UNIVERSITY OF
CAMBRIDGE

Cambridge Working Papers in Economics

Volatility Modeling with a Generalized t-distribution

Andrew Harvey and Rutger-Jan Lange

CWPE 1517

Volatility Modeling with a Generalized t-distribution

Andrew Harvey and Rutger-Jan Lange
Faculty of Economics, Cambridge University

June 5, 2015

Abstract

Beta-t-EGARCH models in which the dynamics of the logarithm of scale are driven by the conditional score are known to exhibit attractive theoretical properties for the t-distribution and general error distribution (GED). The generalized-t includes both as special cases. We derive the information matrix for the generalized-t and show that, when parameterized with the inverse of the tail index, it remains positive definite as the tail index goes to infinity and the distribution becomes a GED. Hence it is possible to construct Lagrange multiplier tests of the null hypothesis of light tails against the alternative of fat tails. Analytic expressions may be obtained for the unconditional moments in the EGARCH model and the information matrix for the dynamic parameters obtained. The distribution may be extended by allowing for skewness and asymmetry in the shape parameters and the asymptotic theory for the associated EGARCH models may be correspondingly extended. For positive variables, the GB2 distribution may be parameterized so that it goes to the generalised gamma in the limit as the tail index goes to infinity. Again dynamic volatility may be introduced and properties of the model obtained. Overall the approach offers a unified, flexible, robust and practical treatment of dynamic scale.
Keywords: Asymmetry; dynamic conditional score (DCS) model; general error distribution; information matrix; partially adaptive estimation; robustness; tail index.

1 Introduction

The generalized Student t-distribution contains the general error distribution (GED), also known as the exponential power distribution (EPD), and the Student $t-$distribution as special cases. It was introduced by McDonald and Newey (1987), who proposed using it for static regression models, and it was subsequently employed by Theodossiou (1998) for financial data. The additional flexibility of the generalized-t enables it to capture a variety of shapes at the peak of the distribution as well as in the tails. This flexibility goes a long way towards meeting the objection that parametric models are too restrictive and hence vulnerable to misspecification. Indeed one of the concerns of McDonald and Newey was to meet these objections and they argued that the flexibility of the generalized-t model made it 'partially adaptive'. They highlighted the robustness of the generalized-t assumption for models of location (or more generally static regression), observing that, as with the t distribution, the score or influence function of the generalized-t is redescending (except in the limiting GED case) and so is resistant to outliers; see also McDonald and White (1993). Of course this is not the case for ordinary least squares; see any book on robustness, such as Maronna, Martin and Yohai (2006). Outliers are particulary prevalent in financial time series yet estimation by quasi-maximum likelihood (QML), that is based on a criterion function appropriate for a Gaussian distribution, is widespread in financial econometrics. As will become apparent shortly, the unsuitability of QML is further enhanced when modeling volatility because a dynamic equation for variance that is driven by squared observations is itself open to distortion by outliers. Thus unlike other researchers, such as Theodossiou (1998), Theodossiou and Savva (2014) and Zhu and Galbraith (2011), who have used GARCH models with generalized-t or t-distributions, we have the dynamics driven by the conditional score: hence the robustness inherent in the generalized-t extends to all parts of the model.

Models constructed using the conditional score were introduced into the literature by Creal, Koopman and Lucas (2011, 2013), where they are called Generalized Autoregressive Score (GAS) models, and Harvey (2013), where they are called Dynamic Conditional Score (DCS) models. The DCS EGARCH model is set up as

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

where the $\varepsilon_t' s$ are independently and identically distributed with location zero and unit scale. The stationary first-order dynamic model for $\lambda_{t|t-1}$, the logarithm of the scale, is

$$\lambda_{t+1|t} = \omega(1-\phi) + \phi\lambda_{t|t-1} + \kappa u_t, \quad |\phi| < 1, \tag{2}$$

where u_t is the score of the conditional distribution of y_t at time t, $\lambda_{1|0} = \omega$ and ϕ and κ are parameters, with ω denoting the unconditional mean. Letting the conditional distribution be Student's t leads to a model known as Beta-t-EGARCH. This model has now been widely applied and shown to be more attractive than the standard GARCH-t model both from the practical and theoretical points of view. The EGARCH formulation has the obvious attraction that scale remains positive and that stationarity conditions are straightforward - in (2) we simply require that $|\phi| < 1$. When combined with conditional score dynamics, the asymptotic theory for the ML estimators is relatively straightforward and other results, such as formuale for moments and autocorrelations, may also be derived. As regards the practical value of DCS models, there is already a good deal of evidence (in the references cited) to show that they tend to outperform standard models. Letting the conditional distribution be GED leads to the Gamma-GED-EGARCH model and sometimes this model fits better than Beta-t-EGARCH. The use of the generalized Student-t distribution gives what we will call Beta-Gen-t-EGARCH. This model has all the theoretical advantages of Beta-t-EGARCH, but it includes Gamma-GED-EGARCH as a limiting case.

A generalized t variable is defined such that its absolute value is distributed as a generalized beta of the second kind (GB2) with two of the shape parameters constrained, so that the mode is zero at the origin, and the scale re-parameterized so that it remains positive and finite for the limiting GED case. An expression for the information matrix may be obtained, but the re-paramerization makes it more complicated than for the GB2. We define the shape parameters, or more specifically the tail index, in such a way that the information matrix remains positive definite as the tail index goes to infinity and the distribution goes towards the GED. This is of some importance if the regularity conditions for a likelihood ratio test of the null hypothesis of thin tails against the alternative of fat tails[1] are to be satisfied. The availability of a limiting positive definite information matrix also allows LM tests against fat tails to be constructed.

Further flexibility in the generalized t can be achieved by introducing skewness and/or creating asymmetry by allowing the shape parameters to be different on either side of the location parameter. The large sample sizes often encountered in financial econometrics make such generalizations a practical proposition[2]. The asymmetry is handled as in Zhu and Zinde-Walsh (2009) and Zhu and Galbraith (2010), where it is introduced into GED and Student-t distributions respectively. These asymmetric distributions therefore emerge as special cases.

The additional flexibility of the distribution has a potential cost in that it may become more difficult to estimate unknown parameters, particularly shape

[1] The following terminology will be adopted. Heavy tails means heavier than exponential tails; all common heavy-tailed distributions are technically sub-exponential; see Embrechts et al (1997, p 49-57). Light tails means exponential tails or more generally super-exponential

Fat tails or power law have the survival function, $\overline{F}(y) = \Pr(Y > y) \sim y^{-\eta}$, where $\eta > 0$ is the tail index. All fat tailed distributions are heavy-tailed, but not the converse.

Thin tailed distributions have smaller kurtosis than a normal distribution. Thick tailed distributions have higher kurtosis than normal, ie excess kurtosis (*leptokurtic).*

[2] Asymmetric distributions have also been fitted to large cross-sectional date sets; a recent example is Holly et al (2013).

parameters. The extent to which this is a problem can be partially determined by examining the correlations between score variables and the way in which this carries over to standard errors. These correlations are available from the information matrix and so having an explicit expressions gives considerable insight. For example, estimating a skew parameter has a relatively small effect on other shape parameters. Notwithstanding such considerations, it could be argued that the estimation of models with several shape parmeters is unreasonably challenging at the best of times. However, for the relatively large sample sizes often encountered in financial econometrics, the estimation of flexible models seems to be feasible.

An exact expression for the information matrix of the dynamic parameters in a Beta-Gen-t-EGARCH model with a first-order dynamic equation can be constructed in much the same way as was done for the Beta-t-EGARCH and GB2 location-scale models in Harvey (2013, sub-sections 4.6 and 5.3). The results extend to skew and/or asymmetric distributions. When simpler models within this class are fitted, Lagrange multiplier (LM) tests can play a valuable diagnostic role and provide an indication as to whether generalizations to more complex specifications are necessary. Much of the theory carries over to ARCH in mean and multivariate models.

The GB2 distribution, which is for positive variables, can be re-parameterized in the same way as the generalized-t so that it goes to a generalized gamma (GG) distribution as the tail index goes to infinity. Thus the theory for dynamic location-scale models with GB2 and GG distributions can also be unified.

The plan of the paper is as follows. Section 2 sets out the theory for generalized-t and derives the limiting score and information matrix as the tail index goes to infinity. The EGARCH model is discussed in Section 3. The generalized-t is extended to include skewness and/or asymmetry in Section 4 and the GB2 theory is set out in Section 5.

2 Generalized Student t-distribution

The generalized t-distribution has been formulated in a number of ways. Our preferred parameterization sets one of the shape parameters equal to the tail index and so the probability density function (PDF) is written as

$$f(y) = \frac{1}{\varphi}K(\eta,\upsilon)\left(1+\frac{1}{\eta}\left|\frac{y-\mu}{\varphi}\right|^{\upsilon}\right)^{-(\eta+1)/\upsilon}, \quad -\infty < y < \infty, \tag{3}$$

where φ is a scale parameter, υ and η are positive shape parameters and

$$K(\eta,\upsilon) = \frac{\upsilon}{2\eta^{1/\upsilon}}\frac{1}{B(\eta/\upsilon,1/\upsilon)} \tag{4}$$

with $B(.,.)$ denoting the beta function. Theodossiou and Savva (2014) also use the tail index, η, but other aspects of their parametrization are slightly different. The range of η is $0 < \eta \leq \infty$, so if we define $\overline{\eta} = 1/\eta$, then $0 \leq \overline{\eta} \leq 1$ for $1 \leq \eta \leq \infty$. The range of υ is $0 < \upsilon < \infty$ but in practice $\upsilon < 1$ (which has heavy, but not fat, tails) is unlikely; see the comments at the end of sub-section 2.2. Note that $f'(0) = 0$ for $\upsilon > 1$, but that the derivative is not continuous at $\upsilon = 1$ and for $\upsilon < 1$, it becomes infinite.

Setting $\upsilon = 2$ gives t_η, a t-distribution with with η degrees of freedom, whereas $GED(\upsilon)$ is obtained when $\eta \to \infty$. The Laplace or double exponential distribution is $GED(1)$, whereas $GED(2)$ is the normal. The form of the GED is as in Zhu and Zinde-Walsh (2009), that is

$$f(y) = \frac{1}{2\varphi\upsilon^{1/\upsilon}\Gamma\left(1+1/\upsilon\right)} \exp\left[-\frac{1}{\upsilon}\left|\frac{y-\mu}{\varphi}\right|^{\upsilon}\right] = \frac{\upsilon^{1-1/\upsilon}}{2\varphi\Gamma\left(1/\upsilon\right)} \exp\left[-\frac{1}{\upsilon}\left|\frac{y-\mu}{\varphi}\right|^{\upsilon}\right]. \tag{5}$$

The standard deviation is $\upsilon^{1/\upsilon}(\Gamma(3/\upsilon)/\Gamma(1/\upsilon))^{1/2}\varphi$. The GED can be seen to be a special case of generalized-t because, as $\eta \to \infty$, using a result in Davis (1964, p. 257) gives

$$K(\nu,\upsilon) = \frac{\upsilon}{2\eta^{1/\upsilon}} \frac{\Gamma(\eta/\upsilon + 1/\upsilon)}{\Gamma(\eta/\upsilon)\Gamma(1/\upsilon)} \to \frac{\upsilon}{2\,\upsilon^{1/\upsilon}\Gamma(1/\upsilon)} \tag{6}$$

while

$$\left(1+\frac{1}{\eta}\left|\frac{y-\mu}{\varphi}\right|^{\upsilon}\right)^{-(\eta+1)/\upsilon} \to \exp\left[-\frac{1}{\upsilon}\left|\frac{y-\mu}{\varphi}\right|^{\upsilon}\right].$$

The extra shape parameter, υ, in the generalized t introduces more flexibility into the distribution, particularly at the peak[3]. The distribution has fat tails for finite η and so moments only exist up to, but not including, η. The absolute moments are

$$E[|y-\mu|^m] = \frac{\Gamma(\frac{1}{\upsilon}+\frac{m}{\upsilon})\Gamma(\frac{\eta}{\upsilon}-\frac{m}{\upsilon})}{\Gamma(\frac{1}{\upsilon})\Gamma(\frac{\eta}{\upsilon})}\eta^{\frac{m}{\upsilon}}\varphi^m, \quad 0 \leq m < \eta.$$

For a t-distribution setting $m = 2$ gives $Var(y) = \varphi^2\eta/(\eta-2)$, $\eta > 2$, whereas for double generalized Pareto,

$$Var(y) = \frac{2\eta^2}{(\eta-1)(\eta-2)}\varphi^2, \quad \eta > 2.$$

[3]For a given tail index and standard deviation, the distribution is more peaked for lower υ whereas, when υ is fixed, the peak becomes higher as the tail index decreases.

As we shall see, many of the properties of the distribution, including the asymptotic distribution of the ML estimators, depend on the fact that

$$b_t = \frac{(|y_t - \mu| /\varphi)^\upsilon/\eta}{(|y_t - \mu| /\varphi)^\upsilon/\eta + 1} = \frac{|\varepsilon_t|^\upsilon/\eta}{|\varepsilon_t|^\upsilon/\eta + 1}, \qquad 0 < \ \eta < \infty,$$

is distributed as $beta(1/\upsilon, \eta/\upsilon)$ at the true parameter values.

The CDF is a function of the CDF of a beta distribution (an incomplete beta function) for $x_t = 1 - b_t$, which is $beta(\eta/\upsilon, 1/\upsilon)$. The corresponding quantile function is readily available. This is relevant for Value at Risk (VaR) and Expected Shortfall (ES), as discussed in Zhu and Galbraith (2011, pp 768-70) and Zhu (2012).

2.1 Score functions

Because one of our primary aims is to model changing volatility, scale will be parameterized using an exponential link function, that is $\varphi = \exp(\lambda)$, where $-\infty < \lambda < \infty$. The score functions for location and (the logarithm of) scale are:

$$\begin{aligned}\frac{\partial \ln f}{\partial \mu} &= (\eta+1)e^{-\lambda}\frac{(|y_t - \mu|\, e^{-\lambda})^{\upsilon-1}/\eta}{(|y_t - \mu|\, e^{-\lambda})^{\upsilon}/\eta + 1} sgn(y_t - \mu) \qquad (7)\\ &= \frac{\eta+1}{\eta e^{\lambda}}(1 - b_t)\,|\varepsilon_t|^{\upsilon-1}.sgn(y_t - \mu)\end{aligned}$$

and

$$\frac{\partial \ln f_t}{\partial \lambda} = (\eta+1)b_t - 1. \qquad (8)$$

The robustness properties of the score function - or influence function - of location are highlighted by McDonald and Newey (1988). Provided η is finite, it is redescending in that it approaches zero as y moves away from zero. The score function for scale has corresponding robustness features in that it is bounded[4].

