Dynamic Models for Volatility and Heavy Tails

Andrew Harvey,

Cambridge University

December 2011

Andrew Harvey, (Cambridge University)

Volatility and Heavy Tails

December 2011 1 / 75

• A unified and comprehensive theory for a class of nonlinear time series models in which the conditional distribution of an observation may be heavy-tailed and the location and/or scale changes over time.

- A unified and comprehensive theory for a class of nonlinear time series models in which the conditional distribution of an observation may be heavy-tailed and the location and/or scale changes over time.
- The defining feature of these models is that the dynamics are driven by the *score* of the conditional distribution.

- A unified and comprehensive theory for a class of nonlinear time series models in which the conditional distribution of an observation may be heavy-tailed and the location and/or scale changes over time.
- The defining feature of these models is that the dynamics are driven by the *score* of the conditional distribution.
- When a suitable *link function* is employed for the dynamic parameter, analytic expressions may be derived for (unconditional) moments, autocorrelations and moments of multi-step forecasts.

- A unified and comprehensive theory for a class of nonlinear time series models in which the conditional distribution of an observation may be heavy-tailed and the location and/or scale changes over time.
- The defining feature of these models is that the dynamics are driven by the *score* of the conditional distribution.
- When a suitable *link function* is employed for the dynamic parameter, analytic expressions may be derived for (unconditional) moments, autocorrelations and moments of multi-step forecasts.
- Furthermore a full asymptotic distributional theory for maximum likelihood estimators can be obtained, including analytic expressions for the asymptotic covariance matrix of the estimators.

 standard linear time series models observed with an error which may be subject to outliers,

- standard linear time series models observed with an error which may be subject to outliers,
- models which capture changing conditional variance, and

- standard linear time series models observed with an error which may be subject to outliers,
- models which capture changing conditional variance, and
- models for non-negative variables.

- standard linear time series models observed with an error which may be subject to outliers,
- models which capture changing conditional variance, and
- models for non-negative variables.
- The last two of these are of considerable importance in financial econometrics.

- standard linear time series models observed with an error which may be subject to outliers,
- models which capture changing conditional variance, and
- models for non-negative variables.
- The last two of these are of considerable importance in financial econometrics.
- (a) Forecasting volatility Exponential GARCH (EGARCH)

- standard linear time series models observed with an error which may be subject to outliers,
- models which capture changing conditional variance, and
- models for non-negative variables.
- The last two of these are of considerable importance in financial econometrics.
- (a) Forecasting volatility Exponential GARCH (EGARCH)
- (b) Duration (time between trades) and volatility as measured by range and realised volatility - Gamma, Weibull, logistic and F-distributions with changing scale and exponential link functions,



Andrew Harvey,

Volatility and Heavy Tails

December 2011



Andrew Harvey ,

(Cambridge University)

Volatility and Heavy Tails

December 2011 5 / 75



A guiding principle is **signal extraction**. When combined with basic ideas of maximum likelihood estimation, the signal extraction approach leads to models which, in contrast to many in the literature, are relatively simple in their form and yield analytic expressions for their principal features.

For estimating location, DCS models are closely related to the unobserved components (UC) models described in Harvey (1989). Such models can be handled using state space methods and they are easily accessible using the STAMP package of Koopman et al (2008).

For estimating scale, the models are close to stochastic volatility (SV) models, where the variance is treated as an unobserved component. The close ties with UC and SV models provides insight into the structure of the DCS models, particularly with respect to modeling trend and seasonality, and into possible restrictions on the parameters.

A simple Gaussian signal plus noise model is

$$y_t = \mu_t + arepsilon_t, \quad arepsilon_t \sim \textit{NID}\left(0, \sigma_arepsilon^2
ight), \quad t = 1, ..., T$$

$$\mu_{t+1} = \phi \mu_t + \eta_t, \quad \eta_t \sim \textit{NID}(0, \sigma_\eta^2),$$

where the irregular and level disturbances, ε_t and η_t , are mutually independent. The AR parameter is ϕ , while the **signal-noise ratio**, $q = \sigma_{\eta}^2 / \sigma_{\varepsilon}^2$, plays the key role in determining how observations should be weighted for prediction and signal extraction. The reduced form (RF) is an ARMA(1,1) process

$$y_t = \phi y_{t-1} + \xi_t - heta \xi_{t-1}, \quad \xi_t \sim \textit{NID}\left(0, \sigma^2\right),$$

but with restrictions on θ . For example, when $\phi = 1$, $0 \le \theta \le 1$. The forecasts from the UC model and RF are the same.

The UC model is effectively in state space form (SSF) and, as such, it may be handled by the Kalman filter (KF). The parameters ϕ and q can be estimated by ML, with the likelihood function constructed from the one-step ahead prediction errors.

