Exponential Conditional Volatility Models*

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

June 6, 2011

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

* Revised version of Working Paper 10-36, Statistics and Econometrics Series 20, Carlos III, Madrid, September 2010.

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-day data rather than daily or weekly observations.

Despite the enormous effort put into developing the theory of GARCH models, there is still 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 a formulation of observation driven volatility models that solves many of the existing difficulties. The first element of the approach is that time-varying parameters (TVPs) are driven by the score of the conditional distribution. 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 for the estimators addressed.

It is shown here that when the conditional score is combined with an exponential link function, the asymptotic distribution of the maximum likelihood estimator of the dynamic parameters can be derived. The theory is much more straightforward than it is for GARCH models; see, for example, Straumann and Mikosch (2006). Furthermore an analytic expression for the asymptotic covariance matrix can be obtained and the conditions for the

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

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

asymptotic theory to be valid are easily checked.

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 and the full conditional distribution is easily simulated.

After introducing the idea of dynamic conditional score (DCS) models in section 2, the main result on the asymptotic distribution is set out in section 3. The conditional distribution of the observations in the Beta-t-EGARCH model, studied 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 4 and the conditions for the asymptotic theory to go through are set out. The complementary Gamma-GED-EGARCH model is also analyzed. Leverage is introduced into the models and the asymptotic theory extended to deal with it.

Section 5 proposes DCS models with an exponential link function for the time-varying mean when the conditional distribution has a Gamma, Weibull, Burr or F- distribution. The results in section 3 yield obtain the asymptotic distribution of the ML estimators. Section 6 reports fitting a Beta-t-EGARCH model to daily stock index returns and compares the analytic standard errors with numerical standard errors.

2 Dynamic conditional volatility models

An observation driven model is set up in terms of a conditional distribution for the $t-th$ observation. Thus

$$p(y_t|\lambda_{t|t-1}, Y_{t-1}), \qquad t = 1, ..., T \tag{1}$$

$$\lambda_{t+1|t} = g(\lambda_{t|t-1}, \lambda_{t-1|t-2}, ..., Y_t)$$

where Y_t denotes observations up to, and including y_t, and $\lambda_{t|t-1}$ is a parameter that changes over time. The second equation in (1) may be regarded as a data generating process or as a way of writing a filter that approximates a nonlinear unobserved components (UC) model. In both cases the notation $\lambda_{t+1|t}$ stresses its status as a parameter of the conditional distribution and as a filter, that is a function of past observations. The likelihood function for an observation driven model is immediately available since the joint density of a set of T observations is

$$L(\boldsymbol{\psi}) = \prod_{t=1}^{T} p(y_t|\lambda_{t|t-1}, Y_{t-1}; \boldsymbol{\psi}),$$

where $\boldsymbol{\psi}$ denotes a vector of unknown parameters.

The first-order Gaussian GARCH model is an observation driven model in which $\lambda_{t|t-1} = \sigma^2_{t|t-1}$. As such it may be written

$$y_t \mid Y_{t-1} \sim NID\left(0, \sigma^2_{t|t-1}\right)$$

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

where $\phi = \alpha + \beta$ and $v_t = y_t^2 - \sigma^2_{t|t-1}$ is a martingale difference (MD).

The distributions of returns typically have heavy tails. Although the GARCH structure induces excess kurtosis in the returns, it is not usually enough to match the data. As a result, it is now customary to assume that the conditional distribution has a Student t_ν-distribution, where ν denotes degrees of freedom. The GARCH-t model, which was originally proposed by Bollerslev (1987), is widely used in empirical work and as a benchmark for other models. The t-distribution is employed in the predictive distribution of returns and used as the basis for maximum likelihood (ML) estimation of the parameters, but it is not acknowledged in the design of the equation for the conditional variance. The specification of the conditional variance as a linear combination of squared observations is taken for granted, but the consequences are that it responds too much to extreme observations and the effect is slow to dissipate. These features of GARCH are well-known and the consequences for testing and forecasting have been explored in a number of papers; see, for example, Franses, van Dijk and Lucas (2004). Other researchers, such as Muler and Yohai (2008), have been prompted to develop procedures for robustification.

In a dynamic conditional score (DCS) model, $\lambda_{t+1|t}$ depends on current and past values of a variable, u_t, that is defined as being proportional to the (standardized) score of the conditional distribution at time t. This variable is a MD by construction. When y_t has a conditional t-distribution with ν degrees of freedom, the DCS modification replaces v_t in the conditional variance equation, (2), by another MD, $v_t = \sigma^2_{t|t-1}u_t$, where

$$u_t = \frac{(\nu+1)y_t^2}{(\nu-2)\sigma^2_{t|t-1} + y_t^2} - 1, \quad -1 \leq u_t \leq \nu, \quad \nu > 2. \tag{3}$$

This model is called Beta-t-GARCH because u_t is a linear function of a variable with a conditional Beta distribution.

Figure 1 plots the conditional score function, u_t, against y_t/σ for $t-$distributions with $\nu = 3$ and 10 and for the normal distribution ($\nu = \infty$). When $\nu = 3$ an extreme observation has only a moderate impact as it is treated as coming from a $t_\nu-$ distribution rather than from a normal distribution with an abnormally high variance. As $|y_t| \to \infty$, $u_t \to \nu$ so $\lambda_{t|t-1}$ is bounded for finite ν, as is the robust conditional variance equation proposed by Muler and Yohai (2008, p 2922).

The use of an exponential link function means that the dynamic equation is set up for $\ln \sigma^2_{t+1|t} = \lambda_{t+1|t}$. The first-order model is

$$\lambda_{t+1|t} = \delta + \phi\lambda_{t|t-1} + \kappa u_t, \qquad t = 1, ..., T \tag{4}$$

and when the conditional distribution is t_ν, (3) is redefined as by replacing $(\nu-2)\sigma^2_{t|t-1}$ by $\nu\exp(\lambda_{t|t-1})$. The class of models obtained by combining the conditional score with an exponential link function is called Beta-t-EGARCH. A complementary class is based on the general error distribution (GED) distribution. The conditional score then has a Gamma distribution, leading to the name Gamma-GED-EGARCH.

The structure of the above model is similar to the stochastic volatility (SV) models where the logarithm of the variance is driven by an unobserved process. The first-order model for y_t, $t = 1, .., T$, is

$$\begin{gathered} y_t = \sigma_t\varepsilon_t, \qquad \sigma_t^2 = \exp\left(\lambda_t\right), \qquad \varepsilon_t \sim IID\left(0, 1\right) \\ \lambda_{t+1} = \delta + \phi\lambda_t + \eta_t, \qquad \eta_t \sim NID\left(0, \sigma^2_\eta\right) \end{gathered} \tag{5}$$

with ε_t and η_t mutually independent. SV models are parameter driven and

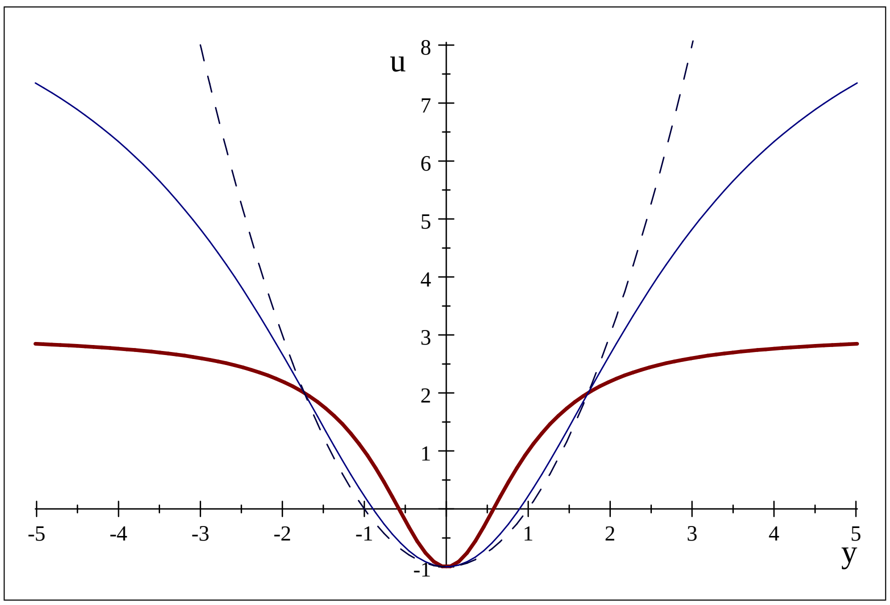

Figure 1: Impact of u_t for t_ν with $\nu = 3$ (thick), $\nu = 10$ (thin) and $\nu = \infty$ (dashed).

unlike GARCH models, which are observation driven, direct ML is not possible. A linear state space form can be obtained by taking the logarithms of the absolute values of the observations to give the following measurement equation:

$$\ln |y_t| = \lambda_t/2 + \ln |\varepsilon_t| \,, \quad t = 1, .., T.$$

The parameters can be estimated by QML, using the Kalman filter, as in Harvey, Ruiz and Shephard (1994). However, there is a loss in efficiency because the distribution of $\ln |\varepsilon_t|$ is far from Gaussian. Efficient estimation can be achived by computer intensive methods, as described in Durbin and Koopman (2001). The exponential DCS model can be regarded as an approximation to the SV model or as a model in its own right.

Similar considerations arise when dealing with location/scale models for non-negative variables. While the DCS approach for a Gamma distribution is consistent with a conditional mean dynamic equation that is linear in the observations, it can suggest a dampening down of the impact of a large observation from a Weibull, Burr and F distributions.

3 ML estimation of DCS models

In DCS models, 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 conditional score vector, $\partial \ln L_t/\partial \boldsymbol{\lambda}$. This vector may be the standardized score - ie divided by the information matrix - or a residual, the choice being largely a matter of convenience. A crucial requirement - though not the only one - 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. The fulfillment of this requirement may require a careful choice of link function for $\boldsymbol{\lambda}$.