The score for the tail index parameter is

$$\begin{aligned}\frac{\partial \ln f}{\partial \eta} &= -\frac{1}{\upsilon}\psi(\eta/\upsilon) + \frac{1}{\upsilon}\psi(\eta/\upsilon + 1/\upsilon) - \frac{1}{\upsilon\eta} - \frac{1}{\upsilon}\ln((|y-\mu|/\varphi)^\upsilon/\eta + 1)\\ &\quad + \frac{\eta+1}{\upsilon\eta}\frac{(|y-\mu|/\varphi)^\upsilon/\eta}{(|y-\mu|/\varphi)^\upsilon/\eta + 1}\\ &= -\frac{1}{\upsilon}\psi(\eta/\upsilon) + \frac{1}{\upsilon}\psi(\eta/\upsilon + 1/\upsilon) - \frac{1}{\upsilon\eta} + \frac{1}{\upsilon}\ln(1 - b_t) + \frac{\eta+1}{\upsilon\eta}b_t, \qquad (9)\end{aligned}$$

[4]Note the general relationship between the location score and the score for the logarithm of scale, namely $\partial \ln f/\partial\lambda = (y-\mu)\partial \ln f/\partial\mu - 1$.

where $\psi(.)$ is the digamma function. For the purposes of taking the limit as $\eta \to \infty$, it is more convenient to work with the reciprocal, $\overline{\eta}$. The score is then

$$\frac{\partial \ln f}{\partial \overline{\eta}} = \frac{\psi(1/\overline{\eta}\upsilon) - \psi(1/\overline{\eta}\upsilon + 1/\upsilon) + \overline{\eta} - \ln(1-b_t) - (\overline{\eta}+1)b_t}{\upsilon \overline{\eta}^2}. \tag{10}$$

As regards the other shape parameter, we find

$$\begin{aligned}\frac{\partial \ln f}{\partial \upsilon} &= \frac{\overline{\eta}\,\upsilon - \ln[1-b_t] - (\overline{\eta}+1)\,b_t\,\ln b_t + \{b_t - \overline{\eta}\,(1-b_t)\}\,\ln[\overline{\eta}\,(1-b_t)]}{\overline{\eta}\,\upsilon^2} \\ &\quad + \frac{\overline{\eta}\,\psi(1/\upsilon) + \psi(1/\overline{\eta}\upsilon) - (\overline{\eta}+1)\psi(1/\overline{\eta}\upsilon + 1/\upsilon)}{\overline{\eta}\,\upsilon^2}\end{aligned}$$

The score function for υ, like that of λ but unlike that of η (or $\overline{\eta}$), is bounded, because as $y \to \pm\infty$, $b_t \to 1$ so

$$\lim_{y_t \to \pm\infty} \Big[-\ln[1-b_t] - (\overline{\eta}+1)\,b_t\,\ln b_t + \{b_t - \overline{\eta}\,(1-b_t)\}\,\ln[\overline{\eta}\,(1-b_t)] \Big] = \ln \overline{\eta}. \tag{11}$$

Also, $b_t = 0$ when $y_t = 0$, so in this case

$$\Big[-\ln[1-b_t] - (\overline{\eta}+1)\,b_t\,\ln b_t + \{b_t - \overline{\eta}\,(1-b_t)\}\,\ln[\overline{\eta}\,(1-b_t)] \Big] = -\overline{\eta}\ln \overline{\eta}.$$

It can be verified from the properties of a beta variable given in Appendix A that the above scores all have zero expectation.

2.2 Information matrix

The static information matrix for the GB2 distribution was derived by Brazauskasin (2002). It can be found in Kleiber and Kotz (2003, p.194). Adapting it to the exponential link function for scale gives the expression in Harvey (2013, p 164). The information matrix for the generalized-t is more complicated because of the re-parameterization[5]. Being able to construct it gives important insight into identifiability and parameter transformations. The derivation is discussed in Appendix A.

[5] The information matrix in Zhu (2012) is slightly simpler because he re-parameterizes the scale by setting it to $\varphi^*/K(\upsilon)$, where $K(\upsilon)$ is given in (6). But if this is done another transformation is needed to get the asymptotic covariance matrix for the ML estimators of the original set of parameters; see also Zhu and Galbraith (2010, p 300) and our Appendix D. Furthermore Zhu (2012) does not parameterize in terms of $\overline{\eta}$ and this is crucial for obtaining the limiting matrix, including the terms involving $\overline{\eta}$, as the tail index goes to infinity.

Proposition 1 *The information matrix for the parameters of the generalized-t distribution, (3), with finite tail index, that is $\overline{\eta} > 0$, takes the form*

$$\mathbf{I}\begin{pmatrix} \mu \\ \lambda \\ \upsilon \\ \overline{\eta} \end{pmatrix} = \begin{pmatrix} I_{\mu\mu} & 0 & 0 & 0 \\ 0 & I_{\lambda\lambda} & I_{\lambda\upsilon} & I_{\lambda\overline{\eta}} \\ 0 & I_{\lambda\upsilon} & I_{\upsilon\upsilon} & I_{\upsilon\overline{\eta}} \\ 0 & I_{\lambda\overline{\eta}} & I_{\upsilon\overline{\eta}} & I_{\overline{\eta}\overline{\eta}} \end{pmatrix} \tag{12}$$

with diagonal elements

$$\begin{aligned}
I_{\mu\mu} &= \frac{(\overline{\eta}+1)\,\upsilon^3\,\overline{\eta}^{2/\upsilon}}{(\overline{\eta}\,\upsilon+\overline{\eta}+1)\,\exp(2\lambda)}\,\frac{\Gamma(2-1/\upsilon)\,\Gamma(2/\upsilon+1/\overline{\eta}\upsilon)}{\Gamma(1/\upsilon)\,\Gamma(1/\overline{\eta}\upsilon)}, \qquad \upsilon > 1/2 \\
I_{\lambda\lambda} &= \frac{\upsilon}{\overline{\eta}\,\upsilon+\overline{\eta}+1} \\
I_{\upsilon\upsilon} &= \frac{\left(\ln\overline{\eta}+(\overline{\eta}-1)\,\upsilon+\psi\left[\frac{1}{\overline{\eta}\upsilon}\right]-\psi\left[\frac{1}{\upsilon}\right]\right)^2\psi'\left[\frac{1}{\overline{\eta}\upsilon}\right]+\psi'\left[\frac{1}{\upsilon}\right]-(1+\overline{\eta}^2)\,\upsilon^2}{\upsilon^3\,(\overline{\eta}\,\upsilon+\overline{\eta}+1)} \\
&\quad +\frac{\overline{\eta}^2\,\psi'\left[\frac{1}{\upsilon}\right]+\psi'\left[\frac{1}{\overline{\eta}\upsilon}\right]-(1+\overline{\eta})^2\,\psi'\left[\frac{\overline{\eta}+1}{\overline{\eta}\,\upsilon}\right]}{\overline{\eta}^2\,\upsilon^4}-\frac{1}{\upsilon^2} \\
I_{\overline{\eta}\overline{\eta}} &= \frac{\psi'(1/\overline{\eta}\upsilon)-\psi'(1/\overline{\eta}\upsilon+1/\upsilon)}{\overline{\eta}^4\,\upsilon^2}-\frac{2\,\overline{\eta}\,\upsilon+\overline{\eta}+1}{\upsilon\,\overline{\eta}^2\,(\overline{\eta}+1)\,(\overline{\eta}\,\upsilon+\overline{\eta}+1)},
\end{aligned}$$

where $\psi'(.)$ is the trigamma function. The off-diagonal elements are

$$I_{\lambda\upsilon} = \frac{\ln\overline{\eta}+\psi\left[1+\frac{1}{\overline{\eta}\upsilon}\right]-\psi\left[1+\frac{1}{\upsilon}\right]}{\upsilon\,(\overline{\eta}\,\upsilon+\overline{\eta}+1)}, \qquad I_{\lambda\overline{\eta}} = \frac{\upsilon}{(\overline{\eta}+1)\,(\overline{\eta}\,\upsilon+\overline{\eta}+1)}$$

$$I_{\upsilon\overline{\eta}} = \frac{\ln\overline{\eta}-\overline{\eta}-1+\psi\left[\frac{1}{\overline{\eta}\upsilon}\right]-\psi\left[1+\frac{1}{\upsilon}\right]}{\upsilon\,(\overline{\eta}+1)\,(\overline{\eta}\,\upsilon+\overline{\eta}+1)}+\frac{\psi'\left[\frac{1}{\overline{\eta}\upsilon}\right]-(1+\overline{\eta})\psi'\left[\frac{\overline{\eta}+1}{\overline{\eta}\,\upsilon}\right]}{\overline{\eta}^3\,\upsilon^3}$$

Remark 1 *For a t-distribution with seven degrees of freedom, that is $\upsilon = 2$ and $\overline{\eta} = 1/7$, the score for the scale is positively correlated with $\overline{\eta}$ and negatively correlated with υ. Further, there is a trade-off between υ and $\overline{\eta}$ in that their scores have a correlation that is large and negative, specifically* -0.915. *On inverting the information matrix, the correlation between the estimates of υ and $\overline{\eta}$ is found to be* 0.842 . *This strong positive correlation means that a higher value of $\overline{\eta}$, corresponding to a heavier tail, will induce a higher value of υ, meaning a lighter tail. The other correlations are smaller: between the estimates of λ and υ it is* 0.124 *and between λ and $\overline{\eta}$ it is* -0.266 . *Estimating υ actually reduces the correlation between λ and $\overline{\eta}$, because when it is fixed the correlation is* -0.693.

When υ is known to be two, the information matrix for the t-distribution, that is

$$\mathbf{I}\begin{pmatrix} \mu \\ \lambda \\ \overline{\eta} \end{pmatrix} = \begin{pmatrix} \frac{e^{-2\lambda}(\overline{\eta}+1)}{3\overline{\eta}+1} & 0 & 0 \\ 0 & \frac{2}{3\overline{\eta}+1} & \frac{2}{(\overline{\eta}+1)(3\overline{\eta}+1)} \\ 0 & \frac{2}{(\overline{\eta}+1)(3\overline{\eta}+1)} & \frac{\psi'\left[\frac{1}{2\overline{\eta}}\right]-\psi'\left[\frac{\overline{\eta}+1}{2\overline{\eta}}\right]}{4\overline{\eta}^4} - \frac{(5\overline{\eta}+1)}{2\overline{\eta}^2(\overline{\eta}+1)(3\overline{\eta}+1)} \end{pmatrix}$$

is obtained. Somewhat less well-known is the information matrix for the double generalized Pareto, obtained by setting $\upsilon = 1$. Specifically

$$\mathbf{I}\begin{pmatrix} \mu \\ \lambda \\ \overline{\eta} \end{pmatrix} = \begin{pmatrix} \frac{e^{-2\lambda}(\overline{\eta}+1)^2}{2\overline{\eta}+1} & 0 & 0 \\ 0 & \frac{1}{(2.0\overline{\eta}+1.0)} & \frac{1}{(2.0\overline{\eta}+1.0)(\overline{\eta}+1.0)} \\ 0 & \frac{1}{(2.0\overline{\eta}+1.0)(\overline{\eta}+1.0)} & \frac{2}{2\overline{\eta}^2+3\overline{\eta}+1} \end{pmatrix}. \tag{13}$$

The simplification in $I_{\overline{\eta}\overline{\eta}}$ comes about because $\psi'(1/\overline{\eta}) - \psi'(1+1/\overline{\eta}) = (1/\overline{\eta})^{-2} = \overline{\eta}^2$.

Remark 2 *The asymptotic distribution theory for ML estimation is not standard for $\upsilon < 2$ because of the singularies in the derivatives of the log-density at μ. In particular the second derivative with respect to μ does not exist at $y = \mu$ and its expectation cannot be found; $I_{\mu\mu}$ in (12) is obtained from the expectation of the square of the first derivative. Neverthless McDonald and Newey (1987) are able to show consistency and asymptotic normality provided $\upsilon > 1$. The problems are essentially the same as those that arise for the GED; see Zhu and Zinde-Walsh (2009, p 91). If μ is known, the usual asymptotics hold for the scale and shape parameter of the GED and this will remain true for the generalized-t. The ML estimators of these parameters are consistent and asymptotically normal when μ is estimated (consistently) by the median. If μ is estimated by ML, the block diagonality of the information matrix suggests that inference for the scale and shape parameters will still be valid for any $\upsilon > 0$. This is borne out by a Monte Carlo study for the GED reported in Bottazzi and Secchi (2011, p 1002-6). They also demonstrate that the ML estimator of μ will be asymptotically efficient for $\upsilon > 1/2$, the condition required to ensure that $I_{\mu\mu}$ exists.*

2.3 General error distribution as a limiting case

As already noted, in the limit $\eta \to \infty$, that is $\overline{\eta} \to 0$, the generalized-t distribution becomes a GED, (5). The limiting scores may similarly be obtained from those in sub-section 2.1 by letting $\overline{\eta} \to 0$. Because $b_t \to 0$ as $\overline{\eta} \to 0$, whereas $b_t/\overline{\eta} \to |\varepsilon_t|^{\upsilon}$, it is easy to see that the limiting scores for location and scale are given by

$$\lim_{\overline{\eta}\to 0}\frac{\partial \ln f_t}{\partial \mu} = e^{-\lambda}\left|\varepsilon_t\right|^{\upsilon-1} sgn(\varepsilon_t) \quad \text{and} \quad \lim_{\overline{\eta}\to 0}\frac{\partial \ln f_t}{\partial \lambda} = \left|\varepsilon_t\right|^{\upsilon} - 1 \tag{14}$$

The limiting scores for the shape parameters are more difficult to derive, but it is shown in appendix B that the problem can be solved by using a known expansion for the digamma function, together with Taylor series expansions. As a result

$$\lim_{\bar{\eta}\to 0}\frac{\partial \ln f_t}{\partial \upsilon} = \frac{|\varepsilon_t|^{\upsilon} - \upsilon\,|\varepsilon_t|^{\upsilon}\ln|\varepsilon_t| + \upsilon + \ln(\upsilon) + \psi(1/\upsilon) - 1}{\upsilon^2} \tag{15}$$

and

$$\lim_{\bar{\eta}\to 0}\frac{\partial \ln f_t}{\partial \bar{\eta}} = \frac{\left(|\varepsilon_t|^{\upsilon} - 1\right)^2}{2\upsilon} - \frac{1}{2}. \tag{16}$$

At the true parameter values, $|\varepsilon_t|^{\upsilon}$ is distributed as $gamma(\upsilon, 1/\upsilon)$ and so it is possible to check that all the scores have zero expectation. The main point to note concerns $E\,|\varepsilon_t|^{\upsilon}\ln|\varepsilon_t|^{\upsilon}$ which is evaluated by taking the expectation of $\ln|\varepsilon_t|^{\upsilon}$ with respect to a $gamma(\upsilon, 1+1/\upsilon)$ variable, giving $\psi(1+1/\upsilon) + \ln\upsilon$.

