

# Time series models with an EGB2 conditional distribution

Michele Caivano and Andrew Harvey\*

Bank of Italy and Faculty of Economics, Cambridge University

May 27, 2014

## **Abstract**

A time series model in which the signal is buried in noise that is non-Gaussian may throw up observations that, when judged by the Gaussian yardstick, are outliers. We describe an observation driven model, based on an exponential generalized beta distribution of the second kind (EGB2), in which the signal is a linear function of past values of the score of the conditional distribution. This specification produces a model that is not only easy to implement, but which also facilitates the development of a comprehensive and relatively straight-

forward theory for the asymptotic distribution of the maximum likelihood estimator. Score driven models of this kind can also be based on conditional t-distributions, but whereas these models carry out what, in the robustness literature, is called a soft form of trimming, the EGB2 distribution leads to a soft form of Winsorizing.

An EGARCH model based on the EGB2 distribution is also developed. This model complements the score driven EGARCH model with a conditional t-distribution. Finally dynamic location and scale models are combined and applied to data on the UK rate of inflation.

KEYWORDS: beta distribution, EGARCH; outlier; robustness; score; Winsorizing.

JEL classification; C22, G17.

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

## 1 Introduction

The changing level, or location, of a time series is usually modelled by an Autoregressive Integrated Moving Average (ARIMA) process or a linear unobserved components model. The statistical treatment of linear Gaussian models is straightforward, with the Kalman filter playing a key role in han-

dling unobserved components. However, time series are often subject to observations that, when judged by the Gaussian yardstick, are outliers.

In an unobserved components signal plus noise model, additive outliers may be captured by letting the noise have a non-normal distribution. Harvey and Luati (2014) provide an alternative approach which is observation-driven in that the conditional distribution of the observations is specified. Their model, which is based on a  $t$ -distribution, belongs to a class of models in which the dynamics are driven by the score of the conditional distribution of the observations; see Creal *et al* (2011, 2013) and Harvey (2013). These Dynamic Conditional Score (DCS) models are relatively easy to implement and their form facilitates the development of a comprehensive and relatively straightforward theory for the asymptotic distribution of the maximum likelihood estimator.

The attractions of using the  $t$ -distribution to guard against outliers in static models is well-documented. Such a parametric approach may be compared and contrasted with the methods in the robustness literature; see, for example, Maronna, Martin and Yohai (2006, ch 8). Robust procedures for guarding against additive outliers typically respond to large observations in one of two ways: either the response function converges to a positive (nega-

tive) constant for observations tending to plus (or minus) infinity or it goes to zero. These two approaches are usually classified as Winsorizing or as trimming. As is well-known, the score for a  $t$ -distribution converges to zero and so can be regarded as a parametric form of trimming. This connection then raises the question as to whether there is a distribution whose score function exhibits some form of Winsorizing and which is amenable to treatment as a DCS model. It turns out that one such distribution is the exponential generalized beta distribution of the second kind (EGB2). This distribution was first analyzed in Prentice (1975) and further explored by McDonald and Xu (1995). The aim of this article is to set out the theory for the DCS location model with an EGB2 distribution and to illustrate its practical value. Because the dynamics are driven by the score, the estimator of location is a weighted moving average of past Winsorized observations. It is worth noting that the need to Winsorize observations is often thought to be desirable in applied work; see Lui, Mitchell and Weale (2011, p. 333-4) where the technique is used on UK data. For more complex econometric modelling, using DCS location models to pre-adjust the data may be a more attractive option than arbitrarily carrying out trimming or Winsorizing.

A DCS EGB2 model for dynamic scale can also be developed. This

model belongs to the exponential generalized autoregressive heteroscedasticity (EGARCH) class and is complementary to the Student- $t$  EGARCH model discussed in Creal et al (2011) and Harvey (2013, Chapter 4). The case for using the EGB2 distribution for modelling volatility was made by Wang et. al. (2001) who fitted GARCH-EGB2 models to exchange rate data.

The article is organized as follows. In Section 2, the DCS location model based on the general form of the EGB2 distribution, which allows for skewness, is introduced and analysed. The properties of the EGB2-EGARCH model are derived in Section 3, while Section 4 fits EGB2 models with time-varying location and scale to UK inflation data. Section 5 concludes.

## 2 Dynamic location model

The stationary first-order DCS location model is

$$\begin{aligned} y_t &= \mu_{t|t-1} + \exp(\lambda)\varepsilon_t, & t = 1, \dots, T, \\ \mu_{t+1|t} &= (1 - \phi)\omega + \phi\mu_{t|t-1} + \kappa u_t, & |\phi| < 1, \end{aligned} \quad (1)$$

where  $\omega$  is the unconditional mean of  $\mu_{t|t-1}$ ,  $\exp(\lambda)$  is scale,  $\varepsilon_t$  is a serially independent, standardized variate and  $u_t$  is proportional to the conditional

score, that is  $\partial \ln f_t / \partial \mu_{t|t-1}$ , where  $f_t = f(y_t | y_{t-1}, y_{t-2}, \dots)$  is the probability density function of  $y_t$  given past observations. More generally, an ARMA-type model may be formulated by adding lags of  $\mu_{t|t-1}$  and  $u_t$ . Nonstationary ARIMA-type models can be constructed, as may time series models with trend and seasonal components. Explanatory variables can be included, as in Harvey and Luati (2014, Section 7).

When  $\varepsilon_t$  has a  $t_\nu$ -distribution, the DCS location model can be set up with  $\mu_{t|t-1}$  generated by a linear function of the (scaled) conditional score

$$u_t = (1 + \nu^{-1} \varphi^{-2}(y_t - \mu_{t|t-1})^2)^{-1} v_t, \quad t = 1, \dots, T, \quad (2)$$

where  $v_t = y_t - \mu_{t|t-1}$  is the prediction error. For low degrees of freedom,  $u_t$  is such that observations that would be seen as outliers for a Gaussian distribution are far less influential. As  $|y| \rightarrow \infty$ , the response tends to zero. Redescending M-estimators, which feature in the robustness literature, have the same property. A redescending M-estimator like Tukey's biweight function implements 'soft trimming', as opposed to metric trimming, in which the response is equal to the standardized observation up to a certain point, but is zero thereafter. In metric Winsorizing, the response after the critical point is

equal to its maximum (or minimum for negative standardized observations).

The EGB2 distribution has exponential tails whereas the  $t$ -distribution has fat tails; see Embrechts *et al* (1997) for a classification of tail properties. The connection between the score for a  $t$ -distribution and redescending M-estimators in the robustness literature is well-known; see Maronna, Martin and Yohai (2006, pp. 29). So far as we are aware, the fact that the EGB2 distribution gives a gentle form of Winsorizing has not been pointed out before, although its robustness properties have been studied by McDonald and White (1993).