The KF can be expressed as a single equation. Writing this equation together with an equation for the one-step ahead prediction error, v_t , gives the innovations form (IF) of the KF:

 $y_t = \mu_{t|t-1} + v_t$ $\mu_{t+1|t} = \phi \mu_{t|t-1} + k_t v_t$

The Kalman gain, k_t , depends on ϕ and q. In the steady-state, k_t is constant. Setting it equal to κ and re-arranging gives the **ARMA(1,1)** model with $\xi_t = v_t$ and $\phi - \kappa = \theta$. Suppose noise is from a heavy tailed distribution, such as Student's t. Outliers.

The RF is still an ARMA(1,1), but allowing the $\xi'_t s$ to have a heavy-tailed distribution does not deal with the problem as a large observation becomes incorporated into the level and takes time to work through the system. An ARMA models with a heavy-tailed distribution is designed to handle *innovations outliers*, as opposed to *additive outliers*. See the **robustness** literature.

But a *model-based approach* is not only simpler than the usual robust methods, but is also more amenable to diagnostic checking and generalization.

Simulation methods, such as MCMC, provide the basis for a direct attack on models that are nonlinear and/or non-Gaussian. The aim is to extend the Kalman filtering and smoothing algorithms that have proved so effective in handling linear Gaussian models. Considerable progress has been made in recent years; see Durbin and Koopman (2001). But simulation-based estimation can be time-consuming and subject to a degree of uncertainty.

Also the statistical properties of the estimators are not easy to establish.

The DCS approach begins by writing down the distribution of the t - th observation, conditional on past observations. Time-varying parameters are then updated by a suitably defined filter. Such a model is *observation driven*, as opposed to a UC model which is *parameter driven*. (Cox's terminology). In a *linear Gaussian UC* model, the KF is driven by the one step-ahead prediction error, v_t . The DCS filter replaces v_t in the KF equation by a variable, u_t , that is proportional to the score of the conditional distribution.

The IF becomes

$$y_t = \mu_{t|t-1} + v_t, \qquad t = 1, ..., T$$

 $\mu_{t+1|t} = \phi \mu_{t|t-1} + \kappa u_t$

where κ is an unknown parameter.

Why the score ?

If the signal in AR(1)+noise model were fixed, that is $\phi = 1$ and $\sigma_{\eta}^2 = 0$, $\mu_{t+1} = \mu$, the sample mean, $\hat{\mu}$, would satisfy the condition

$$\sum_{t=1}^{T} (y_t - \widehat{\mu}) = 0.$$

The ML estimator is obtained by differentiating the log-likelihood function with respect to μ and setting the resulting derivative, the score, equal to zero. When the observations are normal, ML estimator is the same as the sample mean, the moment estimator.

For a non-Gaussian distribution, the moment estimator and the ML estimator differ. Once the signal in a Gaussian model becomes dynamic, its estimate can be updated using the KF. With a non-normal distribution exact updating is no longer possible, but the fact that ML estimation in the static case sets the score to zero provides the rationale for replacing the prediction error, which has mean zero, by the score, which for each individual observation, also has mean zero.

The use of the score of the conditional distribution to robustify the KF was originally proposed by Masreliez (1975). However, it has often been argued that a crucial assumption made by Masreliez (concerning the approximate normality of the prior at each time step) is, to quote Schick and Mitter (1994), '..insufficiently justified and remains controversial.' Nevertheless, the procedure has been found to perform well both in simulation studies and with real data.

The attraction of treating the score-driven filter as a model in its own right is that it becomes possible to derive the asymptotic distribution of the ML estimator and to generalize in various directions.

The same approach can then be used to model scale, using an exponential link function, and to model location and scale for non-negative variables. The justification for the class of DCS models is not that they approximate corresponding UC models, but rather that their statistical properties are both comprehensive and straightforward.

An immediate practical advantage is seen from the response of the score to an outlier.

$$\begin{array}{rcl} \mathbf{y}_t &=& \boldsymbol{\omega} + \boldsymbol{\mu}_{t|t-1} + \mathbf{v}_t = \boldsymbol{\omega} + \boldsymbol{\mu}_{t|t-1} + \exp(\lambda)\varepsilon_t, \\ \boldsymbol{\mu}_{t+1|t} &=& \boldsymbol{\phi}\boldsymbol{\mu}_{t|t-1} + \kappa u_t, \end{array}$$

where ε_t is serially independent, standard t-variate and

$$u_t = \left(1 + \frac{(y_t - \mu_{t|t-1})^2}{\nu e^{2\lambda}}\right)^{-1} v_t,$$

where $v_t = y_t - \mu_{t|t-1}$ is the prediction error and $\varphi = \exp(\lambda)$ is the (time-invariant) scale. Further details in Harvey and Luati (2012).