Suppose initially that there is just one parameter, λ, in the static model. Let k be a finite constant and define

$$u_t = k.\partial \ln L_t/\partial \lambda, \quad t = 1, ..., T.$$

Since u_t is proportional to the score, it has zero mean and finite variance, σ_u^2, when standard regularity conditions hold. The information quantity, I, for a single observation is

$$I = -E(\partial^2 \ln L_t / \partial\lambda^2) = E[(\partial \ln L_t / \partial\lambda)^2] = E(u_t^2)/k^2 = \sigma_u^2/k^2 < \infty. \quad (6)$$

Suppose that, for a particular choice of link function, I does not depend on λ. More generally, consider the following assumption.

Condition 1 *The distribution of u_t in the static model does not depend on λ.*

Now let $\lambda = \lambda_{t|t-1}$ evolve over time as a function of past values of the score. The score can be broken down into two parts:

$$\frac{\partial \ln L_t}{\partial \boldsymbol{\psi}} = \frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}}, \quad (7)$$

where $\boldsymbol{\psi}$ denotes the vector of parameters governing the dynamics. *Since $\lambda_{t|t-1}$ and its derivatives depend only on past information, the distribution of u_t conditional on information at time $t-1$ is the same as its unconditional distribution and so is time invariant.*

The above decomposition carries over into the following lemma.

Lemma 1 *Consider a model with a single time-varying parameter, $\lambda_{t|t-1}$, which satisfies an equation that depends on variables which are fixed at time $t-1$. The process is governed by a set of fixed parameters, $\boldsymbol{\psi}$. If condition 1 holds, then the score for the t-th observation, $\partial \ln L_t / \partial \boldsymbol{\psi}$, is a MD with conditional covariance matrix*

$$E_{t-1}\left[\left(\frac{\partial \ln L_t}{\partial \boldsymbol{\psi}}\right)\left(\frac{\partial \ln L_t}{\partial \boldsymbol{\psi}}\right)'\right] = I.\left(\frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}'}\right), \quad t = 2, ..., T. \quad (8)$$

Proof. The fact that the score in (7) is a MD is confirmed by the fact that $\partial \lambda_{t|t-1} / \partial \boldsymbol{\psi}$ is fixed at time $t-1$ and the expected value of the score in the static model is zero.

Write the outer product as

$$\left(\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}}\right)\left(\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}}\right)' = \left(\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}}\right)^2 \left(\frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}'}\right).$$

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 $t-1$,

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

■

3.1 Information matrix for the first-order model

In theorem 1 below, the unconditional covariance matrix of the score at time t is derived for the first-order model,

$$\lambda_{t|t-1} = \delta + \phi\lambda_{t-1|t-2} + \kappa u_{t-1}, \quad |\phi| < 1, \quad \kappa \neq 0, \quad t = 2, ..., T, \tag{9}$$

and shown to be constant and p.d. when the model is identifiable. Identifiability requires $\kappa \neq 0$. Such a condition is hardly surprising since if κ were zero there would be no dynamics. The assumption that $|\phi| < 1$ enables $\lambda_{t|t-1}$ to be expressed as an infinite moving average in the $u_t's$. Since the $u_t's$ are MDs and hence WN, $\lambda_{t|t-1}$ is weakly stationary with an unconditional mean of $\delta/(1-\phi)$ and an unconditional variance of $\sigma_u^2/(1-\phi^2)$. Note that the process is assumed to have started in the infinite past, though for practical purposes we may set $\lambda_{1|0}$ equal to the unconditional mean, $\delta/(1-\phi)$.

The complications arise because u_{t-1} depends on $\lambda_{t-1|t-2}$ and hence on the parameters in $\boldsymbol{\psi}$. The vector $\partial \lambda_{t|t-1}/\partial \boldsymbol{\psi}$ is

$$\begin{aligned}
\frac{\partial \lambda_{t|t-1}}{\partial \kappa} &= \phi\frac{\partial \lambda_{t-1|t-2}}{\partial \kappa} + \kappa\frac{\partial u_{t-1}}{\partial \kappa} + u_{t-1} \\
\frac{\partial \lambda_{t|t-1}}{\partial \phi} &= \phi\frac{\partial \lambda_{t-1|t-2}}{\partial \phi} + \kappa\frac{\partial u_{t-1}}{\partial \phi} + \lambda_{t-1|t-2} \\
\frac{\partial \lambda_{t|t-1}}{\partial \delta} &= \phi\frac{\partial \lambda_{t-1|t-2}}{\partial \delta} + \kappa\frac{\partial u_{t-1}}{\partial \delta} + 1.
\end{aligned} \tag{10}$$

However,

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

and similarly for the other two derivatives. Therefore

$$\begin{aligned}
\frac{\partial \lambda_{t|t-1}}{\partial \kappa} &= x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \kappa}+u_{t-1} \\
\frac{\partial \lambda_{t|t-1}}{\partial \phi} &= x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}+\lambda_{t-1|t-2} \\
\frac{\partial \lambda_{t|t-1}}{\partial \delta} &= x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \delta}+1.
\end{aligned} \tag{11}$$

where

$$x_t=\phi+\kappa\frac{\partial u_t}{\partial \lambda_{t|t-1}}, \quad t=1,....,T. \tag{12}$$

The next condition, which generalizes condition 1, is needed for the information matrix of $\boldsymbol{\psi}$ to be derived.

Condition 2 *The conditional joint distribution of u_t and u_t', where $u_t'=\partial u_t/\partial \lambda_{t|t-1}$, is time invariant with finite second moment, $E(u_t^{2-k}u_t'^k)<\infty$, $k=0,1,2$, that is, $E(u_t u_t')<\infty$ and $E(u_t'^2)<\infty$ as well as $E(u_t^2)<\infty$.*

The following definitions are needed:

$$\begin{aligned}
a &= E_{t-1}(x_t)=\phi+\kappa E_{t-1}\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right)=\phi+\kappa E\left(\frac{\partial u_t}{\partial \lambda}\right) \\
b &= E_{t-1}(x_t^2)=\phi^2+2\phi\kappa E\left(\frac{\partial u_t}{\partial \lambda}\right)+\kappa^2 E\left(\frac{\partial u_t}{\partial \lambda}\right)^2\geq 0 \\
c &= E_{t-1}(u_t x_t)=\kappa E\left(u_t\frac{\partial u_t}{\partial \lambda}\right)
\end{aligned} \tag{13}$$

The expectations in the above formulae exist in view of condition 2. Because they are time invariant the unconditional expectations can replace conditional ones.

The following lemma is a pre-requisite for theorem 1.

Lemma 2 *When the process for $\lambda_{t|t-1}$ starts in the infinite past and $|a| < 1$,*

$$\begin{aligned} E\left(\frac{\partial \lambda_{t|t-1}}{\partial \kappa}\right) &= 0, \qquad t = 2, ..., T, \\ E\left(\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right) &= \frac{\delta}{(1-a)(1-\phi)}, \\ E\left(\frac{\partial \lambda_{t|t-1}}{\partial \delta}\right) &= \frac{1}{1-a}. \end{aligned} \tag{14}$$

Proof. Applying the law of iterated expectations (LIE) to (11)

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

and

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

Hence, if $|a| < 1$,

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

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

$$E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right) = a\frac{\partial \lambda_{t-1|t-2}}{\partial \phi} + \lambda_{t-1|t-2}. \tag{15}$$

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. Once a solution has been shown to exist, the result can be confirmed by taking unconditional expectations in (15) to give

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

from which

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

As regards δ,

$$E_{t-2}\left(\frac{\partial\lambda_{t|t-1}}{\partial\delta}\right)=a\frac{\partial\lambda_{t-1|t-2}}{\partial\delta}+1 \tag{16}$$

and taking unconditional expectations gives the result. ■

The above lemma requires that $|a|<1$. The result on the information matrix below requires $b<1$ and fulfillment of this condition implies $|a|<1$. That this is the case follows directly from the Cauchy-Schwartz inequality $E_{t-1}(x_t^2)\geq[E_{t-1}(x_t)]^2$.

Theorem 1 *Assume that condition 2 holds and that $b<1$. Then the covariance matrix of the score for a single observation is time-invariant and given by*

$$\mathbf{D}(\boldsymbol{\psi})=\mathbf{D}\begin{pmatrix}\widetilde{\kappa}\\ \widetilde{\phi}\\ \widetilde{\delta}\end{pmatrix}=\frac{1}{1-b}\begin{bmatrix}A & D & E\\ D & B & F\\ E & F & C\end{bmatrix} \tag{17}$$

with

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

and the information matrix for a single observation is

$$\mathbf{I}(\boldsymbol{\psi})=I.\mathbf{D}(\boldsymbol{\psi})=(\sigma_u^2/k^2)\mathbf{D}(\boldsymbol{\psi}). \tag{18}$$

Proof. The information matrix is obtained by taking unconditional expectation of (8) and then combining it with the formula for $\mathbf{D}(\boldsymbol{\psi})$, 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 (11). To evaluate it, first take conditional expectations at time $t-2$, to obtain

$$\begin{aligned} E_{t-2}\left(\frac{\partial\lambda_{t|t-1}}{\partial\kappa}\right)^2 &= E_{t-2}\left(x_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\kappa}+u_{t-1}\right)^2 \\ &= b\left(\frac{\partial\lambda_{t-1|t-2}}{\partial\kappa}\right)^2+2c\frac{\partial\lambda_{t-1|t-2}}{\partial\kappa}+\sigma_u^2. \end{aligned} \tag{19}$$

It was shown in lemma 2 that the unconditional expectation of the second term is zero. Eliminating this term, and taking expectations at $t-3$ gives

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

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

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

Taking unconditional expectations in (19) gives the same result. The derivatives are all evaluated in this way in appendix A. ■

Remark 1 *The condition* $\kappa = 0$ *was imposed on the model at the outset, since otherwise there are no dynamics. If* κ *is zero, then* $\mathbf{D}(\kappa,\phi,\delta)$ *is singular, and the parameters* ϕ *and* δ *are not identified. When* $\kappa \neq 0$, *all three parameters are identified even*[3] *if* $\phi = 0$.

