES(\alpha) - \mu - z_\alpha \sigma,$$

where $y_\alpha = \mu + z_\alpha \sigma$ and z_α is the α quantile for a standard normal variate. From the formula for $ES(\alpha)$ derived in McNeil et al (2005, p 45),

$$e(y_\alpha) = \sigma \left(\frac{\phi(z_\alpha)}{1 - \alpha} - z_\alpha \right),$$

where $\phi(z)$ is the PDF at z of a standard normal variate. Evaluating $1/e(y_\alpha)$ then gives the $p\lim$ of Hill's estimator, which we will denote as H_α. Table 3 shows H_α multiplied by σ for typical values of α. The estimator improves, in the sense that its $p\lim$ gets bigger, as α gets smaller, which is consistent with $k/T \to 0$. As $\sigma \to \infty$, the log-normal tends towards a (degenerate) normal and $H_\alpha \to \infty$. On the other hand, as σ increases, $H_\alpha \to 0$. The fall in H_α corresponds to an increase in excess kurtosis, which is $\exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 6$. For $\sigma = 0.5$, the excess kurtosis is 5.90 whereas the skewness, $(\exp(\sigma^2) + 2)\sqrt{\exp(\sigma^2) - 1}$, is 1.75. For $\sigma = 1$, the skewness is far more pronounced and the excess kurtosis is 110.94. Even with $\sigma = 0.5$, one might conclude, quite erroneously, that, on the basis of the 5% quantile, the existence of fifth moment, and perhaps even the fourth, is in doubt. Even setting α to the unrealistically small value of 0.001 gives a $p\lim$ of only twelve for Hill's estimator.

α	0.10	0.05	0.01	0.001
σH_α	2.10	2.39	3.23	6.00

Table 3 Plim of Hill's estimator (times σ) for data from a lognormal distribution for different quantiles, α.

The above analysis suggest that tail index estimates may be low even when the true index is infinite, with the index estimates being closer to zero the higher is the kurtosis. The average tail indices computed from the logarithms of absolute values of returns of a simulated EGB2 distribution are shown in Figure 6 and the results confirm this conjecture. For example, when $\xi = 1$, the Hill's estimates are centred on $H_{0.05} \simeq 4.5$ while the corresponding figure for the OLS estimator is approximately 5.3.

6.3 Estimates of shape parameters from tail indices of residuals

Tables 4 and 5 compare the ML estimates of the shape parameters for EGB2-EGARCH and Beta-t-EGARCH models obtained in Section 5 with those implied by the Hill's and OLS estimates[10] obtained from the standardized residuals of a GARCH(1,1) model (estimated by QML, assuming normality).

[10] In order to choose the optimal truncation point for the Hill's and OLS estimators a commonly suggested strategy is to plot the estimators for various truncation points and to choose one in a region were the estimator is reasonably stable. A look at Hill's plots showed them to be very unstable in many cases. Nevertheless we report the maximum value obtained in this way (for thresholds less than 20%).

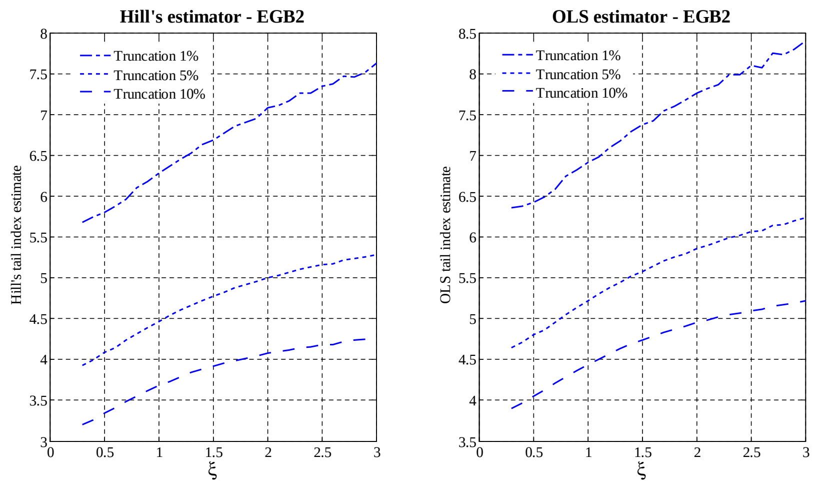

Figure 6: Average estimates of tail index from the logarithms of absolute returns plotted against ξ when data is $EGB2(0, 1, \xi, \xi)$.

