HIGHLIGHTS OF THE SCORE

Andrew Harvey (ach34@cam.ac.uk)

Faculty of Economics, University of Cambridge

December 2013

Introduction to dynamic conditional score (DCS) models

1. 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.
2. The defining feature is that the dynamics are driven by the *score* of the conditional distribution.
3. 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.
4. A full asymptotic distributional theory for ML estimators can be obtained, including analytic expressions for the asymptotic covariance matrix.

***

Harvey, A.C. **Dynamic models for volatility and heavy tails**. CUP. 2013
http://www.econ.cam.ac.uk/DCS
Creal et al (2011, JBES, 2013, JAE).

Introduction to dynamic conditional score (DCS) models

The class of **dynamic conditional score** models includes models for :
1. changing location observed with an error which may be subject to outliers,
2. changing conditional variance,
3. changing location/scale for non-negative variables,
4. changing correlation and association

Introduction to dynamic conditional score (DCS) models

1. Why the score?
2. Robustness, Student's t and EGB2
3. Asymptotic properties
4. Scale
5. Correlation and copulas
6. Kernels and quantiles

Why the score ?

Suppose that at time $t-1$ we have $\widehat{\theta}_{t-1}$ the ML estimate of a parameter, θ. Then the score is zero, that is

$$\frac{\partial \ln L_{t-1}(\theta)}{\partial \theta} = \sum_{j=1}^{t-1} \frac{\partial \ln \ell_j(\theta)}{\partial \theta} = 0 \qquad \text{at} \quad \theta = \widehat{\theta}_{t-1}. \tag{1}$$

where $\ell_j(\theta; y_j) = f(y_j; \theta)$. When a new observation becomes available, a single iteration of the *method of scoring* gives

$$\begin{aligned} \widehat{\theta}_t &= \widehat{\theta}_{t-1} + \frac{1}{I_t(\widehat{\theta}_{t-1})} \frac{\partial \ln L_t(\theta)}{\partial \theta} \\ &= \widehat{\theta}_{t-1} + \frac{1}{I_t(\widehat{\theta}_{t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta} \end{aligned}$$

where $I_t(\widehat{\theta}_{t-1}) = t.I(\widehat{\theta}_{t-1})$ is the information matrix for t observations and the last line follows because of (1).

Why the score ?

For a Gaussian distribution a single update goes straight to the ML estimate at time t (recursive least squares).

Remark

We may choose a link function so that the information quantity does not depend on θ.

As $t \rightarrow \infty$, $I_t(\widehat{\theta}_{t-1}) \rightarrow \infty$ so the recursion becomes closed to new information. If it is thought that θ changes over time, the filter needs to be opened up. This may be done by replacing $1/t$ by a constant, which may be denoted as κ.

Why the score ?

Thus

$$\widehat{\theta}_t = \widehat{\theta}_{t-1} + \kappa \frac{1}{I(\widehat{\theta}_{t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta}$$

With no information about how θ might evolve, the above equation might be converted to the predictive form by letting $\theta_{t+1|t} = \widehat{\theta}_t$ so

$$\theta_{t+1|t} = \theta_{t|t-1} + \kappa \frac{1}{I(\theta_{t|t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta}$$

For a Gaussian distribution in which θ is the mean and the variance is known to be σ^2, the recursion is an EWMA because

$$\frac{1}{I(\theta_{t|t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta} = y_t - \theta_{t|t-1}$$

Why the score ?

If there is reason to think that the parameter tends to revert to an underlying level, ω, the updating scheme might become

$$\theta_{t+1|t} = \omega(1-\phi) + \phi\theta_{t|t-1} + \kappa \frac{1}{I(\theta_{t|t-1})} \frac{\partial \ln \ell_t(\theta)}{\partial \theta}$$

where $|\phi| < 1$. This filter corresponds to a first-order autoregressive process.
More generally we might introduce lags so as to smooth out the changes or allow for periodic effects. This leads to the formulation of a QARMA filter.
The score can also be motivated by a conditional mode argument based on smoothed estimates, $\theta_{t|T}$. See Durbin and Koopman (2012, p 252-3) and Harvey (2013, p 87-9)

Why the score ?

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.

Why the score ?

(1) The attraction of treating the score-driven filter as a model in its own right is that it becomes possible to derive its properties and to find the asymptotic distribution of the ML estimator.
(2) Seen in this way, 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.
(3) An immediate practical advantage comes from the response of the score to an outlier.

Dynamic location model

$$y_t = \omega + \mu_{t|t-1} + v_t = \omega + \mu_{t|t-1} + \exp(\lambda)\varepsilon_t,$$
$$\mu_{t+1|t} = \phi\mu_{t|t-1} + \kappa u_t,$$

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.
$u_t \to 0$ as $|y| \to \infty$. In the robustness literature this is called a redescending M-estimator. It is a gentle form of *trimming*.