Figure: Impact of u_t for t_{ν} (with a scale of one) for $\nu = 3$ (thick), $\nu = 10$ (thin) and $\nu = \infty$ (dashed).



Andrew Harvey,

Volatility and Heavy Tails

December 2011

GARCH(1,1)

$$y_t = \sigma_{t|t-1} z_t, \quad z_t \backsim NID(0,1)$$

with conditional variance

$$\sigma_{t|t-1}^{2} = \gamma + \beta \sigma_{t-1|t-2}^{2} + \alpha y_{t-1}^{2}, \quad \gamma > 0, \beta \ge 0, \alpha \ge 0$$
$$\sigma_{t|t-1}^{2} = \gamma + \phi \sigma_{t-1|t-2}^{2} + \alpha \sigma_{t-1|t-2}^{2} u_{t-1},$$

where $\phi = \alpha + \beta$ and $u_{t-1} = y_{t-1}^2 / \sigma_{t-1|t-2}^2 - 1$ is a martingale difference (MD). Weakly stationary if $\phi < 1$.

イロト イポト イヨト イヨト

3

Observation driven models - parameter(s) of conditional distribution are functions of past observations. Contrast with parameter driven, eg stochastic volatility (SV) models The variance in SV models is driven by an unobserved process. The first-order model is

$$y_{t} = \sigma_{t}\varepsilon_{t}, \qquad \sigma_{t}^{2} = \exp(\lambda_{t}), \qquad \varepsilon_{t} \sim IID(0, 1)$$
$$\lambda_{t+1} = \delta + \phi\lambda_{t} + \eta_{t}, \qquad \eta_{t} \sim NID(0, \sigma_{\eta}^{2})$$

with ε_t and η_t mutually independent.



Andrew Harvey ,

, (Cambridge University)

Volatility and Heavy Tails

December 2011 21 / 75

• Assume that z_t has a Student t_{ν} -distribution, where ν denotes degrees of freedom - GARCH-t model.

- Assume that z_t has a Student t_{ν} -distribution, where ν denotes degrees of freedom GARCH-t model.
- 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.

- Assume that z_t has a Student t_{ν} -distribution, where ν denotes degrees of freedom GARCH-t model.
- 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 $\sigma_{t|t-1}^2$ as a linear combination of squared observations is taken for granted, but the consequences are that $\sigma_{t|t-1}^2$ responds too much to extreme observations and the effect is slow to dissipate.

- Assume that z_t has a Student t_{ν} -distribution, where ν denotes degrees of freedom GARCH-t model.
- 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 σ²_{t|t-1} as a linear combination of squared observations is taken for granted, but the consequences are that σ²_{t|t-1} responds too much to extreme observations and the effect is slow to dissipate.
- Note that QML estimation procedures do not question this linearity assumption. (Also not straightforward for t - see Hall and Yao, 2003)

In the EGARCH model

$$y_t = \sigma_{t|t-1} z_t$$
, z_t is $IID(0, 1)$,

with first-order dynamics

$$\ln \sigma_{t|t-1}^2 = \delta + \phi \ln \sigma_{t-1|t-2}^2 + \theta(|z_{t-1}| - E|z_{t-1}|) + \theta^* z_{t-1}$$

The role of z_t is to capture leverage effects.

Weak and covariance stationary if $|\phi| < 1$. More general infinite MA representation. Moments of $\sigma_{t|t-1}^2$ and y_t exist for the GED(v) distribution with v > 1. The normal distribution is GED(2).

If z_t is t_ν distributed, the conditions needed for the existence of the moments of σ²_{t|t-1} and y_t are rarely (if ever) satisfied in practice.

Weak and covariance stationary if $|\phi| < 1$. More general infinite MA representation. Moments of $\sigma_{t|t-1}^2$ and y_t exist for the GED(v) distribution with v > 1. The normal distribution is GED(2).

- If z_t is t_ν distributed, the conditions needed for the existence of the moments of σ²_{t|t-1} and y_t are rarely (if ever) satisfied in practice.
- No asymptotic theory for ML. See reviews by Linton (2008) and Zivot (2009). For GARCH there is no comprehensive theory.
DCS Volatility Models

What does the assumption of a t_{ν} -distribution imply about the specification of an equation for the conditional variance? The possible inappropriateness of letting $\sigma_{t|t-1}^2$ be a linear function of past squared observations when ν is finite becomes apparent on noting that, if the variance were constant, the sample variance would be an inefficient estimator of it.

Therefore replace u_t in the conditional variance equation

$$\sigma_{t+1|t}^2 = \gamma + \phi \sigma_{t|t-1}^2 + \alpha \sigma_{t|t-1}^2 u_t,$$

by another MD

$$u_t = \frac{(\nu+1)y_t^2}{(\nu-2)\sigma_{t|t-1}^2 + y_t^2} - 1, \quad -1 \le u_t \le \nu, \quad \nu > 2.$$

which is proportional to the score of the conditional variance.