When υ is known to be one, the information matrix is obtained from (13) simply by setting $\bar{\eta} = 0$. For other values of υ, $I_{\lambda\bar{\eta}}$ and $I_{\lambda\lambda}$ are immediately seen to be υ, but finding the limit of $I_{\bar{\eta}\bar{\eta}}$ is less straightforward, even for $\upsilon = 2$. It is shown in appendix C that:

$$\lim_{\bar{\eta}\to 0}\mathbf{I}\begin{pmatrix}\lambda\\ \bar{\eta}\end{pmatrix} = \begin{pmatrix}\upsilon & \upsilon\\ \upsilon & \frac{3\upsilon+1}{2}\end{pmatrix},$$

so

$$\left(\lim_{\bar{\eta}\to 0}\mathbf{I}\right)^{-1} = \begin{pmatrix}\frac{3\upsilon+1}{\upsilon(\upsilon+1)} & -\frac{2}{\upsilon+1}\\ -\frac{2}{\upsilon+1} & \frac{2}{\upsilon+1}\end{pmatrix}.$$

Thus fitting a t-distribution when the true distribution is Gaussian more than doubles the variance of the estimator of λ.

We also note that

$$\lim_{\bar{\eta}\to 0} I_{\mu,\mu} = \frac{\upsilon^{2-2/\upsilon}}{\exp(2\lambda)}\,\frac{\Gamma\left[2 - 1/\upsilon\right]}{\Gamma\left[1/\upsilon\right]} \tag{17}$$

because, from Davis (1964, p. 257),

$$\lim_{\bar{\eta}\to 0}\frac{\Gamma\left(\frac{1}{\upsilon\bar{\eta}} + \frac{2}{\upsilon}\right)}{\left(\frac{1}{\upsilon\bar{\eta}}\right)^{2/\upsilon}\Gamma\left(\frac{1}{\upsilon\bar{\eta}}\right)} = 1.$$

We have not been able to find these limiting expressions in the literature. Although this is somewhat surprising, the lack of results may be due to the complicated nature of trigamma functions and their limits. The existence of these limits is of some importance, however, if the regularity conditions for a likelihood ratio test of the null hypothesis of thin (or exponential) tails are to be satisfied. When the generalized-t distribution is defined in terms of η rather than $\bar{\eta}$, all elements $I_{(\eta,\cdot)} \to 0$ as $\eta \to \infty$ and so the information matrix is singular. The availability of a p.d. information matrix for $\overline{\eta} = 0$ also allows LM tests against fat tails to be constructed.

When υ is estimated, the information matrix[6] of the generalized-t for λ, υ and η, goes to the following limit as $\overline{\eta} \to 0$:

$$\mathbf{I}\begin{pmatrix} \lambda \\ \upsilon \\ \overline{\eta} \end{pmatrix} = \begin{pmatrix} \upsilon & -\frac{\ln\upsilon+\psi\left[1+\frac{1}{\upsilon}\right]}{\upsilon} & \upsilon \\ -\frac{\ln\upsilon+\psi\left[1+\frac{1}{\upsilon}\right]}{\upsilon} & \frac{\left(\upsilon+\ln\upsilon+\psi\left[\frac{1}{\upsilon}\right]\right)^2+\left(\frac{1}{\upsilon}+1\right)\psi'\left[\frac{1}{\upsilon}\right]-\left(\upsilon^2+\upsilon+1\right)}{\upsilon^3} & -\frac{\ln\upsilon+\psi\left[1+\frac{1}{\upsilon}\right]+\frac{1}{2}}{\upsilon} \\ \upsilon & -\frac{\ln\upsilon+\psi\left[1+\frac{1}{\upsilon}\right]+\frac{1}{2}}{\upsilon} & \frac{3\upsilon+1}{2} \end{pmatrix}. \tag{18}$$

The information quantity for μ remains as in (17). The top-left 2×2 submatrix in (18) is the information matrix of the GED. Because $\bar{\eta}$ is estimated, the elements in the bottom row have the effect of increasing the asymptotic standard errors of the estimators of λ and υ. The derivation of the limits of $I_{\upsilon\upsilon}$ and $I_{\upsilon\bar{\eta}}$ is similar to that in Appendix C for $I_{\overline{\eta}\overline{\eta}}$.

Remark 3 *The use of $\bar{\eta}$ rather than η has the practical value of ensuring that ML estimation is stable when the tail index is large.*

2.4 Testing against fat tails

When η is finite the tails of the generalized-t are fat, but this is no longer the case for the limiting GED distribution[7]. Thus, within the generalized-t framework, a test of the null hypothesis that $\eta = \infty$, or equivalently $\overline{\eta} = 0$, against the alternative of $\eta < \infty$, that is $\overline{\eta} > 0$, is a test against fat tails. As has been shown, the regularity condition that the information matrix be positive definite under the null is satisfied when the inverse tail parameterization is used.

A LR test of $\overline{\eta} = 0$ is straightforward to implement but because the alternative, $\overline{\eta} > 0$, is one-sided, the asymptotic distribution of the LR test statistic is a mixture of a χ^2_1 and a degenerate distribution with its mass at the origin. Hence the 5% critical value is the usual 10% one, that is 2.71.

A Lagrange multiplier, or score, test can be implemented using the positive definite information matrix derived in the previous sub-section. The limiting score

[6] The information matrix in Zhu and Zinde-Walsh (2009) is somewhat simpler because they re-parameterize the scale; see the earlier footnote in sub-section 2.2..

[7] The GED has light tails when $\upsilon \geq 1$, but when $\upsilon < 1$ the tails are heavy. However, even for $\upsilon < 1$, they are still not fat; see footnote 1.

for $\overline{\eta}$ remains well-defined and was given in (16). Thus the LM statistic is

$$\begin{aligned} LM &= \frac{(I^{-1})_{\overline{\eta}\overline{\eta}}}{T}\left(\sum \frac{(|\varepsilon_t|^{\upsilon}-1)^2}{2\upsilon}-\frac{1}{2}\right)^2 \qquad (19)\\ &= \frac{(I^{-1})_{\overline{\eta}\overline{\eta}}}{T}\left(\sum \frac{|\varepsilon_t|^{2\upsilon}}{\upsilon}-\frac{2}{\upsilon}|\varepsilon_t|^{\upsilon}+\frac{1}{\upsilon}-1\right)^2, \end{aligned}$$

where $(I^{-1})_{\overline{\eta}\overline{\eta}}$ is the diagonal element in the inverse information matrix corresponding to $\overline{\eta}$ and the parameters μ, υ and λ are replaced by their ML estimators under the null hypothesis that $\overline{\eta}=0$. If υ is known, $(I^{-1})_{\overline{\eta}\overline{\eta}}=2/(\upsilon+1)$. The LM statistic is asymptotically χ^2_1-distributed under the null. For $\upsilon=2$, the test is simply a reformulation of the standard excess kurtosis test. This is no longer the case for $\upsilon\neq 2$. The contrast[8] is then between the moments of $|\varepsilon_t|^{2\upsilon}$ and $|\varepsilon_t|^{\upsilon}$.

Tests of the null hypothesis that υ takes a specific value, υ_0, may also be carried out. The LR statistic of $\upsilon=\upsilon_0$ against $\upsilon\neq\upsilon_0$ is asymptotically χ^2_1, as is the LM statistic. A joint test of $\upsilon=\upsilon_0$ and $\overline{\eta}_0=0$ against $\upsilon\neq\upsilon_0$ and $\overline{\eta}>0$ is also possible. The LM test statistic, which is asymptotically distributed as χ^2_2 under the null, requires the limiting score for υ, as given in (15).

3 Dynamic Scale

The Beta-Gen-t-EGARCH model has time-varying scale, $\exp(\lambda_{t|t-1})$, with the dynamic equation, (2), driven by the score[9]

$$\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} = u_t = (\eta+1)\, b_t - 1, \qquad (20)$$

where

$$b_t = \frac{(|y_t-\mu|\, e^{-\lambda_{t|t-1}})^{\upsilon}/\eta}{(|y_t-\mu|\, e^{-\lambda_{t|t-1}})^{\upsilon}/\eta+1}. \qquad (21)$$

As in the static model b_t is distributed as $beta(1/\upsilon,\eta/\upsilon)$ at the true parameter values. Setting $\upsilon=2$ gives the Beta-t-EGARCH model, with the score as in Harvey (2013, ch 4).

[8] Under the alternative hypothsis, $|\varepsilon_t|^{\upsilon}=\eta b_t/(1-b_t)$ and it can be shown that $E|\varepsilon_t|^{2\upsilon}-2E|\varepsilon_t|^{\upsilon}$ increases as η decreases.

[9] Bollerslev, Engle and Nelson (1994, p 3017-23) propose a related model in which the υ and η parameters in an expression similar in form to u_t are different from the υ and η shape parameters in the conditional distribution. We are grateful to Guiseppe Cavaliere for pointing this out to us. The viability of a model in which υ and η in u_t are different from the υ and η in the conditional distribution is a matter for future investigation.

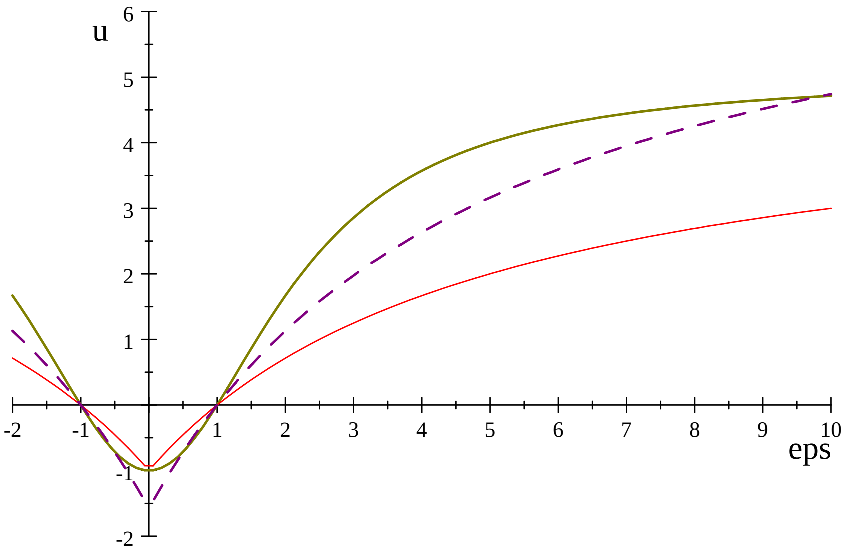

Figure 1: Score functions, u, for $\eta = 5$ and $\upsilon = 2$ (solid) and $\upsilon = 1$ (solid thin) plotted against standardized variable, ε. Dashed line shows $\upsilon = 1$ multiplied by the ratio of its SD to that of the $\upsilon = 2$ variable

Figure 1 shows the score (influence) functions - or, as they are sometimes called in the volatility literature, news impact curves - for $\eta = 5$ and $\upsilon = 1$ or 2. As $|y_t| \to \infty$, $u_t \to \eta$, so the score is bounded for finite η. A bounded score means that the existence of moments is not affected by volatility, although they may become hugely inflated.

3.1 Moments

Because b_t in the score variable, (20), has a beta distribution, Theorem 7 in Harvey (2013, ch 4) generalizes immediately to give exact expressions for the unconditional moments. To be specific, the expectations of powers of absolute values of the observations in the stationary Beta-Gen-t-EGARCH model are

$$E\left(|y_t|^c\right) = \frac{\eta^{c/\upsilon}\Gamma(\frac{c}{\upsilon}+\frac{1}{\upsilon})\Gamma(\frac{-c}{\upsilon}+\frac{\eta}{\upsilon})}{\Gamma(\frac{1}{\upsilon})\Gamma(\frac{\eta}{\upsilon})}\zeta(c;\boldsymbol{\psi}), \quad -1 < c < \eta, \tag{22}$$

where

$$\zeta(c) = e^{c\omega}\prod_{j=1}^{\infty} e^{-\psi_j c}\beta_{\eta,\upsilon}(\psi_j c)$$

with $\psi_j, j = 1, 2, ..,$ denoting the coefficient of u_{t-j} when $\lambda_{t|t-1}$ is expressed in terms of past scores, ω denoting the unconditional expectation of $\lambda_{t|t-1}$ and $\beta_{\eta,\upsilon}(a)$ being

Kummer's (confluent hypergeometric) function, ${}_1F_1(1/\upsilon;\eta/2+1/\upsilon;a(\eta+1))$, that is

$$\beta_{\eta,\upsilon}(a)=1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{1+\upsilon r}{\eta+1+\upsilon r}\right)\frac{a^k(\eta+1)^k}{k!},\qquad 0<\eta<\infty.$$

Conditional moments may be similarly obtained for multi-step forecasts. The full multi-step distribution is easily simulated from beta variates.

3.2 Maximum likelihood estimation

The proposition below for the ML estimator,$\widetilde{\boldsymbol{\psi}}$, of $\boldsymbol{\psi}=(\kappa\ \phi\ \omega)'$ in the first-order stationary dynamic model for volatilty, (2), follows from the fact that the score and its derivative,

$$\frac{\partial u_t}{\partial\lambda_{t|t-1}}=-\upsilon(\eta+1)b_t(1-b_t)$$

depend on the $beta(1/\upsilon,\eta/\upsilon)$ variable, b_t; see Appendix A.