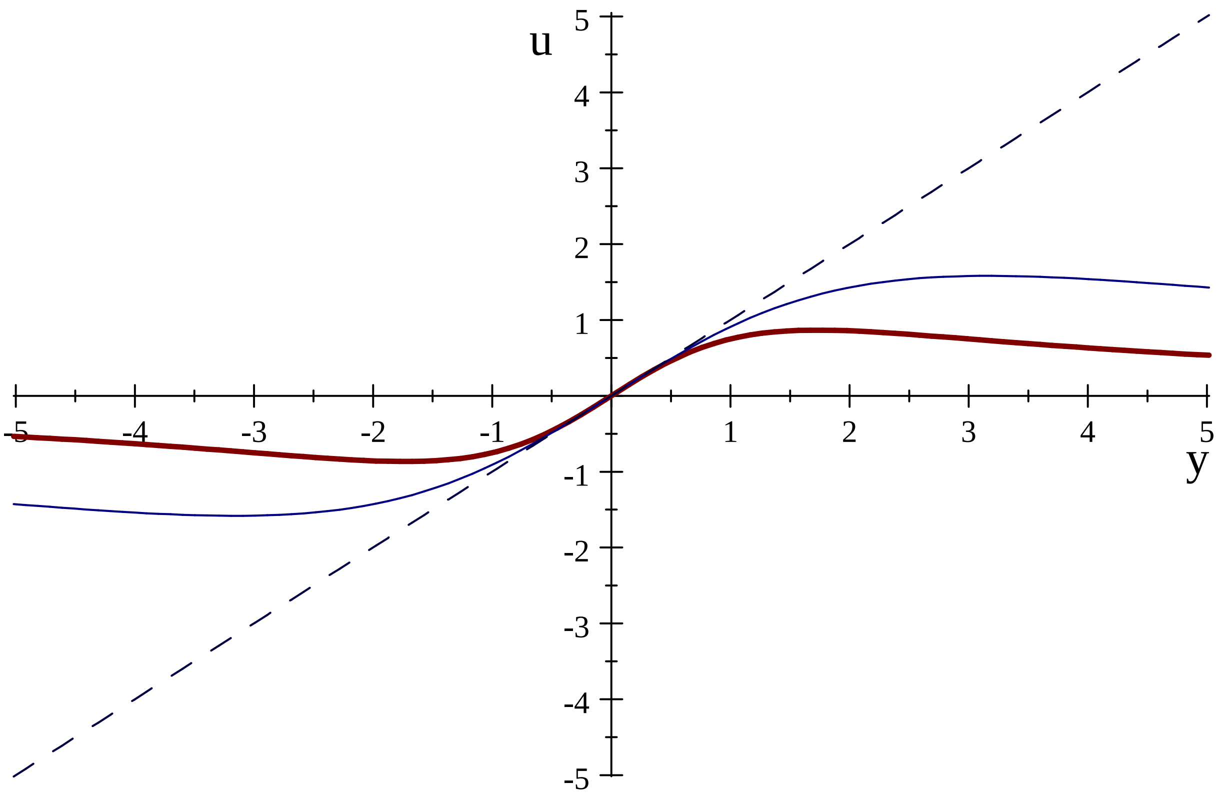


Figure: Impact of u_t for t_ν (with a scale of one) for $\nu = 3$ (thick), $\nu = 10$ (thin) and $\nu = \infty$ (dashed).

For the local level DCS-t filter,

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \\ \mu_{t+1|t} &= \mu_{t|t-1} + \kappa u_t. \end{aligned} \quad (2)$$

Since $u_t = (1 - b_t)(y_t - \mu_{t|t-1})$, re-arranging the dynamic equation gives

$$\mu_{t+1|t} = (1 - \kappa(1 - b_t))\mu_{t|t-1} + \kappa(1 - b_t)y_t, \quad t = 1, ..., T. \quad (3)$$

A sufficient condition for the weights on current and past observations to be non-negative is $0 < \kappa \leq 1$. But the restriction that $\kappa \leq 1$ is much stricter than is either necessary or desirable.
De Rossi and Harvey (2009) show that when the dynamic equation has a unit root, the conditional mode argument leads to smoothed estimates that satisfy

$$\sum_{t=1}^{T} u(y_t, \mu_{t|T}) = 0.$$

Life beyond the score

May have another criterion function, $\rho(\theta) = \sum \rho(\theta; y_t)$ eg sum of squares or a robust function (M-estimator).
Dynamics then driven by its derivative, $\rho'(\theta; y_t) = \partial\rho(\theta; y_t)/\partial\theta$.

Or have an *estimating equation* to be satisfied for estimating θ, that is

$$\sum_{t=1}^{T} g(\theta; y_t) = 0$$

where $g(y_t, \theta) = \rho'(\theta; y_t)$ when it comes from a criterion function.
When θ is dynamic, $g(\theta_{t|t-1}; y_t)$ replaces the score, u_t.

Alternatively consider choosing dynamic parameters so that $\sum_{t=1}^{T} g(\theta_{t|T}; y_t) = 0$, *where* $\theta_{t|T}$ *is a smoothed estimator.*

GARCH

The *generalized autoregressive conditional heteroscedasticity* (GARCH) model, introduced, as ARCH, by Engle (1982) and generalized by Bollerslev (1986) and Taylor (1986), is the classic way of modeling changes in the volatility of returns. It does so by letting the variance be a linear function of past squared observations. The first-order model, $GARCH(1,1)$, is

$$y_t = \sigma_{t|t-1}\varepsilon_t, \quad \varepsilon_t \sim NID(0,1) \tag{4}$$

and

$$\sigma^2_{t|t-1} = \delta + \beta\sigma^2_{t-1|t-2} + \alpha y^2_{t-1}, \quad \delta > 0, \beta \geq 0, \alpha \geq 0. \tag{5}$$

The conditions on α and β ensure that the variance remains positive. The sum of α and β is typically close to one and the *integrated GARCH* (IGARCH) model is obtained when the sum is equal to one. The variance in IGARCH is an exponentially weighted moving average of past squared observations and, as such, is often used by practitioners.

DCS Volatility Models

For a DCS model, replace u_t in the conditional variance equation

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

by another MD

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

which is proportional to the **score** of the conditional variance.