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

where the serially independent, zero mean variable ε_t has a t_{ν} -distribution with degrees of freedom, $\nu > 0$, and the dynamic equation for the log of scale is

$$\lambda_{t|t-1} = \delta + \phi \lambda_{t-1|t-2} + \kappa u_{t-1}.$$

The conditional score is

$$u_t = rac{(
u+1)y_t^2}{
u \exp(\lambda_{t|t-1}) + y_t^2} - 1, \quad -1 \le u_t \le
u, \quad
u > 0$$

NB The variance is equal to the square of the scale, that is $(\nu - 2)\sigma_{t|t-1}^2/\nu$ for $\nu > 2$.



Figure: Impact of u_t for t_v with v = 3 (thick), v = 6 (medium dashed) v = 10 (thin) and $v = \infty$ (dashed).



Andrew Harvey , (Cambridge University)

December 2011 28 / 75

The variable u_t may be expressed as

$$u_t = (\nu + 1)b_t - 1,$$

where

$$b_t = rac{y_t^2 /
u \exp(\lambda_{t \mid t-1})}{1 + y_t^2 /
u \exp(\lambda_{t \mid t-1})}, \qquad 0 \le b_t \le 1, \quad 0 <
u < \infty,$$

is distributed as $Beta(1/2, \nu/2)$, a **Beta distribution**. Thus the $u_t's$ are IID.

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 $2\nu/(\nu+3)$.

• Moments exist and ACF of $|y_t|^c$, $c \ge 0$, can be derived.

___ ▶

э

- Moments exist and ACF of $\left|y_{t}\right|^{c}$, $c\geq$ 0, can be derived.
- Closed form expressions for moments of multi-step forecasts of volatility can be derived and full distribution easily simulated.

- Moments exist and ACF of $\left|y_{t}\right|^{c}$, $c\geq0$, can be derived.
- Closed form expressions for moments of multi-step forecasts of volatility can be derived and full distribution easily simulated.
- Asymptotic distribution of ML estimators with analytic expressions for standard errors.

- Moments exist and ACF of $\left|y_{t}\right|^{c}$, $c\geq0$, can be derived.
- Closed form expressions for moments of multi-step forecasts of volatility can be derived and full distribution easily simulated.
- Asymptotic distribution of ML estimators with analytic expressions for standard errors.
- Can handle time-varying trends (eg splines) and seasonals (eg time of day or day of week).

When the conditional distribution of y_t has a GED(v) distribution, u_t is a linear function of $|y_t|^v$. These variables can be transformed so as to have a **gamma** distribution and the properties of the model are again derived. The normal distribution is a special case of the GED, as is the double exponential, or Laplace, distribution. The conditional variance equation for the Laplace model has the same form as the conditional variance equation in the EGARCH model of Nelson (1991).



Figure: Impact of u_t for GED with v = 1 (thick), v = 0.5 (thin) and v = 2 (dashed).

Theorem

For the Beta-t-EGARCH model $\lambda_{t|t-1}$ 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 < v. Furthermore, for v > 0, $\lambda_{t|t-1}$ and $\exp(\lambda_{t|t-1}/2)$ are strictly stationary and ergodic, as is y_t .

The odd moments of y_t are zero as the distribution of ε_t is symmetric.

The even moments of y_t in the stationary Beta-t-EGARCH model are found from the MGF of a beta:

$$E(y_t^m) = E(|\varepsilon_t|^c) e^{m\gamma/2} \prod_{j=1}^{\infty} e^{-\psi_j m/2} \beta_{\nu}(\psi_j m/2), \qquad m < \nu.$$

= $\frac{\nu^{m/2} \Gamma(\frac{m}{2} + \frac{1}{2}) \Gamma(\frac{-m}{2} + \frac{\nu}{2})}{\Gamma(\frac{1}{2}) \Gamma(\frac{\nu}{2})} e^{m\gamma/2} \prod_{j=1}^{\infty} e^{-\psi_j m/2} \beta_{\nu}(\psi_j m/2)$

where

$$\mathcal{B}_{\nu}(\mathbf{a}) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{1+2r}{\nu+1+2r} \right) \frac{\mathbf{a}^k (\nu+1)^k}{k!}, \quad 0 < \nu < \infty.$$

is Kummer's (confluent hypergeometric) function.

The autocorrelations of the squared observations are given by analytic expressions. These involve gamma and confluent hypergeometric functions. But the ACFs can be computed for the absolute observations raised to any positive power; see Harvey and Chakravary (2009) Heavy-tails tend to weaken the autocorrelations.