Proposition 2 *For a conditional generalized t-distribution with given μ,υ and $\eta<\infty$, and a first-order stationary dynamic model for volatilty, with $\kappa\neq 0$, the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}}-\boldsymbol{\psi}_0)$, where $\boldsymbol{\psi}_0$ denotes the vector of true values of κ, ϕ and ω, is multivariate normal with mean vector zero and a covariance matrix given bythe inverse of the single observation information matrix*

$$\mathbf{I}(\boldsymbol{\psi})=I_{\lambda\lambda}\mathbf{D} \tag{23}$$

where $I_{\lambda\lambda}$ is as in (12) and $\mathbf{D}$ is as defined in Harvey (2013, p. 37) - see Appendix E - with the quantities a,b and c, evaluated from

$$E\left[\frac{\partial u_t}{\partial\lambda}\right]=\frac{-\upsilon\eta}{\eta+\upsilon+1}=\frac{-\upsilon}{\overline{\eta}+1+\upsilon\overline{\eta}} \tag{24}$$

$$\begin{aligned}E\left[\left(\frac{\partial u_t}{\partial\lambda}\right)^2\right] &= \frac{\upsilon^2(\eta+1)(\upsilon+1)(\upsilon+\eta)}{(3\upsilon+\eta+1)(2\upsilon+\eta+1)(\upsilon+\eta+1)}\\ &= \frac{\upsilon^2(\overline{\eta}+1)^2(\upsilon+1)(\upsilon\overline{\eta}+1)}{(3\upsilon\overline{\eta}+\overline{\eta}+1)(2\upsilon\overline{\eta}+\overline{\eta}+1)(\upsilon\overline{\eta}+\overline{\eta}+1)(\overline{\eta}+1)}\end{aligned}$$

$$E\left[u_t\frac{\partial u_t}{\partial\lambda}\right]=\frac{-\upsilon^2(\eta-1)\eta}{(2\upsilon+\eta+1)(\upsilon+\eta+1)}=\frac{-\upsilon^2(1-\overline{\eta})}{(\overline{\eta}+1+2\upsilon\overline{\eta})(\overline{\eta}+1+\upsilon\overline{\eta})}$$

A necessary condition for (23) to be p.d. is

$$b=\phi^2+2\phi\kappa E\left(\frac{\partial u_t}{\partial\lambda}\right)+\kappa^2E\left(\frac{\partial u_t}{\partial\lambda}\right)^2<1. \tag{25}$$

The proof of the above proposition can be constructed by generalizing the result for Beta-t-EGARCH in Harvey (2013, p 37, 116-7, 138). The conditions which need to be verified are those in Jensen and Rahbek (2004) and in the present context the main point is that the boundedness of third derivatives of the log-likelihood is guaranteed by the fact that the terms are functions of bounded variables of the form $b_t^h(1-b_t)^k$, $h,k=1,2,...$

We conjecture that consistency and asymptotic normality of the ML estimator of the full parameter vector, $(\mu,\ \boldsymbol{\psi}', \upsilon, \overline{\eta})'$, will be consistent and asymptotically normal for $\upsilon > 1/2$; see Remark 2. An analytic expression for the information matrix can be derived, but is somewhat intricate. However, it can again be shown that most of the terms in the third derivatives of the log-likelihood are functions of bounded variables. The derivatives with respect to υ involve identities like (11).

3.3 Gamma-GED-EGARCH

Letting $\eta \to \infty$ gives the Gamma-GED-EGARCH model, in which the conditional distribution of y_t is a GED with time-varying scale parameter $\exp(\lambda_{t|t-1})$ and $\lambda_{t|t-1}$ evolving as a linear function of the conditional score variable[10]

$$u_t = \left(\frac{|y_t-\mu|}{\exp(\lambda_{t|t-1})}\right)^{\upsilon} - 1, \qquad t=1,...,T. \tag{26}$$

with $|\varepsilon_t|^{\upsilon}$ distributed as gamma $(\upsilon, 1/\upsilon)$ at the true parameter values. When $\upsilon < 2$, the response is less sensitive to outliers than it is for a normal distribution, but it does not have the robustness of Beta-t-EGARCH because, as is clear from (26), u_t is not bounded. There may therefore be a case for approximating Gamma-GED-EGARCH by a Beta-Gen-t-EGARCH in which η is large but bounded; similar sentiments are expressed by McDonald and Newey (1987) who comment on the attraction of keeping η finite in the generalized-t.

The existence of unconditional moments in Gamma-GED-EGARCH requires constraints on the dynamic parameters. For the first-order model, (2), the $m-th$ order moment exists only for $m < 1/\upsilon\kappa$.

The information matrix, (23), is given by setting $\overline{\eta}=0$ in (24) to yield

$$E\left[\frac{\partial u_t}{\partial \lambda}\right] = -\upsilon, \qquad E\left[\left(\frac{\partial u_t}{\partial \lambda}\right)^2\right] = \upsilon^2(\upsilon+1) \quad \text{and} \quad E\left[u_t\frac{\partial u_t}{\partial \lambda}\right] = -\upsilon^2.$$

The proof of consistency and asymptotic normality of ML estimators for the GED is less straightforward than the proof for generalized t with finite tail index because it requires a bounding argument; see Harvey (2013, p 44-5). It is perhaps simpler

[10] The definition of the GED in Harvey (2013, section 4.4) is slightly different.

just to invoke the proof for generalized t with arbitrariy large, but finite, finite tail index.

When a Gamma-GED-EGARCH model has been estimated, the LM statistic against fat tails takes the form of (19) with $(I^{-1})_{\overline{\eta}\overline{\eta}}$ now obtained from the full information matrix. Note the importance of allowing for dynamic scale because, other things being equal, its presence increases $E\left(\left|\varepsilon_t\right|^{\upsilon}-1\right)^2$.

Remark 4 *The classic EGARCH model, in which $u_t = |\varepsilon_t|$, is obtained by setting $\upsilon = 1$ in (26). The information matrix is easily evaluated for most conditional distributions (including the generalized-t), provided they have finite variance; see Andres (2014). However, absolute values are not robust and when the conditional distribution has fat tails no unconditional moments exist; see Nelson (1991). There seems to be no good reason for not using the conditional score.*

3.4 Asymmetric impact curve (leverage)

Returns may have a different effect on volatility depending on whether they are positive or negative. This effect, sometimes known as leverage, may be modeled by adding another variable to the dynamic equation. Specifically

$$\lambda_{t+1|t} = \omega\left(1-\phi\right) + \phi\,\lambda_{t|t-1} + \kappa\, u_t + \kappa^* u_t^*, \tag{27}$$

where $u_t^* = sgn\ (\mu - y_t)(u_t + 1)$ and κ^* is a parameter[11]. The effect of the extra term is to add or subtract, depending on $sgn(y_t - \mu)$, a fraction of the impact curve plus one. The new impact curve remains continuous at μ.

3.5 Example: Silver returns

Beta-Gen-t-EGARCH model were fitted to daily observations on returns from the iShares Silver Trust from April 28th 2006 to February 11th 2015; see http://finance.yahoo.com/q?s=S This exchange traded fund (ETF) aims to track the price of silver and is traded on the London and New York stock exchanges. The histogram is shown in Figure 2

The three rows in Table 1 show the results for the general model and then for Beta-t-EGARCH, that is $\upsilon = 2$, and Gamma-GED-EGARCH, that is $\overline{\eta} = 0$. As is clear from the estimates and their (numerical) SEs - and as is apparent from the analytic information matrix - the ML estimators of the two shape parameters are strongly correlated. When υ, estimated as 1.34 in the general model, is set to two,

[11] The way in which Nelson (1991) captures leverage in his classic EGARCH model is somewhat different but coincides with our approach in the case when $\eta = \infty$ and $\upsilon = 1$ (Laplace distribution); see also Bollerslev *et al* (1994, p 3019).

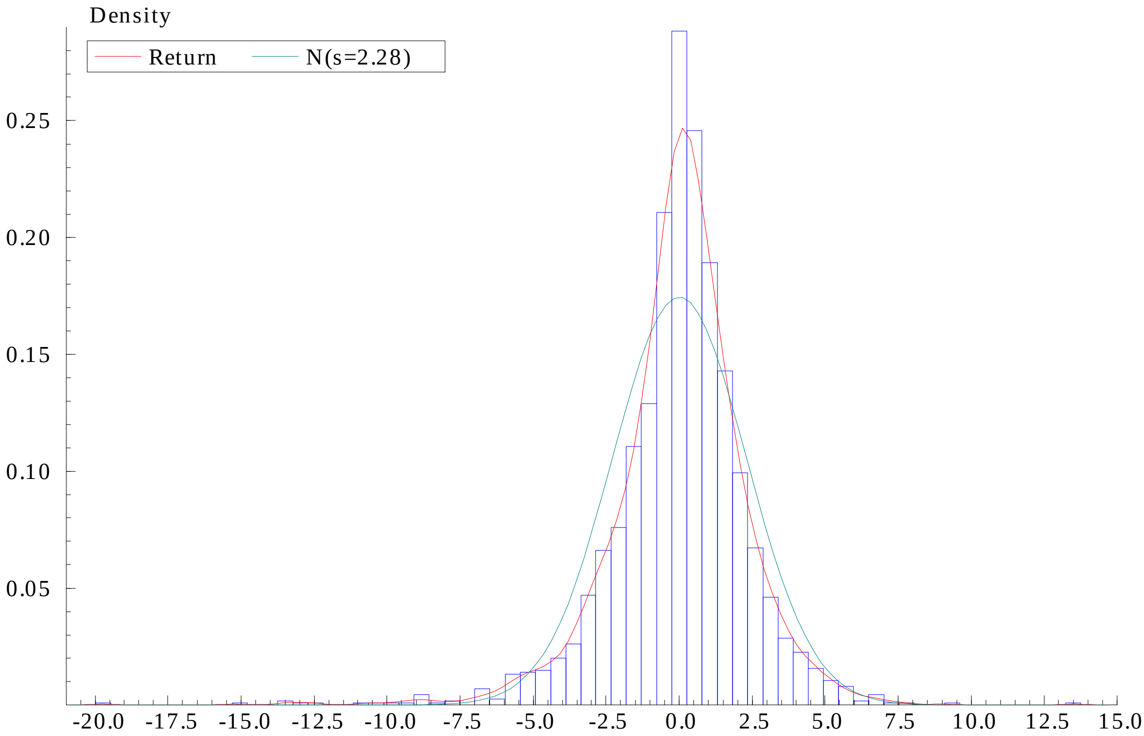

Figure 2: Daily silver returns

the estimate of $\overline{\eta}$ increases from 0.075 to 0.22 and correspondingly η goes from 13.31 to 4.54. The SE of $\overline{\eta}$ is reduced from 0.043 to 0.023. Similarly the estimate of υ decreases when $\overline{\eta}$ is fixed and its SE of falls from 0.13 to 0.05. Leverage effects are estimated but appear not to be present.

All three models fit well according to the Box-Ljung statistics[12], $Q_{\mu}(20)$ and $Q_{\lambda}(20)$, constructed from the first 20 autocorrelations of scores for location and scale respectively. However Beta-t-EGARCH is clearly worse than the other two models on the AIC and BIC goodness of fit criteria. Indeed the hypothesis that $\upsilon = 2$ is convincingly rejected by Wald and LR tests. The situation with regard to $\overline{\eta}$ is less clear cut. The LR statistic (that is minus 2 times the logarithm of the likelihood ratio) of the null hypothesis that $\overline{\eta} = 0$ is 3.60 which is less than 3.84, the 5% significance value for a χ^2_1 distribution. However, as noted in sub-section 2.4, the correct 5% significance value is only 2.70 because of the one-sided alternative. Thus there may be a small gain from using Beta-Gen-t-EGARCH rather than Gamma-GED-EGARCH.

[12] $Q_{\mu}(20)$ and $Q_{\lambda}(20)$ may be tested against χ^2_{20} and χ^2_{18} distributions respectively. Two degrees of freedom are lost for scale because two dynamic parameters are estimated.

Model	κ^*	κ	ϕ	ω	μ	$\bar{\eta}$	υ	LogL	AIC	BIC	$Q_\lambda(20)$	$Q_\mu(20)$
Student-t	0.000	0.038	0.991	0.519	0.070	0.220	$(\upsilon \equiv 2)$	-4648.2	4.210	4.225	22.3	14.6
	0.004	0.006	0.002	0.112	0.036	0.023					0.33	0.80
GED	0.000	0.045	0.987	0.502	0.086	$(\bar{\eta} \equiv 0)$	1.148	-4642.0	4.204	4.220	23.4	19.9
	0.004	0.007	0.003	0.098	0.034		0.046				0.27	0.46
Gen-t	0.000	0.042	0.989	0.493	0.081	0.076	1.342	-4640.1	4.204	4.222	24.0	17.7
	0.004	0.007	0.002	0.106	0.035	0.045	0.138				0.24	0.61

Table 1 Beta-Gen-t-EGARCH fitted to Silver returns - SEs in small typeface

3.6 Extensions

Explanatory variables can also be added to the dynamic equations. As with leverage, the asymptotic theory for ML estimation is readily extended along the lines for Beta-t-EGARCH in Harvey (2013, ch 4).

A DCS EGARCH-M model may be set up as

$$y_t = \mu + \alpha \exp(\lambda_{t|t-1}) + \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, ..., T,$$

where α is a risk-premium parameter. Again it is not difficult to extend the theory for the conditional t-distribution, used in Harvey and Lange (2015), to generalized t.

Arslan (2004) sets out a multivariate generalized t distribution and derives its properties. The distribution generalizes the standard multivariate t and so it can be used to model dynamic volatilities and changing correlations as in Creal et al (2011).

4 Asymmetry and skewness

Skewness can be introduced into the generalized Student-t distribution by means of the Fernandez and Steel (1998) method, as used by Harvey and Sucarrat (2014) for the Student-t distribution. This distribution is similar to, but not quite the same as, the skew generalized t distribution, as used recently in Hansen et al (2010). An equivalent formulation of the skew-t is proposed by Zhu and Galbraith (2010). They make a further extension to asymmetry by allowing different degrees of freedom above and below μ. Zhu and Zinde-Walsh (2009) set up an asymmetric skew GED in a similar way. Extending the generalized Student-t distribution to handle skewness and asymmetry is, in principle, straightforward: we now have υ_1 and υ_2 as well as η_1 and η_2.