## 2.1 The exponential generalized beta distribution of the second kind

The PDF of the EGB2 variate  $y$  is

$$f(y; \mu, \nu, \xi, \varsigma) = \frac{\nu \exp\{\xi(y - \mu)\nu\}}{B(\xi, \varsigma)(1 + \exp\{(y - \mu)\nu\})^{\xi+\varsigma}}, \quad (3)$$

where  $\nu$  is a scale parameter, whereas  $\xi$  and  $\varsigma$  are shape parameters which determine skewness and kurtosis;  $B(\xi, \varsigma)$  denotes the beta function. All moments exist, the mean and variance being  $E(y) = \mu + \nu^{-1}[\psi(\xi) - \psi(\varsigma)]$  and

$\sigma^2 = \nu^{-2}[\psi'(\xi) + \psi'(\varsigma)]$ , where  $\psi$  and  $\psi'$  are digamma and trigamma functions respectively. The EGB2 distribution is positively (negatively) skewed when  $\xi > \varsigma$  ( $\xi < \varsigma$ ) and its kurtosis decreases as  $\xi$  and  $\varsigma$  increase. Skewness ranges between -2 and 2. There is excess kurtosis for finite  $\xi$  and/or  $\varsigma$ ; the maximum is 9, but in the symmetric case it is six, obtained when  $\xi = \varsigma = 0$ .

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

$$\sigma = h/\nu, \quad \text{where} \quad h = \sqrt{\psi'(\xi) + \psi'(\varsigma)}. \quad (4)$$

The following formulae will be used in a number of places when the  $\sigma$  parameterization is adopted.

**Lemma 1** *Let  $\xi = \varsigma$  so that  $h = \sqrt{2\psi'(\xi)}$ . Then (i)  $\xi h^2 = 2$  as  $\xi \rightarrow \infty$ , and  $\xi h \rightarrow \infty$ ; (ii)  $\xi h = \sqrt{2}$  for  $\xi = 0$ . Or equivalently, (i)  $\xi\psi'(\xi) = 1$  and  $\xi\sqrt{\psi'(\xi)} \rightarrow \infty$  as  $\xi \rightarrow \infty$ , (ii)  $\xi\sqrt{\psi'(\xi)} = 1$  for  $\xi = 0$ .*

**Proof.** (i)  $\xi h^2 = 2\xi\psi'(\xi)$ . From formula (6.4.12) in Davis (1964), the



trigamma function can be approximated as

$$\psi'(\xi) \sim \frac{1}{\xi} + \frac{1}{2\xi^2} + \frac{1}{6\xi^3} - \frac{1}{30\xi^5} + \frac{1}{42\xi^7} - \frac{1}{30\xi^9} + \dots$$

for large  $\xi$ . Therefore,  $\xi\psi'(\xi) \rightarrow 1$  as  $\xi \rightarrow \infty$ . Similarly  $\xi\sqrt{\psi'(\xi)} = \sqrt{\xi^2\psi'(\xi)} \rightarrow \infty$ . (ii) Using the recurrence formula (6.4.6) in Davis (1964) we can write  $\psi'(\xi) = \psi'(\xi+1) + 1/\xi^2$ , so  $\xi h = \sqrt{2}\sqrt{\xi^2\psi'(\xi+1) + 1}$ . Hence, for  $\xi = 0$ ,  $\xi h = \sqrt{2}\sqrt{0.(\pi^2/6) + 1} = \sqrt{2}$ . ■

When  $\xi = \varsigma$ , the EGB2 distribution is symmetric; for  $\xi = \varsigma = 1$  it is a logistic distribution and when  $\xi = \varsigma \rightarrow \infty$  it tends to a normal distribution. The case of  $\xi = \varsigma = 0$  is important, but rarely mentioned in the literature.

**Lemma 2** *When  $\xi = \varsigma = 0$  in the EGB2 of (4), the distribution is double exponential or Laplace.*

**Proof.** For simplicity of notation let  $\mu = 0$ . Suppose  $y \leq 0$ . Then, noting that  $\Gamma(kz)/\Gamma(z) = 1/k$  and writing  $\Gamma(\xi) = \xi^{-1}\Gamma(\xi+1)$ ,

$$\begin{aligned} f(y; 0, h/\sigma, \xi, \xi) &= \frac{h\Gamma(2\xi)\xi \exp\{-\xi h |y/\sigma|\}}{\sigma\Gamma(\xi)\Gamma(\xi)(1 + \exp\{-\xi h |y/\sigma|\})^{2\xi}} \\ &= \frac{\xi h\Gamma(2\xi+1)\xi \exp\{-\xi h |y/\sigma|\}}{\sigma\Gamma(\xi+1)\Gamma(\xi+1)2\xi(1 + \exp\{-\xi h |y/\sigma|\})^{2\xi}}. \end{aligned}$$

Cancelling the  $\xi's$ , setting  $\xi = 0$  and noting that  $\xi h = \sqrt{2}$  when  $\xi = 0$  gives the result because  $\sigma = 2\sqrt{2}\varphi$ . When  $y > 0$  we first need to multiply numerator and denominator by  $\exp\{-2\xi hy/\sigma\}$  before invoking the same argument.

■

The symmetric EGB2 distribution therefore provides a continuum between the normal and Laplace distributions. The same is true for the general error distribution, denoted  $GED(v)$ , where  $v$  is the shape parameter; the normal distribution is obtained when  $v = 2$ , whereas setting  $v = 1$  gives the Laplace.

The main difference between the (symmetric) EGB2, GED and Student's  $t$  distributions with the same excess kurtosis (and standard deviation) is in the peak, which is higher and more pointed for the GED. The EGB2 in turn is more peaked than the  $t$ . As the excess kurtosis increases, the peaks of the EGB2 and GED distributions become closer together and much higher than the peak in Student's  $t$ . On the other hand, Student's  $t$  acquires more mass in the tails.

## 2.2 Score function

When the observations in (1) are from the EGB2 distribution, (3), the score function with respect to location is

$$\frac{\partial \ln f_t}{\partial \mu_{t|t-1}} = \nu(\xi + \varsigma)b_t(\xi, \varsigma) - \nu\xi, \quad t = 1, \dots, T,$$

where

$$b_t(\xi, \varsigma) = \frac{e^{(y_t - \mu_{t|t-1})\nu}}{e^{(y_t - \mu_{t|t-1})\nu} + 1}.$$

Because  $0 \leq b_t(\xi, \varsigma) \leq 1$ , it follows that as  $y \rightarrow \infty$ , the score approaches an upper bound of  $\nu\varsigma$ , whereas  $y \rightarrow -\infty$  gives a lower bound of  $\nu\xi$ .

