Exponential Conditional Volatility Models

Andrew Harvey

August 24, 2010

CWPE 1040

Exponential Conditional Volatility Models

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

August 24, 2010

Abstract

The asymptotic distribution of maximum likelihood estimators is derived for a class of exponential generalized autoregressive conditional heteroskedasticity (EGARCH) models. The result carries over to models for duration and realised volatility that use an exponential link function. A key feature of the model formulation is that the dynamics are driven by the score.

KEYWORDS: Duration models; gamma distribution; general error distribution; heteroskedasticity; leverage; score; Student's t.

JEL classification; C22, G17

1 Introduction

Time series models in which a parameter of a conditional distribution is a function of past observations are widely used in econometrics. Such models are termed 'observation driven' as opposed to 'parameter driven'. Leading examples of observation driven models are contained within the class of generalized autoregressive conditional heteroskedasticity (GARCH) models, introduced by Bollerslev (1986) and Taylor (1986). These models contrast with stochastic volatility (SV) models which are parameter driven in that volatility is determined by an unobserved stochastic process. Other examples of observation driven models which are directly or indirectly related to volatility are duration and multiplicative error models (MEMs); see Engle and Russell (1998), Engle (2002) and Engle and Gallo (2006). Like GARCH and SV they are used primarily for financial time series, but for intra-daily data rather than daily or weekly observations.

Despite the enormous effort put into developing the theory of GARCH models, there are still outstanding issues. For example, the parameter restrictions needed to ensure positive variance are not always easy to determine and they can be restrictive. Furthermore, there is no general unified theory for asymptotic distributions of maximum likelihood (ML) estimators. To quote a recent review by Zivot (2009, p 124): 'Unfortunately, verification of the appropriate regularity conditions has only been done for a limited number of simple GARCH models,...'. The class of exponential GARCH, or EGARCH, models proposed by Nelson (1991) takes the logarithm of the conditional variance to be a linear function of the absolute values of past observations and by doing so eliminates the difficulties surrounding parameter restrictions since the variance is automatically constrained to be positive. However, the asymptotic theory remains a problem; see Linton (2008). Apart from some very special cases studied in Straumann (2005), the asymptotic distribution of the ML estimator[1] has not been derived. Furthermore, EGARCH models suffer from a significant practical drawback in that when the conditional distribution is Student's t (with finite degrees of freedom) the observations from stationary models have no moments.

This paper proposes an approach to the formulation of observation driven volatility models that solves many of the existing difficulties. The first ele-

[1]Some progress has been made with quasi-ML estimation applied to the logarithms of squared observations; see Zaffaroni (2010).

ment of the approach is that time-varying parameters (TVPs) are driven by the score. This idea was suggested independently in papers[2] by Creal et al (2010) and Harvey and Chakravarty (2009). Creal et al (2010) went on to develop a whole class of score driven models, while Harvey and Chakravarty (2009) concentrated on EGARCH. However, in neither paper was the asymptotic theory addressed. It is argued here that a key condition for the development of such a theory is that the asymptotic covariance matrix of the ML estimators in the corresponding static model should not depend on parameters that subsequently become time-varying. This condition is not sufficient because certain functions of the score and its derivative must also be independent of TVPs. For the exponential models studied here, this turns out to be the case.

The exponential conditional volatility models considered here have a number of attractions, apart from the fact that their asymptotic properties can be established. In particular, an exponential link function ensures positive scale parameters and enables the conditions for stationarity to be obtained straightforwardly. Furthermore, although deriving a formula for an autocorrelation function (ACF) is less straightforward than it is for a GARCH model, analytic expressions can be obtained and these expressions are more general. Specifically, formulae for the ACF of the (absolute values of) the observations raised to any power can be obtained. Finally, not only can expressions for multi-step forecasts of volatility be derived, but their conditional variances can be also found.

The main result on the asymptotic distribution is set out in section 2. It is shown that the information matrix can be broken down into two parts. One is the information matrix for the static model, while the other is obtained as the expectation of the outer product of first derivatives of the time-varying parameters with respect to the parameters upon which their dynamics depend. Only the first-order dynamic model is considered, but this model corresponds to the GARCH(1,1) specification, which is generally regarded as being adequate for most applications. For the exponential conditional volatility class the outer product matrix depends only on expectations associated with the score and its first derivative in the static model. An analytic expression for the information matrix for the (fixed) parameters in the model is obtained. This expression is independent of TVPs and hence can be shown to be positive definite under clearly defined conditions. The asymptotic distribution

[2] Earlier versions of both papers appeared as discussion papers in 2008.

then follows.

The conditional distribution of the observations in the Beta-t-EGARCH model, introduced by Harvey and Chakravarty (2009), is Student's t with ν degrees of freedom. The volatility is driven by the score, rather than absolute values, and, because the score has a beta distribution, all moments of the observations less than ν exist when the volatility process is stationary. The Beta-t-EGARCH model is reviewed in section 3 and the conditions for the asymptotic theory to go through are set out. The complementary Gamma-GED-EGARCH model is also analyzed.

Section 4 proposes an exponential link function for the conditional mean in gamma and Weibull distributions. As well as setting out the conditions for the asymptotic theory to be valid, expressions for moments, ACFs and multi-step forecasts are derived.

Leverage is introduced into the models in section 5 and the asymptotic results of section 2 are extended to deal with the extended dynamics. Section 6 reports fitting a Beta-t-EGARCH model to daily stock index returns and compares the analytic standard errors with numerical standard errors. The concluding section suggests directions for future research.

2 General model

Let $y_t, t = 1, ..., T$, be a set of time series observations, each of which is drawn from a distribution with probability density function (p.d.f.), $p(y_t; \boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a vector of parameters. When the observations are serially independent, $p(y_t; \boldsymbol{\lambda})$ satisfies the standard regularity conditions for the maximum likelihood estimator, $\widetilde{\boldsymbol{\lambda}}$, to be consistent and asymptotically normal. The information matrix associated with the $t-th$ observation is

$$\mathbf{I}_t(\boldsymbol{\lambda}) = E\left(\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}'}\right) = -E\left(\frac{\partial^2 \ln L_t}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\right), \quad t = 1, ..., T,$$

where $\ln L_t$ is the log-likelihood of the $t-th$ observation. This information matrix is positive definite (p.d.), provided the model is identifiable. The score vector is $\partial \ln L_t/\partial \boldsymbol{\lambda}$.

In the class of models to be considered, some or all of the parameters in $\boldsymbol{\lambda}$ are time-varying, with the dynamics driven by a vector that is equal or proportional to the score. This vector may be the standardized score or a residual, the choice being largely a matter of convenience. The parameters may be connected to more usual parameters by a link function. For example, a parameter may be the logarithm of the variance, rather than the variance itself. Here the link function is exponential in all cases. A crucial requirement for establishing results on asymptotic distributions is that $\mathbf{I}_t(\boldsymbol{\lambda})$ does not depend on parameters in $\boldsymbol{\lambda}$ that are subsequently allowed to be time-varying.

Suppose initially that there is just one parameter, $\lambda = \lambda_{t|t-1}$, which evolves over time as a linear function of past values of the score. Let k be a finite constant and define $u_t = k.\partial \ln L_t / \partial \lambda$. Since u_t is proportional to the score, it is a martingale difference (MD) and it has finite variance because standard regularity conditions hold in the static model. A linear dynamic model of order[3] $(p, q-1)$ is defined as

$$\lambda_{t|t-1} = \delta + \phi_1 \lambda_{t-1|t-2} + ... + \phi_p \lambda_{t-p|t-p-1} + \theta_1 u_{t-1} + ... + \theta_q u_{t-q}, \qquad (1)$$

where $p \geq 0$ and $q \geq 1$ are finite integers and $\delta, \phi_1, .., \phi_p, \theta_1, .., \theta_q$ are (fixed) parameters. The process is assumed to have started in the infinite past. Stationarity (both strict and covariance) of $\lambda_{t|t-1}$ requires that the roots of the autoregressive polynomial lie outside the unit circle, as in an autoregressive-moving average model.

The first-order model,

$$\lambda_{t|t-1} = \delta + \phi \lambda_{t-1|t-2} + \theta u_{t-1}, \qquad (2)$$

is stationary if $|\phi| < 1$, in which case the moving average representation is

$$\lambda_{t|t-1} = \gamma + \theta \sum_{j=1}^{\infty} \phi^{j-1} u_{t-j},$$

where $\gamma = \delta/(1-\phi)$ is the unconditional mean, $E\left(\lambda_{t|t-1}\right)$.

Lemma 1 *Consider a model with a single time-varying parameter,* $\lambda_{t|t-1}$*, which evolves according to a process, such as (1), that depends on variables which are fixed at time* $t-1$*. The process is governed by a set of fixed parameters,* $\boldsymbol{\alpha}$*, which in the case of (1) are* $\delta, \phi_1, .., \phi_p, \theta_1, .., \theta_q$*. Conditional on* $\lambda_{t|t-1}$*, the observations are independently and identically distributed with*

[3] The terminology for the order follows that of Nelson (1991). The notation $\lambda_{t|t-1}$ stresses that $\lambda_{t|t-1}$ is a filter; see also Andersen *et al* (2006).

a positive information scalar, I_t, that in the corresponding static model does not depend on λ. The t-th observation information matrix for $\boldsymbol{\alpha}$ is then

$$\mathbf{I}_t(\boldsymbol{\alpha}) = I_t.\mathbf{D}_t(\boldsymbol{\alpha}), \quad t = 1, ..., T, \tag{3}$$

where

$$\mathbf{D}_t(\boldsymbol{\alpha}) = E\left(\frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}'}\right). \tag{4}$$

Proof. Write the outer product as

$$\mathbf{I}_t(\boldsymbol{\alpha}) = \left(\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}}\right)\left(\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}}\right)' = \left(\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}}\right)^2 \left(\frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}'}\right)$$

Now take expectations conditional on information at time $t-1$. If $E_{t-1}\left(\partial \ln L_t / \partial \lambda_{t|t-1}\right)^2$ does not depend on $\lambda_{t|t-1}$, it is fixed and equal to the unconditional expectation in the static model. Therefore, since $\lambda_{t|t-1}$ is fixed at time $t-1$,

$$E_{t-1}\left[\left(\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}}\right)\left(\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}}\right)'\right] = \left[E\left(\frac{\partial \ln L_t}{\partial \lambda}\right)^2\right] \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\alpha}'}.$$