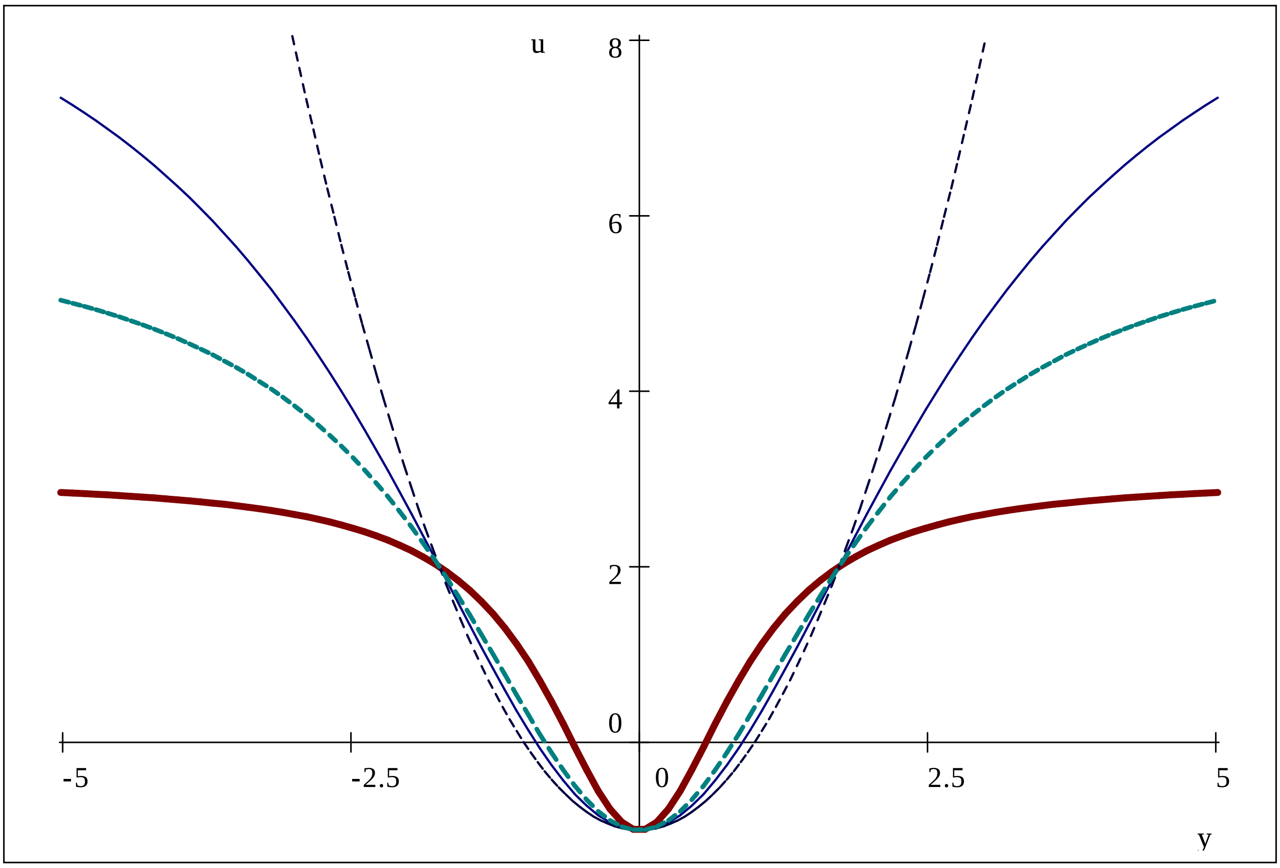


Figure: Impact of u_t for t_ν with $\nu = 3$ (thick), $\nu = 6$ (medium dashed) $\nu = 10$ (thin) and $\nu = \infty$ (dashed).

Exponential DCS Volatility Models

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

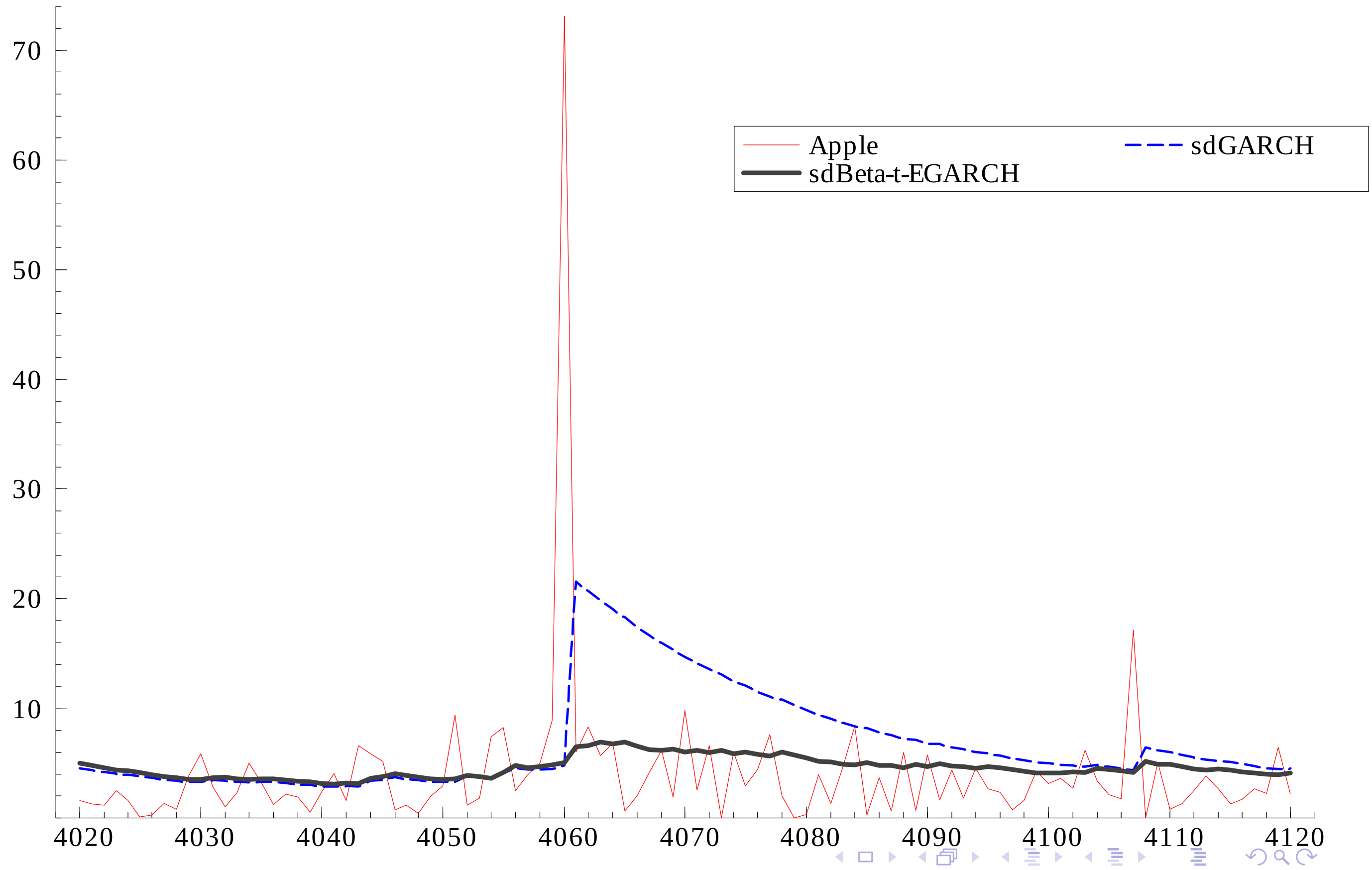
where the serially independent, zero mean variable ε_t has a $t_\nu-$distribution with degrees of freedom, $\nu > 0$, and the dynamic equation for the log of scale is