3.2 Consistency and asymptotic normality of the ML estimator

We now move on to prove consistency and asymptotic normality of the ML estimator for the first-order model.

[3]But if ϕ is set to zero rather than being estimated, ie the lag of $\lambda_{t|t-1}$ does not appear in the dynamic equation, then both κ and δ are identifiable even when $\kappa = 0$.

Theorem 2 *The ML estimator,* $\widetilde{\boldsymbol{\psi}}$, *is consistent when* $\mathbf{D}(\boldsymbol{\psi})$, *and hence* $\mathbf{I}(\boldsymbol{\psi})$, *is p.d.*

Proof. The conditional score is a MD with a constant unconditional covariance matrix given by $\mathbf{D}(\boldsymbol{\psi})$. Hence the weak law of large numbers (WLLN) applies; see[4] Davidson (2000, p.123-4, p 272-3). ∎

Lemma 3 *When condition 1 holds,* u_t *is* $IID(0,\sigma_u^2)$ *and so the process* $\lambda_{t|t-1}$ *in (9) is strictly stationary.*

Lemma 4 *When condition 1 holds and* $|a| < 1$, *the sequences of the derivatives in* $\partial\lambda_{t|t-1}/\partial\boldsymbol{\psi}$, *are strictly stationary.*

Proof. The derivatives, (11), are stochastic recurrence equations and strict stationarity follows from standard results on such equations; see Straumann and Mikosch (2006, p 2450-1) and Vervaat (1979). In fact the necessary condition for strict stationarity is $E(\ln|x_t|) < 0$. This condition is satisfied if $|a| < 1$ because $|a| = |E(x_t)| \leq E(|x_t|)$ and, from Jensen's inequality, $\ln E(|x_t|) \leq E(\ln|x_t|)$. Although it appears that strict stationarity can be achieved without $|a| < 1$, this condition is needed for the first moment to exist. ∎

Remark 2 *Strict stationarity is not actually necessary to prove asymptotic normality of the ML estimator when, as for most of the models considered here, all the moments of the score and its first derivative are finite.*

The next condition is just an extension of condition 2, while the one after is a standard regularity condition.

Condition 3 *The conditional joint distribution of* $(u_t,\, u_t')'$ *is time invariant with finite fourth moment, that is,* $E(u_t^{4-k}u_t^{\prime k}) < \infty$, $k = 0, 1, .., 4$.

Condition 4 *The elements of* $\boldsymbol{\psi}$ *do not lie on the boundary of the parameter space.*

[4]Theorem 6.2.2, which is similar to Khinchine's theorem, can be applied. The moment condition, (ii), holds so it is unnecessary to invoke strict stationarity. Since second moments are finite, Chebyshev's theorem also applies; see p 124.

Theorem 3 *Assume conditions 3 and 4. Define*

$$d = E(\phi + \kappa.\partial u_t/\partial\lambda)^4 \geq 0. \tag{20}$$

Provided that $d < 1$, *the limiting distribution of* $\sqrt{T}\widetilde{\boldsymbol{\psi}}$, *where* $\widetilde{\boldsymbol{\psi}}$ *is the ML estimator of* $\boldsymbol{\psi}$, *is multivariate normal with mean* $\sqrt{T}\boldsymbol{\psi}$ *and covariance matrix*

$$Var(\widetilde{\boldsymbol{\psi}}) = \mathbf{I}^{-1}(\boldsymbol{\psi}) = (k^2/\sigma_u^2)\mathbf{D}^{-1}(\boldsymbol{\psi}). \tag{21}$$

Proof. From lemma 1, the score vector is a MD with conditional covariance matrix, (8). For a single element in the score,

$$\frac{\partial \ln L_t}{\partial \psi_i}, \quad i = 1, 2, 3,$$

where ψ_i is the $i - th$ element of $\boldsymbol{\psi}$, we may write

$$E_{t-1}\left[\left(\frac{\partial \ln L_t}{\partial \psi_i}\right)^2\right] = I.\left(\frac{\partial \lambda_{t|t-1}}{\partial \psi_i}\right)^2 = \sigma_{it}^2, \qquad t = 1, ..., T,$$

From Davidson (2000, pp 271-6), proof of the CLT requires that

$$p\lim T^{-1}\sum \sigma_{it}^2 = \sigma_i^2 < \infty. \tag{22}$$

From theorem 1, each $\sigma_i^2, i = 1, 2, 3$ is finite if $\mathbf{D}(\boldsymbol{\psi})$ is p.d.

In order to simplify notation let $w_{it} = \partial\lambda_{t|t-1}/\partial\psi_i$. (Since I is constant, attention can be concentrated on w_{it} rather than the score). Unlike the $w_{it}'s$ the $w_{it}^{2\prime}s$ are not MDs. However, they are strictly stationary and also weakly stationary provided they have finite unconditional variance. This being the case, the $w_{it}^{2\prime}s$ satisfy the WLLN by Chebychev theorem and (22) is true; see Davidson (2000, p42, p124).

The finite variance condition for the $w_{it}^{2\prime}s$ is fulfilled if the $w_{it}'s$ have finite unconditional fourth moment, that is

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \psi_i}\right)^4 = E\left(w_{it}\right)^4 < \infty. \quad i = 1, 2, 3$$

The first element in (11) is

$$\frac{\partial \lambda_{t|t-1}}{\partial \kappa} = x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \kappa} + u_{t-1}$$

The first subscript in w_{1t} can be dropped without creating any ambiguity, enabling us to write

$$w_t = x_{t-1}w_{t-1} + u_{t-1}, \quad t = 2, ..., T.$$

Hence

$$\begin{aligned} w_t^4 &= (x_{t-1}w_{t-1} + u_{t-1})^4 \\ &= u_{t-1}^4 + 4u_{t-1}^3 w_{t-1}x_{t-1} + 6u_{t-1}^2 w_{t-1}^2 x_{t-1}^2 + 4u_{t-1}w_{t-1}^3 x_{t-1}^3 + w_{t-1}^4 x_{t-1}^4 \end{aligned}$$

As in the earlier proofs, conditional expectations are taken at time $t-2$ to give

$$\begin{aligned} E_{t-2}(w_t^4) &= E_{t-2}(u_{t-1}^4) + 4w_{t-1}E_{t-2}(u_{t-1}^3 x_{t-1}) + 6w_{t-1}^2 E_{t-2}(u_{t-1}^2 x_{t-1}^2) \\ &\quad +4w_{t-1}^3 E_{t-2}(u_{t-1}x_{t-1}^3) + w_{t-1}^4 E_{t-2}(x_t^4) \end{aligned}$$

Now take unconditional expectations so that

$$\begin{aligned} E(w_t^4) &= E(u_{t-1}^4) + 4w_{t-1}E(u_{t-1}^3 x_{t-1}) + 6w_{t-1}^2 E(u_{t-1}^2 x_{t-1}^2) \\ &\quad +4w_{t-1}^3 E(u_{t-1}x_{t-1}^3) + dw_{t-1}^4 \end{aligned}$$

where

$$d = E(x_t^4) = \kappa^4 E(u_t'^4) + 4\kappa^3\phi E(u_t'^3) + 6\kappa^2\phi^2 E(u_t'^2) + 4\kappa\phi^3 E(u_t') + \phi^4, \quad (23)$$

and, as before, u_t' denotes $\partial u_t/\partial\lambda$. Because of condition 3, the terms $E(u_t'^k)$, $k = 1, .., 4$, and $E(u_{t-1}x_{t-1}^3)$, $E(u_{t-1}^2 x_{t-1}^2)$, $E(u_{t-1}^3 x_{t-1})$ are finite unconditional expectations. Hence the unconditional fourth moment of w_t is finite iff $d < 1$. Note that $d < 1$ is sufficient for the first, second and third moments to exist.The above argument is similar to that in Vervaat (1979, p 773-4).

The argument extends to $\partial\lambda_{t|t-1}/\partial\phi$, where $\lambda_{t|t-1}$ replaces u_t, because $\lambda_{t|t-1}$ is stationary and, since it depends on u_t, the necessary moments exist. ■

The condition $d < 1$ implicitly imposes constraints on the range of κ. The nature of the constraints will be investigated for the various models. On the whole they do not appear to present practical difficulties.

3.3 Nonstationarity

If $\phi = 1$, the matrix $\mathbf{D}(\boldsymbol{\psi})$, and hence $\mathbf{I}(\boldsymbol{\psi})$, is no longer p.d. The usual asymptotic theory does not apply as the model contains a unit root. However, if the unit root is imposed, so that ϕ is set equal to unity, then standard asymptotics apply. The following result is a corollary to theorems 1, 2 and 3.

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

$$\mathbf{I}(\widetilde{\kappa}, \widetilde{\delta}) = \frac{\sigma_u^2}{k^2(1-b)} \begin{bmatrix} \sigma_u^2 & \frac{c}{1-a} \\ \frac{c}{1-a} & \frac{1+a}{1-a} \end{bmatrix}, \tag{24}$$

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

$$b = 1 - 2\kappa\sigma_u^2/k + \kappa^2 E[(\partial u_t/\partial\lambda)^2], \tag{25}$$

and the ML estimators of $\widetilde{\kappa}$ and $\widetilde{\delta}$ are consistent. Furthermore $\sqrt{T}(\widetilde{\kappa}, \widetilde{\delta})'$ has a limiting normal distribution with mean $\sqrt{T}(\kappa, \delta)'$ and covariance matrix $\mathbf{I}^{-1}(\widetilde{\kappa}, \widetilde{\delta})$ provided that $d < 1$.

It can be seen from (25) that $\kappa > 0$ is a necessary condition for $b < 1$. Hence it is also necessary for $d < 1$.