Taking unconditional expectations gives (3). ■

Lemma 2 *If $\mathbf{D}_t(\boldsymbol{\alpha})$ is time-invariant and p.d., the limiting distribution of $\sqrt{T}\widetilde{\boldsymbol{\alpha}}$, where $\widetilde{\boldsymbol{\alpha}}$ is the ML estimator of $\boldsymbol{\alpha}$, is multivariate normal with mean $\sqrt{T}\boldsymbol{\alpha}$ and covariance matrix*

$$Var(\widetilde{\boldsymbol{\alpha}}) = \mathbf{I}_t^{-1}(\boldsymbol{\alpha}).$$

Proof. See Davidson (2000, pp 271-6), but note that only first derivatives of $\ln L_t$ are needed; see the discussion in van der Vaart (1998). The score, $\partial \ln L_t / \partial \boldsymbol{\alpha}$, is a MD because the score in the static model is a MD and $\partial \lambda_{t|t-1} / \partial \boldsymbol{\alpha}$ is fixed at time $t-1$. ■

In theorem 1 below, the $\mathbf{D}_t(\boldsymbol{\alpha})$ matrix is derived for the first-order model, (2), and shown to be p.d. when the model is identifiable. The complications arise because u_{t-1} depends on $\lambda_{t-1|t-2}$ and hence on the parameters in $\boldsymbol{\alpha}$.

The vector $\partial\lambda_{t|t-1}/\partial\boldsymbol{\alpha}$ is

$$\begin{aligned}
\frac{\partial\lambda_{t|t-1}}{\partial\theta} &= \phi\frac{\partial\lambda_{t-1|t-2}}{\partial\theta} + \theta\frac{\partial u_{t-1}}{\partial\theta} + u_{t-1} \\
\frac{\partial\lambda_{t|t-1}}{\partial\phi} &= \phi\frac{\partial\lambda_{t-1|t-2}}{\partial\phi} + \theta\frac{\partial u_{t-1}}{\partial\phi} + \lambda_{t-1|t-2} \\
\frac{\partial\lambda_{t|t-1}}{\partial\delta} &= \phi\frac{\partial\lambda_{t-1|t-2}}{\partial\delta} + \theta\frac{\partial u_{t-1}}{\partial\delta} + 1.
\end{aligned} \tag{5}$$

However,

$$\frac{\partial u_t}{\partial\theta} = \frac{\partial u_t}{\partial\lambda_{t|t-1}}\frac{\partial\lambda_{t|t-1}}{\partial\theta},$$

and similarly for the other two derivatives. Therefore

$$\begin{aligned}
\frac{\partial\lambda_{t|t-1}}{\partial\theta} &= x_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\theta} + u_{t-1} \\
\frac{\partial\lambda_{t|t-1}}{\partial\phi} &= x_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\phi} + \lambda_{t-1|t-2} \\
\frac{\partial\lambda_{t|t-1}}{\partial\delta} &= x_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\delta} + 1.
\end{aligned} \tag{6}$$

where

$$x_t = \phi + \theta\frac{\partial u_t}{\partial\lambda_{t|t-1}}, \quad t = 1, ..., T. \tag{7}$$

Evaluation of the above derivatives and their squares requires taking conditional expectations of functions of u_t and its first derivative. In the class of exponential conditional volatility models these quantities are independent of $\lambda_{t|t-1}$ and the expectational formulae are as in the corresponding static model. These results, coupled with the fact that the elements of $\partial\lambda_{t|t-1}/\partial\boldsymbol{\alpha}$ are fixed at time $t-1$, enable the information matrix of $\boldsymbol{\alpha}$ to be found.

The following definitions are needed for theorem 1:

$$\begin{aligned}
a &= E_{t-1}(x_t) = E_{t-1}\left(\phi + \theta\frac{\partial u_t}{\partial\lambda_{t|t-1}}\right) = \phi + \theta E_{t-1}\left(\frac{\partial u_t}{\partial\lambda_{t|t-1}}\right) \\
b &= E_{t-1}(x_t^2) = \phi^2 + 2\phi\theta E_{t-1}\left(\frac{\partial u_t}{\partial\lambda_{t|t-1}}\right) + \theta^2 E_{t-1}\left(\frac{\partial u_t}{\partial\lambda_{t|t-1}}\right)^2 \\
c &= E_{t-1}(u_t x_t) = \theta E_{t-1}\left(u_t\frac{\partial u_t}{\partial\lambda_{t|t-1}}\right).
\end{aligned} \tag{8}$$

Note that the first derivative of u_t is (proportional to) the Hessian and so

$$E_{t-1}\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right) = -k.I_t = -\sigma_u^2/k < 0, \tag{9}$$

where $\sigma_u^2 = E_{t-1}(u_t^2)$.

The following lemma is a pre-requisite for theorem 1. The formulae also appear directly in the information matrix when there is a second parameter in the static model that is not allowed to be time-varying; see (17).

Lemma 3 *Suppose that the process for $\lambda_{t|t-1}$ starts in the infinite past. Then, provided $|a| < 1$ and $|\phi| < 1$,*

$$\begin{aligned} E\left(\frac{\partial \lambda_{t|t-1}}{\partial \theta}\right) &= 0, \qquad t = 1, ..., T, && (10) \\ E\left(\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right) &= \frac{\delta}{(1-a)(1-\phi)}, \\ E\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\right) &= \frac{1}{1-a}. && (11) \end{aligned}$$

Proof. Applying the law of iterated expectations (LIE) to (6)

$$E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \theta}\right) = E_{t-2}\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \theta} + u_{t-1}\right) = a\frac{\partial \lambda_{t-1|t-2}}{\partial \theta} + 0$$

and

$$\begin{aligned} E_{t-3}E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \theta}\right) &= aE_{t-3}\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \theta}\right) \\ &= aE_{t-3}\left(x_{t-2}\frac{\partial \lambda_{t-2|t-3}}{\partial \theta} + u_{t-2}\right) = a^2\frac{\partial \lambda_{t-2|t-3}}{\partial \theta} \end{aligned}$$

Hence, if $|a| < 1$,

$$\lim_{n\to\infty} E_{t-n}\left(\frac{\partial \lambda_{t|t-1}}{\partial \theta}\right) = 0, \quad t = 1, ..., T.$$

Taking conditional expectations of $\partial\lambda_{t|t-1}/\partial\phi$ at time $t-2$ gives

$$E_{t-2}\left(\frac{\partial\lambda_{t|t-1}}{\partial\phi}\right) = a\frac{\partial\lambda_{t-1|t-2}}{\partial\phi} + \lambda_{t-1|t-2}. \tag{12}$$

We can continue to evaluate this expression by substituting for $\partial\lambda_{t-1|t-2}/\partial\phi$, taking conditional expectations at time $t-3$, and then repeating this process. However, if a solution is assumed to exist, taking unconditional expectations in (12) gives

$$E\left(\frac{\partial\lambda_{t|t-1}}{\partial\phi}\right) = aE\left(\frac{\partial\lambda_{t-1|t-2}}{\partial\phi}\right) + \frac{\delta}{1-\phi},$$

from which

$$E\left(\frac{\partial\lambda_{t|t-1}}{\partial\phi}\right) = \frac{\delta}{(1-a)(1-\phi)}.$$

As regards δ,

$$E_{t-2}\left(\frac{\partial\lambda_{t|t-1}}{\partial\delta}\right) = a\frac{\partial\lambda_{t-1|t-2}}{\partial\delta} + 1 \tag{13}$$

and taking unconditional expectations gives the result. ∎

Theorem 1 *Assume $\theta \neq 0$, $|\phi| < 1$, $|a| < 1$ and $|b| < 1$. Then $\mathbf{D}_t(\boldsymbol{\alpha})$ is p.d. and the limiting distribution of $\sqrt{T}(\widetilde{\theta},\, \widetilde{\phi},\, \widetilde{\delta})'$ is normal with mean $\sqrt{T}(\theta,\, \phi,\, \delta)'$ and covariance matrix*

$$Var\begin{pmatrix}\widetilde{\theta}\\ \widetilde{\phi}\\ \widetilde{\delta}\end{pmatrix} = \frac{k^2(1-b)}{\sigma_u^2}\begin{bmatrix} A & D & E\\ D & B & F\\ E & F & C\end{bmatrix}^{-1} \tag{14}$$

where

$$
\begin{aligned}
A &= \sigma_u^2 \\
B &= \frac{2a\delta(\delta+\theta c)}{(1-\phi)(1-a)(1-a\phi)} + \frac{1+a\phi}{(1-a\phi)(1-\phi)}\left(\frac{\delta^2}{1-\phi}+\frac{\theta^2\sigma_u^2}{1+\phi}\right) \\
C &= (1+a)/(1-a) \\
D &= \frac{c\delta}{(1-\phi)(1-a)} + \frac{a\theta\sigma_u^2}{1-a\phi} \\
E &= c/(1-a) \\
F &= \frac{\delta - a\delta\phi + a\delta - a^2\delta\phi + a\theta c - a\theta c\phi}{(1-\phi)(1-a)(1-a\phi)}.
\end{aligned}
$$

Proof. First note that I_t in (3) is given by

$$I_t = E_{t-1}(u_t^2)/k^2 = E(u_t^2)/k^2 = \sigma_u^2/k^2 < \infty.$$

This expression is then combined with the formula for $\mathbf{D}_t(\boldsymbol{\alpha})$ which is derived in appendix A. The derivation of the first term, A, is given here to illustrate the method. This term is the unconditional expectation of the square of the first derivative in (6). To evaluate it, first take conditional expectations at time $t-2$, to obtain

$$
\begin{aligned}
E_{t-2}\left(\frac{\partial\lambda_{t|t-1}}{\partial\theta}\right)^2 &= E_{t-2}\left(x_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\theta}+u_{t-1}\right)^2 \\
&= b\left(\frac{\partial\lambda_{t-1|t-2}}{\partial\theta}\right)^2 + 2c\frac{\partial\lambda_{t-1|t-2}}{\partial\theta} + \sigma_u^2. \qquad (15)
\end{aligned}
$$