It will prove more convenient<sup>1</sup> to replace  $\nu$  by  $h/\sigma$  and to define  $u_t$  as

$$u_t = \sigma^2 \frac{\partial \ln f_t}{\partial \mu_{t|t-1}} = \sigma h[(\xi + \varsigma)b_t(\xi, \varsigma) - \xi]. \quad (5)$$

We note that the upper and lower bounds are  $\sigma\sqrt{2}$  and  $-\sigma\sqrt{2}$  respectively when  $\varsigma = \xi = 0$ . On the other hand, there is no upper (lower) bound for  $\varsigma$  (or  $\xi$ )  $\rightarrow \infty$  because  $h\varsigma \rightarrow \infty$  (as does  $h\xi$ ). As  $\varsigma = \xi \rightarrow \infty$ , the distribution becomes normal and so for large  $\varsigma$  and  $\xi$ ,  $u_t \simeq y_t - \mu_{t|t-1}$ .

---

<sup>1</sup>Much the same effect is obtained by dividing by the information quantity for  $\mu$ , that

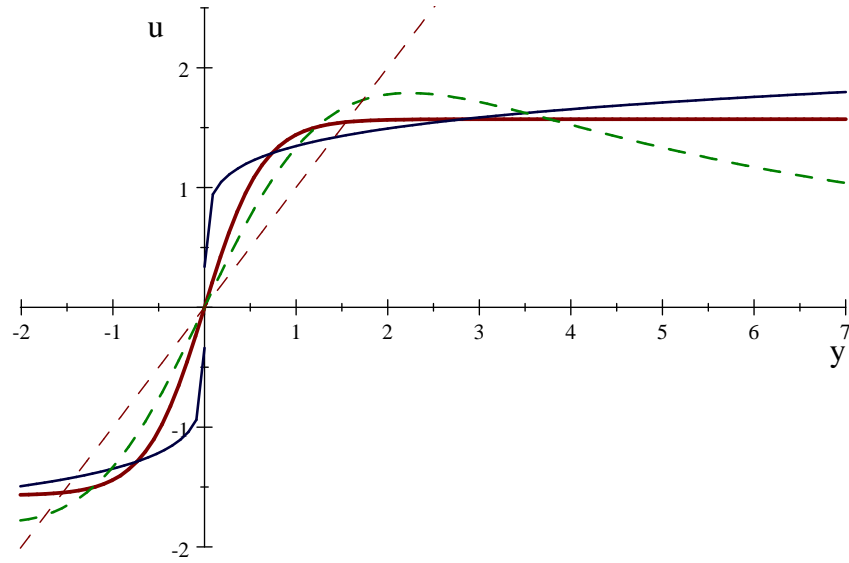


Figure 1: Score functions for EGB2 (thick line), GED (medium line) and t (thick dash), all with excess kurtosis of 2. Thin line shows normal score. (Note that  $\sigma = 1$  and  $u(-y) = -u(y)$  for  $y > 0$ ).

Figure 1 shows the score functions for standardized ( $\sigma = 1$ ) EGB2, GED and  $t$  distributions, all with excess kurtosis of two. The shape parameters for the three distributions are  $\xi = 0.5$ ,  $\nu = 1.148$  and  $\nu = 7$ . The difference in the behaviour of the score functions is striking. The score for the  $t$  distribution is redescending, reflecting the fact that it has fat tails. There is no upper bound with GED, except when it becomes a Laplace distribution and the score is  $\pm\sqrt{2}$  for  $y \neq 0$ . Neither the EGB2 nor the GED distribution has heavy tails but the EGB2 distribution has exponential tails, whereas the GED distribution is super-exponential for  $\nu > 1$ . Hence the EGB2 score is bounded and what we get is a gentle form of Winsorizing.

### 2.3 Maximum likelihood estimation

The analysis of the properties of a dynamic location model with an EGB2 distribution is essentially the same as for a dynamic model for the logarithm of scale with a GB2 distribution; see Harvey (2013, pp 164-5). Thus the variable  $b_t(\xi, \varsigma)$  in the score, (5), is independently and identically distributed with a  $beta(\xi, \varsigma)$  distribution at the true parameter values. It is easy to confirm that  $E(u_t) = 0$  and to obtain the variance of  $u_t$  as  $\sigma_u^2 = \sigma^2 h^2 \xi \varsigma / (\xi +$

---

is  $\nu^2 \xi \varsigma / (1 + \xi + \varsigma)$ .

$\varsigma + 1$ ). Note that, as  $\xi, \varsigma \rightarrow \infty$ ,  $\sigma_u^2 \rightarrow \sigma^2$ .

The asymptotic distribution for the EGB2 dynamic location model is most conveniently presented using an exponential link function for the scale, as in (1); thus  $\nu$  is replaced by  $\exp(-\lambda)$ . Such a link function ensures that estimates of scale remain positive. More importantly it is necessary when the scale is allowed to be time-varying later in the paper<sup>2</sup>.

For a first-order DCS model, the asymptotic distribution of the ML estimator of the parameters,  $\boldsymbol{\psi} = (\kappa, \phi, \omega)'$ , upon which the time-varying parameter,  $\theta_{t|t-1}$ , depends, is derived in Harvey (2013, Chapter2). The crucial point is that the expectations of the derivative of the score, its square and the product of the score and its derivative should be independent of the time-varying parameter. We can then define

$$\begin{aligned} a &= \phi + \kappa E(\partial u_t / \partial \theta), \quad c = \kappa E[u_t (\partial u_t / \partial \theta)], \\ b &= \phi^2 + 2\phi\kappa E(\partial u_t / \partial \theta) + \kappa^2 E[(\partial u_t / \partial \theta)^2] \geq 0, \end{aligned} \tag{6}$$

---

<sup>2</sup>For many puposes, it is better to parameterize the scale in terms of the standard deviation in (4) and so  $\nu$  is replaced by  $h \exp(-\lambda_\sigma)$ . Unfortunately, the presence of  $h = h(\xi, \varsigma)$  complicates the information matrix, as shown in Caivano and Harvey (2013). Thus it is simpler to just replace  $\nu$  by  $\exp(-\lambda)$ , where  $\lambda = \lambda_\sigma - \ln h$ , if asymptotic standard errors are to be computed. The likelihood function can still be maximized with respect to  $\lambda_\sigma$  and, in fact, this turns out to be much better for stability and convergence of the numerical optimization. Standard errors are of little practical importance for scale parameters and the standard errors of the other parameters do not depend on the parameterization of the scale.