3.4 Extensions

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 5 *Suppose that there are two parameters in $\boldsymbol{\lambda}$, but that $\lambda_{j,t|t-1} = f(\boldsymbol{\psi}_j), j = 1, 2$ with the vectors $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ having no elements in common.*

When the information matrix in the static model does not depend on λ_1 and λ_2

$$\mathbf{I}(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) = E\left[\left(\begin{array}{c} \frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1} \\ \frac{\partial \ln L_t}{\partial \lambda_2}\frac{\partial \lambda_2}{\partial \boldsymbol{\psi}_2} \end{array}\right)\left(\begin{array}{c} \frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1} \\ \frac{\partial \ln L_t}{\partial \lambda_2}\frac{\partial \lambda_2}{\partial \boldsymbol{\psi}_2} \end{array}\right)'\right] \tag{26}$$

$$= \left[\begin{array}{cc} E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\right)^2 E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1}\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1'}\right) & E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \lambda_2}\right) E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1}\frac{\partial \lambda_2}{\partial \boldsymbol{\psi}_2'}\right) \\ E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \lambda_2}\right) E\left(\frac{\partial \lambda_2}{\partial \boldsymbol{\psi}_2}\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1'}\right) & E\left(\frac{\partial \ln L_t}{\partial \lambda_2}\right)^2 E\left(\frac{\partial \lambda_2}{\partial \boldsymbol{\psi}_2}\frac{\partial \lambda_2}{\partial \boldsymbol{\psi}_2'}\right) \end{array}\right].$$

The above matrix is p.d. if $\mathbf{I}(\boldsymbol{\lambda})$ and $\mathbf{D}(\boldsymbol{\psi}_1, \boldsymbol{\psi}_1)$ 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 6 *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,*

$$\mathbf{I}(\boldsymbol{\psi}_1, \boldsymbol{\lambda}_2) = \left[\begin{array}{cc} E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\right)^2 E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1}\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1'}\right) & E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1}\right) E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2'}\right) \\ E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2}\right) E\left(\frac{\partial \lambda_1}{\partial \boldsymbol{\psi}_1'}\right) & E\left(\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2'}\right) \end{array}\right]. \tag{27}$$

The conditions for asymptotic normality are as in theorem 3.

When $n = 2$, the information matrix for the first-order model is

$$\mathbf{I}(\boldsymbol{\psi}_1, \boldsymbol{\lambda}_2) = \left[\begin{array}{cc} E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\right)^2 \mathbf{D}(\boldsymbol{\psi}_1) & E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \lambda_2}\right)\left(\begin{array}{c} 0 \\ \frac{\delta}{(1-a)(1-\phi)} \\ 1/(1-a) \end{array}\right) \\ E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \lambda_2}\right)\left(\begin{array}{ccc} 0 & \frac{\delta}{(1-a)(1-\phi)} & \frac{1}{1-a} \end{array}\right) & E\left(\frac{\partial \ln L_t}{\partial \lambda_2}\right)^2 \end{array}\right]$$

where $\mathbf{D}(\boldsymbol{\psi}_1)$ is the matrix in (17).

4 Exponential GARCH

In the EGARCH model

$$y_t = \sigma_{t|t-1}\varepsilon_t, \qquad t = 1, ..., T, \tag{28}$$

where ε_t is serially independent with unit variance. The logarithm of the conditional variance in (28) is given by

$$\ln \sigma^2_{t|t-1} = \gamma + \sum_{k=1}^{\infty} \psi_k g\left(\varepsilon_{t-k}\right), \quad \psi_1 = 1, \tag{29}$$

where γ and $\psi_k, k = 1, .., \infty$, are real and nonstochastic. The model may be generalized by letting γ be a deterministic function of time, but to do so complicates the exposition unnecessarily. The analysis in Nelson (1991), and in almost all subsequent research, focusses on the specification

$$g\left(\varepsilon_t\right) = \alpha^* \varepsilon_t + \alpha\left[\left|\varepsilon_t\right| - E\left|\varepsilon_t\right|\right], \tag{30}$$

where α and α^* are parameters: the first-order model was given in (**??**). By construction, $g\left(\varepsilon_t\right)$ has zero mean and so is a MD. Indeed the $g\left(\varepsilon_t\right)' s$ are IID.

Theorem 2.1 in Nelson (1991, p. 351) states that for model (28) and (29), with $g\left(\cdot\right)$ as in (30), $\sigma^2_{t|t-1}, y_t$ and $\ln \sigma^2_{t|t-1}$ are strictly stationary and ergodic, and $\ln \sigma^2_{t|t-1}$ is covariance stationary if and only if $\sum_{k=1}^{\infty} \psi_k^2 < \infty$. His theorem 2.2 demonstrates the existence of moments of $\sigma^2_{t|t-1}$ and y_t for the $GED(\upsilon)$ distribution with $\upsilon > 1$. The normal distribution is included as it is $GED(2)$. Nelson notes that if ε_t is t_ν distributed, the conditions needed for the existence of the moments of $\sigma^2_{t|t-1}$ and y_t are rarely satisfied in practice.

4.1 Beta-t-EGARCH

When the observations have a conditional t-distribution,

$$y_t = \varepsilon_t \exp(\lambda_{t|t-1}/2), \quad t = 1, ..., T, \tag{31}$$

where the serially independent, zero mean variable ε_t has a $t_\nu-$distribution with positive degrees of freedom, ν. Note that ε_t differs from ε_t in (28) in that the variance is not unity.

The (conditional score) variable

$$u_t = \frac{(\nu+1)y_t^2}{\nu \exp(\lambda_{t|t-1}) + y_t^2} - 1, \quad -1 \le u_t \le \nu, \quad \nu > 0. \tag{32}$$

may be expressed as

$$u_t = (\nu+1)b_t - 1, \tag{33}$$

where

$$b_t = \frac{y_t^2/\nu \exp(\lambda_{t|t-1})}{1 + y_t^2/\nu \exp(\lambda_{t|t-1})}, \qquad 0 \le b_t \le 1, \quad 0 < \nu < \infty, \tag{34}$$

is distributed as $Beta(1/2, \nu/2)$, a Beta distribution of the first kind; see Stuart and Ord (1987, ch 2). Since $E(b_t) = 1/(\nu+1)$ and $Var(b_t) = 2\nu/\{(\nu+3)(\nu+1)^2\}$, u_t has zero mean and variance $2\nu/(\nu+3)$.

The properties of Beta-t-EGARCH may be derived by writing $\lambda_{t|t-1}$ as

$$\lambda_{t|t-1} = \gamma + \sum_{k=1}^{\infty} \psi_k u_{t-k}, \tag{35}$$

where the $\psi_k's$ are parameters, as in (29), but ψ_1 is not constrained to be unity. Since

$$u_t = \frac{(\nu+1)\varepsilon_t^2}{\nu + \varepsilon_t^2} - 1,$$

it is a function only of the IID variables, ε_t, and hence is itself an IID sequence. When $\Sigma\psi_k^2 < \infty$ and $0 < \nu < \infty$, $\lambda_{t|t-1}$ is covariance stationary, the moments of the scale, $\exp(\lambda_{t|t-1}/2)$, always exist and the $m-th$ moment of y_t exists for $m < \nu$. Furthermore, for $\nu > 0$, $\lambda_{t|t-1}$ and $\exp(\lambda_{t|t-1}/2)$ are strictly stationary and ergodic, as is y_t. The proof is straightforward. Since u_t has bounded support for finite ν, all its moments exist; see Stuart and Ord (1987 p215). Similarly its exponent has bounded support for $0 < \nu < \infty$ and so $E[\exp(au_t)] < \infty$ for $|a| < \infty$. Strict stationarity of $\lambda_{t|t-1}$ follows immediately from the fact that the $u_t's$ are IID. Strict stationarity and ergodicity of y_t holds for the reasons given by Nelson (1991, p92) for the EGARCH model.

Analytic expression for the moments and the autocorrelations of the absolute values of the observations raised to any power can be derived as in Harvey and Chakravary (2009). Analytic expressions for the $\ell - step$ ahead conditional moments can be similarly obtained. However, it is easy to simulate the $\ell - step$ ahead predictive distribution. When $\lambda_{t|t-1}$ has a moving average representation in MDs, the optimal estimator of

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

is its conditional expectation

$$\lambda_{T+\ell|T} = \gamma + \sum_{k=0}^{\infty} \psi_{\ell+k} u_{T-k}, \quad \ell = 2, 3, .. \tag{36}$$

Hence the difference between $\lambda_{T+\ell|T+\ell-1}$ and its estimator is $\sum_{j=1}^{\ell-1} \psi_j u_{T+\ell-j}$. Hence the distribution of $y_{T+\ell}$, $\ell = 2, 3, ...$, conditional on the information at time T, is the distribution of

$$y_{T+\ell} = \varepsilon_{T+\ell} \exp(\lambda_{T+\ell|T+\ell-1}/2) = \varepsilon_{T+\ell} \left[\prod_{j=1}^{\ell-1} e^{\psi_j((\nu+1)b_{T+\ell-j}-1)/2} \right] e^{\lambda_{T+\ell|T}/2}$$

Simulating the predictive distribution of the scale and observations is striagtforward as the term in square brackets is made up of $\ell - 1$ independent Beta variates and these can be combined with a draw from a t-distribution to obtain a value of $\varepsilon_{T+\ell}$ and hence $y_{T+\ell}$. In contrast to convential GARCH models, it is not necessary to simulate the full sequence of observations from y_{T+1} to $y_{T+\ell}$; see the discussion in Andersen et al (2006, p 810-811).

4.2 Maximum likelihood estimation and inference

The log-likelihood function for the Beta-t-EGARCH model is

$$\begin{aligned} \ln L &= T \ln \Gamma\left((\nu+1)/2\right) - \frac{T}{2} \ln \pi - T \ln \Gamma\left(\nu/2\right) - \frac{T}{2} \ln \nu \\ & -\frac{1}{2} \sum_{t=1}^{T} \lambda_{t|t-1} - \frac{(\nu+1)}{2} \sum_{t=1}^{T} \ln \left(1 + \frac{y_t^2}{\nu e^{\lambda_{t|t-1}}}\right). \end{aligned}$$

It is assumed that $u_j = 0$, $j \leq 0$ and that $\lambda_{1|0} = \gamma$ (though, as is common practice, $\lambda_{1|0}$ may be set equal to the logarithm of the sample variance minus

$\ln(\nu/(\nu-2)$, assuming that $\nu > 2$.) The ML estimates are obtained by maximizing $\ln L$ with respect to the unknown parameters, which in the first-order model are ν, ϕ, κ and δ.