For Beta-t-EGARCH the tail estimates are computed from the logarithms of absolute values. The implied EGB2 shape parameter is given by solving the equation $\eta = \xi\sqrt{2\psi'(\xi)}$, whereas the degrees of freedom for the t is the tail index. As might be expected from Figure 4, both Hill's and OLS estimates tend to be much smaller than the ML estimates of ν in the t-distribution. This is not true of the estimates for the EGB2 distribution, for the reasons given earlier. However, for emerging countries there are a number of missing entries for EGB2 because the corresponding tail index estimate was below the theoretical lower bound of $\sqrt{2}$; it comes as no surprise that most of these occur when the ML procedure failed to converge. In such cases there is a clear indication of fat tails.

Developed countries

	Hill's			OLS			EGB2-EGARCH	
	5%	10%	max	5%	10%	max	ξ	Implied kurtosis
AUD	0.93	1.08	1.11	-	0.20	0.56	1.29	0.94
CAD	1.73	1.60	1.93	1.16	1.41	1.45	1.71	0.70
CHF	1.24	1.36	1.40	-	0.57	0.82	1.05	1.15
DKK	1.46	1.12	1.47	0.46	0.80	0.95	1.12	1.08
EUR	2.10	1.44	2.10	1.50	1.58	1.60	1.46	0.83
GBP	2.03	1.55	2.11	2.00	1.85	2.35	2.30	0.51
JPY	0.35	0.53	0.73	-	-	0.28	0.73	1.56
NOK	1.19	1.15	3.28	1.79	1.34	2.88	1.54	0.78
NZD	0.56	0.86	0.98	0.17	0.44	0.65	1.03	1.17
SEK	1.53	1.23	2.57	2.20	1.61	3.29	1.91	0.62

Table 4a Shape parameter estimates implied by tail index estimates for EGB2-EGARCH.

Developed countries

	Hill's			OLS			Beta-t-EGARCH	
	5%	10%	max	5%	10%	max	ν	Implied kurtosis
AUD	4.52	3.88	4.70	4.30	4.23	4.33	9.22	1.15
CAD	5.20	4.30	6.60	5.54	5.07	6.55	12.64	0.69
CHF	4.79	4.17	5.90	4.98	4.68	5.01	8.47	1.34
DKK	5.11	3.88	6.19	5.19	4.72	5.20	8.87	1.23
EUR	5.68	4.16	6.47	5.81	5.14	6.73	11.13	0.84
GBP	5.49	4.24	7.17	6.13	5.24	8.05	16.01	0.50
JPY	3.89	3.33	4.26	4.01	3.72	4.33	6.14	2.80
NOK	4.64	3.92	8.99	5.92	4.65	9.10	11.03	0.85
NZD	4.01	3.67	4.67	4.39	4.06	4.95	7.77	1.59
SEK	4.98	3.95	7.52	6.33	4.85	9.32	13.18	0.65

Table 4b Tail index (degrees of freedom) estimates for Student's t

Emerging countries

	Hill's			OLS			EGB2-EGARCH	
	5%	10%	max	5%	10%	max	ξ	Implied kurtosis
BRL	0.68	0.85	1.13	0.12	0.45	0.70	1.08	1.12
CNY	-	-	-	-	-	-	0.35	2.35
HKD	-	-	-	-	-	-	-	-
INR	-	-	-	-	-	-	-	-
KRW	0.22	0.35	0.50	-	-	0.03	0.33	1.04
LKR	-	-	-	-	-	-	-	0.93
MXN	0.47	0.63	0.85	-	0.28	0.50	1.31	1.03
MYR	-	-	-	-	-	-	1.16	1.33
SGD	0.53	0.70	0.73	-	0.15	0.42	0.89	2.40
THB	-	0.10	0.34	-	-	-	0.31	-
TWD	-	-	-	-	-	-	-	-
ZAR	0.84	0.99	1.27	0.92	0.89	1.25	1.18	2.45

Table 5a Shape parameter estimates implied by tail index estimates for EGB2-EGARCH.