because unconditional and conditional expectations are the same. If the vector  $\boldsymbol{\theta}$  denotes unknown fixed parameters and the terms in the information matrix of the static model that involve the time-varying parameter  $\theta$ , including cross-products, do not depend on  $\theta$ , then, from expression (2.56) of Harvey (2013),

$$\mathbf{I} \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{bmatrix} E \left( \frac{\partial \ln f_t}{\partial \theta} \right)^2 \mathbf{D}(\boldsymbol{\psi}) & \mathbf{d} E \left( \frac{\partial \ln f_t}{\partial \theta} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}'} \right) \\ E \left( \frac{\partial \ln f_t}{\partial \theta} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}'} \right) \mathbf{d}' & E \left( \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}} \frac{\partial \ln f_t}{\partial \boldsymbol{\theta}'} \right) \end{bmatrix}, \quad (7)$$

where  $\mathbf{d} = (0, 0, (1 - \phi)/(1 - a))'$  and

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

with

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

For an EGB2 DCS location model, the information matrix is as given

in (7) with  $\theta = \mu$  and  $\boldsymbol{\theta}' = (\lambda, \xi, \varsigma)$ . When  $u_t$  is defined as in (5) with  $\sigma = h \exp(\lambda)$ , using (6) to evaluate  $a, b$  and  $c$  gives

$$a = \phi - \kappa \frac{h^2 \xi \varsigma}{\xi + \varsigma + 1}, \quad c = \kappa \frac{e^{-\lambda} h^3 \varsigma \xi (\xi - \varsigma)}{(\varsigma + \xi + 2)(\varsigma + \xi + 1)} \quad \text{and}$$

$$b = \phi^2 - 2\phi\kappa \frac{h^2 \xi \varsigma}{\xi + \varsigma + 1} + \kappa^2 \frac{h^4 (\xi + \varsigma) \xi \varsigma (\varsigma + 1) (\xi + 1)}{(\varsigma + \xi + 3)(\varsigma + \xi + 2)(\varsigma + \xi + 1)}.$$

Provided  $b < 1$  and  $\kappa \neq 0$ , the limiting distribution of  $\sqrt{T}(\tilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \tilde{\lambda} - \lambda, \tilde{\xi} - \xi, \tilde{\varsigma} - \varsigma)'$  is multivariate normal with covariance matrix given by the inverse of (7) which is

$$\mathbf{I} \begin{pmatrix} \boldsymbol{\psi} \\ \lambda \\ \xi \\ \varsigma \end{pmatrix} = \begin{bmatrix} \frac{e^{-2\lambda} \xi \varsigma}{1 + \xi + \varsigma} \mathbf{D}(\boldsymbol{\psi}) & I_{12} \mathbf{d} & \frac{e^{-\lambda} \varsigma}{\xi + \varsigma} \mathbf{d} & \frac{-e^{-\lambda} \xi}{\xi + \varsigma} \mathbf{d} \\ I_{21} \mathbf{d}' & I_{22} & I_{23} & I_{24} \\ \frac{e^{-\lambda} \varsigma}{\xi + \varsigma} \mathbf{d}' & I_{23} & \psi'(\xi) - \psi'(\xi + \varsigma) & -\psi'(\xi + \varsigma) \\ \frac{-e^{-\lambda} \xi}{\xi + \varsigma} \mathbf{d}' & I_{24} & -\psi'(\xi + \varsigma) & \psi'(\varsigma) - \psi'(\xi + \varsigma) \end{bmatrix},$$

where  $\mathbf{d} = (0, 0, (1 - \phi)/(1 - a))'$  and

$$I_{21} = I_{12} = \frac{-e^{-\lambda} (\xi - \varsigma - \xi \varsigma (\psi(\xi) - \psi(\varsigma)))}{1 + \xi + \varsigma}, \quad I_{23} = I_{32} = \frac{\varsigma (\psi(\xi) - \psi(\varsigma)) - 1}{\xi + \varsigma},$$

$$I_{24} = I_{42} = \frac{\xi (\psi(\varsigma) - \psi(\xi)) - 1}{\xi + \varsigma}$$



and

$$I_{22} = \frac{\xi\varsigma}{1 + \xi + \varsigma} \left[ \left( \psi'(\xi) + \psi'(\varsigma) + \left( \psi(\varsigma) - \psi(\xi) + \frac{\xi - \varsigma}{\xi\varsigma} \right)^2 \right) - \left( \frac{\xi^2 + \varsigma^2}{\xi^2\varsigma^2} \right) \right] + 1.$$

Consistency and asymptotic normality follow by verifying the conditions in Jensen and Rahbek (2004). The boundedness of the score, (5), and its derivatives makes this task relatively straightforward; see Harvey (2013, pp. 40-4).

## 2.4 Estimation for a symmetric distribution

In the symmetric EGB2 distribution, denoted EGB2sym,  $\xi$  and  $\varsigma$  are constrained to be equal. The information matrix is then

$$\mathbf{I} \begin{pmatrix} \boldsymbol{\psi} \\ \lambda \\ \xi \end{pmatrix} = \begin{bmatrix} \frac{e^{-2\lambda\xi^2}}{1+2\xi} \mathbf{D}(\boldsymbol{\psi}) & 0 & 0 \\ 0 & \frac{2\xi+2\xi^2\psi'(\xi)-1}{1+2\xi} & -1/\xi \\ 0 & -1/\xi & 2\psi'(\xi) - 4\psi'(2\xi) \end{bmatrix}. \quad (8)$$

The expression for  $b$  can be simplified to

$$b = \phi^2 - 2\phi\kappa \frac{h^2\xi^2}{2\xi + 1} + \kappa^2 \frac{h^4\xi^3(\xi + 1)}{4\xi^2 + 8\xi + 3},$$

whereas  $a = \phi - \kappa h^2\xi^2/(2\xi + 1)$  and  $c = 0$ .

For small  $\xi$ ,  $b \simeq \phi^2 - 4\phi\kappa + (4/3)\kappa^2/\xi$  so the condition  $b < 1$  will be violated if  $\xi$  is too close to zero when  $\kappa$  is non-zero. (For  $\phi$  close to one, the lower bound for  $\xi$  is approximately  $\kappa/3$ .) On the other hand, letting  $\xi \rightarrow \infty$  yields  $b = \phi^2 - 2\phi\kappa + \kappa^2 = (\phi - \kappa)^2 = a^2$  and  $|\phi - \kappa| < 1$  is the standard invertibility condition for the Gaussian ARMA(1,1) model.

The null hypothesis that  $\xi = \varsigma$  is easily tested with a likelihood ratio (LR) statistic. A Wald or Lagrange multiplier (LM) test is also an option.