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

The conditional score is

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

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



Beta-t-EGARCH

The variable u_t may be expressed as

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

where

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

is distributed as $Beta(1/2, \nu/2)$.
The $u_t's$ are IID.

Beta-t-EGARCH: moments

The existence of unconditional moments of the observations, y_t, depends only on the existence of moments of the conditional distribution, that is the distribution of ε_t.

The moments of the scale always exist and hence the volatility process does not affect the existence of unconditional moments.
Analytic expressions for the unconditional moments can be derived for $|y_t|^c$, $c \geq 0$.
Can also find expresions for autocorrelations of $|y_t|^c$.

Gamma-GED-EGARCH

General Error distribution (GED) leads to Gamma-GED-EGARCH model. We have

$$\begin{aligned}\ln L(\lambda, \upsilon) &= -T\left(1+\upsilon^{-1}\right)\ln 2 - T\ln\Gamma(1+\upsilon^{-1}) - \sum_{t=1}^{T}\lambda_{t|t-1} \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\left|y_t \exp(-\lambda_{t|t-1})\right|^{\upsilon},\end{aligned}$$

where υ is a shape parameter.
The score

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

is **gamma** distributed.

Gamma-GED-GARCH*

Assume υ is known and let $\xi = \varphi^{\upsilon}$. Setting the score to zero yields the ML estimator

$$\widetilde{\xi} = \frac{\upsilon}{2} \frac{\sum |y_t|^{\upsilon}}{T},$$

The form of this estimator suggests that, if the scale changes over time, a weighting scheme should be applied to $|y_t|^{\upsilon}$, giving a variant of Gamma-GED-GARCH:

$$\xi_{t+1|t} = \delta + \phi \xi_{t|t-1} + \theta \xi_{t|t-1} u_t,$$

Re-writing $\xi_{t|t-1}$ in terms of $\sigma^{\upsilon}_{t|t-1}$ yields the power ARCH or APARCH class of models of Ding, Granger and Engle (1993), except that there the conditional distribution of the observations is normal. (Leverage effects are also handled differently). Setting $\upsilon = 2$ gives GARCH whereas $\upsilon = 1$ gives the equation proposed by Taylor (1986).

Asymptotic distribution of ML estimator*

Now let $\theta = \theta_{t|t-1}$ evolve over time as a function of past observations and past values of the score of the conditional distribution. Since the conditional score depends on past observatons through $\theta_{t|t-1}$, it can be broken down into two parts: where the notation $f_t(y_t; \theta_{t|t-1})$ indicates that the distribution of y_t depends on the time-vary

$$\frac{\partial \ln f_t(y_t \mid Y_{t-1}; \boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = \frac{\partial \ln f_t(y_t; \theta_{t|t-1})}{\partial \theta_{t|t-1}} \frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}},$$

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

The above decomposition of the conditional score leads to the following result.

Asymptotic distribution of ML estimator*

Consider a model with a single time-varying parameter, $\theta_{t|t-1}$, 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, $\boldsymbol{\psi}$. Under cerytain conditions, the conditional score for the $t-$th observation, $\partial \ln f_t(y_t \mid Y_{t-1}; \boldsymbol{\psi})/\partial \boldsymbol{\psi}$, is a MD at $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, with conditional covariance matrix

$$E_{t-1}\left(\frac{\partial \ln f_t(y_t \mid Y_{t-1}; \boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right)\left(\frac{\partial \ln f_t(y_t \mid Y_{t-1}; \boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right)' = I.\left(\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}}\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}'}\right), \quad t$$

where the information quantitly, I, is constant over time and independent of $\boldsymbol{\psi}$.

Asymptotic distribution of ML estimator*

Follows because, the conditional covariance matrix of the score is found by writing its outer product as

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

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

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

Asymptotic distribution of ML estimator*

The information matrix is

$$\mathbf{I}(\boldsymbol{\psi}) = I.\mathbf{D}(\boldsymbol{\psi}), \quad \textit{where} \quad \mathbf{D}(\boldsymbol{\psi}) = E\left(\frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}} \frac{\partial \theta_{t|t-1}}{\partial \boldsymbol{\psi}'}\right).$$

The following definitions are needed for $\mathbf{D}(\boldsymbol{\psi})$:

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

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

Asymptotic distribution of ML estimator*

Then

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

with

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

$$D = \frac{a\kappa\sigma_u^2}{1-a\phi}, \qquad E = \frac{c(1-\phi)}{1-a} \qquad \textit{and} \qquad F = \frac{ac\kappa(1-\phi)}{(1-a)(1-a\phi)}.$$

Asymptotic distribution of ML estimator*

Provided that $b < 1$ and that the elements of $\boldsymbol{\psi}$ do not lie on the boundary of the parameter space, the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi})$, where $\widetilde{\boldsymbol{\psi}}$ is the ML estimator of $\boldsymbol{\psi}$, is multivariate normal with mean zero and covariance matrix

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

If the unit root is imposed, so that $\phi = 1$, then standard asymptotics apply.