Emerging countries

	Hill's			OLS			Beta-t-EGARCH	
	5%	10%	max	5%	10%	max	ν	Implied kurtosis
BRL	4.00	3.44	4.83	4.11	3.89	5.37	8.30	1.40
CNY	2.60	2.02	2.92	2.58	2.37	2.58	5.37	4.38
HKD	2.42	2.14	2.85	2.39	2.33	2.39	2.29	-
INR	2.98	2.54	3.24	3.09	2.85	3.18	2.97	-
KRW	3.86	3.23	4.79	4.20	3.76	4.48	4.83	1.06
LKR	2.53	2.24	2.65	2.39	2.42	2.45	1.89	1.13
MXN	4.05	3.50	5.27	4.40	4.02	5.02	9.32	1.26
MYR	3.22	2.53	3.97	3.30	3.05	3.32	9.64	1.97
SGD	4.07	3.52	5.01	4.33	3.97	4.63	7.04	7.23
THB	3.47	3.08	4.25	4.00	3.51	4.60	4.26	-
TWD	3.05	2.65	3.33	3.12	2.93	3.18	2.72	-
ZAR	4.25	3.74	6.18	5.00	4.30	6.91	8.75	23.08

Table 5b Tail index (degrees of freedom) estimates for Student's t

6.4 Tail index estimates for raw data

Although the tail index estimators are consistent when computed from raw data, they are typically much lower the corresponding estimates obtained from residuals. This is certainly true of the tail index estimates of the exchange rates of Section 5 as reported by Ibragimov et al. (2013). Even for the developed economies, the tail index estimates are mostly less than four, implying infinite fourth moments; see also Loretan and Phillips (1994).

There is some work to suggest that for fat-tailed conditional distributions, a GARCH(1,1) process can lower the tail index when it is close to IGARCH; see Mikosch and Starica (2000) and Huisman et al (2001, p212). However, some calculations in McNeil et al (2005, p 296-7) suggest that, for plausible values of the parameters, the reduction may be small. For a stationary Beta-t-EGARCH the situation is perhaps more clear-cut in that a basic property of the model is that the existence of moments, and hence the tail index of the conditional distribution, is not changed by changing volatility. What does change, for EGB2 EGARCH as well as Beta-t-EGARCH, is that the excess kurtosis increases. The increase can be worked out and Table 6 shows the (proportional) increase for normal, Laplace and t-distributions. Perhaps surprisingly the increase is bigger for Laplace, and to a lesser extent normal, than it is for t when $\phi = 0.999$. It was noted in the previous sub-section that tail index estimates can be quite low even when EGB2 fits better than t, and so the fact that the increase in kurtosis can be very large for a Laplace distribution with persistent volatility is of some significance.

	Kurtosis	Increase in kurtosis, K					
κ	-	.03			.06		
ϕ	-	.98	.99	.999	.98	.99	.999
normal	3	1.05	1.10	2.35	1.24	1.54	43.38
Laplace	6	1.10	1.28	5.51	1.54	2.36	1881
t	6	1.25	1.34	1.64	1.74	2.69	8.52

Table 6 Increase in kurtosis induced by changing volatility

7 Conclusions and extensions

Most financial returns time series exhibit non-normal behavior, which is often modeled by a Student t distribution. This choice is strongly supported by tail index estimates, which almost invariably point to fat-tailed distributions. We argue here that the fat-tailed distributions are not always appropriate and that for many returns series a leptokurtic distribution which is light-tailed can give a better fit. The EGB2 distribution provides a bridge between the normal and t-distribution in that it exhibits excess kurtosis without having heavy tails. Unlike the general error distribution (apart from Laplace), it has a score function that is bounded when divided by the variable. This property corresponds to the gentle form of Winsorizing that is a feature of the EGB2 score for location. Both EGB2 and a modified version of the t-distribution are able to handle asymmetric distributions.