**Remark 1** *For the GED( $v$ ) distribution, the usual asymptotic properties of the ML estimator of  $\mu$  can be shown to hold, though the proof is non-standard for  $v < 2$  because the score function is not continuous at  $y = \mu$ ; see Zhu and Zinde-Walsh (2009). However, for the DCS model, the asymptotic theory runs into difficulties when  $v \leq 1.5$  because the higher order moments upon which  $b$  depends do not exist. Specifically,  $E(\partial u_t / \partial \mu)^2$  only exists for  $v > 1.5$ . The ability of the model to capture leptokurtic behaviour is therefore limited, because the excess kurtosis for  $v = 1.5$  is only 0.762.*

## 2.5 Example: US investment

Dynamic location models were fitted to the growth rates of US gross fixed private investment using EGB2, Student's  $t$  and normal distributions. The

observations on the level of investment were quarterly, ranging from 1947q1 to 2012q4. The data are seasonally adjusted and taken from the Federal Reserve Economic Data (FRED) database of the Federal Reserve of St. Louis.

Table 1 shows the results; corresponding results for GDP and industrial production can be found in our original discussion paper, Caivano and Harvey (2013). The asymptotic and numerical standard errors, the latter computed from the estimated Hessian, are reasonably close given that the sample size is only  $T = 268$ . The Student- $t$  and symmetric  $EGB2$  models easily outperform the Gaussian model with the Gaussian model being convincingly rejected by likelihood ratio tests. The  $EGB2_{sym}$  model has a bigger likelihood than the  $t$  model. Finally, a likelihood ratio test of  $\xi = \varsigma$  is unable to reject  $EGB2_{sym}$  against the more general  $EGB2$ .

Table 1 **US gross fixed private investment**

	$\kappa$	$\phi$	$\omega$	$\lambda$	$\xi$ ( <i>or</i> $\nu$ )	$\zeta$	$\ln L$
EGB2	0.17	0.41	0.019	−4.22	0.39	0.50	
NSE	(0.056)	(0.155)	(0.008)	(0.453)	(0.201)	(0.299)	416.90
ASE	(0.047)	(0.232)	(0.008)	(0.483)	(0.223)	(0.301)	
EGB2sym	0.14	0.41	0.011	−4.57	0.29		
NSE	(0.057)	(0.176)	(0.004)	(0.826)	(0.269)	—	416.00
ASE	(0.043)	(0.261)	(0.004)	(0.614)	(0.200)		
t	0.35	0.38	0.011	−3.24	4.20		
NSE	(0.114)	(0.171)	(0.004)	(0.085)	(1.275)	—	411.91
ASE	(0.104)	(0.257)	(0.004)	(0.060)	(0.881)		
Gaussian	0.21	0.36	0.009	−2.95			
NSE	(0.059)	(0.161)	(0.004)	(0.044)	—	—	401.95
ASE	(0.062)	(0.257)	(0.004)	(0.044)			

### 3 EGB2-EGARCH

Dynamic scale models can be constructed for conditional t and GED distributions; see Harvey (2013, Chapter4). In the former case the score has a

beta distribution, whereas in the latter it has a gamma distribution. Just as in the dynamic location case, the EGB2 distribution offers an alternative to the GED for capturing responses between the normal and Laplace.

The first-order dynamic scale model with EGB2 distributed errors is

$$y_t = \mu + \exp(\lambda_{t|t-1})\varepsilon_t, \quad t = 1, \dots, T, \quad (9)$$

where  $\varepsilon_t$  is a standardized ( $\mu = 0, \nu = 1$ ) EGB2, that is  $\varepsilon_t \sim EGB2(0, 1, \xi, \varsigma)$ .

The dynamic equation is

$$\lambda_{t+1|t} = \omega(1 - \phi) + \phi\lambda_{t|t-1} + \kappa u_t, \quad (10)$$

where  $u_t$  is now the score with respect to  $\lambda_{t|t-1}$ . The conditional distribution is

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

where  $\boldsymbol{\psi}$  denotes the parameters in (10), and so

$$u_t = \partial \ln f_t / \partial \lambda_{t|t-1} = (\xi + \varsigma)\varepsilon_t b_t - \xi \varepsilon_t - 1, \quad (11)$$

with

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

At the true parameters values,  $b_t \sim \text{beta}(\xi, \varsigma)$  as in the score for the dynamic location model.

The model may be parameterized in terms of the standard deviation,  $\sigma_{t|t-1}$ , by defining  $\epsilon_t = \varepsilon_t/h$ , where  $h$  was defined in (4). Then

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

with the only difference between  $\lambda_{\sigma,t|t-1}$  and  $\lambda_{t|t-1}$  being in the constant term which in  $\lambda_{\sigma,t|t-1}$  is  $\omega_\sigma = \omega + \ln h$ ; see the discussion in footnote 2. Note that the variance of  $\epsilon_t$  is unity.

Writing the score, (11), as

$$u_t = h(\xi + \varsigma)\epsilon_t b_t - h\xi\epsilon_t - 1, \quad (12)$$

it can be seen<sup>3</sup> that when  $\xi = \varsigma = 0$ ,  $u_t = \sqrt{2}|\epsilon_t| - 1$  and, when  $\xi = \varsigma \rightarrow \infty$ ,

---

<sup>3</sup>When  $\xi = 0$ ,  $\xi h = \sqrt{2}$  and  $b_t$  degenerates to a Bernoulli variable such that  $b_t = 0$  when  $\epsilon_t < 0$  and  $b_t = 1$  when  $\epsilon_t > 0$ . Then  $2b_t - 1 = 1$  ( $-1$ ) for  $\epsilon_t > 0$  ( $\epsilon_t < 0$ ) and the score can be written as:  $u_t = \sqrt{2}|\epsilon_t| - 1$ .

As regards  $\xi \rightarrow \infty$ , note that because  $\partial b_t / \partial \epsilon_t = h b_t (1 - b_t)$ , a first order Taylor ex-

$$u_t = \epsilon_t^2 - 1.$$

Figure 2 compares the way observations are weighted by the score of a EGB2 distribution with  $\xi = \varsigma = 0.5$ , a Student's  $t_7$  distribution and a  $GED(1.148)$ . These are the same distributions as were used in Figure 1; all have excess kurtosis of two. Consistent with the Winsorizing of the location score, dividing (12) by  $\epsilon_t$  gives a bounded function as  $|\epsilon_t| \rightarrow \infty$ . Note that the score function is often called the news-impact curve in the GARCH literature and that it becomes asymmetric when a leverage term is introduced into the dynamics; see Harvey (2013, pp 105-7).