Robust estimation and tail behaviour

The Gaussian distribution has kurtosis of three and a distribution is said to exhibit *excess kurtosis* if its kurtosis is greater than three.
A distribution is said to be *heavy-tailed* if

$$\lim_{y \to \infty} \exp(y/\beta)\overline{F}(y) = \infty \qquad \text{for all } \beta > 0, \tag{6}$$

where $\overline{F}(y) = \Pr(Y > y) = 1 - F(y)$ is the survival function.
When y has an exponential distribution, $\overline{F}(y) = \exp(-y/\alpha)$, so when $\beta = \alpha$, $\exp(y/\alpha)\overline{F}(y) = 1$ for all y.
The exponential distribution is not heavy-tailed. Nor is GED.

Robust estimation and tail behaviour

A distribution is said to be *fat-tailed* if, for a fixed positive value of η,

$$\overline{F}(y) = cL(y)y^{-\eta}, \qquad \eta > 0, \tag{7}$$

where c is a non-negative constant and $L(y)$ is slowly varying, that is $\lim_{y\to\infty}(L(ky)/L(y)) = 1$. eg Pareto distribution - $\overline{F}(y) = y^{-\eta}$ for $y > 1$.

The parameter η is the **tail index**. The implied PDF is a *power law* PDF

$$f(y) \sim cL(y)\eta y^{-\eta-1}, \qquad \eta > 0, \tag{8}$$

The m-th moment exists if $m < \eta$.

The complement to the power law PDF is

$$f(y) \sim cL(y)\overline{\eta}y^{\overline{\eta}-1} \quad as \quad y \to 0, \quad 0 < y < 1, \quad \overline{\eta} > 0.$$

EGB2 (work with Michele Caivano)

The exponential generalized beta distribution of the second kind (EGB2) is obtained by taking the logarithm of a variable with a GB2 distribution.

The PDF of a GB2 is

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

where α is the scale parameter, ν, ξ and ς are shape parameters and $B(\xi,\varsigma)$ is the beta function; see Kleiber and Kotz (2003, ch6).

The GB2 distribution contains many important distributions as special cases, including the Burr ($\xi = 1$) and log-logistic ($\xi = 1, \varsigma = 1$).

GB2 distributions are fat tailed for finite ξ and ς with upper and lower tail indices of $\eta = \varsigma\nu$ and $\overline{\eta} = \xi\nu$ respectively.

The absolute value of a t_f variate is GB2$(\varphi, 2, 1/2, f/2)$ with tail index is $\eta = \overline{\eta} = f$.

EGB2

If x is distributed as GB2$(\alpha, \nu, \xi, \varsigma)$ and $y = \ln x$, the PDF of the EGB2 variate y is

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

What was the logarithm of scale in GB2 now becomes location in EGB2, that is $\ln \alpha$ becomes μ. Furthermore ν is now a scale parameter, but ξ and ς are still shape parameters and they determine skewness and kurtosis. All moments exist.

EGB2

When $\xi = \varsigma$, the distribution is symmetric; for $\xi = \varsigma = 1$ it is a logistic distribution and when $\xi = \varsigma \to \infty$ it tends to a normal distribution. *For $\xi = \varsigma = 0$, the distribution is double exponential or Laplace.* Similar coverage to GED but asymptotics more standard. Also general form allows skewness,

Dynamic location model

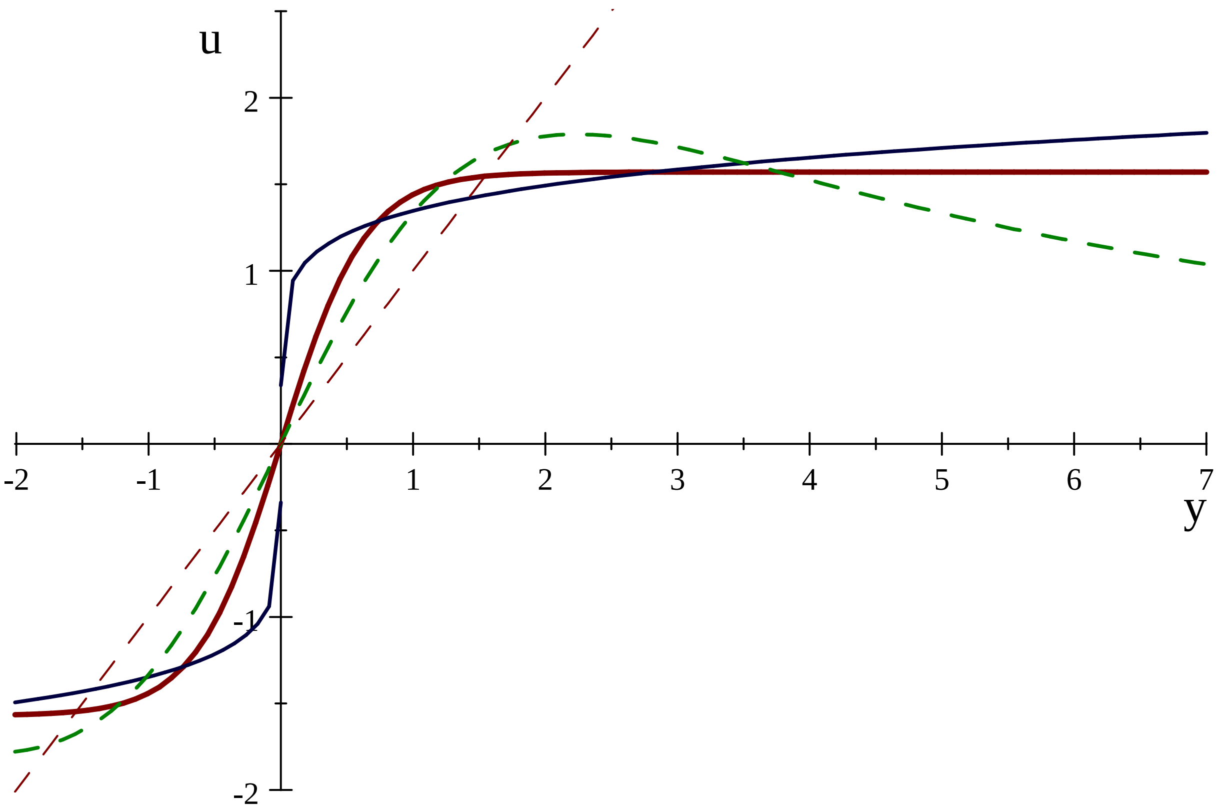
The score function for the EGB2 distribution with respect to location is