The EGB2 and Beta-t-EGARCH models were fitted to data on exchange rates and stock returns. For the exchange rates of developed countries, the evidence for fat-tails is unconvincing. On the whole the EGB2 fits better than the t, with the tail indices computed for both residuals and raw data being entirely consistent with the kind of values indicated by our simulations. The case for fat-tails in the distributions of developing country exchange rates is more persuasive. Similarly for most stock prices a t-distribution seems to fit better than EGB2.

The raw tail indices are very misleading when the conditional distribution is not fat-tailed. Even when the conditional distribution is best modeled by a Student's t, tail index estimates are typically much smaller than the degrees of freedom estimated by maximum likelihood, probably because of the increase in kurtosis which changing volatility induces. The low tail indices should be treated with caution if conclusions about the existence of moments are to be drawn. On a more positive note, they can be useful as an indicator of fat-tailed distributions with very small tail indices. Similarly they can provide sensible starting values for shape parameters in the EGB2 distribution, because these parameters are typically quite small.

In summary, while it is undeniable that low tail estimates are a feature of financial returns, we argue that this does not, in itself, provide strong evidence of fat-tailed distributions. Our findings lend support to the cautionary note sounded by Clauset et al (2009) on this matter. Placing too much store on nonparametric estimates, particulary from raw data is unwise.

Acknowledgement

We are grateful to Paul Kattuman for providing us with the exchange rate data.

REFERENCES

Asmussen, S. (2003). Applied Probability and Queues, Berlin, Springer Verlag.

Caivano, M and A.C. Harvey (2013). Time series models with an EGB2 conditional distribution, Cambridge Working paper in Economics, CWPE 1325.

Clauset, A., Shalizi, C. R. and M. E. J. Newman (2009). Power-Law Distributions in Empirical Data. SIAM Review, 51, 661–703.

Creal, D., Koopman, S.J., and A. Lucas (2013). Generalized autoregressive score models with applications. *J of Applied Econometrics* (to appear).

Creal, D., Koopman, S.J. and A. Lucas (2011). A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations, Journal of Business and Economic Statistics, 29, 552-63.

Embrechts, P., Kluppelberg, C. and T. Mikosch (1997). Modelling Extremal Events. Berlin: Springer Verlag.

Fernandez, C. and M. F. J. Steel (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* **99**. 359-371.

Gabaix, X. and R. Ibragimov (2011). Rank-1/2: a simple way to improve the OLS estimation of tail exponents. *Journal of Business and Economic Statistics,* 29, 24-39.

Harvey, A. C. (2013). *Dynamic models for Volatility and Heavy Tails.* Econometric Society Monograph. Cambridge University Press.

Harvey, A. C. and G. Sucarrat (2012). EGARCH Models with Fat Tails, Skewness and Leverage. Cambridge Working paper in Economics, CWPE 1236. Submitted.

Huisman, R., Koedijk, K.G. Kool, C.J.M. and F. Palm (2001). Tail-index estimates in small samples. *Journal of Business and Economic Statistics,* 19, 208-16.

Ibragimov, M., Ibragimov, R. and P. Kattuman. (2013). Emerging markets and heavy tails, *Journal of Banking & Finance,* 37, 2546-59.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences.* New York: Wiley.

Loretan, M. and P. Phillips (1994). Testing the covariance structure of heavy-tailed time series. *Journal of Empirical Finance*, 1, 211-48.

Maronna, R., Martin, D. and Yohai, V. (2006). *Robust Statistics: Theory and Methods*. John Wiley & Sons Ltd.

McDonald J. B. and Y. J. Xu (1995). A generalization of the beta distribution with applications. Journal of Econometrics 66: 133-152.

McNeil, A.J., Frey, R. and P. Embrechts (2005). Quantitative Risk Management. Princeton University Press, Princeton.

Mikosch, T. and C. Starica (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Staistics*, 28, 1427-51.

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

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347-370.

Prentice, R. L. (1975). Discrimination among some parametric models. *Biometrika*, 62, 607-614.

Resnick, S. and C. Starica (1995). Consistency of Hill's Estimator for Dependent Data. Journal of Applied Probability, 32, 139-167.

Wang, K-L, Fawson, C., Barrett, C.B. and J. B. McDonald.(2001). *Journal of Applied Econometrics*, 16, 521-536.