The standard EGARCH model readily delivers the optimal ℓ -step ahead forecast - in the sense of minimizing the mean square error - of future logarithmic conditional variance. Unfortunately, as Andersen *et al* (2006, p804-5, p810-11) observe, the optimal forecast of the conditional variance, that is $E_T(\sigma_{T+\ell_1T+\ell-1}^2)$, where E_T denotes the expectation based on information at time T, generally depends on the entire ℓ -step ahead forecast distribution and this distribution is not usually available in closed form for EGARCH.

The exponential conditional volatility models overcome this difficulty because an analytic expression for the conditional scale and variance can be obtained from the law of iterated expectations. Expressions for higher order moments may be similarly derived. The full distribution is easy to simulate. In DCS models, some or all of the parameters in λ are time-varying, with the dynamics driven by a vector that is equal or proportional to the conditional score vector, $\partial \ln L_t / \partial \lambda$. 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 $\mathbf{I}_t(\lambda)$ does not depend on parameters in λ that are subsequently allowed to be time-varying. The fulfillment of this requirement may require a careful choice of link function for λ .

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

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

Although

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

is the conventional formulation of a first-order dynamic model, it turns out that the information matrix takes a simpler form if the paramerization is in terms of ω rather than δ . Thus

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

Re-writing the above model in a similar way to (1) gives

$$\lambda_{t|t-1} = \omega(1-\phi) + \phi \lambda_{t-1|t-2} + \kappa u_{t-1}.$$
(3)

The following definitions are needed:

$$a = E_{t-1}(x_t) = \phi + \kappa E_{t-1} \left(\frac{\partial u_t}{\partial \lambda_{t|t-1}} \right) = \phi + \kappa E \left(\frac{\partial u_t}{\partial \lambda} \right)$$
(4)
$$b = E_{t-1}(x_t^2) = \phi^2 + 2\phi \kappa E \left(\frac{\partial u_t}{\partial \lambda} \right) + \kappa^2 E \left(\frac{\partial u_t}{\partial \lambda} \right)^2 \ge 0$$

$$c = E_{t-1}(u_t x_t) = \kappa E \left(u_t \frac{\partial u_t}{\partial \lambda} \right)$$

Because they are time invariant the unconditional expectations can replace conditional ones.

Information matrix for the first-order model

Assume that the pdf is continuous and twice differentiable and that the joint distribution of $(u_t, u'_t)'$ does not depend on λ and is time invariant with finite second moment, that is, $E(u_t^{2-k}u'_t^k) < \infty$, k = 0, 1, 2. Provided that b < 1,

$$\mathbf{I}(\boldsymbol{\psi}) = I \cdot \mathbf{D}(\boldsymbol{\psi}) = (\sigma_u^2 / k^2) \mathbf{D}(\boldsymbol{\psi}),$$

where

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

with

$$A = \sigma_{u}^{2}, \qquad B = \frac{\kappa^{2} \sigma_{u}^{2} (1 + a\phi)}{(1 - \phi^{2})(1 - a\phi)}, \qquad C = \frac{(1 - \phi)^{2} (1 - \phi)^{2}}{1 - \phi^{2}}, \qquad C = \frac{(1 - \phi)^{2} (1 - \phi)^{2}}{1 - \phi^{2}}, \qquad C = \frac{(1 - \phi)^{2} (1 - \phi)^{2}}{1 - \phi^{2}}, \qquad C = \frac{a\kappa\sigma_{u}^{2}}{(1 - \phi)^{2}}, \qquad C = \frac{a\kappa\sigma_{u}^{2}}{1 - \phi^{2}}, \qquad C = \frac{a\kappa\sigma_{u}^{2}}{$$

Theorem

The ML estimator of ψ , denoted $\tilde{\psi}$, is consistent and the limiting distribution of $\sqrt{T}(\tilde{\psi} - \psi_0)$ multivariate normal with mean vector zero and covariance matrix

$$Var(\widetilde{\boldsymbol{\psi}}_0) = \mathbf{I}^{-1}(\boldsymbol{\psi}_0) = I^{-1}\mathbf{D}^{-1}(\boldsymbol{\psi}_0).$$
(5)

The proof follows along the lines of Jensen and Rahbek (ET, 2004), but the details are much simpler. Assume that $\ln L$ is three times continuously differentiable for ψ , ψ_0 is an interior point of the compact parameter space, and the following are true.

(i) As
$$T \to \infty$$
, $(1/\sqrt{T})\partial \ln L(\psi_0)/\partial \psi \to N(0, \mathbf{I}(\psi_0))$, where $\mathbf{I}(\psi_0)$ is p.d.
(ii) As $T \to \infty$, $(-1/T)\partial^2 \ln L(\psi_0)/\partial \psi \partial \psi' \xrightarrow{P} \mathbf{I}(\psi_0)$

Asymptotic theory for the first-order model

(iii) If ψ_i , i = 1, ..., n denote the parameters in ψ ,

$$\max_{h,i,j=1,\ldots,n} \sup_{\psi \in \mathcal{N}(\psi_0)} \left| \frac{\partial^3 \ln L}{T \cdot \partial \psi_h \partial \psi_i \partial \psi_j} \right| \leq c_T,$$

where $N(\psi_0)$ is a neighbourhood of the true parameter value, ψ_0 , and $0 \le c_T \xrightarrow{P} c$, $0 < c < \infty$.