Rather than formulate the asymmetric skew generalized-t distribution as in Zhu (2012), which follows the method Zhu and Galbraith (2010) and Zhu and

Zinde-Walsh (2009), we write the PDF. as

$$f(y)=\frac{K_{12}}{\exp(\lambda)}\times\begin{cases}\left(1+\frac{1}{\eta_1}\left|\frac{y-\mu}{2\alpha e^{\lambda}}\right|^{\upsilon_1}\right)^{-(\eta_1+1)/\upsilon_1} & y\leq\mu,\\ \left(1+\frac{1}{\eta_2}\left|\frac{y-\mu}{2(1-\alpha)e^{\lambda}}\right|^{\upsilon_2}\right)^{-(\eta_2+1)/\upsilon_2}, & y>\mu.\end{cases}, \qquad (28)$$

where $0<\alpha<1$ and

$$K_{12}=\frac{1}{\alpha/K_1+(1-\alpha)/K_2},$$

with $K_i=K(\eta_i,\upsilon_i), i=1,2$, defined as in (4). The derivation of (28) and the way in which the approach differs from that in the papers by Zhu and co-authors is set out in Appendix D. The distribution is constructed in such a way that both $f(y)$ and $f'(y)$ are continuous through $y=\mu$. Under symmetry, $K_{12}=K_1=K_2=K(\eta,\upsilon)$, irrespective of α. When $\alpha=1/2$, the distribution is not skewed, even though it may be asymmetric.

The probability that $y<\mu$ is given by

$$\alpha^{\dagger}\;=\;\frac{\alpha/K_1}{\alpha/K_1+(1-\alpha)/K_2}. \qquad (29)$$

Clearly $\alpha^{\dagger}$ is affected by both skewness and asymmetry. Indeed, all quantiles are affected by both asymmetry and skewness. Under symmetry, $K_1=K_2$ and so $\alpha^{\dagger}=\alpha$. The moments of the skew asymmetric distribution are given by

$$E[|y-\mu|^m;\upsilon_1,\eta_1\mid y<\mu]\alpha^{\dagger}+E[|y-\mu|^m;\upsilon_2,\eta_2\mid y>\mu](1-\alpha^{\dagger})$$

Skewness can be seen as asymmetry in the scale[13], but rather than having two distinct values[14] for scale (or, strictly speaking, its logarithm), it is better to have a single scale and $\alpha^{\dagger}$. Not only does $\alpha^{\dagger}$ have the convenient property of being equal to $\Pr(y<\mu)$, but more importantly in the present context, a single scale parameter allows a straightforward extension to dynamic volatility.

Remark 5 *The introduction of skewness and/or asymmetry means that although μ is still the mode, it is no longer the mean or the median when $\alpha^{\dagger}=1/2$.*

EGARCH effects are introduced with the scores for scale being $u_{it}=(\eta_i+1)b_{it}-1$, $i=1,2$, where the variables

$$b_{1t}\;=\;\frac{(|y_t-\mu|/2\alpha e^{\lambda_{t|t-1}})^{\upsilon_1}/\eta_1}{(|y_t-\mu|/2\alpha e^{\lambda_{t|t-1}})^{\upsilon_1}/\eta_1+1},\quad y\leq\mu,$$
$$b_{2t}\;=\;\frac{(|y_t-\mu|/2(1-\alpha)e^{\lambda_{t|t-1}})^{\upsilon_2}/\eta_2}{(|y_t-\mu|/2(1-\alpha)e^{\lambda_{t|t-1}})^{\upsilon_2}/\eta_2+1},\quad y>\mu$$

[13] The logarithm of scale could be written as $\lambda+\ln 2\alpha$ or $\lambda+\ln 2(1-\alpha)$.

[14] See, for example, the formulation of the asymmetric GED - or AEPD - in Bottazzi and Secchi (2011).

are (independently) distributed as $beta(1/\upsilon_i, \eta_i/\upsilon_i)$, $i = 1, 2$, at the true parameter values.

Remark 6 *For a distribution which is asymmetric in shape parameters and/or skewed, u_t is asymmetric, but this asymmetry should not be confused with the asymmetric effects (leverage) introduced into the impact curve in (27).*

4.1 Skewness

When there is skewness but no asymmetry in the shape parameters, $\alpha = \alpha^{\dagger}$. Under skewness, the multiplicative factors 2α and $2(1-\alpha)$ in the position of the scale are different from unity. When $\alpha > 1/2$, the left-hand side scale, that is for $y < \mu$, will be amplified, whereas the right-hand side scale will be diminished. The opposite is true when $\alpha < 1/2$.

The score for α is uncorrelated with the score for λ and the two shape parameters, υ and η. However, it is correlated with μ. The sub-matrix for μ and α in the information matrix, which is obtained from the more general result in the next sub-section, is

$$\mathbf{I}\begin{pmatrix} \mu \\ \alpha \end{pmatrix} = \begin{pmatrix} \dfrac{I_{\mu\mu}}{4\alpha(1-\alpha)} & \dfrac{-I^{+}_{\mu\lambda}}{2\alpha(1-\alpha)} \\ \dfrac{-I^{+}_{\mu\lambda}}{2\alpha(1-\alpha)} & \dfrac{I_{\lambda\lambda}+1}{\alpha(1-\alpha)} \end{pmatrix}, \tag{30}$$

where $I_{\mu\mu}$ and $I_{\lambda\lambda}$ are as in (12) and

$$I^{+}_{\mu\lambda} = \frac{e^{-\lambda}\,\overline{\eta}^{1/\upsilon}}{B(1/\overline{\eta}\upsilon + 1/\upsilon)}\,\frac{(\overline{\eta}+1)\,\upsilon^2}{\overline{\eta}\,\upsilon + \overline{\eta} + 1}. \tag{31}$$

Letting $\overline{\eta} \to 0$ gives $I^{+}_{\mu\lambda} = e^{-\lambda}\,\upsilon^{2-1/\upsilon}/\Gamma(1/\upsilon)$, which is the expression for the GED.

The information sub-matrix for λ, υ and η is unchanged. Thus the introduction of skewness has no effect on the asymptotic distribution of the ML estimators of λ, υ and η. (Subject to the qualifications outlined in Remark 2 - these qualifications extend to any υ because of the discontinuity coming from the model for skewness.)

For the EGARCH model, the fact that b_{1t} and b_{2t} have the same beta distribution means that $\mathbf{D}(\boldsymbol{\psi})$ is as in (23).

If the standard generalized t of the previous section is estimated, an LM test against skewness may be carried out. The score for α at $\alpha = 0.5$ is $2\,(\eta+1)\,b_t.sgn(y_t - \mu)$. This is similar to the score for λ, but it is an odd function because of the sign. This is perhaps not surprising as skewness is formulated as asymmetry in the scale. (In fact it is -2 times the leverage term, u_t^*, in (27).) The LM statistic,

$$LM = \frac{(I^{-1})_{\alpha\alpha}}{T}\left(2\,(\eta+1)\sum b_t.sgn(y_t-\mu)\right)^2, \tag{32}$$

is asymptotically χ_1^2 under the null hypothesis that $\alpha = 0.5$.

4.2 Asymmetry

Asymmetry adds an additional set of shape parameters. An expression for the full information matrix for a skew, asymmetric distribution with shape parameters $\boldsymbol{\theta}_i, i = 1, 2$, is derived in Appendix D. The information matrix for a skew, asymmetric generalized-t distribution is a special case with $\boldsymbol{\theta}_i = (\eta_i, \upsilon_i)$, or $(\overline{\eta}_i, \upsilon_i)$, $i = 1, 2$.

Proposition 3 *Let $\mathbf{I}_1(\mu, \lambda, \boldsymbol{\theta}_1)$ and $\mathbf{I}_2(\mu, \lambda, \boldsymbol{\theta}_2)$ denote the information matrices for the distributions with shape parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ respectively. Further, let $\mathbf{I}_1^-(\mu, \lambda, \boldsymbol{\theta}_1)$ and $\mathbf{I}_2^+(\mu, \lambda, \boldsymbol{\theta}_2)$ denote the information matrices for the same two distributions, but conditioned on the observation being below μ, for $\boldsymbol{\theta}_1$, or above μ, for $\boldsymbol{\theta}_2$. Then the information matrix of the skewed and asymmetric random variable Y constructed as in (42) can be written as:*

$$\mathbf{I}\begin{pmatrix}\mu\\ \alpha\\ \lambda\\ \boldsymbol{\theta}_1\\ \boldsymbol{\theta}_2\end{pmatrix} = \alpha^{\dagger}\begin{pmatrix} \frac{I_{1,\mu\mu}}{4\alpha^2} & \frac{I^-_{1,\mu\lambda}}{2\alpha^2} & \frac{I^-_{1,\mu\lambda}}{2\alpha} & \frac{I^-_{1,\mu\theta_1}}{2\alpha} & 0 \\ \frac{I^-_{1,\lambda\mu}}{2\alpha^2} & \frac{I_{1,\lambda\lambda}+1}{\alpha^2} & \frac{I_{1,\lambda\lambda}}{\alpha} & \frac{I_{1,\lambda\theta_1}}{\alpha} & 0 \\ \frac{I^-_{1,\lambda\mu}}{2\alpha} & \frac{I_{1,\lambda\lambda}}{\alpha} & I_{1,\lambda\lambda} & I_{1,\lambda\theta_1} & 0 \\ \frac{I^-_{1,\theta_1\mu}}{2\alpha} & \frac{I_{1,\theta_1\lambda}}{\alpha} & I_{1,\theta_1\lambda} & I_{1,\theta_1\theta_1} + \frac{\partial^2 \ln K_1}{\partial\theta_1\,\partial\theta_1'} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$
$$+(1-\alpha^{\dagger})\begin{pmatrix} \frac{I_{2,\mu\mu}}{4(1-\alpha)^2} & -\frac{I^+_{2,\mu\lambda}}{2(1-\alpha)^2} & \frac{I^+_{2,\mu\lambda}}{2(1-\alpha)} & 0 & \frac{I^+_{2,\mu\theta_2}}{2(1-\alpha)} \\ -\frac{I^+_{2,\lambda\mu}}{2(1-\alpha)^2} & \frac{I_{2,\lambda\lambda}+1}{(1-\alpha)^2} & -\frac{I_{2,\lambda\lambda}}{(1-\alpha)} & 0 & \frac{I_{2,\lambda\theta_2}}{(1-\alpha)} \\ \frac{I^+_{2,\lambda\mu}}{2(1-\alpha)} & -\frac{I_{2,\lambda\lambda}}{(1-\alpha)} & I_{2,\lambda\lambda} & 0 & I_{2,\lambda\theta_2} \\ 0 & 0 & 0 & 0 & 0 \\ \frac{I^+_{2,\theta_2\mu}}{2(1-\alpha)} & -\frac{I_{2,\theta_2\lambda}}{\alpha} & I_{2,\theta_2\lambda} & 0 & I_{2,\theta_2\theta_2} + \frac{\partial^2 \ln K_1}{\partial\theta_2\,\partial\theta_2'} \end{pmatrix}$$
$$-\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\partial^2 \ln K_{12}}{\partial\alpha^2} & 0 & \frac{\partial^2 \ln K_{12}}{\partial\alpha\partial\theta_1'} & \frac{\partial^2 \ln K_{12}}{\partial\alpha\partial\theta_2'} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\partial^2 \ln K_{12}}{\partial\theta_1\partial\alpha} & 0 & \frac{\partial^2 \ln K_{12}}{\partial\theta_1\,\partial\theta_1'} & \frac{\partial^2 \ln K_{12}}{\partial\theta_1\partial\theta_2'} \\ 0 & \frac{\partial^2 \ln K_{12}}{\partial\theta_2\partial\alpha} & 0 & \frac{\partial^2 \ln K_{12}}{\partial\theta_2\,\partial\theta_1'} & \frac{\partial^2 \ln K_{12}}{\partial\theta_2\,\partial\theta_2'} \end{pmatrix} \quad (33)$$

Corollary 1 *When* $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ *there is skewness but no asymmetry in the shape parameters and the information matrix of (30) is obtained. All terms involving second derivatives cancel out or are zero.*

Corollary 2 *When there is asymmetry but no skewness the row and column corresponding to* α *in (33) disappear and* $\alpha^{\dagger} = 1/(1+K_1/K_2)$*. (In this case, the PDF of* Y *below* μ *is* $2\alpha^{\dagger}$ *times the PDF for a variable with shape parameters* $\boldsymbol{\theta}_1$ *whereas above* Y *it is the* $2(1-\alpha^{\dagger})$ *times the PDF for a variable with shape parameters* $\boldsymbol{\theta}_2$*.)*

Remark 7 *As with the symmetric distribution, the usual asymptotics hold for the scale and shape parameters in a static asymmetric generalized t model when* μ *is known. The ML estimators of these parameters are consistent and asymptotically normal when* μ *is estimated by the mode, but they may not be efficient; see Bickel (2002). A Monte Carlo study for the asymmetric GED reported*[15] *in Bottazzi and Secchi (2011, p 1002-6) suggests that, when* μ *is estimated by ML, inference for the scale and shape parameters remains valid for any* $\upsilon > 0$*, despite the fact that the information matrix is no longer block diagonal. They also demonstrate that the ML estimator of* μ *will be asymptotically efficient for* $\upsilon > 1/2$*, the condition required to ensure that* $I_{\mu\mu}$ *exists.*

In the EGARCH model, the moments may again be found analytically with $E(e^{m\lambda_{t|t-1}})$ depending on whether $y < \mu$ or $y \geq \mu$. The information matrix takes on board the different shape parameters in the evaluation of u_t and its derivatives. Thus the block corresponding to the dynamic parameters in the volatility equation is

$$\mathbf{I}(\boldsymbol{\psi}) = \alpha^{\dagger} I_{1,\lambda\lambda}\mathbf{D}_1 + (1-\alpha^{\dagger}) I_{2,\lambda\lambda}\mathbf{D}_2, \quad \text{with} \quad I_{i,\lambda\lambda} = \frac{\upsilon_i \eta_i}{\eta_i + \upsilon_i + 1} = \frac{\upsilon_i}{\overline{\eta}_i + 1 + \upsilon_i \overline{\eta}_i}, \; i = 1, 2.$$

LM tests against asymmetry can be carried out if a symmetric model has been fitted. The test statistic will be based on a contrast between the shape parameter score above and below μ. For the tail index, the difference in the two scores near the tails will carry a good deal of influence.