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

where

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

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



EGB2-EGARCH

As in the dynamic location case, the EGB2 distribution offers an alternative to the GED for capturing responses between the normal and Laplace. *Wang et al (2001) fitted GARCH-EGB2 models to daily dollar exchange rates and found evidence to favour them over GARCH-t alternatives.* The first-order dynamic scale model with EGB2 distributed errors is

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

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

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

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

EGB2-EGARCH

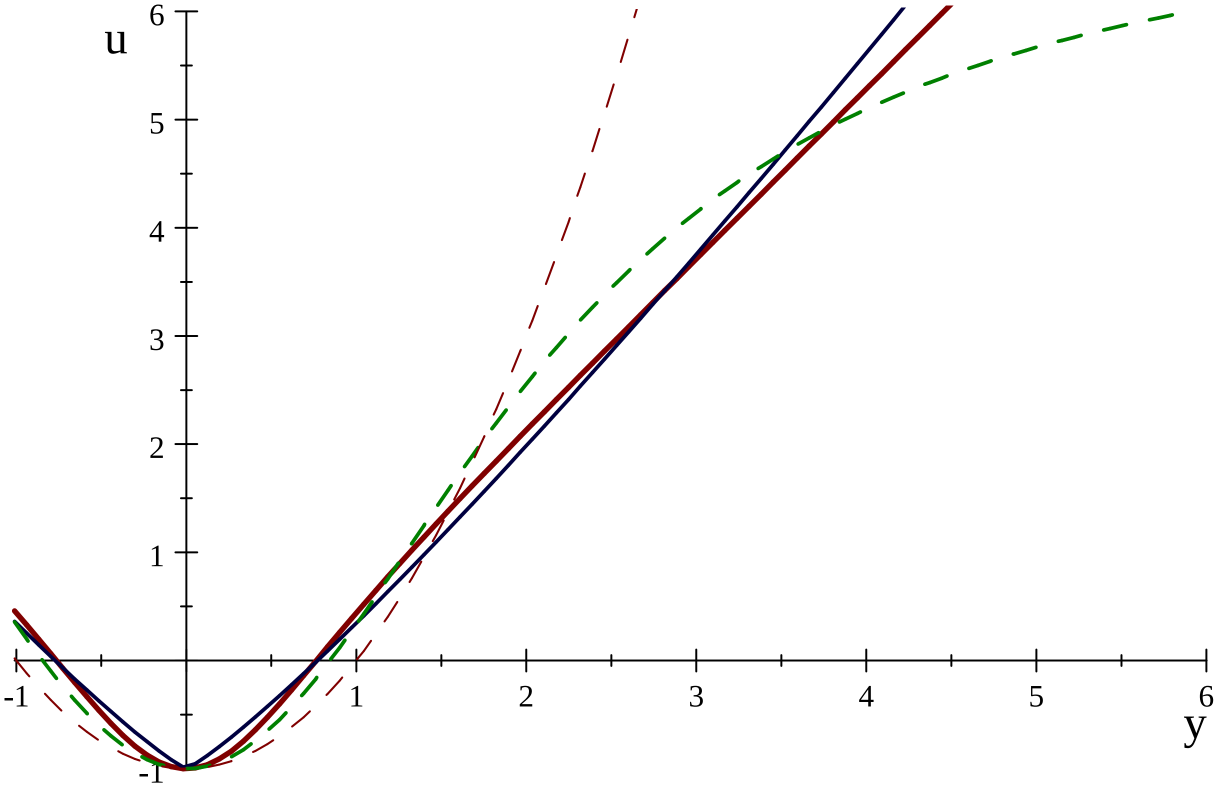
The conditional score is

$$u_t = \frac{\partial \ln f(y_t)}{\partial \lambda_{t|t-1}} = (\xi + \varsigma)\varepsilon_t b_t - \xi\varepsilon_t - 1,$$

and

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

At the true parameters values, $b_t \sim beta(\xi, \varsigma)$ as in the score for the dynamic location model.
Figure compares the way observations are weighted by the score of a EGB2 distribution with $\xi = \varsigma = 0.5$, a Student's t_7 distribution and a $GED(1.148)$.
Consistent with this relationship and the Winsorizing of the location score, dividing u_t by ϵ_t gives a bounded function as $|\epsilon_t| \to \infty$