Apart from a special case ($\phi = 0$), analyzed in Straumann (2005, p125), no formal theory of the asymptotic properties of ML for EGARCH models has been developed. Nevertheless, ML estimation has been the standard approach to estimation of EGARCH models ever since it was proposed by Nelson (1991).

Straumann and Mikosch (2006) give a definitive treatment of the asymptotic theory for GARCH models. The mathematics are complex. The emphasis is on quasi-maximum likelihood and on p 2452 they state 'A final treatment of the QMLE in EGARCH is not possible at the time being, and one may regard this open problem as one of the limitations of this model.' As will be shown below the problem lies with the classic formulation of the EGARCH model and the attempt to estimate it by QML.

Straumann and Mikosch (2006, p 2490) also note the difficulty of deriving analytic formulae for asymptotic standard errors: 'In general, it seems impossible to find a tractable expression for the asymptotic covariance matrix...even for GARCH(1,1)'. They suggest the use of numerical expressions for the first and second derivatives, as computed by recursions[5].

It was noted below (35) that the $u_t's$ are IID. Differentiating (32) gives

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

and since, like u_t, this depends only on a Beta variable, it is also IID. All moments of u_t and $\partial u_t/\partial \lambda$ exist and this is more than enough to satisfy condition 3. The expression for d is

$$d = \phi^4 - 4\phi^3\kappa(\nu+1)b(1,1) + 6\phi^2\kappa^2(\nu+1)^2 b(2,2) - 4\phi\kappa^3(\nu+1)^3 b(3,3) + \kappa^4(\nu+1)^4 b(4,4) \tag{37}$$

where $b(h,k) = E(b^h(1-b)^k)$, as defined in Appendix B.

[5] QML also requires that an estimate the fourth moment of the standardized disturbances be computed.

Proposition 1 *Let $k = 2$ in the Beta-t-EGARCH model and define*

$$\begin{aligned} a &= \phi - \kappa \frac{\nu}{\nu+3} \\ b &= \phi^2 - 2\phi\kappa \frac{\nu}{\nu+3} + \kappa^2 \frac{3\nu(\nu+1)(\nu+2)}{(\nu+7)(\nu+5)(\nu+3)} \\ c &= \kappa \frac{2\nu(1-\nu)}{(\nu+5)(\nu+3)}, \qquad \nu > 0. \end{aligned}$$

Provided that $d < 1$, the limiting distribution of $\sqrt{T}$ times the ML estimators of the parameters in the stationary first-order model, (4), is multivariate normal with covariance matrix

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

where $\mathbf{D}(\boldsymbol{\psi}_1)$ is the matrix in (17).

Proof. 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,

$$E_{t-1}\left[\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right)^2\right] = (\nu+1)^2 E(b_t^2(1-b_t)^2) = \frac{3\nu(\nu+1)(\nu+2)}{(\nu+7)(\nu+5)(\nu+3)}$$

and

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

These formulae are then substituted in (13). The ML estimators of ν and λ

are not asymptotically independent in the static model. Hence the expression below (27) is used with $\lambda_2 = \nu$, together with (14). ■

The above result does not require the existence of moments of the conditional $t-$distribution. However, a model with $\nu \leq 1$ has no mean and so would probably be of little practical value.

Corollary 2 *When ϕ is set to unity in the first-order model, it follows, as a corollary to proposition 1, that, provided that $d < 1$, $\sqrt{T}(\widetilde{\kappa}, \widetilde{\delta})'$ has a limiting normal distribution with mean $\sqrt{T}(\kappa, \delta)'$ and covariance matrix $\mathbf{I}^{-1}(\widetilde{\kappa}, \widetilde{\delta})$,as in (35), with $a = 1 - \kappa\nu/(\nu+3)$ and*

$$b = 1 - 2\kappa\frac{\nu}{\nu+3} + \kappa^2\frac{3\nu(\nu+1)(\nu+2)}{(\nu+7)(\nu+5)(\nu+3)}.$$

4.3 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, $I(y_t < 0)$, taking the value one for $y_t < 0$ and zero otherwise; see Taylor (2005, p 220-1). In the Beta-t-EGARCH model this additional variable is constructed by multiplying $(\nu+1)b_t = u_t + 1$ by the indicator. Alternatively, the sign of the observation may be used, so the first-order model, (4), becomes

$$\lambda_{t|t-1} = \delta + \phi\lambda_{t-1|t-2} + \kappa u_{t-1} + \kappa^* sgn(-y_{t-1})(u_t + 1). \tag{38}$$

Taking the sign of *minus* y_t means that the parameter κ^* is normally non-negative for stock returns. With the above parameterization $\lambda_{t|t-1}$ is driven by a MD, as is apparent by writing (38) as

$$\lambda_{t|t-1} = \delta + \phi\lambda_{t-1|t-2} + g(u_{t-1}), \tag{39}$$

where $g(u_t) = \kappa u_t + \kappa^* sgn(-y_t)(u_t + 1)$. The mean of $\lambda_{t|t-1}$ is as before, but

$$E(\lambda_{t|t-1}^2) = \delta^2/(1-\phi)^2 + \theta^2\sigma_u^2/(1-\phi^2) + \theta^{*2}(\sigma_u^2 + 1)/(1-\phi^2). \tag{40}$$

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

Proposition 2 *Provided that $d^* < 1$, and the parameter ν is known, the limiting distribution of $\sqrt{T}$ times the ML estimators of the parameters in the stationary first-order model, (38), is multivariate normal with covariance matrix*

$$Var\begin{pmatrix} \widetilde{\kappa} \\ \widetilde{\phi} \\ \widetilde{\delta} \\ \widetilde{\kappa}^* \end{pmatrix} = \frac{k^2(1-b^*)}{\sigma_u^2}\begin{bmatrix} A & D & E & 0 \\ D & B^* & F^* & D^* \\ E & F^* & C & E^* \\ 0 & D^* & E^* & A^* \end{bmatrix}^{-1} \tag{41}$$

where A, C, D and E are as in (21), F^* is F with κc expanded to become $\kappa c + \kappa^* c^*$,

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

with a as in proposition 1,

$$b^* = \phi^2 - 2\kappa\phi\kappa\frac{\nu}{\nu+3} + (\kappa^2+\kappa^{*2})\frac{3\nu(\nu+1)(\nu+2)}{(\nu+7)(\nu+5)(\nu+3)},$$

$$c^* = \kappa^* E\left(u_t\frac{\partial u_t}{\partial \lambda}\right) + \kappa^* E\left(\frac{\partial u_t}{\partial \lambda}\right) = \kappa^*\frac{2\nu(1-\nu)}{(\nu+5)(\nu+3)} + \kappa^*\frac{\nu}{\nu+3}$$

and

$$d^* = \phi^4 - 4\phi^3\kappa(\nu+1)b(1,1) + 6\phi^2(\kappa^2+\kappa^{*2})(\nu+1)^2 b(2,2) - 4\phi\kappa^3(\nu+1)^3 b(3,3) + (\kappa^4+\kappa^{*4})(\nu+1)^4 b(4,4),$$

where the notation is as in (37).

Proof.

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

where

$$x_t^* = \phi + (\kappa + \kappa^* sgn(-y_t)) \frac{\partial u_t}{\partial \lambda_{t|t-1}}$$

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

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

The derivatives in (10) are similarly modified by the addition of the derivatives of the leverage term, so x_t^* replaces x_t in all cases. However

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

and the formulae for the expectations in (14) 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

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

so

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

The formulae for b^* and d^* are similarly derived. Further details can be found in appendix F. ■

Corollary 3 *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 modifying the covariance matrix in a similar way to that in proposition 1.*

4.4 Gamma-GED-EGARCH

The probability density function (pdf) of the general error distribution, denoted $GED(\upsilon)$, is

$$f(y;\varphi,\upsilon)=\left[2^{1+1/\upsilon}\varphi\Gamma(1+1/\upsilon)\right]^{-1}\exp(-\left|(y-\mu)/\varphi\right|^{\upsilon}/2),\quad \varphi>0,\upsilon>0, \tag{42}$$

where φ is a scale parameter, related to the standard deviation by the formula $\sigma=2^{1/\upsilon}(\Gamma\left(3/\upsilon\right)/\Gamma\left(1/\upsilon\right))^{1/2}\varphi$, and υ is a tail-thickness parameter. Let $\lambda_{t|t-1}$ in (4) evolve as a linear function of u_t defined as

$$u_t=(\upsilon/2)\left|y_t/\exp(\lambda_{t|t-1})\right|^{\upsilon}/-1,\quad t=1,...,T, \tag{43}$$

and let $y_t\mid Y_{t-1}$ have a GED, (42), with parameter $\varphi_{t|t-1}=\exp(\lambda_{t|t-1})$, that is (31) becomes $y_t=\varepsilon_t\exp(\lambda_{t|t-1})$, where $\varepsilon_t\sim GED(\upsilon)$. When $\upsilon=1$, u_t is a linear function of $|y_t|$. The response is less sensitive to outliers than it is for a normal distribution, but it is far less robust than is Beta-t-EGARCH with small degrees of freedom.

The name Gamma-GED-EGARCH is adopted because $u_t=(\upsilon/2)g_t-1$, where g_t has a $Gamma(1/2,\ 1/\upsilon)$ distribution. The variance of u_t is υ. Moments, ACFs and predictions can be made in much the same way as for Beta-t-EGARCH; see Harvey and Chakravary (2009).

The conditional joint distribution of u_t and its derivative

$$\frac{\partial u_t}{\partial\lambda_{t|t-1}}=-(\upsilon^2/2)\left|y_t\right|^{\upsilon}/\exp(\lambda_{t|t-1}\upsilon)=-(\upsilon^2/2)g_t \tag{44}$$

is time invariant. Hence the asymptotic distribution of the ML estimators is easily obtained.