A test of whether just one of the tails is light may also be relevant.

4.3 Example: Silver

Table 2 shows the results from fitting a Beta-Asymmetric-Gen-t-EGARCH model to Silver returns. A model with skewness was estimated as well but the α parameter

[15] They also note, on p 2011, that the likelihood may possess several local maxima and that this situation becomes more severe as υ decreases.

was not significantly different from 0.5. Asymmetry is present in the tail index but not in the υ shape parameter. As can be seen, the left hand tail is fat, with a tail index of $1/0.136 = 7.35$, whereas right tail is light because $\widetilde{\overline{\eta}}_2 = 0$. The likelihood ratio test statistic that they are the same, which is asymptotically distributed as χ_1^2 under the null hypothesis, is $-2(-4640.1 + 4628.4) = 23.4$. Hence the value of 0.076 reported for the symmetric model in Table 1 is now spread between $\overline{\eta}_1$ and $\overline{\eta}_2$. The SEs for the separate estimates are not a great deal bigger than the SE for obtained when it is assumed that the parameters are the same. The estimate of υ is virtually unchanged, as is its SE. The residuals are shown in Figure **??**. The asymmetry and the sharp peak induced by a value of υ well below two are clearly evident.

	κ^*	κ	ϕ	ω	μ	$\overline{\eta}_1$	$\overline{\eta}_2$	υ	LogL	AIC	BIC	$Q_\lambda(20)$	$Q_\mu(20)$
Estimate	0.004	0.040	0.990	0.499	0.098	0.136	0.000	1.349	-4628.4	4.194	4.215	27.1	20.1
SE	0.004	0.007	0.002	0.104	0.034	0.052	0.048	0.144				0.13	0.45

Table 2 Beta-Gen-t-EGARCH fitted to Silver returns

Leverage effects are again insignificant, but the asymmetry in the distribution imparts an asymmetry to the score function, as shown in Figure 3. The expectation of the score is still, of course, zero.

4.4 Martingale difference formulation

There is arguably a problem with imposing skewness and asymmetry in the way we have done, in that y_t cannot be a martingale difference. This is because its conditional expectation, $\mu + \mu_\varepsilon \exp(\lambda_{t|t-1})$, where $\mu_\varepsilon = E(\varepsilon_t)$, is time-varying. This issue is well-known for skewed distributions; see the discussion in Harvey (2013, p 145-7). With asymmetry the problem again arises. The general solution is to set up the model as

$$y_t = \mu + (\varepsilon_t - \mu_\varepsilon)\exp(\lambda_{t|t-1}),$$

where

$$\mu_\varepsilon = -\alpha^\dagger\,(2\alpha)\,\overline{\eta}_1^{1/\upsilon}\,\frac{\Gamma\left(\frac{2}{\upsilon}\right)\,\Gamma\left(\frac{1-\overline{\eta}_1}{\overline{\eta}_1\,\upsilon}\right)}{\Gamma\left(\frac{1}{\upsilon}\right)\,\Gamma\left(\frac{1}{\overline{\eta}_1\,\upsilon}\right)} + (1-\alpha^\dagger)\,2(1-\alpha)\,\overline{\eta}_2^{1/\upsilon}\,\frac{\Gamma\left(\frac{2}{\upsilon}\right)\,\Gamma\left(\frac{1-\overline{\eta}_2}{\overline{\eta}_2\,\upsilon}\right)}{\Gamma\left(\frac{1}{\upsilon}\right)\,\Gamma\left(\frac{1}{\overline{\eta}_2\,\upsilon}\right)}$$

The score is now

$$u_t = \begin{cases} \left(1+\dfrac{1}{\overline{\eta}_1}\right)\left(1-\dfrac{\mu_\varepsilon}{\varepsilon_t^-}\right)\dfrac{|\varepsilon_t^-|^\upsilon}{|\varepsilon_t^-|^\upsilon + 1/\overline{\eta}_1} - 1, & \varepsilon_t \le 0 \\ \left(1+\dfrac{1}{\overline{\eta}_2}\right)\left(1-\dfrac{\mu_\varepsilon}{\varepsilon_t^+}\right)\dfrac{|\varepsilon_t^+|^\upsilon}{|\varepsilon_t^+|^\upsilon + 1/\overline{\eta}_2} - 1, & \varepsilon_t > 0 \end{cases}$$

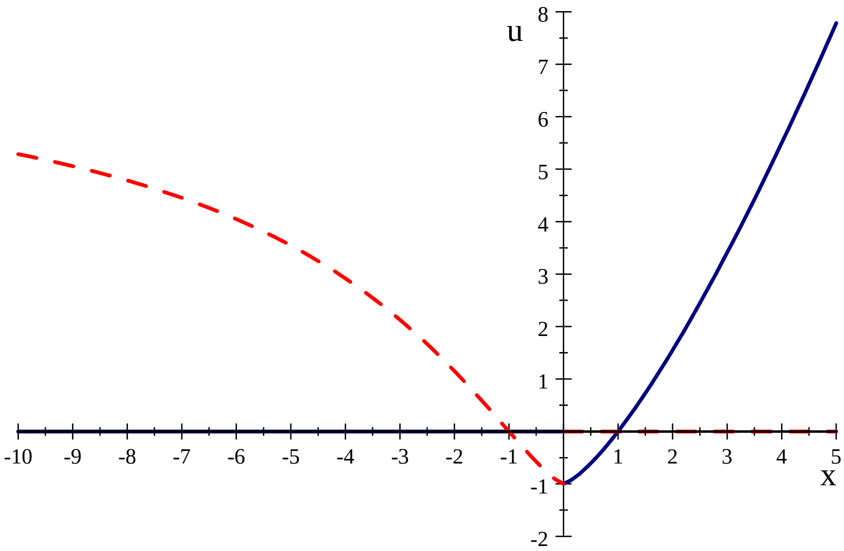

Figure 3: Score function (news impact curve) for asymmetric generalized-t EGARCH model (without leverage) fitted to Silver returns.

where $\varepsilon_t^- = \varepsilon_t/2\alpha$ and $\varepsilon_t^+ = \varepsilon_t/2(1-\alpha)$. The score is still a martingale difference as is

$$u_t^* = \begin{cases} \left(1+\frac{1}{\bar{\eta}_1}\right)\frac{|\varepsilon_t^-|^\upsilon}{|\varepsilon_t^-|^\upsilon + 1/\bar{\eta}_1} - \frac{\alpha/K_1-(1-\alpha)/K_2}{\alpha/K_1+(1-\alpha)/K_2}, & \varepsilon_t \leq 0 \\ -\left(1+\frac{1}{\bar{\eta}_2}\right)\frac{|\varepsilon_t^+|^\upsilon}{|\varepsilon_t^+|^\upsilon + 1/\bar{\eta}_2} - \frac{\alpha/K_1-(1-\alpha)/K_2}{\alpha/K_1+(1-\alpha)/K_2}, & \varepsilon_t > 0 \end{cases},$$

the additional variable in the dynamic leverage equation (27).

The above formulation was used to estimate EGARCH models with leverage for daily returns on SP500 from 2 Jan 2004 to 31 Dec 2013. The results are shown in Table 3 and the histogram of residuals is presented in Figure 4. The skew parameter is significantly different from 0.5. On the other hand, asymmetry in the tail index, while telling the story of a fatter left-hand tail, is not statistically significant when allowance is made for skewness: the LR test statistic is only 1.0. When the tail indices are the constrained to be same they take the value 32.26 (that is 1/0.031). Without skewness, the common tail index is 8.26, which is more typical for daily returns. Finally υ is significantly below two, just as it is for silver. On the other hand, when a model for weekly SP500 returns was estimated, υ was around 2.4.

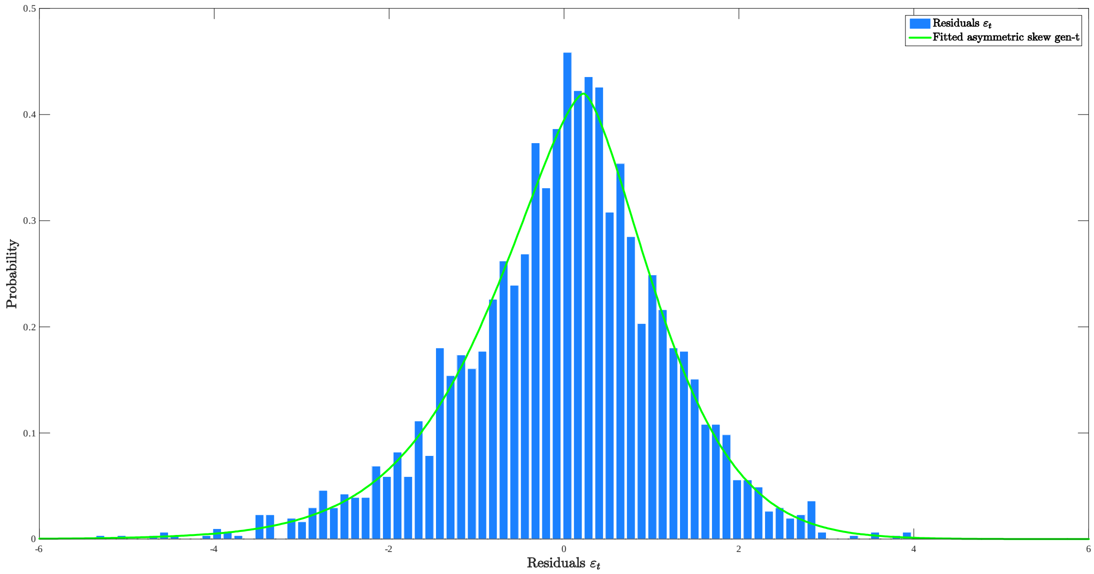

Figure 4: Standardized residuals from a skewed asymmetric generalized t EGARCH model fitted to daily SP500 returns

κ^*	κ	ϕ	ω	μ	α	$\bar{\eta}_1$	$\bar{\eta}_2$	υ	LogL	AIC	BIC
0.045	0.026	0.988	-0.535	0.040		0.041		1.497	-3361.5	2.6766	2.6928
0.005	0.005	0.001	0.102	0.014		0.024		0.114			
0.049	0.036	0.986	-0.376	0.023		0.121	0.004	1.617	-3347.5	2.6663	2.6848
0.005	0.005	0.001	0.100	0.013		0.039	0.044	0.149			
0.050	0.030	0.986	-0.233	0.015	0.573	0.031		1.478	-3342.0	2.6619	2.6805
0.005	0.005	0.001	0.115	0.014	0.011	0.029		0.122			
0.047	0.032	0.987	-0.323	0.021	0.542	0.068	0.004	1.561	-3341.5	2.6623	2.6832
0.004	0.005	0.001	0.116	0.014	0.023	0.032	0.056	0.153			

Table 3 SP500 daily excess returns from 2 Jan 2004 to 31 Dec 2013 ($T = 2,517$).

5 Dynamic location/scale model with a GB2 distribution

The usual form of the GB2 density, as given in Kleiber and Kotz (2003, p.187) and Harvey (2013, ch 5), is

$$f(y) = \frac{\upsilon (y/\alpha)^{\upsilon\xi - 1}}{\alpha B(\xi, \varsigma) \left[(y/\alpha)^{\upsilon} + 1 \right]^{\xi + \varsigma}}, \qquad \alpha, \upsilon, \xi, \varsigma > 0, \tag{34}$$

where α is the scale parameter, υ, ξ and ς are shape parameters and $B(\xi, \varsigma)$ is the beta function. The tail index is $\upsilon\varsigma$. Setting $\upsilon = \xi = 1$ gives the generalized Pareto distribution. The tail index can replace ς, that is $\eta = \upsilon\varsigma$, without further complicating the information matrix. To get the generalized gamma as a limiting case it is necessary to replace α by $\varphi\eta^{1/\upsilon}$. Thus the reparameterized GB2 has PDF

$$f(y) = \frac{\upsilon (y/\varphi)^{\upsilon\xi - 1}}{\varphi \eta^{\xi} B(\xi, \eta/\upsilon) \left[(y/\varphi)^{\upsilon} / \eta + 1 \right]^{\xi + \eta/\upsilon}}, \qquad \varphi, \upsilon, \xi, \eta > 0, \tag{35}$$

which is similar in form to the generalized-t PDF in (3).

The generalized gamma,

$$f(y; \alpha, \gamma, \upsilon) = \frac{\upsilon^{1-\gamma}}{\varphi \Gamma(\gamma)} \left(\frac{y}{\varphi} \right)^{\upsilon\gamma - 1} \exp\left(-(y/\varphi)^{\upsilon} / \upsilon \right), \quad 0 \leq y < \infty, \quad \varphi, \gamma, \upsilon > 0,$$

is now a limiting case of (35) as $\eta \to \infty$ with $\xi = \gamma$; compare the GED in (5). There is a slight difference in parameterization as compared with the GG in Kleiber and Kotz (2003) or Harvey (2013, sub-section 5.2). However, the standard gamma distribution, obtained by setting $\upsilon = 1$, is the same, as is the exponential distribution where $\upsilon = \gamma = 1$.

As noted earlier, the information matrix for (34) was found by Brazauskasin (2002); see also Kleiber and Kotz (2003, p.194). The information matrix for (35), like that of the generalized-t, is more complicated, but with the reparametrization $\overline{\eta} = 1/\eta$, it remains finite as the tail index goes to infinity, that is $\overline{\eta} \to 0$. The derivation is similar to that for the generalized-t. The main difference is the introduction of ξ, but it turns out that the elements $I_{\xi\xi}$ and $I_{\xi\overline{\eta}}$ in the information matrix are almost the same as in the information matrix for (34) because the score (which is bounded as $y \to \infty$) is

$$\frac{\partial \ln f_t}{\partial \xi} = \upsilon \ln(y_t/\varphi) - \ln \eta - \ln((y_t/\varphi)^{\upsilon} + 1) - \psi(\xi) + \psi(\eta/\upsilon + \xi).$$

Hence $I_{\xi\xi} = \psi'(\xi) - \psi'(\eta/\upsilon + \xi)$ and $I_{\xi\overline{\eta}} = (1/\upsilon)\psi'(\eta/\upsilon + \xi)$. The other terms involving η (or equivalently $\overline{\eta}$), and those for υ, are still relatively complex but no more so than the corresponding terms for information matrix of the generalized-t.