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

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

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

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

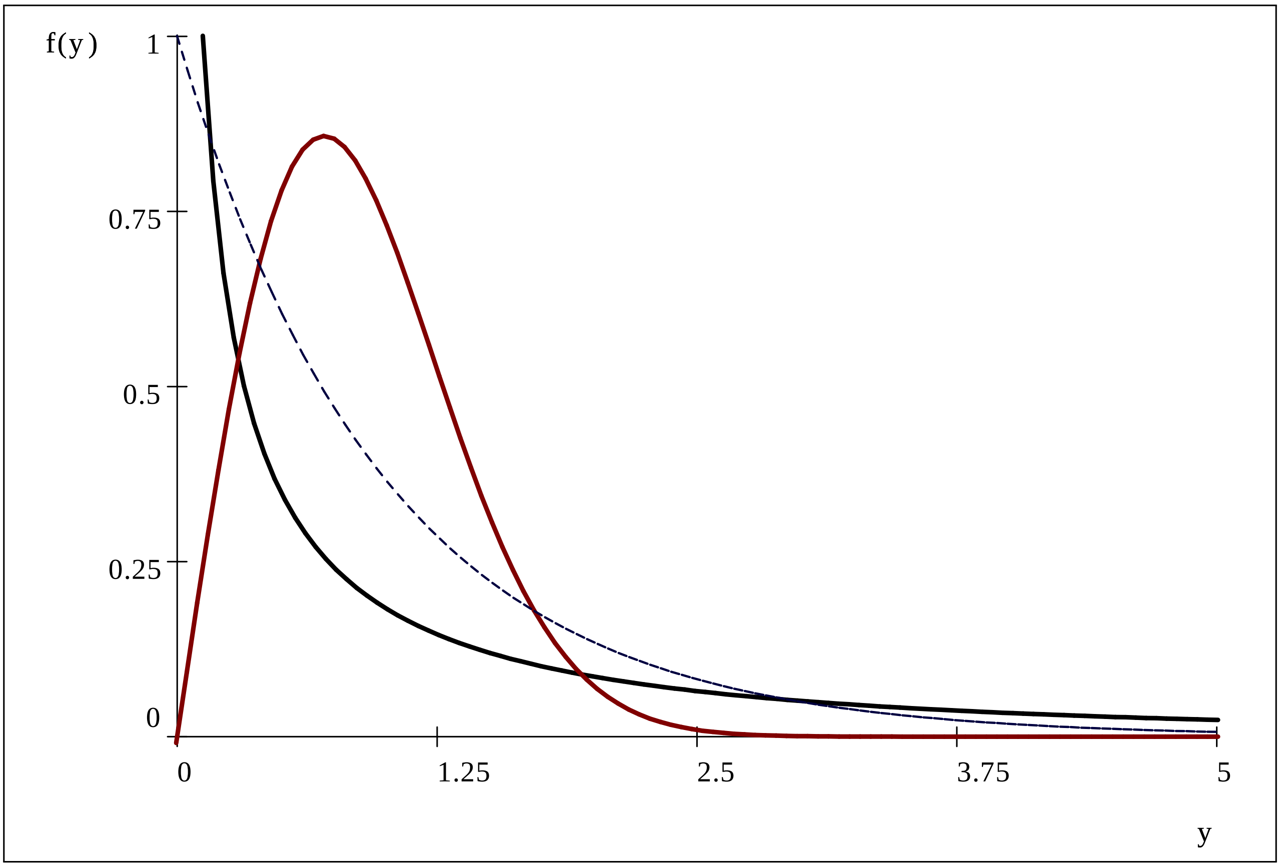
where, for a Gamma distribution

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

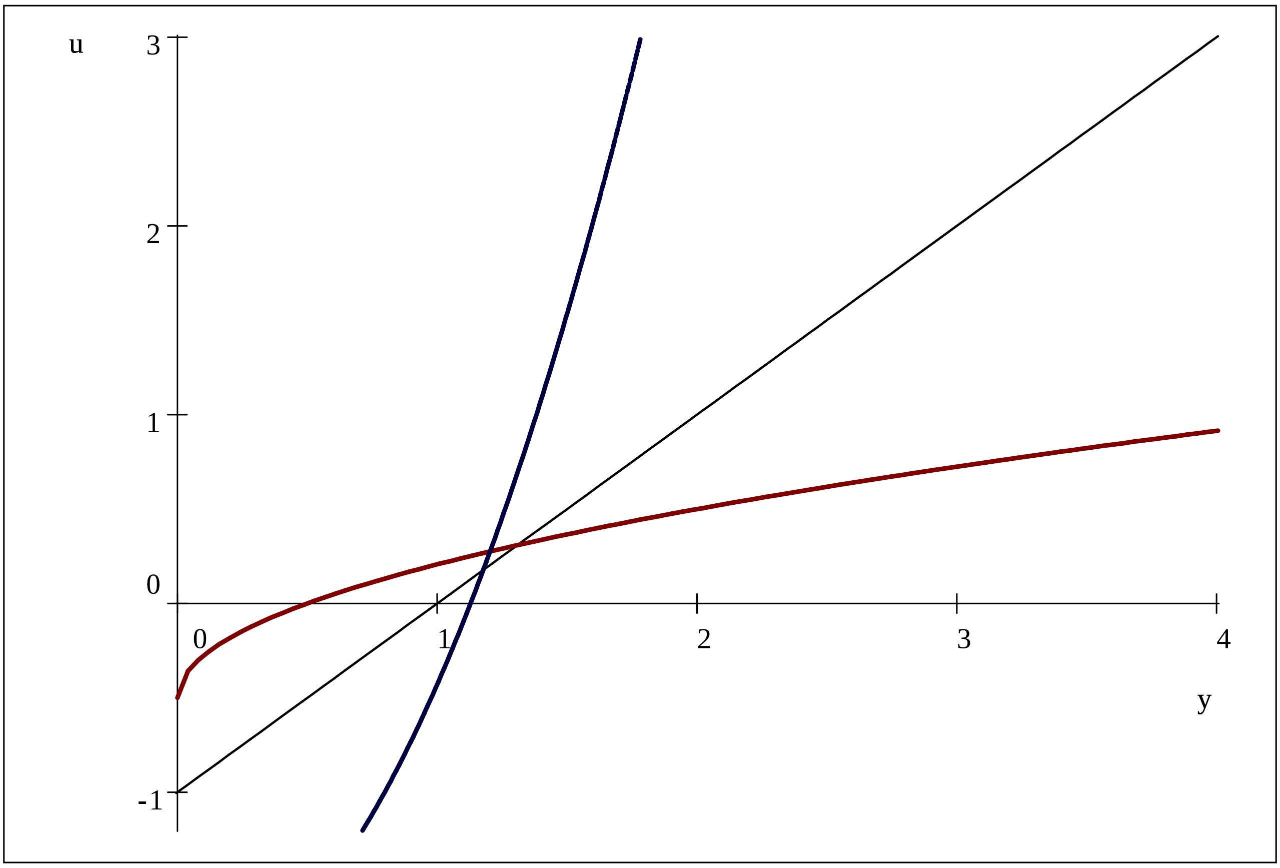
The response is linear but this is not the case for Weibull. The PDF is