It is not difficult to show that (iii) holds for IID observations from Student's t as the score and its first two derivatives are **bounded**. The argument needs to be combined with the following result on the stationarity and ergodicity of the derivatives of $\lambda_{t|t-1}$.

Proposition The first three derivatives of $\lambda_{t|t-1}$ wrt κ , ϕ and ω are stochastic recurrence equations (SREs) and the condition b < 1 is sufficient to ensure that they are strictly stationarity and ergodic at the true parameter value.

The $u'_t s$ are IID. Differentiating gives

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

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.

The condition b < 1 implicitly imposes constraints on the range of κ . But the constraint does not present practical difficulties.

Proposition

For a given value of v, the asymptotic covariance matrix of the dynamic parameters has

$$a = \phi - \kappa \frac{2\nu}{\nu+3}$$

$$b = \phi^2 - 4\phi \kappa \frac{\nu}{\nu+3} + \kappa^2 \frac{12\nu(\nu+1)}{(\nu+5)(\nu+3)}$$

$$c = \kappa \frac{4\nu(1-\nu)}{(\nu+5)(\nu+3)}.$$



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

Proposition

The asymptotic distribution of the dynamic parameters changes when v is estimated because the ML estimators of v and λ are not asymptotically independent in the static model. Specifically

$$I(\lambda, \nu) = rac{1}{2} \left[egin{array}{c} rac{
u}{
(
u+3)} & rac{-1}{
(
u+3)(
u+1)} \\ rac{-1}{
(
u+3)(
u+1)} & h(
u) \end{array}
ight]$$

where

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

and $\psi'\left(.
ight)$ is the trigamma function



Andrew Harvey , (Ca

(Cambridge University)

Volatility and Heavy Tails

December 2011 47 / 75

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, $I(y_t < 0)$. GJR. In the Beta-t-EGARCH model this additional variable is constructed by multiplying $(\nu + 1)b_t = u_t + 1$ by $I(y_t < 0)$.

Alternatively, the sign of the observation may be used, so

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

and hence $\lambda_{t|t-1}$ is driven by a **MD**.

(Taking the sign of *minus* y_t means that κ^* is normally non-negative for stock returns.)

Results on moments, ACFs and asymptotics may be generalized to cover leverage.

Dow-Jones from 1st October 1975 to 13th August 2009, giving T = 8548 returns.

Hang Seng from 31st December 1986 to 10th September 2009, giving T = 5630.

As expected, the data have heavy tails and show strong serial correlation in the squared observations.

$\frac{\delta}{\phi}$	Estimates (SE) 0.006 (0.002)	Asy. SE 0.0018	Estimates (SE) -0.005 (0.001)	Asy. SE
	()	0.0018	-0.005 (0.001)	0.0000
φ	0.002 (0.002)		0.000 (0.001)	0.0026
	0.993 (0.003)	0.0017	0.989 (0.002)	0.0028
к	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
а	.931		.946	
Ь	.876		.898	

Estimates with numerical and asymptotic standard errors

э



Andersen and Bollerslev (1998) - intra-day returns with explanatory variables eg time of day effects Beta-t-EGARCH model is

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

where

$$\begin{aligned} \lambda_{t|t-1} &= \mathbf{w}_t' \gamma + \lambda_{t|t-1}^{\dagger}, \\ \lambda_{t|t-1}^{\dagger} &= \phi_1 \lambda_{t-1|t-2}^{\dagger} + \kappa u_{t-1} \end{aligned}$$

No pre-adjustments needed.

Asymptotics work and extend to time-varying trends and seasonals

A non-zero location can be introduced into the t-distribution without complicating the asymptotic theory.

More generally the location may depend linearly on a set of static exogenous variables,

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + \varepsilon_t \exp(\lambda_{t|t-1}/2), \quad t = 1, ..., T,$$

in which case the ML estimators of β are asymptotically independent of the estimators of ψ and ν .

Engle and Lee (1999) proposed a GARCH model in which the variance is broken into a long-run and a short-run component. The main role of the short-run component is to pick up the temporary increase in variance after a large shock. Another feature of the model is that it can approximate *long memory* behaviour.

EGARCH models can be extended to have more than one component:

$$\lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}$$

where

Formulation - and properties - much simpler. Asymptotics hold for ML.