It was shown in lemma 3 that the unconditional expectation of the second term is zero. Eliminating this term, and taking expectations at $t-3$ gives

$$
\begin{aligned}
E_{t-3}\left(\frac{\partial\lambda_{t|t-1}}{\partial\theta}\right)^2 &= bE_{t-3}\left(x_{t-2}\frac{\partial\lambda_{t-2|t-3}}{\partial\theta}+u_{t-2}\right)^2 + \sigma_u^2 \\
&= b^2\left(\frac{\partial\lambda_{t-2|t-3}}{\partial\theta}\right)^2 + 2cb\frac{\partial\lambda_{t-2|t-3}}{\partial\theta} + b\sigma_u^2 + \sigma_u^2.
\end{aligned}
$$

Again the second term can be eliminated and it is clear that

$$\lim_{n\to\infty} E_{t-n}\left(\frac{\partial\lambda_{t|t-1}}{\partial\theta}\right)^2 = \frac{\sigma_u^2}{1-b}.$$

Taking unconditional expectations in (15) gives the same result. The derivatives are all evaluated in this way in appendix A ■

Remark 1 *Note that a, b and c depend on the model. However, if, as is usually the case, ϕ and θ are positive, then $a < \phi$.*

Corollary 1 *The information matrix when $\phi = \theta = 0$ is*

$$\mathbf{I}_t(\theta,\phi,\delta) = \frac{\sigma_u^2}{k^2}\begin{bmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \delta^2 & \delta \\ 0 & \delta & 1 \end{bmatrix}$$

and so ϕ and δ are not identified. When $\theta \neq 0$, all three parameters are identified even if $\phi = 0$.

Corollary 2 *When ϕ is taken to be unity, but $|b| < 1$ and $|a| < 1$, the information matrix for $\widetilde{\theta}$ and $\widetilde{\delta}$ is*

$$\mathbf{I}(\widetilde{\theta},\widetilde{\delta}) = \frac{\sigma_u^2}{k^2(1-b)}\begin{bmatrix} \sigma_u^2 & \frac{c}{1-a} \\ \frac{c}{1-a} & \frac{1+a}{1-a} \end{bmatrix}$$

with $a = 1 - \theta\sigma_u^2/k$ and

$$b = 1 - 2\theta\sigma_u^2/k + \theta^2 E_{t-1}\left(\frac{\partial u_t}{\partial\lambda_{t|t-1}}\right)^2.$$

Since $I(\widetilde{\theta},\,\widetilde{\delta})$ is p.d., $\sqrt{T}(\widetilde{\theta},\widetilde{\delta})'$ has a limiting normal distribution with mean $\sqrt{T}(\theta,\delta)'$ and covariance matrix $\mathbf{I}^{-1}(\widetilde{\theta},\widetilde{\delta})$. Note that $\theta > 0$ is a necessary condition for $|b| < 1$ and a sufficient condition for $a < 1$.

Lemma 1 can be extended to deal with n parameters in $\boldsymbol{\lambda}$ and a generalization of theorem 1 then follows. The lemma below is for $n = 2$ but this is simply for notational convenience.

Lemma 4 *Suppose that there are two parameters in $\boldsymbol{\lambda}$, but that $\lambda_{j,t|t-1} = f(\boldsymbol{\alpha}_j), j = 1,2$ with the vectors $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ having no elements in common. When the information matrix in the static model does not depend on λ_1 and λ_2*

$$\begin{aligned}\mathbf{I}_t(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) &= E\left[\begin{pmatrix} \frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1} \\ \frac{\partial \ln L_t}{\partial \lambda_2}\frac{\partial \lambda_2}{\partial \boldsymbol{\alpha}_2} \end{pmatrix}\begin{pmatrix} \frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1} \\ \frac{\partial \ln L_t}{\partial \lambda_2}\frac{\partial \lambda_2}{\partial \boldsymbol{\alpha}_2} \end{pmatrix}'\right] \qquad (16)\\ &= \begin{bmatrix} E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\right)^2 E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1}\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1'}\right) & E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \lambda_2}\right) E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1}\frac{\partial \lambda_2}{\partial \boldsymbol{\alpha}_2'}\right) \\ E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \lambda_2}\right) E\left(\frac{\partial \lambda_2}{\partial \boldsymbol{\alpha}_2}\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1'}\right) & E\left(\frac{\partial \ln L_t}{\partial \lambda_2}\right)^2 E\left(\frac{\partial \lambda_2}{\partial \boldsymbol{\alpha}_2}\frac{\partial \lambda_2}{\partial \boldsymbol{\alpha}_2'}\right) \end{bmatrix}.\end{aligned}$$

This above matrix is p.d. if $\mathbf{I}_t(\boldsymbol{\lambda})$ and $\mathbf{D}_t(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)$ are both p.d.

The conditions for the above lemma will rarely be satisfied. A more useful result concerns the case when $\boldsymbol{\lambda}$ contains some fixed parameters. As in theorem 1, it will be assumed that there is only one TVP, but if there are more it is straightforward to combine this result with the previous one.

Lemma 5 *When $\boldsymbol{\lambda}_2$ contains $n-1 \geq 1$ fixed parameters and the terms in the information matrix of the static model that involve λ_1, including cross-products, do not depend on λ_1,*

$$\mathbf{I}_t(\boldsymbol{\alpha}_1, \boldsymbol{\lambda}_2) = \begin{bmatrix} E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\right)^2 E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1}\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1'}\right) & E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2'}\right) E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1}\right) \\ E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2}\right) E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\alpha}_1'}\right) & E\left(\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2'}\right) \end{bmatrix}. \qquad (17)$$

3 Exponential GARCH

The Beta-t-EGARCH and Gamma-GED-EGARCH models are studied in Harvey and Chakravarty (2009). Theoretical properties, such as moments and ACFs, are derived and the first-order model is shown to provide a good fit to daily data on stock indices.

3.1 Beta-t-EGARCH

In the Beta-t-EGARCH model the observations can be written as

$$y_t = \varepsilon_t \exp(\lambda_{t|t-1}/2), \quad t = 1, ..., T, \qquad (18)$$

where the serially independent, zero mean variable ε_t has a $t_\nu-$distribution with positive degrees of freedom, ν.

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

$$u_t = \frac{(\nu+1)y_t^2}{\nu\exp(\lambda_{t|t-1})+y_t^2} - 1, \quad -1 \leq u_t \leq \nu, \quad \nu > 0. \tag{19}$$

This variable is the score multiplied by two. It may be expressed as

$$u_t = (\nu+1)b_t - 1, \tag{20}$$

where, for finite degrees of freedom,

$$b_t = \frac{y_t^2/\nu\exp(\lambda_{t|t-1})}{1+y_t^2/\nu\exp(\lambda_{t|t-1})}, \qquad 0 \leq b_t \leq 1, \quad 0 < \nu < \infty, \tag{21}$$

is distributed as $Beta(1/2, \nu/2)$. Since $E(b_t) = 1/(\nu+1)$ and $Var(b_t) = 2\nu/\{(\nu+3)(\nu+1)^2\}$, u_t has zero mean and variance $2\nu/(\nu+3)$.

Proposition 1 *For a given value of* ν*, the asymptotic covariance matrix of the dynamic parameters is as in (14) with*

$$\begin{aligned} a &= \phi - \theta\frac{\nu}{\nu+3} \\ b &= \phi^2 - 2\phi\theta\frac{\nu}{\nu+3} + \theta^2\frac{3\nu(\nu+1)}{(\nu+5)(\nu+3)} \\ c &= \theta\frac{2\nu(1-\nu)}{(\nu+5)(\nu+3)}. \end{aligned}$$

and $k = 2$.

Proof. Differentiating (19) gives

$$\frac{\partial u_t}{\partial\lambda_{t|t-1}} = \frac{-(\nu+1)y_t^2\nu\exp(\lambda_{t|t-1})}{(\nu\exp(\lambda_{t|t-1})+y_t^2)^2} = -(\nu+1)b_t(1-b_t).$$

From appendix B,

$$E_{t-1}\left[\left(\frac{\partial u_t}{\partial\lambda_{t|t-1}}\right)\right] = -(\nu+1)E(b_t(1-b_t)) = \frac{-\nu}{\nu+3},$$

which is $-\sigma_u^2/2$. For b and c,

$$E_{t-1}\left[\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right)^2\right] = (\nu+1)^2 E(b_t^2(1-b_t)) = \frac{3\nu(\nu+1)}{(\nu+5)(\nu+3)}$$

and

$$\begin{aligned} E_{t-1}\left[u_t\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right)\right] &= -E_{t-1}\left[((\nu+1)b_t-1)(\nu+1)b_t(1-b_t)\right] \\ &= -(\nu+1)^2 E_{t-1}(b_t^2(1-b_t)) + (\nu+1)E_{t-1}(b_t(1-b_t)) \\ &= \frac{-3\nu(\nu+1)}{(\nu+5)(\nu+3)} + \frac{\nu}{\nu+3} = \frac{2\nu(1-\nu)}{(\nu+5)(\nu+3)}. \end{aligned}$$

■

Proposition 2 *The asymptotic distribution of the dynamic parameters given in proposition 1 changes when ν is estimated because the ML estimators of ν and λ are not asymptotically independent in the static model. Specifically*

$$I(\lambda,\nu) = \frac{1}{2}\begin{bmatrix} \frac{\nu}{(\nu+3)} & \frac{1}{(\nu+3)(\nu+1)} \\ \frac{1}{(\nu+3)(\nu+1)} & h(\nu) \end{bmatrix},$$

where

$$h(\nu) = \frac{1}{2}\psi'(\nu/2) - \frac{1}{2}\psi'((\nu+1)/2) - \frac{\nu+5}{\nu(\nu+3)(\nu+1)}$$

and $\psi'(.)$ is the trigamma function; see Taylor and Verbyla (2004). In (17), $\lambda_2 = \alpha_2 = \nu$ and, using (10),

$$I(\boldsymbol{\alpha}_1,\nu) = \begin{bmatrix} \mathbf{I}(\boldsymbol{\alpha}_1) & \frac{1}{2(\nu+3)(\nu+1)}\begin{pmatrix} 0 \\ \frac{\delta}{(1-a)(1-\phi)} \\ 1/(1-a) \end{pmatrix} \\ \frac{1}{2(\nu+3)(\nu+1)}\begin{pmatrix} 0 & \frac{\delta}{(1-a)(1-\phi)} & \frac{1}{1-a} \end{pmatrix} & h(\nu)/2 \end{bmatrix}$$

where $\mathbf{I}(\boldsymbol{\alpha}_1)$ is the inverse of the matrix in (14) with a, b and c as in proposition 1.