The unconditional mean is given by  $E(y_t) = \mu + E(\epsilon_t) E(e^{\lambda_t|t-1})$ , whereas the  $m$ -th unconditional moment about the mean is  $E(\epsilon_t^m) E(e^{m\lambda_t|t-1})$ ,  $m > 1$ . In the Beta-t-EGARCH model, the expression  $E(\exp(m\lambda_{t|t-1}))$  depends on the moment generating function (MGF) of a beta variate which has a known form; see Harvey (2013, Chapter 4). For EGB2-EGARCH, the unconditional moments depend on the MGF of  $u_t$ , ie  $E_{EGB2(\xi, \varsigma)}[mu_t]$ , where  $u_t$  is defined in (11). The limiting normal and Laplace cases of the EGB2 have score functions, and hence unconditional moments, which are the same as for  $v = 2$

---

pansion of  $b_t$  around  $\epsilon_t = 0$  yields  $b_t \simeq 0.5 + (h/4)\epsilon_t$ . Therefore  $2b_t - 1 \simeq (h/2)\epsilon_t$  and  $u_t \simeq (\xi h^2/2)\epsilon_t^2 - 1$ . As  $\xi \rightarrow \infty$ ,  $\xi h^2 \rightarrow 2$ .

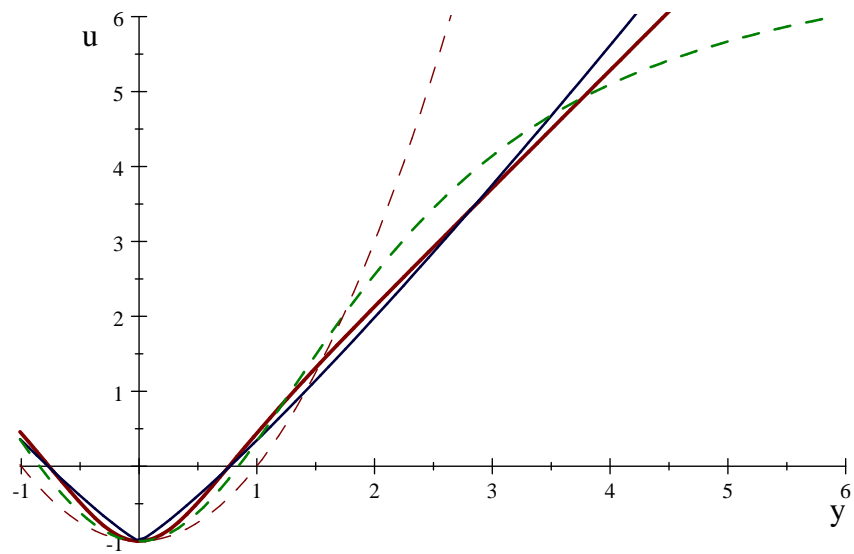


Figure 2: Score functions for EGB2 (thick line), GED (medium line) and  $t$  (thick dash), all with excess kurtosis of two. Thin line shows normal score.



and  $v = 1$  in Gamma-GED-EGARCH; see Harvey (2013, sub-section 4.2.2). When  $v = 1$  it is necessary to have  $m\kappa < 1$  in the first-order model for the  $m - th$  moment to exist, whereas for  $v = 2$  the condition is  $m\kappa < 1/2$ . For  $0 < \xi, \varsigma < \infty$  having the last condition hold is therefore sufficient for the existence of the unconditional moments. This being the case, we can at least assert, using Jensen's inequality, that the unconditional moments exceed the conditional moments and that the kurtosis increases; see Harvey (2013, p. 102).

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

**Remark 2** *There is a problem with modelling returns with a skewed distribution because the conditional expectation,  $E_{t-1}y_t = \mu_\varepsilon \exp(\lambda_{t|t-1})$ , is not constant. Therefore  $y_t$  cannot be a martingale difference. Following Harvey and Sucarrat (2014), the model can be reformulated as  $y_t = (\varepsilon_t - \mu_\varepsilon) \exp(\lambda_{t|t-1})$ ,  $t = 1, \dots, T$ , where  $\mu_\varepsilon = \psi(\xi) - \psi(\varsigma)$ , and the score adapted accordingly.*

### 3.1 Maximum likelihood estimation

The information matrix of the parameters in a dynamic scale model with an EGB2 distribution is given below; further details can be found in the appendix. In the general asymmetric case, it is assumed that  $\mu$  is given, because the cross-terms of the information matrix associated with it and the other parameters depend on scale.

**Proposition 1** *Consider the model defined by (9) and (10) with  $|\phi| < 1$  and  $\mu$  assumed to be known. Define  $a, b$  and  $c$  as in (6) with*

$$\begin{aligned}
E(u'_t) &= - \left( \frac{1}{\xi + \varsigma + 1} \left( \psi'(\xi) + \psi'(\varsigma) + \left( \psi(\xi) - \psi(\varsigma) - \frac{\xi - \varsigma}{\xi \varsigma} \right)^2 - \frac{\xi^2 + \varsigma^2}{\xi^2 \varsigma^2} \right) + 1 \right) = -\sigma_u^2 \\
E(u_t'^2) &= \frac{\xi \varsigma (\xi + 1) (\varsigma + 1) (\xi + \varsigma)}{(\xi + \varsigma + 3) (\xi + \varsigma + 2) (\xi + \varsigma + 1)} \gamma_1 (\xi + 2, \varsigma + 2) \\
&\quad + \frac{2\xi \varsigma (\xi + 1) (\xi + \varsigma)}{(\xi + \varsigma + 2) (\xi + \varsigma + 1)} \gamma_2 (\xi + 2, \varsigma + 1) - \frac{2\xi^2 \varsigma}{\xi + \varsigma + 1} \gamma_2 (\xi + 1, \varsigma + 1) + \sigma_u^2 + 1 \\
E(u_t u'_t) &= - \frac{\xi \varsigma (\xi + 1) (\xi + \varsigma)}{(\xi + \varsigma + 2) (\xi + \varsigma + 1)} \gamma_2 (\xi + 2, \varsigma + 1) + \frac{\xi^2 \varsigma}{\xi + \varsigma + 1} \gamma_2 (\xi + 1, \varsigma + 1) - 1,
\end{aligned}$$

where

$$\begin{aligned}
\gamma_1(p, q) &= \psi'''(p) + \psi'''(q) + 3(\psi'(p) + \psi'(q))^2 + 4(\psi''(p) - \psi''(q))(\psi(p) - \psi(q)) \\
&\quad + 6(\psi'(p) + \psi'(q))(\psi(p) - \psi(q))^2 + (\psi(p) - \psi(q))^4, \\
\gamma_2(p, q) &= \psi''(p) - \psi''(q) + 3(\psi(p) - \psi(q))(\psi'(p) + \psi'(q)) + (\psi(p) - \psi(q))^3
\end{aligned}$$

Let  $\boldsymbol{\psi} = (\kappa, \phi, \omega)'$ , where  $\omega = \omega_\sigma - \ln h$ . Assuming that  $b < 1$ , the information matrix is

$$\mathbf{I} \begin{pmatrix} \boldsymbol{\psi} \\ \xi \\ \varsigma \end{pmatrix} = \begin{bmatrix} I_{22}\mathbf{D}(\boldsymbol{\psi}) & I_{23}\mathbf{d} & I_{24}\mathbf{d} \\ I_{23}\mathbf{d}' & \psi'(\xi) - \psi'(\xi + \varsigma) & -\psi'(\xi + \varsigma) \\ I_{24}\mathbf{d}' & -\psi'(\xi + \varsigma) & \psi'(\varsigma) - \psi'(\xi + \varsigma) \end{bmatrix}$$

with  $I_{22}, I_{23}$  and  $I_{24}$ , together with  $\mathbf{D}(\boldsymbol{\psi})$  and  $\mathbf{d}$ , defined as in (7).