Proposition 3 *For a given value of υ, and provided that $d<1$, the limiting distribution of $\sqrt{T}$ times the ML estimators of the parameters in the stationary first-order model, Gamma-GED-EGARCH model, (43), is multivariate normal with covariance matrix as in (21) with $k=1$ and*

$$\begin{aligned}a&=\phi-\kappa\upsilon\\ b&=\phi^2-2\phi\kappa\upsilon+4\kappa^2(\upsilon+1)\\ c&=-\kappa\upsilon^2.\end{aligned}$$

Proof. Taking conditional expectations of (44) gives $-\upsilon$, which is $-\sigma_u^2$. In addition,

$$E_{t-1}\left[\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right)^2\right] = (\upsilon^2/2)^2 E_{t-1}\left(g_t\right) = 4(\upsilon+1)$$

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.

Remark 3 *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 $[|\varepsilon_t| - E|\varepsilon_t|]$ where $\varepsilon_t = y_t/\sigma_{t|t-1}$. The difficulties arise because, unless $\upsilon = 1$, the conditional expectation of $[|\varepsilon_t| - E|\varepsilon_t|]$ depends on $\sigma_{t|t-1}$.*

Remark 4 *If the location is non-zero, or more generally dependent on a set of exogenous explanatory variables, and the whole model is estimated by ML, the asymptotic distribution for the dynamic scale parameters are unaffected; see Zhu and Zinde-Walsh (2010). The same is true if a constant location is first estimated by the mean or median. These results require $\upsilon \geq 1$.*

The formula for d is

$$d = \phi^4 - 4\phi^3\kappa(\upsilon^2/2)E(g_t) + 6\phi^2\kappa^2(\upsilon^2/2)^2E(g_t) - 4\phi\kappa^3(\upsilon^2/3)^3E(g_t) + \kappa^4(\upsilon^2/4)^4E(g_t)$$

where $E(g_t) = 2^k\Gamma(k+\upsilon^{-1})/\Gamma(\upsilon^{-1})$. For a Gaussian distribution

$$d = 105\kappa^4 - 142.22\kappa^3\phi + 72\kappa^2\phi^2 - 8\kappa\phi^3 + \phi^4$$

For the Laplace distribution,

$$d = 1.5\kappa^4 - 7.11\kappa^3\phi + 12\kappa^2\phi^2 - 4\kappa\phi^3 + \phi^4$$

which, perhaps surprisingly, permits a wider range for κ, even though the Laplace distribution has heavier tails than does the normal.

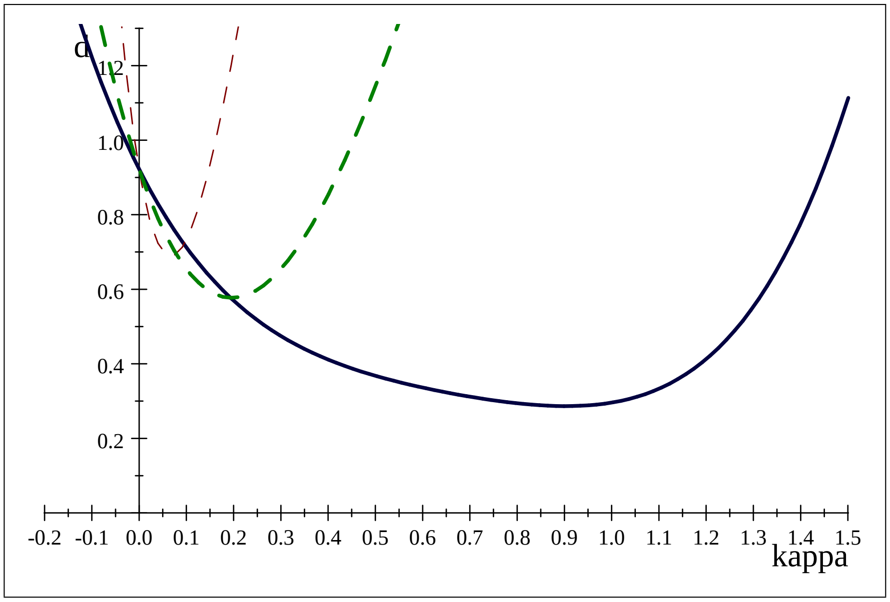

Figure 2: d against κ for $\phi = 0.98$ and (i) $t - distribution$ with $\nu = 6$ (solid), (ii) normal (thin dash), (iii) Laplace (thick dash).

Figure 2 shows d plotted against κ for $\phi = 0.98$ and t_6, normal and Laplace distributions. Since κ is typically less than 0.1, the constraint imposed by d is unlikely ever to be violated, though for low degrees of freedom the t-distribution can clearly accomodate much bigger values of κ.

5 Non-negative variables

Engle (2002) introduced a class of multiplicative error models (MEMs) 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.

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

$$y_t = \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, ..., T, \tag{45}$$

with dynamics as in (4).

5.1 Gamma distribution

The *pdf* of a $Gamma(\alpha, \gamma)$ variable, $g_t(\alpha, \gamma)$, is

$$f(y) = \alpha^\gamma y^{\gamma-1} e^{-\alpha y}/\Gamma(\gamma), \quad 0 \leq y < \infty, \quad \alpha, \gamma > 0, \tag{46}$$

where γ is the shape parameter and α is the scale. The pdf can be parameterized in terms of the mean, $\mu = \gamma/\alpha$, by writing

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

see, for example, Engle and Gallo (2006). The variance is μ^2/γ. The exponential distribution is a special case in which $\gamma = 1$.

For a dynamic model, the log-likelihood function for the $t-th$ observation is

$$\ln L_t(\boldsymbol{\psi}, \gamma) = \gamma \ln \gamma - \gamma \ln \mu_{t|t-1} + (\gamma - 1) \ln y_t - \gamma y_t/\mu_{t|t-1} - \ln \Gamma(\gamma)$$

and the exponential link function gives a conditional score of γu_t, where

$$u_t = (y_t - \exp(\lambda_{t|t-1}))/\exp(\lambda_{t|t-1}) = y_t \exp(-\lambda_{t|t-1}) - 1, \tag{47}$$

with $\sigma_u^2 = 1/\gamma$. Note that u_t is the standardized conditional score.

Expressions for moments, ACFs and predictions may be obtained in the same way as for the Beta-t-EGARCH model.

The asymptotic distribution of the ML estimators is easily established since $u_t = \varepsilon_t - 1$ and

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

where, as in (45), ε_t is Gamma (γ, γ) distributed.

Proposition 4 *Consider the first-order model, (4). Provided that* $|\phi| < 1$ *and* $d < 1$, *where*

$$d = \kappa^4\gamma^{-3}(1+\gamma)(2+\gamma)(3+\gamma) - 4\kappa^3\phi\gamma^{-2}(1+\gamma)(2+\gamma) + 6\kappa^2\phi^2\gamma^{-1}(1+\gamma) - 4\kappa\phi^3 + \phi^4,$$

the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi},\ \widetilde{\gamma} - \gamma)$ *is multivariate normal with covariance matrix*

$$Var\begin{pmatrix} \widetilde{\kappa} \\ \widetilde{\phi} \\ \widetilde{\delta} \\ \widetilde{\gamma} \end{pmatrix} = \begin{bmatrix} \gamma^{-1}\mathbf{D}^{-1}(\boldsymbol{\psi}) & 0 \\ 0 & \psi'(\gamma) - 1/\gamma \end{bmatrix}$$

where $\psi'(\gamma)$ *is the trigamma function and* $\mathbf{D}_t(\boldsymbol{\psi})$ *is as in (17) with*

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

and $\sigma_u^2 = 1/\gamma$. *The asymptotic distribution of* $\widetilde{\boldsymbol{\psi}}$ *is the same whether or not* γ *is estimated.*

Proof. Since $\partial u_t/\partial\lambda_{t|t-1}$ is Gamma(γ,γ) distributed, its conditional expectation is minus one. Furthermore

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

and

$$E_{t-1}\left[u_t\left(\frac{\partial u_t}{\partial\lambda_{t|t-1}}\right)\right] = E\left[\varepsilon_t^2 - \varepsilon_t\right] = -1/\gamma.$$

Note that $b = (\phi-\kappa)^2 + \kappa^2/\gamma$.

The expression for d is rather simple here as

$$u_t'^k = (-1)^k(y_t\exp(-\lambda))^k = (-1)^k\varepsilon_t^k, \quad k = 1, .., 4$$

Thus

$$E(u_t'^k) = (-1)^k E(\varepsilon_t^k) = (-1)^k\frac{\Gamma(k+\gamma)}{\gamma^k\Gamma(\gamma)}, \quad k = 1, .., 4$$

The independence of the ML estimators of λ and γ follows on noting that $E(\partial^2 \ln L_t/\partial\lambda\partial\gamma) = 0$. Indeed this must be the case because the ML estimator of λ in the static model is just the logarithm of the sample mean. The derivation of $Var(\widetilde{\gamma})$ is left to the reader. ■

Corollary 4 *The limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi})$ for the exponential distribution can be obtained by setting $\gamma = 1$.*

For the exponential

$$d = 24\kappa^4 - 24\kappa^3\phi + 12\kappa^2\phi^2 - 4\kappa\phi^3 + \phi^4$$

and with $\phi = 0.98$ and $\kappa = 0.1$ the value of d is 0.640.

Engle and Gallo (2006) estimate their MEM models with leverage. Information on the direction of the market is available from previous returns. Such effects may be introduced into the models of this section using (38). The asymptotic distribution of the ML estimators is obtained in the same way as for Beta-t-EGARCH.

5.2 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 $\alpha^2\Gamma(1+2/\nu) - \mu^2$. Re-arranging the pdf gives

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

where

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

when the scale is time-varying.

The exponential link function, $\alpha_{t|t-1} = \exp(\lambda_{t|t-1})$, yields the following log-likelihood function for the $t-th$ observation:

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

Hence the score is

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

A convenient choice for u_t in the equation for $\lambda_{t|t-1}$ is

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

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

Furthermore

$$\frac{\partial u_t}{\partial \lambda_{t|t-1}} = -\nu[(y_t e^{-\lambda_{t|t-1}})^{\nu}] = -\nu w_t$$

and so condition 3 is fulfilled.