Remark 2 *A non-zero median can be introduced into the t-distribution without complicating the asymptotic theory. More generally the median may depend linearly on a set of static exogenous variables, in which case the ML* *estimators of the associated parameters are asymptotically independent of the estimators of $\boldsymbol{\alpha}_1$ and ν.*

Remark 3 *Instead of (18), let*

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

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

$$\sigma^2_{t|t-1} = \delta + \phi\sigma^2_{t-1|t-2} + \theta\sigma^2_{t-1|t-2}u_{t-1}, \quad \delta > 0,\ \phi \geq 0,\ \theta \geq 0$$

where u_t is as in (19); see Harvey and Chakravarty (2009). The model can be re-written as

$$\sigma^2_{t|t-1} = \delta + \beta\sigma^2_{t-1|t-2} + \alpha\sigma^2_{t-1|t-2}(\nu+1)b_{t-1}, \quad \delta > 0, \beta \geq 0,\ \alpha \geq 0.$$

where $\alpha = \theta$ and $\beta = \phi - \theta$. In the limit as $\nu \to \infty$, $(\nu+1)b_t = y_t^2$ leading to the standard $GARCH(1,1)$ specification. The asymptotic result in Theorem 1 does not apply here as the information matrix in the static model depends on σ^2. For a recent discussion of the asymptotics of the Gaussian GARCH(1,1) model see Fiorentina et al. (1996). Note that an analytic expression for the information matrix cannot be obtained.

3.2 Gamma-GED-EGARCH

In the Gamma-GED-EGARCH model[4], $y_t = \varepsilon_t \exp(\lambda_{t|t-1})$ and ε_t has a general error distribution (GED) with positive shape (tail-thickness) parameter υ and scale $\lambda_{t|t-1}$. The log-likelihood function, $\ln L(\boldsymbol{\alpha}, \upsilon)$, is

$$-T\left(1+\upsilon^{-1}\right)\ln 2 - T\ln\Gamma(1+\upsilon^{-1}) - \sum_{t=q+1}^{T}\lambda_{t|t-1} - \frac{1}{2}\sum_{t=q+1}^{T}|y_t|^{\upsilon}\exp(-\lambda_{t|t-1}\upsilon),$$

leading to a model in which $\lambda_{t|t-1}$ evolves as a linear function of the score,

$$u_t = (\upsilon/2)(|y_t|^{\upsilon}/\exp(\lambda_{t|t-1}\upsilon) - 1, \quad t = 1, ..., T. \tag{22}$$

[4]Harvey and Chakravarty (2009) have $y_t = \varepsilon_t \exp(\lambda_{t|t-1}/\upsilon)$.

Hence $\sigma_u^2 = \upsilon$. When $\lambda_{t|t-1}$ is stationary, the properties of the Gamma-GED-EGARCH model can be obtained in much the same way as those of Beta-t-EGARCH. The name Gamma-GED-EGARCH is adopted because $u_t = (\upsilon/2)\varsigma_t - 1$, where $\varsigma_t = |y_t|^{\upsilon} / \exp(\lambda_{t|t-1}\upsilon)$ has a *gamma*$(1/2,\ 1/\upsilon)$ distribution; see expression (37) in appendix D.

Proposition 3 *For a given value of υ, the asymptotic[5] covariance matrix of the dynamic parameters is as in (14) with $k = 1$ and*

$$\begin{aligned} a &= \phi - \theta\upsilon \\ b &= \phi^2 - 2\phi\theta\upsilon + \theta^2(\upsilon + 1) \\ c &= -\theta\upsilon^2. \end{aligned}$$

Proof. The derivative of u_t with respect to $\lambda_{t|t-1}$ is

$$\frac{\partial u_t}{\partial \lambda_{t|t-1}} = -(\upsilon^2/2)\,|y_t|^{\upsilon} / \exp(\lambda_{t|t-1}\upsilon) = -(\upsilon^2/2)\varsigma_t$$

and taking conditional expectations gives $-\upsilon$, which is $-\sigma_u^2$. In addition,

$$E_{t-1}\left[\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right)^2\right] = -(\upsilon^2/2)^2 E_{t-1}\left(\varsigma_t^2\right) = -(\upsilon + 1)$$

and

$$E_{t-1}\left[u_t\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right)\right] = -\upsilon(\upsilon + 1) + \upsilon = -\upsilon^2.$$

Important special cases are the normal distribution, $\upsilon = 2$, and the Laplace distribution, $\upsilon = 1$. However, as with the Beta-t-EGARCH model, the asymptotic distribution of the dynamic parameters changes when υ is estimated since the ML estimators of υ and λ are not asymptotically independent in the static model. ■

[5] When $\upsilon < 2$ the pdf at $y = 0$ does not satisfy all the regularity conditions for the usual ML properties to hold; see Varanasi and Aazhang (1989, p 1408-9). However, this irregularity only affects the mean which here is assumed to be given. Note that Varanasi and Aazhang refer to the GED as the generalized Gaussian distribution.

Proposition 4 *The information matrix for the GED distribution is* [6]

$$\mathbf{I}(\lambda,\upsilon)=\begin{bmatrix} \upsilon & -\upsilon^{-1}(1+0.5\Gamma(2/\upsilon)\psi(2/\upsilon)/\Gamma(1/\upsilon)) \\ -\upsilon^{-1}(1+0.5\Gamma(2/\upsilon)\psi(2/\upsilon)/\Gamma(1/\upsilon)) & \upsilon^{-3}\ln 2+g(\upsilon)+h(\upsilon) \end{bmatrix},$$

where $\psi(.)$ is the digamma function,

$$g(\upsilon)=2\upsilon^{-3}\psi\left(1+1/\upsilon\right)+\upsilon^{-4}\psi'\left(1+1/\upsilon\right)$$

and

$$h(\upsilon)=\frac{\Gamma(2/\upsilon)((\psi(2/\upsilon))^2-\psi'(2/\upsilon))}{2\upsilon^2\Gamma(1/\upsilon)}.$$

When $\lambda_2=\alpha_2=\upsilon$, $I(\boldsymbol{\alpha}_1,\upsilon)$ has the same form as $I(\boldsymbol{\alpha}_1,\nu)$ in proposition 2.

Proof. See appendix D. ■

Remark 4 *In the equation for the logarithm of the conditional variance, $\sigma^2_{t|t-1}$, in the Gaussian EGARCH model (without leverage) of Nelson (1991), u_t is replaced by $[|z_t|-E\,|z_t|]$ where $z_t=y_t/\sigma_{t|t-1}$. The difficulties arise because, unless $\upsilon=1$, the conditional expectation of $[|z_t|-E\,|z_t|]$ depends on $\sigma_{t|t-1}$.*

4 Intra-daily data: realized volatility and duration

Engle (2002) introduced a class of multiplicative error models for modeling non-negative variables, such as duration, realized volatility and spreads. In these models, the conditional mean, $\mu_{t|t-1}$, and hence the conditional scale, is a GARCH-type process and the observations can be written

$$y_t=\varepsilon_t\mu_{t|t-1},\qquad 0\leq y_t<\infty,\quad t=1,....,T,$$

where ε_t has a distribution with mean one. The leading cases are the gamma and Weibull distributions. Both include the exponential distribution as a special case.

[6] Varanasi and Aazhang (1989) give some formulae but not for the exponential link.

The use of an exponential link function, $\mu_{t|t-1} = \exp(\lambda_{t|t-1})$, not only ensures that $\mu_{t|t-1}$ is positive, but also allows theorem 1 to be applied. The model can be written

$$y_t = \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, ..., T, \tag{23}$$

with dynamics as in (1).

4.1 Gamma distribution

The *pdf* of a *gamma* variable can be written as

$$f(y; \mu, \gamma) = \gamma^\gamma \mu^{-\gamma} y^{\gamma-1} e^{-\gamma y/\mu} / \Gamma(\gamma), \quad 0 \leq y < \infty, \quad \mu, \gamma > 0,$$

where γ is the shape parameter, μ is the mean and the variance is μ^2/γ; see, for example, Engle and Gallo (2006). The exponential distribution is a special case in which $\gamma = 1$. The scale parameter in the parameterization of appendix C is $\alpha = \gamma/\mu$.

The exponential link function gives a score of

$$u_t = (y_t - \exp(\lambda_{t|t-1})) / \exp(\lambda_{t|t-1}) \tag{24}$$

with $\sigma_u^2 = 1/\gamma$. Hence the asymptotic variance of the ML estimator of μ in the static model is independent of μ.

Proposition 5 *The ML estimators of the parameters δ, ϕ and θ in (23) and (2) are asymptotically normal with covariance matrix as in (14) with*

$$\begin{aligned} a &= \phi - \theta \\ b &= \phi^2 - 2\phi\theta + \theta^2(1+\gamma)/\gamma \\ c &= -\theta/\gamma \end{aligned}$$

and $k = 1/\gamma$. The result is unchanged if γ is estimated.