The information matrix simplifies considerably in the symmetric case.

**Corollary 1** When it is known that  $\xi = \varsigma$ , the expressions needed to obtain

$a, b$  and  $c$  are:

$$E(u'_t) = \frac{1 - 2\xi^2\psi'(\xi) - 2\xi}{2\xi + 1} = -\sigma_u^2, \quad E(u_t u'_t) = -1 \quad (13)$$

and

$$E(u_t'^2) = \frac{\xi^3 (\xi + 1)}{(2\xi + 3)(2\xi + 1)} (2\psi'''(\xi + 2) + 12\psi'^2(\xi + 2)) + \sigma_u^2 + 1. \quad (14)$$

When  $\xi = 0$ , so that the distribution is Laplace,  $E(u_t') = -1$ . Similarly as  $\xi \rightarrow \infty$ ,  $E(u_t') = -2$ , which is the correct result for a Gaussian distribution. In addition, when  $\xi = 0$  both  $\psi'(\xi + 2)$  and  $\psi'''(\xi + 2)$  are finite so  $E(u_t'^2) = 2$ . Hence  $b = \phi^2 - 2\phi\kappa + 2\kappa^2$ , which is the same as given by the expression in Harvey (2013, p 120) for  $b$  in Gamma-GED-EGARCH when  $v = 1$ . (Also  $c = -1$ .) Similarly for  $\xi \rightarrow \infty$ ,  $b = \phi^2 - 4\phi\kappa + 12\kappa^2$ .

**Remark 3** *As can be seen from (8), the information matrix is block diagonal in the symmetric model, so  $\mu$  can be included in the set of parameters to be estimated by ML without affecting the asymptotic distribution of  $(\tilde{\psi}', \tilde{\xi})$ .*

### 3.2 Example: exchange rates

Table 2 reports the full ML estimates of the (symmetric) EGB2-EGARCH and Beta-t-EGARCH models for the returns of daily exchange rates against the US dollar. The currencies are the Australian dollar (AUD), the Canadian dollar (CAD), the Swiss franc (CHF), the Denmark krone (DKK), the Euro

(EUR), the Pound sterling (GBP), the Japanese yen (JPY), the Norwegian krone (NOK), the New Zealand dollar (NZD) and the Swedish krona (SEK) and the observations run from 4th January 1999 to 15th March 2013.

The EGB2 gives a better fit for five countries, whereas the  $t$  is best for four<sup>4</sup>. In the case of Switzerland, the exchange rate experienced a sudden fall on 6th September 2011 when the Swiss National Bank announced its intention to enforce a ceiling on the exchange rate of the euro against the Swiss franc. If the resulting outlier is removed from the returns series, the EGB2 performs better than the Student's  $t$ .

---

<sup>4</sup>Although it is not the purpose of this exercise to compare DCS EGARCH models with standard GARCH - there is already a good deal of evidence in Creal et al (2011), Harvey and Sucarrat (2014) and elsewhere to suggest that DCS EGARCH tends to be better- we did fit GARCH- $t$  models and found that in only 7 out of 23 cases did they beat Beta- $t$ -EGARCH in terms of goodness of fit.

Table 2 **Exchange rates**

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

\* CHF series without the outlier corresponding to the Swiss national bank

intervention on September 6, 2011

## 4 Changing location and changing scale: an example with UK inflation

The DCS model for time-varying location may be combined with a DCS EGARCH model to give

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

For symmetric distributions, the structure of the information matrix in the static model is such that the form of the dynamic equations for  $\mu_{t|t-1}$  and  $\lambda_{t|t-1}$  is essentially unchanged, except that both scores now contain  $\lambda_{t|t-1}$  and  $\mu_{t|t-1}$ . Estimation by ML is straightforward. However, the presence of  $\lambda_{t|t-1}$  in the part of the information matrix associated with  $\mu_{t|t-1}$  means that it cannot be evaluated analytically.

A number of DCS models were fitted to the quarterly rate of CPI inflation in the UK (expressed in annualized percentage terms) from 1956q1 to 2013q1. The data were taken from the FRED database and were seasonally adjusted using the STAMP package of Koopman, Harvey, Doornik and

Shephard (2009). A local level location model,

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa^\dagger u_t^\dagger, \quad t = 1, \dots, T,$$

was first estimated. (The dagger is used to differentiate the score and its coefficient from the score of scale and its coefficient,  $\kappa$ ). Tests based on the first-order autocorrelations of squared residuals show that the null hypothesis of homoscedasticity is convincingly rejected. (The  $\chi_1^2$  statistics are 11.60, 11.78 and 12.12 for residuals from EGB2,  $t$  and Gaussian models respectively).

Table 3 shows results for the local level model with time-varying scale. Numerical standard errors are given in parenthesis. Likelihood ratio tests easily reject the null hypothesis of Gaussianity against the Student  $t$  and *EGB2sym* alternatives. When the unrestricted EGB2 is fitted, the null hypothesis of symmetry is clearly rejected by a likelihood ratio test. The same is true for the Beta- $t$ -EGARCH model, where the  $t$ -distribution is skewed as in Harvey and Sucarrat (2014). The maximized log-likelihood function for EGB2 is larger than that for skew- $t$ , but the difference is very small.