LR and LM tests of the null of a light tailed GG distribution against the alternative of fat tails can be carried out in much the same way as tests of GED against generalized-t.

Dynamic volatility can be modeled in the same way as for EGARCH. An exponential link function for the scale, that is $\varphi_{t|t-1} = \exp(\lambda_{t|t-1})$, gives the score

$$\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} = u_t = (\upsilon\xi + \eta)b_t(\xi, \eta) - \upsilon\xi, \tag{36}$$

where

$$b_t(\xi, \eta) = \frac{(y_t e^{-\lambda_{t|t-1}})^\upsilon/\eta}{(y_t e^{-\lambda_{t|t-1}})^\upsilon/\eta + 1}, \qquad t = 1, ..., T.$$

Because $0 \leq b_t(\xi, \eta) \leq 1$, it follows that as $y \to \infty$, the score approaches an upper bound of η. The GG score is obtained in the limit as $\overline{\eta} \to 0$, that is

$$\lim_{\overline{\eta}\to 0} \frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} = (y_t e^{-\lambda_{t|t-1}})^\upsilon - \upsilon\gamma = \varepsilon_t^\upsilon - \upsilon\gamma. \tag{37}$$

As in the GED, ε_t^υ is distributed as gamma $(\upsilon, 1/\upsilon)$ at the true parameter values. If η is large but finite, the distribution is approximately GG but retains the properties of GB2, in that the score for λ is bounded.

If $\xi \to \infty$ in the GG then we get the lognormal distribution, provided additional conditions are put on the other parameters, eg $\upsilon \to 0$ and $\varphi \to \infty$, Kleiber and Kotz (2003, p.149). More generally taking the logarithm of a GB2, that is $\ln y_t$, gives the class of exponential GB2 (EGB2) distributions; see McDonald and Xu (1995). Because $\ln y_t$ is normal in the limiting case when y_t is lognormal, the GB2 provides two distinct routes for specializing to the normal case. Furthermore when $\xi = \eta \to 0$, the Laplace distribution is obtained. Hence the GED and the EGB2 are both light tailed distributions covering the space between normal and Laplace. The EGB2 is skewed if $\xi \neq \eta$. A case for using the EGB2 distribution for modeling volatility is made by Wang et. al. (2001) who fitted GARCH-EGB2 models to exchange rate data. Caivano and Harvey (2014) develop the theory for DCS-EGARCH models. The EGB2 distribution may be better for modeling changing location than GED because the GED requires $\upsilon > 1.5$ for the DCS asymptotic theory to work. The same condition applies to generalized-t.

6 Conclusion

An EGARCH model in which the dynamic equation for the logarithm of scale is driven by the conditional score can be set up with a generalized-t distribution which can be extended to accomodate skewness and/or asymmetry. Properties such as unconditional moments may be derived and the asymptotic distribution of the maximum likelihood estimators worked out. For a finite tail index, the score, or influence, function is bounded, so mitigating the effect of outliers. Empirical evidence shows the practical value of the generalized-t.

It was shown that the information matrix remains positive definite as the tail index goes to infinity - and the generalized t goes to a GED - by working with the inverse tail index. The proof uses Taylor series and expansions for the trigamma function. Similarly digamma function expansions are used to obtain limiting expressions for the score with respect to the inverse tail index. These results can be used to construct Lagrange multiplier tests of light tails against fat tails.

The full asymmetric model is very flexible but will often simplify. Parameter restrictions may be tested by likelihood ratio, Wald or Lagrange multiplier tests. Given the generality of the full model, Lagrange multiplier tests may be particularly useful in practice. The form of the score in the cases examined indicates their plausibility.

The dynamics can include leverage effects and explanatory variables. Again the theory goes through. ARCH in mean effects may also be incorporated in the general model and the multivariate generalized t distribution can be used to model dynamic volatilities and changing correlations.

For positive variables the GB2 distribution may be adapted in a similar way to the generalized-t so as to include generalized gamma as a special case. Hence the null hypothesis of light tails against the alternative of fat tails may be tested within this framework. The overall theory parallels that of the generalized-t thereby providing a fully integrated approach to modeling volatility.

Acknowledgements

The work was carried out while Rutger-Jan Lange was a Post-Doctoral Research Associate on the project Dynamic Models for Volatility and Heavy Tails at the Economics Faculty, Cambridge. We are grateful to the Keynes Fund for financial support. Earlier versions of the paper were presented at Aix-Marseille School of Economics (GREQAM), at the Cambridge INET *Workshop on Developments in Time Series* in March 2015, at the EUI in Florence and at the Department of Statistics in Bologna. We are grateful to Christian Brownlees, Umberto Cherubini, Guiseppe Cavaliere, Russell Davidson, Simone Giannerini, Peter Hansen, Ryoko Ito, Sebastien Laurent, Alessandra Luati and Stephen Thiele for helpful comments.

APPENDIX

A Beta variables and the information matrix

For a $beta(\alpha, \beta)$ distribution,

$$E(b^h(1-b)^k) = \frac{B(\alpha+h, \beta+k)}{B(\alpha, \beta)}, \quad h > -\alpha, k > -\beta \tag{38}$$

Note $(|y-\mu|/\varphi)^{\upsilon}/\eta = b/(1-b)$. Also $E(\ln b) = \psi(\alpha) - \psi(\alpha+\beta)$; $E(b\ln b)$ can be found as $((\alpha+\beta)/\alpha)E(\ln b)$ with the expectation taken with respect to a $beta(\alpha+1,\beta)$ distribution.

When $beta(1/\upsilon, \eta/\upsilon)$,

$$E(b^h(1-b)^k) = \frac{\Gamma(1/\upsilon+h)\Gamma(\eta/\upsilon+k)}{\Gamma(1/\upsilon+h+\eta/\upsilon+k)}\frac{\Gamma(1/\upsilon+\eta/\upsilon)}{\Gamma(1/\upsilon)\Gamma(\eta/\upsilon)}$$

$$E(b) = \frac{1/\upsilon}{1/\upsilon+\eta/\upsilon} = \frac{1}{1+\eta}, \quad E(1-b) = \frac{\eta}{1+\eta}, \quad \textit{and} \quad E(b(1-b)) = \frac{\eta}{(\eta+1+\upsilon)(\eta+1)}$$

Expressions for $E(b^2(1-b)^2)$ and $E(b^2(1-b))$ may be similarly obtained.

Remark 8 *When $\overline{\eta}$ is used, b is $beta(1/\upsilon, 1/\upsilon\overline{\eta})$ and*

$$E(b^h(1-b)^k) = \frac{\Gamma(1/\upsilon+h)\Gamma(1/\upsilon\overline{\eta}+k)}{\Gamma(1/\upsilon+h+1/\upsilon\overline{\eta}+k)}\frac{\Gamma(1/\upsilon+1/\upsilon\overline{\eta})}{\Gamma(1/\upsilon)\Gamma(1/\upsilon\overline{\eta})}$$

$$E(b) = \frac{\overline{\eta}}{1+\overline{\eta}}, \quad E(1-b) = \frac{1}{1+\overline{\eta}}, \quad \textit{and} \quad E(b(1-b)) = \frac{\overline{\eta}}{(\overline{\eta}+1+\upsilon\overline{\eta})(\overline{\eta}+1)}$$

The above results can be used to evaluate the expectations of the elements of the Hessian. Details are available on request. Note that if η is replaced by $\overline{\eta}$ then the first derivative of η is multiplied by η^2, that is divided by $\overline{\eta}^2$, whereas the second is divided by $\overline{\eta}^4$.

B Derivation of the scores as the tail index goes to infinity

The following approximation of the digamma function ψ is used to evaluate scores when $\overline{\eta} \to 0$:

$$\psi(x) \sim \ln(x) - \frac{1}{2x} - \frac{1}{12x^2} + \mathcal{O}\left(\frac{1}{x^3}\right)$$

where the approximation[16] holds for large x; see Davis (1964, p 259). Using this approximation, the digamma terms in (10), the score with respect to $\overline{\eta}$, can be

[16] The displayed approximation is actually correct up to (and including) order x^{-3}, so, had it been necessary, we could have put $\mathcal{O}(x^{-4})$ rather than $\mathcal{O}(x^{-3})$.

written as

$$
\begin{aligned}
&\frac{1}{\upsilon\overline{\eta}^2}\psi\left(1/\overline{\eta}\upsilon\right) - \frac{1}{\upsilon\overline{\eta}^2}\psi\left(1/\overline{\eta}\upsilon + 1/\upsilon\right) \\
= \; & \frac{\ln\left(\frac{1}{\upsilon\overline{\eta}}\right) - \frac{\upsilon\overline{\eta}}{2} - \frac{\upsilon^2\overline{\eta}^2}{12}}{\overline{\eta}^2\upsilon} - \frac{\ln\left(\frac{1}{\upsilon\overline{\eta}}+\frac{1}{\upsilon}\right) - \frac{1}{2}\left(\frac{1}{\upsilon\overline{\eta}}+\frac{1}{\upsilon}\right)^{-1} - \frac{1}{12}\left(\frac{1}{\upsilon\overline{\eta}}+\frac{1}{\upsilon}\right)^{-2}}{\overline{\eta}^2\upsilon} + \mathcal{O}\left(\overline{\eta}\right) \\
= \; & \frac{1}{\overline{\eta}^2\upsilon}\left[\ln\left(\frac{1}{\overline{\eta}}\right) - \frac{\upsilon\overline{\eta}}{2} - \frac{\upsilon^2\overline{\eta}^2}{12} - \ln\left(\frac{1}{\overline{\eta}}+1\right) + \frac{\upsilon}{2}\left(\frac{1}{\overline{\eta}}+1\right)^{-1} + \frac{\upsilon^2}{12}\left(\frac{1}{\overline{\eta}}+1\right)^{-2}\right] + \mathcal{O}\left(\overline{\eta}\right)
\end{aligned}
$$

Taylor expansions of the above terms around small $\overline{\eta}$ up to second order, together with similar expansions of $b_t = \overline{\eta}|\varepsilon_t|^\upsilon/(1+\overline{\eta}|\varepsilon_t|^\upsilon)$ and $\ln(1-b_t) = -\ln(1+|\varepsilon_t|^\upsilon\overline{\eta})$ yield an expression in which all terms that are not $\mathcal{O}\left(1\right)$ or $\mathcal{O}\left(\overline{\eta}\right)$ cancel. As a result we obtain (15). We make use of the following approximations:

$$
\begin{aligned}
\overline{\eta}^{-1}b_t \ln b_t &\sim |\varepsilon_t|^\upsilon \ln\overline{\eta} + |\varepsilon_t|^\upsilon \ln|\varepsilon_t|^\upsilon + \mathcal{O}(\overline{\eta}), \\
(1-b_t)\ln[\overline{\eta}(1-b_t)] &\sim \ln\overline{\eta} + \mathcal{O}(\overline{\eta}) \\
\frac{\ln(1-b_t)}{\overline{\eta}^2} &\sim -\frac{|\varepsilon_t|^\upsilon}{\overline{\eta}} + \frac{|\varepsilon_t|^{2\upsilon}}{2} + \mathcal{O}(\overline{\eta}), \\
\frac{b_t}{\overline{\eta}}\ln[\overline{\eta}\left(1-b_t\right)] &\sim |\varepsilon_t|^\upsilon \ln\overline{\eta} + \mathcal{O}(\overline{\eta}),
\end{aligned} \tag{39}
$$

C Information matrix as the tail index goes to infinity

In order to calculate the limit of $I_{\overline{\eta}\overline{\eta}}$ in (12) as $\overline{\eta}$ goes to zero we will use the following approximation of the trigamma function given by Davis (1964, p. 260):

$$
\psi'(x) \sim \frac{1}{x} + \frac{1}{2x^2} + \frac{1}{6x^3} + \mathcal{O}\left(\frac{1}{x^5}\right),
$$

where the approximation holds for large x and is correct up to and including order x^{-4} because in the expansion the coefficient of the term involving x^{-4} is zero. For the trigamma functions appearing in $I_{\overline{\eta}\overline{\eta}}$ we have

$$\begin{aligned}
&\tfrac{1}{\upsilon^2\overline{\eta}^4}\left[\psi'\left(1/\upsilon\overline{\eta}\right)-\psi'\left(1/\upsilon\overline{\eta}+1/\upsilon\right)\right]\\
&\approx \tfrac{1}{\upsilon^2\overline{\eta}^4}\left(\left[\upsilon\overline{\eta}+\tfrac{\upsilon^2\overline{\eta}^2}{2}+\tfrac{\upsilon^3\overline{\eta}^3}{6}\right]-\left[\left(\tfrac{1}{\upsilon\overline{\eta}}+\tfrac{1}{\upsilon}\right)^{-1}+\tfrac{1}{2}\left(\tfrac{1}{\upsilon\overline{\eta}}+\tfrac{1}{\upsilon}\right)^{-2}+\tfrac{1}{6}\left(\tfrac{1}{\upsilon\overline{\eta}}+\tfrac{1}{\upsilon}\right)^{-3}\right]\right)+\mathcal{O}\left(\overline{\eta}\right)\\
&= \tfrac{1}{\upsilon^2\overline{\eta}^4}\left(\left[\upsilon\overline{\eta}+\tfrac{\upsilon^2\overline{\eta}^2}{2}+\tfrac{\upsilon^3\overline{\eta}^3}{6}\right]-\left[\upsilon\overline{\eta}\left(\overline{\eta}+1\right)^{-1}+\tfrac{\upsilon^2\,\overline{\eta}^2}{2}\left(\overline{\eta}+1\right)^{-2}+\tfrac{\upsilon^3\,\overline{\eta}^3}{6}\left(\overline{\eta}+1\right)^{-3}\right]\right)+\mathcal{O}\left(\overline{\eta}\right)\\
&= \tfrac{1}{\upsilon^2\overline{\eta}^4}\left(\upsilon\overline{\eta}\left[1-\tfrac{1}{\overline{\eta}+1}\right]+\tfrac{\upsilon^2\overline{\eta}^2}{2}\left[1-\tfrac{1}{(\overline{\eta}+1)^2}\right]+\tfrac{\upsilon^3\overline{\eta}^3}{6}\left[1-\tfrac{1}{(\overline{\eta}+1)^3}\right]\right)+\mathcal{O}\left(\overline{\eta}\right)\\
&= \tfrac{1}{\upsilon^2\overline{\eta}^4}\left(\upsilon\overline{\eta}\left[\overline{\eta}-\overline{\eta}^2+\overline{\eta}^3+\ldots\right]+\tfrac{\upsilon^2\overline{\eta}^2}{2}\left[2\overline{\eta}-3\overline{\eta}^2+\ldots\right]+\tfrac{\upsilon^3\overline{\eta}^3}{6}\left[3\overline{\eta}+\ldots\right]\right)+\mathcal{O}\left(\overline{\eta}\right)\\
&= \tfrac{1}{\upsilon^2\overline{\eta}^4}\left(\upsilon\,\overline{\eta}^2+\left[-\upsilon+\upsilon^2\right]\overline{\eta}^3+\left[\upsilon-\tfrac{3\upsilon^2}{2}+\tfrac{\upsilon^3}{2}\right]\overline{\eta}^4\right)+\mathcal{O}\left(\overline{\eta}\right)\\
&= \frac{1}{\upsilon\overline{\eta}^2}+\frac{\upsilon-1}{\upsilon\,\overline{\eta}}+\left[\frac{\upsilon}{2}-\frac{3}{2}+\frac{1}{\upsilon}\right]+\mathcal{O}\left(\overline{\eta}\right).
\end{aligned}$$

Because the trigamma functions in $I_{\overline{\eta}\overline{\eta}}$ are pre-multiplied by $1/\overline{\eta}^4$, the fact that the expansion is correct up to and including the fourth order in $\overline{\eta}^4$ is crucial. In the above result, the first two terms are singular in the limit as $\overline{\eta}$ goes to zero, but, as we shall see, they cancel with other terms in the score.