Stochastic location and stochastic scale

The Student t model for time-varying location may be combined with one for the scale.

$$y_t = \omega + \mu_{t|t-1} + \exp(\lambda_{t|t-1})\varepsilon_t$$
$$\mu_{t+1|t} = \phi_{\mu}\mu_{t|t-1} + \kappa_{\mu}u_{\mu t},$$
$$\lambda_{t+1|t} = \delta + \phi_{\lambda}\lambda_{t|t-1} + \kappa_{\lambda}u_{\lambda t}$$

where

$$u_{\mu t} = \left(1 + \frac{(y_t - \mu_{t|t-1})^2}{\nu e^{2\lambda_{t|t-1}}}\right)^{-1} (y_t - \mu_{t|t-1})$$

and the score in Beta-t-EGARCH becomes

$$u_{\lambda t} = \frac{(\nu+1)(y_t - \mu_{t|t-1})^2}{\nu \exp(2\lambda_{t|t-1}) + (y_t - \mu_{t|t-1})^2} - 1$$

Seasonally adjusted rate of inflation in the United States. The rate of inflation is often taken to follow a random walk plus noise and so the estimator of the level is an exponentially weighted moving average of current and past observations. Thus

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa_{\mu} v_t,$$

Fitting a Gaussian model with the STAMP8 package of Koopman *et al* (2007), gives an estimate of 0.579 for the parameter corresponding to κ_{μ} . The plot of the filtered level, $\mu_{t+1|t}$, shows it to be sensitive to extreme values, while the ACF of the absolute values of the residuals provides strong evidence of serial correlation in variance.


ML estimates (with standard errors in parentheses): for location

$$\widetilde{\kappa}_{\mu}=$$
 0.699(0.097),

for scale

$$\widetilde{\delta}=-$$
0.370(0.214), $\widetilde{\phi}=$ 0.912(0.051), $\widetilde{\kappa}=$ 0. 118(0.041)

and $\tilde{\nu} = 11.71(4.58)$.

The filtered estimates respond less to extreme values than those from the homoscedastic Gaussian model.



Non-negative variables: duration, realized volatility and range

Engle (2002) introduced a class of multiplicative error models (MEMs) for modeling non-negative variables, such as duration, realized volatility and range.

The conditional mean, $\mu_{t:t-1},$ and hence the conditional scale, is a GARCH-type process. Thus

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

where ε_t has a distribution with mean one and, in the first-order model,

$$\mu_{t|t-1} = \beta \mu_{t-1|t-2} + \alpha y_{t-1}.$$

The leading cases are the gamma and Weibull distributions. Both include the exponential distribution.

Non-negative variables: duration, realized volatility and range

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 the asymptotic distribution to be derived. The model can be written

$$y_t = \varepsilon_t \exp(\lambda_{t|t-1})$$

with dynamics

$$\lambda_{t|t-1} = \delta + \phi \lambda_{t-1|t-2} + \kappa u_{t-1},$$

where, for a Gamma distribution

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

The response is linear but this is not the case for Weibull, Log-logistic and F.





The DCS location model is

$$\mathbf{y}_t = \boldsymbol{\omega} + \boldsymbol{\mu}_{t|t-1} + \mathbf{v}_t, \quad \boldsymbol{\nu}_t \sim t_{\boldsymbol{\nu}} \left(\mathbf{0}, \Omega
ight), \quad t = 1, ..., T$$
 $\boldsymbol{\mu}_{t+1|t} = \Phi \boldsymbol{\mu}_{t|t-1} + \mathbf{K} \mathbf{u}_t.$

A direct extension of Beta-t-EGARCH to model changing scale, $\Omega_{t|t-1}$, is difficult. Matrix exponential is $\Omega_{t|t-1} = \exp \Lambda_{t|t-1}$. As a result, $\Omega_{t|t-1}$ is always p.d. and if $\Lambda_{t|t-1}$ is symmetric then so is $\cdot_{t|t-1}$; see Kawakatsu (2006, JE). Unfortunately, the relationship between the elements of $\Omega_{t|t-1}$ and those of $\Lambda_{t|t-1}$ is hard to disentangle. Can't separate scale from association.

Issues of interpretation aside, differentiation of the matrix exponential is needed to obtain the score and this is not straightforward.

A better way forward is to follow the approach in Creal, Koopman and Lucas (2011, JBES) and let

$$\Omega_{t|t-1} = \mathbf{D}_{t|t-1} \mathbf{R}_{t|t-1} \mathbf{D}_{t|t-1},$$

where $\mathbf{D}_{t|t-1}$ is *diagonal* and $\mathbf{R}_{t|t-1}$ is a pd correlation matrix with diagonal elements equal to unity. An exponential link function can be used for the volatilities in $\mathbf{D}_{t|t-1}$.

If only the volatilities change, ie $\mathbf{R}_{t|t-1} = \mathbf{R}$, it is possible to derive the asymptotic distribution of the ML estimator.