Table 3 **UK quarterly CPI inflation: 1956q1-2013q1**

	$\kappa^\dagger$	$\kappa$	$\phi$	$\omega$	$\xi$ (or $\nu$ )	$\varsigma$ (or $\gamma$ )	$\ln L$
EGB2	0.244	0.091	0.996	0.583	1.528	0.718	−514.1
	(0.051)	(0.027)	(0.017)	(0.801)	(1.066)	(0.342)	
EGB2sym	0.276	0.105	0.991	0.205	0.715	—	−518.8
	(0.054)	(0.027)	(0.018)	(0.550)	(0.270)		
Skew-t	0.501	0.100	0.991	0.784	5.538	1.231	−514.7
	(0.098)	(0.029)	(0.020)	(0.514)	(1.681)	(0.094)	
t	0.501	0.107	0.989	0.787	5.363	—	−518.0
	(0.100)	(0.030)	(0.020)	(0.489)	(1.542)		
Gaussian	0.304	0.053	0.989	1.084	—	—	−536.8
	(0.058)	(0.011)	(0.012)	(0.341)			

## 5 Conclusions and extensions

This article has shown how DCS models with changing location can be extended to cover EGB2 conditional distributions. Most of the theoretical results on the properties of DCS-t models, including the asymptotic distribution of ML estimators, carry over to EGB2 models. However, whereas the t-distribution has fat-tails, and hence subjects outliers to a form of trim-

ming, the EGB2 distribution has exponential tails (but excess kurtosis) and so gives a gentle form of Winsorizing. Like the EGB2, the GED distribution also includes distributions between the normal and Laplace, but its score function is unbounded for distributions with lighter tails than Laplace.

The statistical properties of the EGB2 distribution means that it nicely complements the  $t$ -distribution and the examples show that it provides a viable alternative for modelling changing scale as well as changing location. Another attraction of the EGB2 is that its extra parameter allows for skewness. Extensions to handle multivariate series may be possible by following the approach in Yang et al (2011).

## APPENDIX

The following result, which is related to Lemma 1 of Harvey (2013, p. 23), is useful for deriving the information matrix for the EGB2 dynamic scale model. (It can also be used to confirm that  $E(u_t) = 0$ ; note that  $\psi(\xi + 1) = \psi(\xi) + 1/\xi$ ).

**Lemma 3** *If  $\epsilon_t \sim EGB2(0, 1, \xi, \varsigma)$ , then for  $h$  and  $k \geq 0$ ,*

$$E_{EGB2(0,1,\xi,\varsigma)}[\varepsilon_t^r b_t^h (1-b_t)^k] = \frac{B(\xi + h, \varsigma + k)}{B(\xi, \varsigma)} E_{EGB2(0,1,\xi+h,\varsigma+k)}[\varepsilon_t^r], \quad r = 1, 2, 3, \dots$$

**Proof.** The result follows from writing

$$\begin{aligned}
E_{EGB2(0,1,\xi,\varsigma)}[\varepsilon_t^r b_t^h (1 - b_t)^k] &= \int \frac{\nu \varepsilon_t^r}{B(\xi, \varsigma)} \frac{\exp h \varepsilon_t}{(1 + \exp \varepsilon_t)^{h+k}} \frac{\exp \xi \varepsilon_t}{(1 + \exp \varepsilon_t)^{\xi+\varsigma}} d\varepsilon_t \\
&= \frac{B(\xi + h, \varsigma + k)}{B(\xi, \varsigma)} \int \frac{\nu \varepsilon_t^r}{B(\xi + h, \varsigma + k)} \frac{\exp(\xi + h) \varepsilon_t}{(1 + \exp \varepsilon_t)^{\xi+\varsigma+h+k}} d\varepsilon_t
\end{aligned}$$

■

The information matrix in Proposition 1 is derived as follows. The first derivative with respect to  $\lambda_{t|t-1}$  is:

$$u'_t = -(\xi + \varsigma)[\varepsilon_t^2 b_t(1 - b_t) + \varepsilon_t b_t] + \xi \varepsilon_t$$

Lemma 3 can be used to evaluate

$$E(u'_t) = -(\xi + \varsigma)[E(\varepsilon_t^2 b_t(1 - b_t)) - E(\varepsilon_t b_t) + \frac{\xi}{\xi + \varsigma} E(\varepsilon_t)]$$

$$\begin{aligned}
E(u_t'^2) &= (\xi + \varsigma)^2 [E(\varepsilon_t^2 b_t^2) + E(\varepsilon_t^4 b_t^2 (1 - b_t)^2) + 2E(\varepsilon_t^3 b_t^2 (1 - b_t))] \\
&\quad + \xi^2 E(\varepsilon_t^2) - 2\xi(\xi + \varsigma)[E(\varepsilon_t^3 b_t(1 - b_t)) + E(\varepsilon_t^2 b_t)]
\end{aligned}$$

and  $E(u'_t u_t)$ .

## Acknowledgments

We would like to thank the editor and referees for their helpful comments. Thanks also to Paul Kattuman for supplying the exchange rate data. We are very grateful to the Keynes fund for financial support.

## REFERENCES

- Caivano, M. and Harvey, A.C. (2013) *Time series models with an EGB2 conditional distribution*, Cambridge Working Papers in Economics, CWPE 1325
- Creal, D., Koopman, S.J. and Lucas, A. (2013) Generalized autoregressive score models with applications. *Journal of Applied Econometrics* **28**, 777-95.
- Creal, D., Koopman, S.J. and Lucas, A. (2011) A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations, *Journal of Business and Economic Statistics* **29**, 552-63.
- Davis, P.J. (1964) Gamma function and related functions, in Abramowitz, M. and Stegun, I. A. (eds.). *Handbook of Mathematical Functions*, 253-66, New York: Dover Publications Inc.

- Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events*. Berlin: Springer Verlag.
- Harvey, A. C. (2013) *Dynamic Models for Volatility and Heavy Tails*. Econometric Society Monograph. Cambridge University Press.
- Harvey, A. C. and Luati, A. (2014) Filtering with heavy tails. *Journal of the American Statistical Association* (forthcoming)
- Harvey, A. C. and Sucarrat, G. (2014). EGARCH Models with fat tails, skewness and leverage. *Computational Statistics and Data Analysis* **26**, 320-338.
- Jensen, S. T. and Rahbek, A. (2004) Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**, 1203-26.
- Koopman, S. J. Harvey, A. C., Doornik, J. A. and Shephard, N. (2009) *STAMP 8.2 Structural Time Series Analysis Modeller and Predictor*. London: Timberlake Consultants Ltd.
- Lui, S., Mitchell, J. and Weale, M. (2011) Qualitative business surveys: signal or noise? *J. R. Statist. Soc. A* **174**, 327–348.

- Maronna, R., Martin, D. and Yohai, V. (2006) *Robust Statistics: Theory and Methods*. John Wiley & Sons Ltd.
- 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.
- McNeil, A.J., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management*. Princeton: Princeton University Press.
- Prentice, R. L. (1975) Discrimination among some parametric models. *Biometrika* **62**, 607-614.
- Wang, K-L, Fawson, C., Barrett, C.B. and McDonald, J. B. (2001) A flexible parametric GARCH model with an application to exchange rates. *Journal of Applied Econometrics* **16**, 521-536.
- Yang, X., Frees, E.W. and Zhang, Z. (2011) A generalized beta copula with applications in modeling multivariate long-tailed data. *Insurance: Mathematics and Economics* **49**, 265-84.

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