Taylor-expanding the fraction in $I_{\overline{\eta}\overline{\eta}}$ for $\overline{\eta}$ close to zero gives

$$-\frac{2\,\overline{\eta}\,\upsilon+\overline{\eta}+1}{\upsilon\,\overline{\eta}^2\left(\overline{\eta}+1\right)\left(\overline{\eta}\,\upsilon+\overline{\eta}+1\right)}\approx-\frac{1}{\upsilon\,\overline{\eta}^2}-\frac{\upsilon-1}{\upsilon\,\overline{\eta}}+\left[\upsilon+2-\frac{1}{\upsilon}\right].$$

Again, the calculation is tedious but straightforward and details have been suppressed. As before, the first two terms are singular in the limit as $\overline{\eta}\rightarrow 0$. Crucially, however, they are equal in magnitude, but opposite in sign, to the singular terms obtained above in the expansion of the trigamma functions. Hence the singular terms in $I_{\overline{\eta}\overline{\eta}}$ cancel out leaving

$$\lim_{\overline{\eta}\rightarrow 0} I_{\overline{\eta}\overline{\eta}} = \left[\frac{\upsilon}{2}-\frac{3}{2}+\frac{1}{\upsilon}\right]+\left[\upsilon+2-\frac{1}{\upsilon}\right]=\frac{3\upsilon+1}{2}.$$

Despite its complicated derivation, the final expression for $\lim I_{\overline{\eta}\overline{\eta}}$ is remarkably simple. The quantities $\lim I_{\upsilon\upsilon}$ and $\lim I_{\overline{\eta}\upsilon}$ follow from similar calculations.

D Skewness and asymmetry

Suppose that a standardized (location zero and unit scale) random variable X is symmetrically distributed. It will be convenient to write its probability density function f as follows:

$$f(x) = K(\theta)\,p\left(|x|,\theta\right),\quad p(0,\boldsymbol{\theta})\equiv 1,\quad f'(0)=0, \tag{40}$$

where K is a normalizing constant that depends on shape parameters, contained in the vector $\boldsymbol{\theta}$. We assume that $f(x)$ is symmetric around zero, so that can be written as some function of $|x|$. We also assume[17] $f'(0) = 0$, since this ensures continuity of $f'(x)$.

A random variable Y with both skewness and asymmetry can be constructed by drawing from two standardize symmetric random variables X_1 and X_2 with shape parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ as follows:

$$Y = \begin{cases} \text{with probability} & \alpha^\dagger & \text{take a draw from } -2\alpha\,|X_1|, \\ \text{with probability} & 1-\alpha^\dagger & \text{take a draw from } 2(1-\alpha)\,|X_2|, \end{cases} \tag{41}$$

where $0 < \alpha < 1$ and

$$\alpha^\dagger \quad = \quad \frac{\alpha/K_1}{\alpha/K_1 + (1-\alpha)/K_2}, \qquad \text{and} \qquad K_i = K(\theta_i),\ i = 1, 2.$$

The resulting PDF, after the introduction of location μ and scale $\exp\lambda$, is

$$f(y) = \frac{K_{12}}{\exp(\lambda)} \times \begin{cases} p\left(\dfrac{|y-\mu|}{2\alpha\ \exp\lambda}, \theta_1\right), & y \leq \mu, \\ p\left(\dfrac{|y-\mu|}{2(1-\alpha)\ \exp\lambda}, \theta_2\right), & y > \mu. \end{cases}, \tag{42}$$

where

$$K_{12} = \frac{1}{\alpha/K_1 + (1-\alpha)/K_2}.$$

Both $f(y)$ and $f'(y)$ are continuous through $y = \mu$, but $f''(y)$ generally contains a jump. Under symmetry, $K_{12} = K(\theta)$, irrespective of α. If, furthermore, $\alpha = 1/2$, we obtain the symmetric PDF.

By skewness, we shall mean that $\alpha \neq 1/2$. Under skewness, the multiplicative factors 2α and $2(1-\alpha)$ in the position of the scale are different from one. By asymmetry, we shall mean that $\theta_1 \neq \theta_2$. Asymmetry is the result of drawing from different distributions.

D.1 Information matrix

This information matrix shown in (33) is for all asymmetric and skewed random variables Y that are constructed according to (42) The result may be derived using the following considerations:

[17] It is possible to relax this restriction.

- All elements of I_1^- and I_2^+ which are proportional to $\exp(-\lambda)$, that is the off-diagonel elements of the the first row and column, get an extra division by 2α and $2(1-\alpha)$, respectively.
- The elements $I_{1,\mu\mu}^-$ and $I_{2,\mu\mu}^+$, which scale with $\exp(-2\lambda)$, get an extra division by $4\alpha^2$ and $4(1-\alpha)^2$, respectively.
- Single derivatives with respect to α can be transformed into derivatives with respect to λ by the chain rule, i.e.

$$\frac{\partial \ln f(y)}{\partial \alpha} = \begin{cases} \dfrac{1}{\alpha}\left(\dfrac{\partial \ln f(y)}{\partial \lambda} + 1\right), & y \leq \mu, \\ \dfrac{-1}{1-\alpha}\left(\dfrac{\partial \ln f(y)}{\partial \lambda} + 1\right), & y > \mu. \end{cases}$$

  In the context of another derivative, the additive term $+1$ disappears. For $y > \mu$, the minus sign remains.
- For the double derivative with respect to α, another application of the chain rule gives

$$\frac{\partial^2 \ln f(y)}{\partial \alpha^2} = \begin{cases} \dfrac{1}{\alpha^2}\left(\dfrac{\partial^2 \ln f(y)}{\partial \lambda^2} - \dfrac{\partial \ln f(y)}{\partial \lambda} - 1\right), & y \leq \mu, \\ \dfrac{1}{(1-\alpha)^2}\left(\dfrac{\partial^2 \ln f(y)}{\partial \lambda^2} - \dfrac{\partial \ln f(y)}{\partial \lambda} - 1\right), & y > \mu. \end{cases}$$

  The single derivatives with respect to λ combine to give the score with respect to λ. The expectation of the score is zero, and thus these terms disappear in the information matrix.
- For the element $\tilde{J}_{\alpha\mu}$, we must use first the chain rule to change the derivative from α into one with respect to λ, and, then, use that the resulting elements scale like $\exp(-\lambda)$. Thus we finally get multiplicative factors $1/(2\alpha^2)$ for $y \leq \mu$ and $-1/(2(1-\alpha)^2)$ for $y > \mu$.
- For derivatives with respect to α, θ_1 and θ_2, we must take into account the absence of prefactors K_1 and K_2, and, instead, the presence the prefactor $\ln K_{12}$.

D.2 Alternative parameterization

Zhu and Zinde-Walsh (2009), Zhu and Galbraith (2010) and Zhu (2012) propose a different construction:

$$Y = \begin{cases} \text{with probability} & \alpha & \text{take a draw from } -2\alpha^* \left|X_1\right|, \\ \text{with probability} & (1-\alpha) & \text{take a draw from } 2(1-\alpha^*) \left|X_2\right|. \end{cases} \tag{43}$$

where

$$\alpha^* = \frac{\alpha K_1}{\alpha K_1 + (1-\alpha) K_2}. \tag{44}$$

Introducing location μ and scale $\exp\lambda$, the associated PDF is

$$f(y) = \frac{\alpha K_1 + (1-\alpha) K_2}{\exp\lambda} \times \begin{cases} p\left(\dfrac{|y-\mu|}{2\alpha^* \exp\lambda}, \theta_1\right) & , \quad y \leq \mu, \\ p\left(\dfrac{|y-\mu|}{2(1-\alpha^*) \exp\lambda}, \theta_2\right) & , \quad y > \mu. \end{cases} \tag{45}$$

The main difference between their contruction and ours can be summarized as follows: In our case, the probablity of drawing a negative number is equal to $\alpha^\dagger$, which is affected by both by asymmetry and skewness, whereas the asymmetry in the scale is affected only by α. In their approach the opposite is true: the probability of drawing a negative number is affected only by α, whereas the asymmetry in the scale is affected by α^*, that is, both by asymmetry and by skewness.

Our construction has some advantages. First, it avoids the shape parameters appearing in the position of the scale. This clean separation between scale and shape is desirable from the point of view of estimation. Second, we are able to derive, in all generality, the information matrix associated with the asymmetric PDF. in (42). In contrast, Zhu and co-authors derive the information matrix only for a simplified version of (42). The full information matrix requires a further transformation, as noted in Zhu and Galbraith (2010, p 300). Third, the modeling of dynamic scale, which is the focus of this paper, is somewhat more natural using our definition of asymmetry, because only α affects the asymmetry in the scale.

E Definition of D matrix

The matrix $\mathbf{D}$ which appears in the information matrix for the block associated with the parameters, $\boldsymbol{\psi}$, governing the time-varying parameter, λ, is

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

where

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

with $a = \phi + \kappa E\left(\partial u_t/\partial\lambda\right)$, b is as in (25), $c = \kappa E\left(u_t.\partial u_t/\partial\lambda\right)$ and $\sigma_u^2 = Var(u_t) = -\partial u_t/\partial\lambda$.

References

Andres, P. (2014). Essays on Dynamic Conditional Score Models for Volatility. Unpublished PhD thesis, Faculty of Economics, Cambridge University.

Arslan, O. (2004) Family of multivariate generalized t distributions. *Journal of Multivariate Analysis,* 89, 329-37.

Bickel, D.R., (2002). Robust estimators of the mode and skewness of continuous data. *Comput. Statist. Data Anal.* 39, 153-63.

Bollerslev, T., Engle, R.F. and D.B. Nelson (1994). ARCH models, in *Handbook of Econometrics, Volume 4*, 2959-3038. Engle, R.F. and D.L.McFadden (eds). Amsterdam: North-Holland.

Bottazzi, G. and A. Secchi (2011). A new class of asymmetric exponential power densities with applications to economics and finance. *Industrial and Corporate Change*, 20, 991–1030.

Brazauskasin, V. (2002). Fisher information matrix for the Feller–Pareto distribution. *Statistics and Probability Letters,* 59, 159–167

Blasques, F., Koopman, S. J. and A. Lucas (2014). Maximum likelihood estimation for generalized autoregressive score models. Tinbergen Institute Discussion Paper, TI 2014-029/III, Amsterdam.

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.

Creal, D., Koopman, S. J. and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics,* 28, 777-95.

Davis, P.J. (1964). Gamma function and related functions, in Abramowitz, M. and I. A. Stegun (eds.). *Handbook of Mathematical Functions*, 253-66, New York: Dover Publications Inc.

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

Hansen, J. V., McDonald, J.B., Theodossiou, P. and B. J. Larsen (2010). Partially Adaptive Econometric Methods for Regression and Classification. *Comput Econ.* 36,153–169

Harvey, A.C. (2013). *Dynamic Models for Volatility and Heavy Tails: with applications to financial and economic time series.* Econometric Society Monograph, Cambridge University Press.

Harvey, A.C. and R-J. Lange (2015). Modeling the Interactions between Volatility and Returns. Faculty of Economics, Cambridge University

Harvey, A. C. and G. Sucarrat (2014). EGARCH Models with Fat Tails, Skewness and Leverage. *Computational Statistics and Data Analysis*, 26, 320-338.

Holly, S., Petrella, I. and E. Santoro (2013) Aggregate fluctuations and the cross-sectional dynamics of firm growth. *J. R. Statist. Soc. A,* 176, 459–479.

Jensen, S. T. and A. Rahbek (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory* 20, 1203-26.

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

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

McDonald, J.B. and W.K. Newey (1988). Partially Adaptive Estimation of Regression Models via the Generalized t Distribution, *Econometric Theory*, 4, 428-457.

McDonald,. J.B. and White, S. B. (1993) A comparison of some robust, adaptive, and partially adaptive estimators of regression models. *Econometric Reviews* **12**, 103-124

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

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

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution, *Management Science*, 44, 1650-1661.

Theodossiou, P. and C.S. Savva (forthcoming), Skewness and the relation between risk and return, *Management Science.*

Wang, K-L, Fawson, C., Barrett, C.B. and J. B. McDonald (2001). A flexible parametric GARCH model with an application to exchange rates. J*ournal of Applied Econometrics*, **16**, 521-536.

Zhu, D. (2012). Asymmetric Parametric Distributions and a New Class of Asymmetric Generalized t Distribution. Working paper. School of Economics, Shanghai University of Finance and Economics. http://papers.ssrn.com/sol3/papers.cfm?abstract_i

Zhu, D. and Galbraith, J.W. (2010). A generalized asymmetric Student-t distribution with application to financial econometrics. *Journal of Econometrics* 157, 297–305.

Zhu, D. and Galbraith, J.W. (2011). Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions. *Journal of Finance*,18, 765-78.

Zhu, D. and V. Zinde-Walsh (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics,* 148, 86-99.