Proof. Since

$$\frac{\partial u_t}{\partial \lambda_{t|t-1}} = -y_t \exp(-\lambda_{t|t-1})$$

is gamma distributed, its conditional expectation is minus one (as given by

$-\sigma_u^2/k)$. Furthermore

$$E_{t-1}\left[\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right)^2\right] = E_{t-1}\left[(-y_t \exp(-\lambda_{t|t-1}))^2\right] = (1+\gamma)/\gamma,$$

and

$$\begin{aligned} & E_{t-1}\left[u_t\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right)\right] = E_{t-1}\left[-y_t \exp(-\lambda_{t|t-1})((y_t - \exp(\lambda_{t|t-1}))/\exp(\lambda_{t|t-1}))\right] \\ = \ & E_{t-1}\left[-y_t^2 \exp(-2\lambda_{t|t-1})) + (y_t - \exp(\lambda_{t|t-1})\right] = -1/\gamma. \end{aligned}$$

The independence of the ML estimators of λ and γ follows on noting that $E(\partial^2 \ln L_t/\partial\lambda\partial\gamma) = 0$. ■

4.1.1 Moments and ACF

Proposition 6 *For the gamma model defined by (23) and $\lambda_{t|t-1}$ generated by a stationary process with mean ω, that is*

$$\lambda_{t|t-1} = \omega + \sum_{j=1}^{\infty} \psi_j u_{t-j}, \tag{25}$$

with $\psi_j, j = 1, 2, ..$ fixed and u_t as in (24), the $m-th$ moment exists if and only if $\psi_j < \gamma/2m$, for all $j = 1, 2, ...$, and is given by the expression

$$E\left(y_t^m\right) = \frac{\Gamma(m+\gamma)}{\gamma^m \Gamma(\gamma)} e^{m(\omega - \Sigma\psi_j)} \prod_{j=1}^{\infty} (1 - m\psi_j/\gamma)^{-\gamma}, \qquad \psi_j < \gamma/m, \qquad m > 0. \tag{26}$$

Proof.

$$E_{t-1}\left(y_t^m\right) = E_{t-1}\left(\varepsilon_t^m\right).e^{m\lambda_{t|t-1}} = E_{t-1}\left(\varepsilon_t^m\right).e^{m(\omega - \Sigma\psi_j)} \prod_{j=1}^{\infty} e^{m\psi_j \varepsilon_{t-j}}$$

as $u_t = \varepsilon_t - 1$. It follows from the formula for the moment generating function of a standardized gamma distribution that, if k is a constant,

$$E(e^{k\varepsilon}) = (1 - k/\gamma)^{-\gamma}, \quad -\infty < k < \gamma, \quad \gamma > 0. \tag{27}$$

Hence the result is given by taking iterated expectations of the product term and substituting for $E_{t-1}(\varepsilon_t^m) = E(\varepsilon_t^m)$ from the formula in appendix D. ■

Corollary 3 *The level increases by a factor of* $\Pi_{j=1}^{\infty} e^{-\psi_j}(1-\psi_j/\gamma)^{-\gamma}$, *while the variance increases by* $\Pi_{j=1}^{\infty} e^{-2\psi_j}(1-2\psi_j/\gamma)^{-\gamma}$. *Just as the increase in kurtosis minus one can be taken as a measure of volatility in a GARCH model, so one subtracted from the increase in variance divided by the square of the increase in level can be taken as a measure of volatility here.*

Proposition 7 *When* $\lambda_{t|t-1}$ *is covariance stationary and* $\psi_j < \gamma/2c$, $j = 1, 2, ...,$ *the ACF of* y_t^c *is*

$$\rho(\tau) = \frac{G_\gamma(\tau) - 1}{\kappa_c V_\gamma - 1}, \quad \tau = 1, 2, ..., \tag{28}$$

where

$$\kappa_c = \frac{E(\varepsilon_t^{2c})}{(E(\varepsilon_t^c))^2} = \frac{\Gamma(2c+\gamma)\Gamma(\gamma)}{(\Gamma(c+\gamma))^2}, \qquad c > 0,$$

$$V_\gamma = \left(\prod_{j=1}^{\infty}(1 - c\psi_j/\gamma)^{-\gamma}\right)^{-2} \prod_{j=1}^{\infty}(1 - 2c\psi_j/\gamma)^{-\gamma},$$

and $G_v(\tau, c)$, $\tau = 2, .,$ *is*

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

or, for $\tau = 1$,

$$G_v(1, c) = (1-c\psi_1/\gamma)^{-(c+1)\gamma}\left(\prod_{j=1}^{\infty}(1 - c\psi_j/\gamma)^{-\gamma}\right)^{-2} \prod_{i=1}^{\infty}(1-c(\psi_{\tau+i}+\psi_i)/\gamma)^{-\gamma}.$$

Proof. The autocorrelations of the powers of a stationary model are

$$\rho(\tau; y_t^c) = \frac{E\left(y_t^c y_{t-\tau}^c\right) - E(y_t^c)\, E\left(y_{t-\tau}^c\right)}{E(y_t^{2c}) - E(y_t^c)\, E(y_{t-\tau}^c)}.$$

Using (26) with $m = c$ gives all the terms except $E\left(y_t^c y_{t-\tau}^c\right)$. Evaluating this expression is not straightforward because of the dependence between $e^{\lambda_{t|t-1}}$

and $\varepsilon^c_{t-\tau}$. Following the argument in Harvey and Chakravarty (2009) gives the result. ■

The autocorrelations for $\ln y_t$, which corresponds to $c = 0$, are easily derived because $\ln y_t = \lambda_{t|t-1} + \ln \varepsilon_t$.

4.1.2 Forecasts

When $\lambda_{t|t-1}$ has a moving average representation, as in (25), it follows from standard prediction theory that the optimal estimator of

$$\lambda_{T+\ell|T+\ell-1} = \omega + \sum_{j=1}^{\ell-1} \psi_j u_{T+\ell-j} + \sum_{k=0}^{\infty} \psi_{\ell+k} u_{T-k}$$

is its conditional expectation

$$\lambda_{T+\ell|T} = \omega + \sum_{k=0}^{\infty} \psi_{\ell+k} u_{T-k}, \quad \ell = 1, 2, 3, .. \tag{29}$$

Proposition 8 *The optimal (MMSE) predictor of the level, assuming that* $\psi_j < \gamma$, $j = 1, 2, ...,$ *is*

$$\mu_{T+\ell|T} = E_T\left(e^{\lambda_{T+\ell|T+\ell-1}}\right) = e^{\lambda_{T+\ell|T}} \prod_{j=1}^{\ell-1} e^{-\psi_j}(1 - \psi_j/\gamma)^{-\gamma}, \quad \ell = 2, 3, .. \tag{30}$$

The volatility of the volatility, what Engle (2002, sect 5) calls the VoV*, is, for* $\psi_j < \gamma/2$, $j = 1, 2, ...,$

$$\begin{aligned} VoV(\ell) &= E_T\left(e^{2\lambda_{T+\ell|T+\ell-1}}\right) - \left(E_T\left(e^{\lambda_{T+\ell|T+\ell-1}}\right)\right)^2, \quad \ell = 1, 2, 3, .. \\ &= e^{2\lambda_{T+\ell|T}} \prod_{j=1}^{\ell-1} e^{-2\psi_j}(1 - 2\psi_j/\gamma)^{-\gamma} - \mu^2_{T+\ell|T}. \end{aligned}$$

The optimal (MMSE) predictor of the observation at $T+\ell$*, that is* $E_T(y_{T+\ell})$*, is the same as* $\mu_{T+\ell|T}$*, since* $E_{T+\ell-1}(\varepsilon_{T+\ell}) = 1$*. The optimal predictor of the variance of* $y_{T+\ell}, \ell = 2, 3, ..,$ *is*

$$Var_T(y_{T+\ell}) = (1+\gamma)\gamma^{-1} e^{2\lambda_{T+\ell|T}} \prod_{j=1}^{\ell-1} e^{-2\psi_j}(1 - 2\psi_j/\gamma)^{-\gamma} - \left(E_T(y_{T+\ell})\right)^2.$$

Remark 5 *Since $e^{-\psi_j}(1-\psi_j/\gamma)^{-\gamma} > 1$ for $\psi_j \neq 0$, the forecast function, $\mu_{T+\ell|T}$, will converge to a level above $e^{\lambda_{T+\ell|T}}$.*

4.2 Weibull distribution

The pdf of a Weibull distribution is

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

where α is the scale parameter and ν is the shape parameter. The mean is $\mu = \alpha\Gamma(1+1/\nu)$ and the variance is $\alpha^2\Gamma(1+2/\nu) - \mu^2$. Again it is convenient to parameterize in terms of the mean so that when the scale is time-varying

$$f(y_t) = (\nu/y_t)w_t \exp\left(-w_t\right), \quad 0 \leq y_t < \infty, \quad \nu > 0,$$

where

$$w_t = \left(\frac{y_t\Gamma(1+1/\nu)}{\mu_{t|t-1}}\right)^{\nu}, \quad t = 1,...,T.$$

The exponential link function, $\mu_{t|t-1} = \exp(\lambda_{t|t-1})$, yields the log-likelihood function

$$\ln L = \ln\nu - \ln y_t \; - \nu\ln(y_t\Gamma(1+1/\nu)e^{-\lambda_{t|t-1}}) - (y_t\Gamma(1+1/\nu)e^{-\lambda_{t|t-1}})^{\nu}.$$

Hence the score is

$$\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} = -\nu + \nu(y_t\Gamma(1+1/\nu)e^{-\lambda_{t|t-1}})^{\nu} = -\nu + \nu w_t.$$

A convenient choice for u_t in (1) is

$$u_t = w_t - 1, \quad t = 1,...,T,$$

and since w_t has a standard exponential distribution, $E(u_t) = 0$ and $\sigma_u^2 = 1$.

The first-order dynamic model for $\lambda_{t|t-1}$ is of the same form as (2) and the following result applies.