Proposition 5 *For the first-order model with $|\phi| < 1$ and $d < 1$, where*

$$d = 24\kappa^4\nu^4 - 24\kappa^3\nu^3\phi + 12\kappa^2\nu^2\phi^2 - 4\kappa\nu\phi^3 + \phi^4$$

the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi},\ \widetilde{\nu} - \nu)$ is multivariate normal with covariance matrix as in (21) with

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

and $k = 1/\nu$.

Proof. *Since w_t has an exponential distribution,*

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

while

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

and

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

More generally $E(u'^k) = (-1)^k \nu^k E(w_t^k)$ and the formula for d follows easily from that for the exponential distribution (and reduces to it when $\nu = 1$). ■

In contrast to the Gamma case, estimation of the shape parameter does make a difference to the asymptotic distribution of the ML estimators of the dynamic parameters, since the information matrix for λ and ν in the static model is not diagonal.

5.3 Burr distribution

The generalized Burr distribution with scale parameter α has pdf

$$p(y) = \gamma\alpha^{-1}y^{\gamma-1}\left[\left(\frac{y}{\nu\alpha}\right)^{\gamma} + 1\right]^{-\nu-1}, \qquad \alpha, \gamma, \nu > 0 \tag{48}$$

To model changing scale we can let $\alpha_{t|t-1} = \nu^{-1/\gamma}\exp(\lambda_{t|t-1})$. Then

$$\ln L_t = \ln\nu + \ln\gamma + \nu\gamma\lambda_{t|t-1} + (\gamma - 1)\ln y_t - (\nu - 1)\ln(y_t^{\gamma} + \exp(\gamma\lambda_{t|t-1})$$

and so

$$\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} = \gamma\nu - \gamma(\nu + 1)\frac{\exp(\gamma\lambda_{t|t-1})}{y_t^{\gamma} + \exp(\gamma\lambda_{t|t-1})} = \gamma u_t$$

where

$$u_t = \nu - (\nu + 1)b_t(\nu, 1) \tag{49}$$

and

$$b_t(\nu, 1) = \frac{\exp(\gamma\lambda_{t|t-1})}{y_t^{\gamma} + \exp(\gamma\lambda_{t|t-1})}$$

is distributed as $Beta(\nu, 1)$; the result may be shown directly by change of variable. The formula for the mean of the Beta confirms that $E(u_t) = 0$.

A plot of u_t is shown in figure 3 for $\nu = 2$ and $\gamma = 1$. The dashed line shows the response for Gamma. The scale parameter has been set so that the mean is one in all cases. The Weibull response for a mean of one is $\nu(x\Gamma(1 + 1/\nu))^{\nu} - 1)$; it coincides with the Gamma response when $\nu = 1$, in which case it is an exponential distribution. The graph shows $\nu = 0.5$. Note that the Weibull and Burr responses are less sensitive to large values of the standardized observations. In the case of this particular Burr distribution, the second moment does not exist.

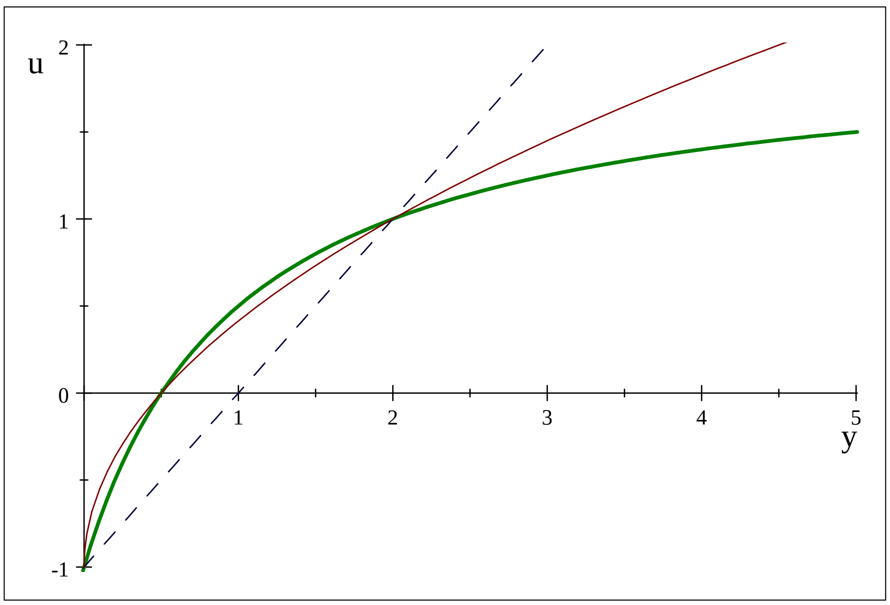

Figure 3: Plot of u_t against a standardized observation for Burr with $\nu = 2$ and $\gamma = 1$ (thick) and Weibull for $\nu = 0.5$, together with gamma (dashed).

Differentiating the score gives

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

and so the asymptotic theory is very similar to that for the Beta-t-EGARCH model.

Proposition 6 *For a conditional Burr distribution with ν and γ fixed and a first-order dynamic model with $|\phi| < 1$ and $d < 1$, the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi})$ is multivariate normal with covariance matrix as in (21) with*

$$\begin{aligned} a &= \phi - \kappa\frac{\nu}{\nu+2} \\ b &= \phi^2 - 2\phi\kappa\frac{\nu}{\nu+2} + \kappa^2\frac{2\nu(\nu+1)^2}{(\nu+4)(\nu+3)(\nu+2)} \\ c &= -\kappa\frac{\nu(\nu-1)}{(\nu+3)(\nu+2)} \end{aligned} \tag{50}$$

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

$$E[b^h(1-b)^k] = \frac{\nu k}{\nu+h+k}\frac{\Gamma(\nu+h)\Gamma(k)}{\Gamma(\nu+h+k)},$$

taking conditional expectations gives

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

while

$$E_{t-1}\left[\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right]^2 = \frac{2\nu(\nu+1)^2}{(\nu+4)(\nu+3)(\nu+2)}$$

and

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

■

The formula for d is obtained by noting that

$$E(u_t'^k) = (-1)^k(\nu+1)^k \frac{\nu k\Gamma(k+\nu)\Gamma(k)}{(\nu+2k)\Gamma(\nu+2k)}, \quad k = 1,..,4 \tag{51}$$

When ν is estimated, the information matrix for a given value of γ is easily shown to be

$$\mathbf{I}(\boldsymbol{\psi}, \nu; \gamma) = \begin{bmatrix} \frac{\nu}{\nu+2}\mathbf{D}(\boldsymbol{\psi}) & \frac{-\gamma}{\nu+1} \\ \frac{-\gamma}{\nu+1} & \nu^{-2} \end{bmatrix}$$

Remark 5 *Grammig and Maurier (2000) give first derivatives for dynamics of the GARCH form. These are relatively complex.*

5.4 F-distribution

If centered returns have a t_ν-distribution, their squares will be distributed as $F(1, \nu)$. This observation suggests the F - distribution as a candidate for modeling various measures of daily volatility. In general, the F - distribution depends on two degrees of freedom parameters and is denoted $F(\nu_1, \nu_2)$.

The log-likelihood function for the t-th observation from an $F(\nu_1, \nu_2)$ distribution is

$$\begin{aligned} \ln L_t \ &= \ \frac{\nu_1}{2} \ln \nu_1 y_t e^{-\lambda_{t|t-1}} + \frac{\nu_2}{2} \ln \nu_1 \ + \frac{\nu_1+\nu_2}{2} \ln(\nu_1 y_t e^{-\lambda_{t|t-1}} + \nu_2) \\ &\quad - \ln y_t e^{-\lambda_{t|t-1}} - \ln B(\nu_1/2, \nu_2/2) \end{aligned}$$

Hence the score is

$$\frac{\partial \ln L_t}{\partial \lambda_{t|t-1}} = \frac{\nu_1+\nu_2}{2} b_t(\nu_1/2, \nu_2/2) - \frac{\nu_1}{2},$$

where

$$b_t(\nu_1/2, \nu_2/2) = \frac{\nu_1 y_t e^{-\lambda_{t|t-1}}/\nu_2}{1+\nu_1 y_t e^{-\lambda_{t|t-1}}/\nu_2} = \frac{\nu_1 \varepsilon_t/\nu_2}{1+\nu_1\varepsilon_t/\nu_2}$$

is distributed as $Beta(\nu_1/2, \nu_2/2)$. Taking expectations confirms that the score has zero mean since $E(b_t(\nu_1/2, \nu_2/2)) = \nu_1/(\nu_1+\nu_2)$.

The moments and ACF can be found from the properties of the Beta distribution. As regards the asymptotic distribution, differentiating the score gives

$$\frac{\partial u_t}{\partial \lambda_{t|t-1}} = -\frac{\nu_1 + \nu_2}{2} b_t(1 - b_t)$$

and so a, b, c and d are easily found. The formulae are like those for the Burr distribution, except that $(\nu_1 + \nu_2)/2$ replaces $(\nu + 1)$.

6 Daily Hang-Seng and Dow-Jones returns

The estimation procedures were programmed in `Ox 5` and the sequential quadratic programming (SQP) maximization algorithm for nonlinear functions subject to nonlinear constraints, `MaxSQP()`in Doornik (2007), was used throughout. The conditional variance or scale was initialized using the sample variance of the returns as in the G@RCH package of Laurent (2007). Standard errors were obtained from the inverse of the Hessian matrix computed using numerical derivatives.

The parameter estimates are presented in tables 1 to 5, together with the maximized log-likelihood and the Akaike information criterion (AIC), defined as $(-2\ln L + 2\times$ number of parameters$)/T$. The Bayes (Schwartz) information criteria were also calculated but they are very similar and so are not reported.

Table 1 reports estimates for Beta-t-EGARCH (1,0), with leverage. The ML estimates and associated numerical standard errors (SEs) were reported in Harvey and Chakravarty (2009). The asymptotic SEs are close to the numerical SEs. For both series the leverage parameter, κ^*, has the expected (positive) sign. The likelihood ratio statistic is large in both cases and it easily rejects[6] the null hypothesis that $\kappa^* = 0$ against the alternative that $\kappa^* > 0$. The same conclusion is reached if the standard errors are used to construct (one-sided) tests based on the standard normal distribution. (A Wald test). The values of d are also given in table 1: for both series they are well below unity.