Assume a bivariate model with a conditional Gaussian distribution. Zero means and variances time-invariant.

How should we drive the dynamics of the filter for changing correlation, $\rho_{t|t-1}$, and with what link function ? Specify the standard deviations with an exponential link function so $Var(y_i) = \exp(2\lambda_i)$, i = 1, 2.

A simple moment approach would use

$$\frac{y_{1t}}{\exp(\lambda_1)}\frac{y_{2t}}{\exp(\lambda_2)} = x_{1t}x_{2t},$$

to drive the covariance, but the effect of $x_1 = x_2 = 1$ is the same as $x_1 = 0.5$ and $x_2 = 4$.

Better to transform $\rho_{t|t-1}$ to keep it in the range, $-1 \leq \rho_{t|t-1} \leq 1.$ The link function

$$\rho_{t|t-1} = \frac{\exp(2\gamma_{t|t-1}) - 1}{\exp(2\gamma_{t|t-1}) + 1}$$

allows $\gamma_{t|t-1}$ to be unconstrained. The inverse is the **arctanh** transformation originally proposed by Fisher to create the *z*-transform (his *z* is our γ) of the correlation coefficient, *r*, which has a variance that depends on ρ .

But $\tanh^{-1} r$ is asymptotically normal with mean $\tanh^{-1} \rho$ and variance 1/T.

The dynamic equation for correlation is defined as

$$\gamma_{t+1|t} = (1-\phi)\omega + \phi\gamma_{t|t-1} + \kappa u_t, \quad t = 1, \dots, T.$$

Setting $x_i = y_i \exp(-\lambda_i)$, i = 1, 2, as before gives the score as

$$\frac{\partial \ln L}{\partial \gamma} = \frac{1}{2} (x_1 + x_2)^2 \exp(-\gamma_{t|t-1}) - \frac{1}{2} (x_1 - x_2)^2 \exp(\gamma_{t|t-1}),$$

The score reduces to x_1x_2 when $\rho = 0$, but more generally the second term makes important modifications. It is zero when $x_1 = x_2$ while the first term gets larger as the correlation moves from being strongly positive, that is $\gamma_{t|t-1}$ large, to negative. In other words, $x_1 = x_2$ is evidence of strong positive correlation, so little reason to change $\gamma_{t|t-1}$ when $\rho_{t|t-1}$ is close to one but a big change is needed if $\rho_{t|t-1}$ is negative. Opposite effect if $x_1 = -x_2$.

The ML estimators of γ and the $\lambda's$ are asymptotically independent. The condititional score also provides guidance on dynamics for a copula - Creal et al (2011).

Is specifying the conditional variance in a GARCH-t model as a linear combination of past squared observations appropriate? The score of the *t*-distribution is an alternative to squared observations. **

The score transformation can also be used to formulate an equation for the logarithm of the conditional variance, in which case no restrictions are needed to ensure that the conditional variance remains positive. **

Since the score variables have a beta distribution, we call the model Beta-t-EGARCH. The transformation to beta variables means that *all* moments of the observations exist when the equation defining the logarithm of the conditional variance is stationary.

Furthermore, it is possible to obtain analytic expressions for the kurtosis and for the autocorrelations of powers of absolute values of the observations.

**

Volatility can be nonstationary, but an attraction of the EGARCH model is that, when the logarithm of the conditional variance is a *random walk*, it does not lead to the variance collapsing to zero almost surely, as in IGARCH.

Closed form expressions may be obtained for multi-step forecasts of volatility from Beta-t-EGARCH models, including nonstationary models and those with leverage. *There is a closed form expression for the mean square error of these forecasts.* (Or indeed the expectation of any power). **

When the conditional distribution is a GED, the score is a linear function of absolute values of the observations raised to a positive power. These variables have a gamma distribution and the properties of the model, Gamma-GED-EGARCH, can again be derived. For a Laplace distribution, it is equivalent to the standard EGARCH specification.

Beta-t-EGARCH and Gamma-GED-GARCH may both be modified to include **leverage** effects.

ML estimation of these EGARCH models seems to be relatively straightforward, avoiding some of the difficulties that can be a feature of the conventional EGARCH model. **

Unlike EGARCH models in general, a formal proof of the asymptotic properties of the ML estimators is possible. The main condition is that the score and its first derivative are independent of the TVP and hence time-invariant as in the static model.

Extends to (1) Two-component model; (2) Explanatory variables in the level or scale. (3) Higher-order models. (4) Nonstationary components (5) Skew distributions

Class of *Dynamic Conditional Score* models includes changing location and changing scale/location in models for non-negatve variables. **

Provides a solution to the specification of dynamics in multivariate models, including copulas.

*** THE END ***

Slides available at

http://www.econ.cam.ac.uk/faculty/harvey/volatility.pdf

э