Proposition 9 *For a given value of ν, the ML estimators of the parameters*

δ, ϕ and θ are asymptotically normal with covariance matrix as in (14) with

$$\begin{aligned} a &= \phi - \theta\nu \\ b &= \phi^2 - 2\phi\theta\nu + 2\theta^2\nu^2 \\ c &= -\theta\nu \end{aligned}$$

and $k = 1/\nu$.
Proof.

$$E_{t-1}\left[\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right] = -\nu E_{t-1}[(y_t e^{-\lambda_{t|t-1}}\Gamma(1+1/\nu))^{\nu}] = -\nu E_{t-1}(w_t) = -\nu$$

while

$$E_{t-1}\left[\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right]^2 = \nu^2 E_{t-1}\left[(y_t e^{-\lambda_{t|t-1}}\Gamma(1+1/\nu))^{2\nu}\right] = \nu^2 E_{t-1}(w_t^2) = 2\nu^2$$

and

$$E_{t-1}\left[u_t \frac{\partial u_t}{\partial \lambda_{t|t-1}}\right] = -\nu E_{t-1}\left[w_t^2 - w_t\right] = -\nu.$$

■

In contrast to the gamma case, estimation of the shape parameter does make a difference to the asymptotic distribution since the information matrix for λ and ν in the static model is not diagonal. The inverse of the information matrix for λ and ν, derived in appendix E, is

$$I^{-1}(\lambda, \nu) = \begin{bmatrix} 1.1087\nu^{-2} & -0.2570 \\ -0.2570 & 0.6079\nu^2 \end{bmatrix}.$$

Expressions for the moments, ACF and forecasts are obtained as follows.

Proposition 10 *Assuming that, in (25), $\psi_j < 1/m$, $j = 1, 2, ...$,*

$$E(y_t^m) = \frac{\Gamma(1+m/\nu)}{(\Gamma(1+1/\nu))^m} e^{m(\omega - \Sigma\psi_j)} \prod_{j=1}^{\infty}(1 - m\psi_j)^{-1}, \quad m > 0.$$

Proof. Taking conditional expectations

$$E_{t-1}(y_t^m) = \frac{\Gamma(1+m/\nu)}{(\Gamma(1+1/\nu))^m} e^{m\lambda_{t|t-1}}.$$

Since $\lambda_{t|t-1}$ depends on exponential variables its unconditional expectation is as in (27) with $\gamma = 1$. ■

Proposition 11 *The ACF of y_t^c is as in (28) with $\gamma = 1$ and*

$$\kappa_c = \frac{E\left(\varepsilon_t^{2c}\right)}{\left(E\left(\varepsilon_t^c\right)\right)^2} = \frac{\Gamma(1+2c/\nu)}{(\Gamma(1+c/\nu))^2}, \qquad c > 0.$$

Proposition 12 *The MMSE of the level and the $VoV(\ell)$ are as in proposition 8 with $\gamma = 1$. The MMSE of $y_{T+\ell}, \ell = 2, 3, ..,$ is the same as the level while*

$$Var_T(y_{T+\ell}) = \frac{\Gamma(1+2/\nu)}{\left(\Gamma(1+1/\nu)\right)^2} e^{2\lambda_{T+\ell|T}} \prod_{j=1}^{\ell-1} e^{-2\psi_j}(1-2\psi_j)^{-1} - \left(E_T\left(y_{T+\ell}\right)\right)^2.$$

5 Leverage

The standard way of incorporating leverage effects into GARCH models is by including a variable in which the squared observations are multiplied by an indicator taking the value one for $y_t < 0$ and zero otherwise; see Taylor (2005, pp. 220-1). In the Beta-t-EGARCH and Gamma-GED-EGARCH models this additional variable is constructed by multiplying $u_t + 1$ by the indicator. Alternatively, the sign of the observation may be used, so the first-order model, (2), becomes

$$\lambda_{t|t-1} = \delta + \phi\lambda_{t-1|t-2} + \theta u_{t-1} + \theta^* sgn(-y_{t-1})(u_{t-1}+1). \tag{31}$$

Taking the sign of *minus* y_t means that the parameter θ^* is normally non-negative for stock returns[7]. With the above parameterization $\lambda_{t|t-1}$ is driven by a MD. The mean of $\lambda_{t|t-1}$ is as before, but

$$E(\lambda_{t|t-1}^2) = \delta^2/(1-\phi)^2 + \theta^2\sigma_u^2/(1-\phi^2) + \theta^{*2}(\sigma_u^2+1)/(1-\phi^2). \tag{32}$$

Engle and Gallo (2006) estimate their MEM models with leverage. Such effects may be introduced into gamma and Weibull exponential models using

[7]Although the statistical validity of the model does not require it, the restriction $\theta \geq \theta^* \geq 0$ may be imposed in order to ensure that an increase in the absolute values of a standardized observation does not lead to a decrease in volatility.

(31). The results on ACFs may be extended to deal with leverage in the same way as for the Beta-t-EGARCH and Gamma-GED-EGARCH models; see Harvey and Chakravarty (2009, sect 3.4, sect. 4)

Theorem 2 *The asymptotic covariance matrix of the ML estimator of the parameters in (31), assuming that the parameter ν or υ is known, is*

$$Var\begin{pmatrix}\widetilde{\theta}\\ \widetilde{\phi}\\ \widetilde{\delta}\\ \widetilde{\theta}^{*}\end{pmatrix}=\frac{k^{2}(1-b^{*})}{\sigma_{u}^{2}}\begin{bmatrix}A & D & E & 0\\ D & B^{*} & F^{*} & D^{*}\\ E & F^{*} & C & E^{*}\\ 0 & D^{*} & E^{*} & A^{*}\end{bmatrix}^{-1} \tag{33}$$

where A, C, D and E are as in (14), F^{*} is F with θc expanded to become $\theta c+\theta^{*}c^{*}$,

$$\begin{aligned}
A^{*} &= \sigma_{u}^{2}+1\\
B^{*} &= \frac{2a\delta(\delta+\theta c)}{(1-\phi)(1-a)(1-a\phi)}+\frac{1+a\phi}{(1-a\phi)(1-\phi)}\left(\frac{\delta^{2}}{1-\phi}+\frac{\theta^{2}\sigma_{u}^{2}}{1+\phi}+\frac{\theta^{*2}(\sigma_{u}^{2}+1)}{1+\phi}\right)\\
E^{*} &= c^{*}/(1-a)\\
D^{*} &= \frac{\delta c^{*}}{(1-\phi)(1-a)}+\frac{a\theta^{*}(\sigma_{u}^{2}+1)}{1-a\phi}
\end{aligned}$$

with

$$b^{*}=\phi^{2}-2\theta\phi\sigma_{u}^{2}/k+(\theta^{2}+\theta^{*2})E_{t-1}\left(\frac{\partial u_{t}}{\partial\lambda_{t|t-1}}\right)^{2},$$

and

$$c^{*}=\theta^{*}E_{t-1}\left(u_{t}\frac{\partial u_{t}}{\partial\lambda_{t|t-1}}\right)+\theta^{*}E_{t-1}\left(\frac{\partial u_{t}}{\partial\lambda_{t|t-1}}\right).$$

Proof.

$$\begin{aligned}
\frac{\partial\lambda_{t|t-1}}{\partial\theta^{*}} &= \phi\frac{\partial\lambda_{t-1|t-2}}{\partial\theta^{*}}+\theta\frac{\partial u_{t-1}}{\partial\theta^{*}}+\theta^{*}sgn(-y_{t-1})\frac{\partial u_{t-1}}{\partial\theta^{*}}+sgn(-y_{t-1})(u_{t-1}+1)\\
&= x_{t-1}^{*}\frac{\partial\lambda_{t-1|t-2}}{\partial\theta^{*}}+sgn(-y_{t-1})(u_{t-1}+1)
\end{aligned}$$

where

$$x_{t}^{*}=\phi+(\theta+\theta^{*}sgn(-y_{t}))\frac{\partial u_{t}}{\partial\lambda_{t|t-1}}$$

Since y_t is symmetric and u_t depends only on y_t^2, $E(sgn(-y_{t-1})(u_{t-1}+1)) = 0$, and so

$$E(\frac{\partial\lambda_{t|t-1}}{\partial\theta^*}) = 0.$$

The derivatives in (5) are similarly modified by the addition of the derivatives of the leverage term, so x_t^* replaces x_t in all cases. However

$$E_{t-1}(x_t^*) = \phi + E_{t-1}\left((\theta + \theta^* sgn(-y_t))\frac{\partial u_t}{\partial\lambda_{t|t-1}}\right) = a$$

and the formulae for the expectations in (10) are unchanged.

The expected values of the squares and cross-products in the extended information matrix are obtained in much the same way as in appendix A. Note that

$$\begin{aligned} x_t^* sgn(-y_t)(u_t+1) &= (\phi + ((\theta + \theta^* sgn(-y_t))\frac{\partial u_t}{\partial\lambda_{t|t-1}})(sgn(-y_t)(u_t+1)) \\ &= (\phi + \theta\frac{\partial u_t}{\partial\lambda_{t|t-1}})(sgn(-y_t)(u_t+1) + \theta^*\frac{\partial u_t}{\partial\lambda_{t|t-1}}(u_t+1) \end{aligned}$$

so

$$c^* = E_{t-1}(x_t^* sgn(-y_t)(u_t+1)) = \theta^* E_{t-1}\left(\frac{\partial u_t}{\partial\lambda_{t|t-1}}(u_t+1)\right).$$

The formula for b^* is similarly derived. Further details can be found in appendix F. ■

Corollary 4 *When ν is estimated by ML in the Beta-t-EGARCH model, the asymptotic covariance matrix of the full set of parameters is given by proposition 2 with $\mathbf{I}(\boldsymbol{\alpha}_1)$ as in (33). The asymptotic covariance matrices for the Gamma-GED-EGARCH model with υ estimated and the Weibull model with ν estimated are similarly obtained.*

6 Daily Hang-Seng and Dow-Jones returns

First-order Beta-t-EGARCH models were estimated by Harvey and Chakravarty (2009) for the daily de-meaned returns of two stock market indices, the Dow Jones Industrial Average and the Hang Seng. The Dow-Jones data runs from 1st October 1975 to 13th August 2009, giving $T = 8548$ returns. The

Hang Seng runs from 31st December 1986 to 10th September 2009, giving $T = 5630$. The ML estimates and associated numerical standard errors (SEs) reported in Harvey and Chakravarty (2009) are shown in table 1 below. The asymptotic SEs are close to the numerical SEs. A Wald test on the leverage parameter, θ^*, shows it to be significantly different from zero for both series. The values of a and b are also given in table 1: both are well below the corresponding value of ϕ (which is less than one).