[6]For a two-sided alternative the LR statistic is asymptotically χ^2_1, while for a one-sided alternative the distibution is a mixture of χ^2_1 and χ^2_0 leading to a test at the $\alpha\%$ level of significance being carried out with the $2\alpha\%$ significance points.

	Hang Seng		**DOW-JONES**	
Parameter	Estimates (num. SE)	Asy. SE	Estimates (num. SE)	Asy. SE
δ	0.006 (0.0020)	0.0018	-0.005 (0.001)	0.0026
ϕ	0.993 (0.003)	0.0017	0.989 (0.002)	0.0028
κ	0.093 (0.008)	0.0073	0.060 (0.005)	0.0052
κ^*	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
$d \quad (a, b)$	0.775	(0.931, 0.876)	0.815	(0.946,0.898)
$\ln L$	-9747.6		-11180.3	
AIC	3.474		2.629	

Table 1 Parameter estimates for Beta-t-EGARCH models with leverage.

Table 2 gives the estimates and standard errors for the benchmark GARCH-t model obtained with the G@RCH program of Laurent (2007). The leverage is captured by the indicator variable, but using the sign gives essentially the same result. If it is acknowledged that the conditional distribution is not Gaussian then the estimates computed under the assumption that it is conditionally Gaussian - denoted in table 2 simply as GARCH - are best described as quasi-maximum likelihood (QML). The columns at the end show the estimates for Beta-t-GARCH converted from the parameterization used in this article. The maximized log-likelihoods for Beta-t-GARCH are greater than those for GARCH-t.

While the condition for covariance stationarity of the GARCH-t fitted to Hang Seng is satisfied, as the estimate of $\beta + \alpha + \alpha^*/2 = \phi$ is 0.982, the condition for the existence of the fourth moment is violated because the relevant statistic takes a value of 1.021 and so is greater than one. On the other hand, the fourth moment condition for the Beta-t-GARCH model, (**??**), is 0.997.

	Hang Seng			DOW-JONES		
Parameter	GARCH-t	GARCH	B-t-G	GARCH-t	GARCH	B-t-G
δ	0.048 (0.011)	0.081	0.035	0.011(0.002)	0.017	0.010
β	.888 (.014)	.845	.884	0.936 (0.007)	0.919	0.927
α	.051 (0.008)	.053	.050	0.027(0.004)	0.027	0.050
α^*	.087 (0.021)	0.151	.108	0.052(0.010)	0.076	0.076
ν	5.87 (0.54)	-	5.97	7.21 (0.62)	-	7.64
$LogL$	-9770.1	-9991.0	-9748.6	-11192.8	-11428.8	-11182.4
AIC	3.473	3.548	3.465	2.620	2.673	2.618

Table 2 Parameter estimates for GARCH-t and GARCH with leverage, together with Beta-t-GARCH estiamtes

Plots of the conditional standard deviations (SDs) produced by the Beta-t-GARCH and Beta-t-EGARCH models are difficult to distinguish and the only marked differences between their conditional standard deviations and those obtained from conventional EGARCH and GARCH-t models are after extreme values. Figure 4 shows the SDs of Dow-Jones produced by Beta-t-EGARCH and GARCH-t, both with leverage effects, around the Great Crash of 1987. (The largest value is 22.5 but the y axis has been truncated). The GARCH-t filter reacts strongly to the extreme observations and then returns slowly to the same level as Beta-t-EGARCH.

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

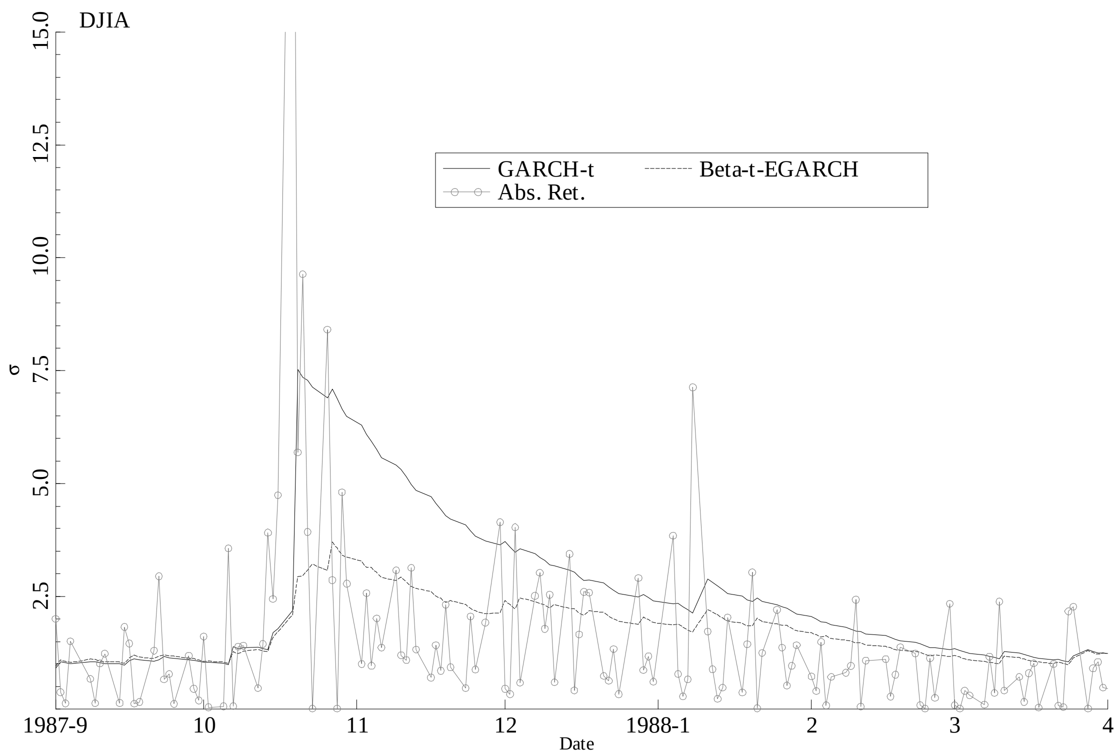

Figure 4: Dow-Jones absolute (de-meaned) returns around the great crash of October 1987, together with estimated conditional standard deviations for Beta-t-EGARCH and GARCH-t, both with leverage. The horizontal axis gives the year and month.

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.

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. 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.

Acknowledgements

This paper was written while I was a visiting professor at the Department of Statistics, Carlos III University, Madrid. I am grateful to Tirthankar Chakravarty, Mardi Dungey, Stan Hurn, Siem-Jan Koopman, Gloria Gonzalez-Rivera, Esther Ruiz, Richard Smith, Genaro Sucarrat, Abderrahim Taamou 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 $\mathbf{D}(\boldsymbol{\psi})$ matrix, as in (17). The formula for κ was derived in the main text. For ϕ

$$\begin{aligned} E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right)^2 &= E_{t-2}\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}+\lambda_{t-1|t-2}\right)^2 \qquad (52) \\ &= b\left(\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\right)^2+\lambda_{t-1|t-2}^2+2a\frac{\partial \lambda_{t-1|t-2}}{\partial \phi}\lambda_{t-1|t-2} \end{aligned}$$

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

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

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

$$E\left(\frac{\partial\lambda_{t|t-1}}{\partial\phi}\lambda_{t|t-1}\right)=\frac{\phi E(\lambda^2_{t|t-1})}{1-a\phi}+\frac{\gamma(\delta+\kappa c)}{(1-a)(1-a\phi)} \tag{53}$$

Taking unconditional expectations in (52) and substituting from (53) gives

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

which leads to B on substituting for

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

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

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

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

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

Thus

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

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

$$E\left(\frac{\partial\lambda_{t|t-1}}{\partial\delta}\frac{\partial\lambda_{t|t-1}}{\partial\phi}\right)=bE\left(\frac{\partial\lambda_{t-1|t-2}}{\partial\delta}\frac{\partial\lambda_{t-1|t-2}}{\partial\phi}\right)+\gamma+\frac{a\gamma}{1-a}+aE\left(\frac{\partial\lambda_{t-1|t-2}}{\partial\delta}\lambda_{t-1|t-2}\right) \tag{54}$$

but we require

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

Taking unconditional expectations in the above expression yields

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

and so

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

and substituting in (54) gives F (divided by $1-b$).

Finally

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

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 Proof of proposition 2

To derive B^*, first observe that the conditional expectation of the last term in expression (52), that is $E_{t-2}\left(\lambda_{t|t-1}.\partial\lambda_{t|t-1}/\partial\phi\right)$, is now

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

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

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

Substituting in (52) and noting that $E(\lambda_{t|t-1}^{2})$ is now given by (40) gives B^{*}.

REFERENCES

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

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

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

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.

Durbin, J., Koopman, S.J., 2001. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

Engle. R.F. (2002). New frontiers for ARCH models. Journal 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.

Franses, P.H., van Dijk, D. and Lucas, A., 2004. Short patches of outliers, ARCH and volatility modelling, Applied Financial Economics 14, 221-231.

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

Grammig, J. and Maurer, K-O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal* 3, pp. 16-38.

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

Harvey A.C., E. Ruiz and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies* 61: 247-64.

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

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

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

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, In: Lecture Notes in Statistics, vol. 181, Springer, Berlin.

Straumann, D. and T. Mikosch (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. Annals of Statistics, 34, 2449-2495.

Stuart, A. and J.K. Ord (1987). Kendall's Advanced Theory of Statistics, originally by Sir Maurice Kendall. Fifth Edition of Volume 1, Distribution theory. Charles Griffin & Company Limited, London.

Tadikamalla, P. R. (1980) A look at the Burr and related distributions. International Statististical Review, 48, 337-344.

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

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

Vervaart, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible variables. Advances in Applied Probability, 11,750-83.

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

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

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.