	Hang Seng		**DOW-JONES**	
Parameter	Estimates (num. SE)	Asy. SE	Estimates (num. SE)	Asy. SE
δ	0.006 (0.002)	0.0018	-0.005 (0.001)	0.0026
ϕ	0.993 (0.003)	0.0017	0.989 (0.002)	0.0028
θ	0.093 (0.008)	0.0073	0.060 (0.005)	0.0052
θ^*	0.042 (0.006)	0.0054	0.031 (0.004)	0.0038
ν	5.98 (0.45)	0.355	7.64 (0.56)	0.475
a	.931		.946	
b	.876		.898	

Table 1 Estimates and standard errors for Beta-t-EGARCH models

Harvey and Chakravarty (2009) also estimate the Gamma-GED-EGARCH model. However, the fit is not as good. In general the Beta-t-EGARCH model seems to be a better choice.

7 Conclusions

This article has established the asymptotic distribution of maximum likelihood estimators for a class of exponential volatility models and provided an analytic expression for the asymptotic covariance matrix. The models include a modification of EGARCH that retains all the advantages of the original EGARCH while eliminating disadvantages such as the absence of moments for a conditional t-distribution. The asymptotics carry over to models for duration and realized volatility by simply employing an exponential link function. The unified theory is attractive in its simplicity. Only the first-order model has been analyzed, but this model is the one used in most situations. Clearly there is work to be done to extend the results to more general dynamics.

The analysis shows that stationarity of the (first-order) dynamic equation is not sufficient for the asymptotic theory to be valid. However, it will be sufficient in most situations and the other conditions are easily checked. If a unit root is imposed on the dynamic equation the asymptotic theory can still be established.

A key feature of the model formulation is that the dynamics are driven by the score. The associated Lagrange multiplier portmanteau tests for the presence of volatility similarly depend on residuals derived from the score. The properties of such tests are currently under investigation. The residuals derived from the Beta-t-EGARCH model have the advantage of being more robust to outliers than tests based on squares, or even absolute values, and this property may well translate into increased power in many practical situations.

The analytic expression obtained for the information matrix establishes that it is positive definite. This is crucial in demonstrating the validity of the asymptotic distribution of the ML estimators. In practice, numerical derivatives may be used for computing ML estimates and at present this is necessary for higher order models. However, the analytic information matrix for the first-order model may be of value in enabling ML estimates to be computed rapidly, by the method of scoring, as well as in providing accurate estimates of asymptotic standard errors; see the comments made by Fiorentini *et al* (1996) in the context of GARCH estimation.

When observations are from a Gamma–GED-EGARCH model, the standardized observations, $y_t \exp(-\lambda_{t|t-1})$, have a $gamma(1/2, 1/v)$ distribution when their absolute values are raised to the power v. This link suggests a rationale for using the gamma distribution for certain types of non-negative observations, for example those derived from squares or absolute values. Given the attractions of Beta-t-EGARCH and the fact that $t_\nu^2 = F_{1,\nu}$, a model for non-negative observations in which ε_t in (23) is distributed as $F_{1,\nu}$ may be worthy of consideration. In other words if a variable is similar to the square of an observation from Beta-t-EGARCH, a conditional $F_{1,\nu}$ is appropriate. More generally ε_t may be modeled as an F-distribution with (ν_1, ν_2) degrees of freedom. The score has a beta distribution[8] and theorem 1 still applies. The theory for deriving moments, ACFs and forecasts is similar to that for Beta-t-EGARCH. Although the argument from squaring Beta-t-EGARCH

[8] For an $F(\nu_1, \nu_2)$ distribution, the score is $(\nu_1 + \nu_2)b_t/2 - \nu_1/2$, where b_t has a $beta(\nu_1/2, \nu_2/2)$ distribution.

observations suggests that the F-distribution be applied in MEM models, the versatility of the F-distribution may make it applicable to duration data as well; Gonzalez-Rivera et al (2010, sect 5) have recently highlighted the failure of the usual distributions to capture all features of such data.

The fact that we have a formula for the expectation of (the absolute values of) observations raised to any positive power (not just integers) offers the possibility of approximating the unconditional distribution of the observations; see Stuart and Ord (1987, ch 4). The conditional distribution of the forecasts of volatility may be similarly approximated. The importance of being able to make such approximations is discussed in Engle (2002, sect 5). Investigation of this issue is a topic for future research as is the possibility of extending the asymptotics to other models.

Acknowledgements I am grateful to Tirthankar Chakravarty, Siem-Jan Koopman, Gloria Gonzalez-Rivera, Richard Smith, Esther Ruiz and Paolo Zaffaroni for helpful discussions and comments. Of course any errors are mine.

APPENDIX

A Derivation of the formulae for theorem 1

The LIE is used to evaluate the outer product form of the $\mathbf{D}_t(\boldsymbol{\alpha})$ matrix, as in (4). The formula for θ was derived in the main text. For ϕ

$$\begin{aligned} E_{t-2}\left(\frac{\partial\lambda_{t|t-1}}{\partial\phi}\right)^2 &= E_{t-2}\left(x_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\phi}+\lambda_{t-1|t-2}\right)^2 \qquad (34) \\ &= b\left(\frac{\partial\lambda_{t-1|t-2}}{\partial\phi}\right)^2+\lambda_{t-1|t-2}^2+2a\frac{\partial\lambda_{t-1|t-2}}{\partial\phi}\lambda_{t-1|t-2} \end{aligned}$$

The unconditional expectation of the last term is found by writing (shifted forward one period)

$$\begin{aligned} E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \phi}\lambda_{t|t-1}\right) &= E_{t-2}\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}+\lambda_{t-1|t-2}\right)(\phi\lambda_{t-1|t-2}+\delta+\theta u_{t-1}) \\ &= \phi E_{t-2}\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\lambda_{t-1|t-2}\right)+\phi\lambda^2_{t-1|t-2}+\delta E_{t-2}\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\right) \\ & \quad +\delta\lambda_{t-1|t-2}+\theta E_{t-2}\left(u_{t-1}x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\right)+\theta E_{t-2}(u_{t-1}\lambda_{t-1|t-2}) \end{aligned}$$

The last term is zero. Taking unconditional expectations and substituting for $E(\lambda_{t|t-1})$ gives

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \phi}\lambda_{t|t-1}\right)=\frac{\phi E(\lambda^2_{t|t-1})}{1-a\phi}+\frac{\gamma(\delta+\theta c)}{(1-a)(1-a\phi)} \tag{35}$$

Taking unconditional expectations in (34) and substituting from (35) gives

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right)^2=bE\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\right)^2+E(\lambda^2_{t|t-1})+\frac{2a\phi E(\lambda^2_{t|t-1})}{1-a\phi}+\frac{2a\delta(\delta+\theta c)}{(1-a)(1-\phi)(1-a\phi)}$$

which leads to B on substituting for

$$E\left(\lambda^2_{t|t-1}\right)=\delta^2/(1-\phi)^2+\sigma_u^2\theta^2/(1-\phi^2).$$

Now consider δ

$$E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\right)^2=b\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\right)^2+2a\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\right)+1.$$

Unconditional expectations give

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\right)^2=\frac{1+a}{(1-a)(1-b)}$$

As regards the cross-products

$$
\begin{aligned}
E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \theta}\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right) &= E_{t-2}\left[\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \theta}+u_{t-1}\right)\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}+\lambda_{t-1|t-2}\right)\right] \\
&= E_{t-2}\left[x_{t-1}^2\frac{\partial \lambda_{t-1|t-2}}{\partial \theta}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\right]+E_{t-2}\left[\left(x_{t-1}u_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\right)\right] \\
&\quad +E_{t-2}\left[\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \theta}\lambda_{t-1|t-2}\right)\right]+E_{t-2}\left[\lambda_{t-1|t-2}u_{t-1}\right] \\
&= b\left[\frac{\partial \lambda_{t-1|t-2}}{\partial \theta}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\right]+c\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}+a\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \theta}\lambda_{t-1|t-2}\right)+0
\end{aligned}
$$

The unconditional expectation of the last (non-zero) term is found by writing (shifted forward one period)

$$
\begin{aligned}
E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \theta}\lambda_{t|t-1}\right) &= E_{t-2}\left[\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \theta}+u_{t-1}\right)\left(\phi\lambda_{t-1|t-2}+\delta+\theta u_{t-1}\right)\right] \\
&= a\phi E\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \theta}\lambda_{t-1|t-2}\right)+\theta\sigma_u^2
\end{aligned}
$$

Thus

$$
E\left(\frac{\partial \lambda_{t|t-1}}{\partial \theta}\lambda_{t|t-1}\right)=\frac{\theta\sigma_u^2}{1-a\phi}
$$

leading to D.

$$
\begin{aligned}
E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right) &= E_{t-2}\left[\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}+1\right)\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}+\lambda_{t-1|t-2}\right)\right] \\
&= b\left[\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\right]+\lambda_{t-1|t-2}+a\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}+a\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\lambda_{t-1|t-2}
\end{aligned}
$$

For δ and ϕ, taking unconditional expectations gives

$$
E\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right)=bE\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\right)+\gamma+\frac{a\gamma}{1-a}+aE\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\lambda_{t-1|t-2}\right) \tag{36}
$$

but we require

$$
\begin{aligned}
E_{t-1}\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\lambda_{t|t-1}\right) &= E_{t-1}\left[\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}+1\right)(\delta+\phi\lambda_{t-1|t-2}+\theta u_{t-1})\right] \\
&= a\phi\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\lambda_{t-1|t-2}\right)+\delta a\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}+\delta+\phi\lambda_{t-1|t-2} \\
&\quad +E_{t-1}\left(x_{t-1}u_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\right)+\theta E_{t-1}(u_{t-1}) \\
&= a\phi\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\lambda_{t-1|t-2}\right)+\delta a\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}+\delta+\phi\lambda_{t-1|t-2}+\theta c\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}+0
\end{aligned}
$$

Taking unconditional expectations in the above expression yields

$$
\begin{aligned}
E\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\lambda_{t|t-1}\right) &= a\phi E\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\lambda_{t-1|t-2}\right)+\frac{\delta a}{1-a}+\delta+\phi\gamma+\frac{\theta c}{1-a} \\
&= a\phi E\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}\lambda_{t-1|t-2}\right)+\frac{\delta-a\phi\delta+\theta c-\phi\theta c}{(1-a)(1-\phi)}
\end{aligned}
$$

and so

$$
E\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\lambda_{t|t-1}\right)=\frac{\delta-a\phi\delta+\theta c-\phi\theta c}{(1-a\phi)(1-a)(1-\phi)}
$$

and substituting in (36) gives F (divided by $1-b$).

Finally

$$
E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\frac{\partial \lambda_{t|t-1}}{\partial \theta}\right)=E_{t-2}\left[\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}+1\right)\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \theta}+u_{t-1}\right)\right]
$$

Expanding and taking unconditional expectations gives E.

B Functions of beta

When b has a $Beta(1/2, \nu/2)$ distribution, the pdf is

$$
f(b)=\frac{1}{B(1/2,\nu/2)}b^{-1/2}(1-b)^{\nu/2-1},
$$

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

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

Now $B(\alpha,\beta)=\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$. Thus

$$
\begin{aligned}
E(b(1-b)) &= \frac{B(1/2+1,\nu/2+1)}{B(1/2,\nu/2)} = \frac{\Gamma(1/2+1)\Gamma(\nu/2+1)}{\Gamma(1/2+\nu/2+2)}\frac{\Gamma(1/2+\nu/2)}{\Gamma(1/2)\Gamma(\nu/2)} \\
&= \frac{(1/2)(\nu/2)}{(1/2+\nu/2+1)(1/2+\nu/2)} = \frac{\nu}{(3+\nu)(\nu+1)}
\end{aligned}
$$

and

$$
E(b^2(1-b)) = \frac{B(1/2+2,\nu/2+1)}{B(1/2,\nu/2)} = \frac{3\nu}{(\nu+3)(\nu+1)(\nu+5)}.
$$

C Information matrix for GED

Differentiating the log-likelihood function for the GED gives

$$
\frac{\partial \ln L_t}{\partial \upsilon} = \frac{\ln 2}{\upsilon^2} - \upsilon^{-2}\psi\left(1+1/\upsilon\right) - \frac{1}{2}\left|y_t \exp(-\lambda_{t|t-1})\right|^{\upsilon} \ln\left|y_t \exp(-\lambda_{t|t-1})\right|
$$

or

$$
\frac{\partial \ln L_t}{\partial \upsilon} = \frac{\ln 2}{\upsilon^2} - \upsilon^{-2}\psi\left(1+1/\upsilon\right) - \frac{1}{2}\left|y_t \exp(-\lambda_{t|t-1})\right|^{\upsilon} (\ln|y_t| - \lambda_{t|t-1}).
$$

Hence

$$
\frac{\partial^2 \ln L_t}{\partial \upsilon^2} = \frac{-2\ln 2}{\upsilon^3} - g(\upsilon) - \frac{1}{2}\left|y_t \exp(-\lambda_{t|t-1})\right|^{\upsilon} (\ln|y_t \exp(-\lambda_{t|t-1})|)^2
$$

and

$$\frac{\partial^2 \ln L_t}{\partial \upsilon \partial \lambda_{t|t-1}} = (1/2) \left| y_t \exp(-\lambda_{t|t-1}) \right|^{\upsilon} (1 + \ln(|y_t \exp(-\lambda_{t|t-1})|)^{\upsilon})$$

Taking expectations, and recalling that $\varsigma_t = |y_t \exp(-\lambda_{t|t-1})|^{\upsilon}$ is gamma distributed, gives

$$\begin{aligned} E\left(\frac{\partial^2 \ln L_t}{\partial \upsilon^2}\right) &= \frac{-2\ln 2}{\upsilon^3} - g(\upsilon) - \frac{1}{2} E(\varsigma_t (\ln \varsigma_t)^2) \\ &= \frac{-2\ln 2}{\upsilon^3} - g(\upsilon) - \frac{\Gamma(2/\upsilon)((\psi(2/\upsilon))^2 - \acute{\psi}(2/\upsilon))}{2\upsilon^2 \Gamma(1/\upsilon)} \end{aligned}$$

and

$$E\left(\frac{\partial^2 \ln L_t}{\partial \upsilon \partial \lambda_{t|t-1}}\right) = (1/2)(E(\varsigma_t) + E(\varsigma_t \ln \varsigma_t)) = \upsilon^{-1} + \upsilon^{-1} 0.5 \Gamma(2/\upsilon) \psi(2/\upsilon)/\Gamma(1/\upsilon)).$$

D Gamma distribution

The *pdf* of a *gamma*(α, γ) variable is

$$f(y) = \alpha^{\gamma} y^{\gamma - 1} e^{-\alpha y}/\Gamma(\gamma), \quad 0 \leq y < \infty, \quad \alpha, \gamma > 0, \tag{37}$$

and the raw moments are given by $E(y^c) = \alpha^{-c}\Gamma(c + \gamma)/\Gamma(\gamma)$, $c > 0$.

E Weibull information matrix

$$\frac{\partial^2 \ln L_t}{\partial \lambda_{t|t-1} \partial \nu} = -1 + (y_t e^{-\lambda_{t|t-1}} \Gamma(1+1/\nu))^{\nu} - \nu (y_t e^{-\lambda_{t|t-1}} \Gamma(1+1/\nu))^{\nu} \ln(y_t e^{-\lambda_{t|t-1}})$$

and so

$$E\left(\frac{\partial^2 \ln L_t}{\partial \lambda_{t|t-1} \partial \nu}\right) = -E\left((y_t e^{-\lambda_{t|t-1}} \Gamma(1 + 1/\nu))^{\nu} \ln(y_t e^{-\lambda_{t|t-1}})^{\nu}\right) = -0.422\,8,$$

since, if x has a standard exponential distribution, $\int_0^\infty x \ln x \exp(-x)dx = 1-\gamma$, where γ is Euler's constant. Furthermore

$$\frac{\partial^2 \ln L_t}{\partial \nu^2} = -\nu^{-2} + \nu^{-2}(\ln(y_t e^{-\lambda_{t|t-1}}\Gamma(1+1/\nu))^\nu)^2(y_t e^{-\lambda_{t|t-1}}\Gamma(1+1/\nu))^\nu$$

giving

$$E\left(\frac{\partial^2 \ln L_t}{\partial \nu^2}\right) = -\nu^{-2} - \nu^{-2}(\gamma^2 - 2\gamma + \pi^2/6) = -1.823\,7/\nu^2.$$

F Proof of theorem 2

To derive B^*, first observe that the conditional expectation of the last term in expression (34), that is $E_{t-2}\left(\lambda_{t|t-1}.\partial\lambda_{t|t-1}/\partial\phi\right)$, is now

$$\begin{aligned}
& E_{t-2}\left(x^*_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\phi} + \lambda_{t-1|t-2}\right)(\phi\lambda_{t-1|t-2} + \delta + \theta u_{t-1} + \theta^* sgn(-y_{t-1})(u_{t-1}+1)) \\
= \; & \phi E_{t-2}\left(x^*_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\phi}\lambda_{t-1|t-2}\right) + \phi\lambda^2_{t-1|t-2} + \delta E_{t-2}\left(x^*_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\phi}\right) \\
& + \delta\lambda_{t-1|t-2} + \theta E_{t-2}\left(u_{t-1}x^*_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\phi}\right) + \theta E_{t-2}(u_{t-1}\lambda_{t-1|t-2}) \\
& + \theta^* E_{t-2}\left(x^*_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\phi}sgn(-y_{t-1})(u_{t-1}+1)\right) + \theta^* E_{t-2}(sgn(-y_{t-1})(u_{t-1}+1)\lambda_{t-1|t-2})
\end{aligned}$$

The last term is zero, but the penultimate term is not. Taking unconditional expectations, and substituting for $E(\lambda_{t-1|t-2})$, which is unchanged, gives

$$E\left(\frac{\partial\lambda_{t|t-1}}{\partial\phi}\lambda_{t|t-1}\right) = \frac{\phi E(\lambda^2_{t|t-1})}{1-a\phi} + \frac{\gamma(\delta + \theta c + \theta^* c^*)}{(1-a)(1-a\phi)}$$

Substituting in (34) and noting that $E(\lambda^2_{t|t-1})$ is now given by (32) gives B^*.

REFERENCES

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

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

Creal, D., Koopman, S.J., and A. Lucas (2010). Generalized autoregressive score models with applications. Working paper. Earlier version appeared as: 'A general framework for observation driven time-varying parameter models', Tinbergen Institute Discussion Paper, TI 2008-108/4, Amsterdam.

Davidson, J. (2000). *Econometric Theory.* Blackwell: Oxford.

Engle. R.F. (2002). New frontiers for ARCH models. J*ournal of Applied Econometrics*, 17, 425-46.

Engle, R.F. and J.R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica,* 66, 1127-1162.

Engle, R.F. and G.M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131, 3-27.

Fiorentini, G., G. Calzolari and I. Panattoni (1996). Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics,* 11, 399-417.

Gonzalez-Rivera, G., Senyuz, Z. and E. Yoldas (2010). Autocontours: dynamic specification testing. *Journal of Business and Economic Statistics,* (to appear).

Harvey, A.C. and T. Chakravarty (2009). Beta-t-EGARCH. Working paper. Earlier version appeared in 2008 as Cambridge Working paper in Economics, CWPE 0840.

Linton, O. (2008). ARCH models. In *The New Palgrave Dictionary of Economics*, 2nd Ed.

Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6, 318-24.

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

Straumann, D. (2005). *Estimation in Conditionally Heteroskedastic Time Series Models*, Lecture Notes in Statistics, 181, Springer, Berlin.

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

Taylor, S. J., (1986). *Modelling Financial Time Series.* Wiley: Chichester.

Taylor, S. J., (2005). *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press: Princeton.

van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge: CUP.

Varanasi, M.K. and B. Aazhang (1989). Parametric generalized Gaussian density estimation. *Journal of the Acoustical Society of America,* 86, 1404-15.

Zaffaroni, P. (2009). Whittle estimation of EGARCH and other exponential volatility models. *Journal of Econometrics*, 151, 190-200.

Zivot, E. (2009). Practical issues in the analysis of univariate GARCH models. In Anderson, T.G. et al. *Handbook of Financial Time Series*, 113-155. Berlin: Springer-Verlag.