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

where α is the scale and υ is the shape parameter. When $\upsilon < 1$ it is long-tailed. Score is concave.



Weibull density functions for $\upsilon = 0.5$, exponential ($\upsilon = 1$, dashes) and $\upsilon = 2$ (humped shape)



Weibull score functions for $\upsilon = 0.5$, exponential ($\upsilon = 1$, thin) and $\upsilon = 2$ (dashes)

Log-logistic distribution

$$f(y) = (\nu/\alpha)(y/\alpha)^{\nu-1}(1+(y/\alpha)^{\nu})^{-2}, \qquad \nu, \alpha > 0.$$

A time-varying scale with an exponential link function, ie $\alpha_{t|t-1} = \exp \lambda_{t|t-1}$, gives

$$\ln f_t(\boldsymbol{\psi}, \nu) = \ln \nu - \nu\lambda_{t|t-1} + (\nu - 1)\ln y_t - 2\ln(1 + (y_t e^{-\lambda_{t|t-1}})^{\nu}),$$

and so

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

where

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

is distributed as $beta(1,1)$. Since a $beta(1,1)$ distribution is a standard uniform distribution, it is immediately apparent that the expectation of u_t is zero.

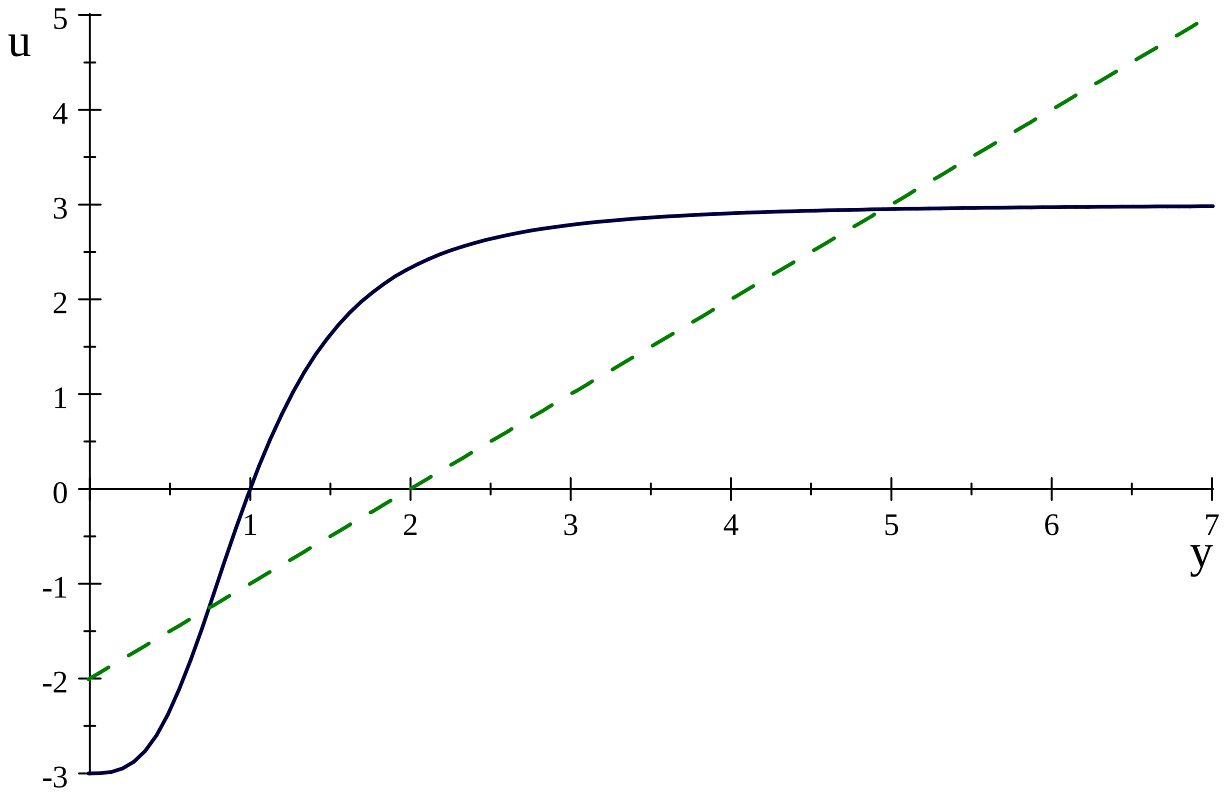


Figure: Impact of u for a log-logistic distribution and a gamma (dashed), with shape parameters $\nu = 3$ and $\gamma = 2$ respectively.

Log-logistic distribution

The asymptotic theory is not complicated. Differentiating the score gives

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

Proposition

Provided that $b < 1$, the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \widetilde{\nu} - \nu)'$ is multivariate normal with zero mean and covariance matrix

$$Var\begin{pmatrix} \widetilde{\boldsymbol{\psi}} \\ \widetilde{\nu} \end{pmatrix} = \begin{bmatrix} (3/\nu^2)\mathbf{D}^{-1}(\boldsymbol{\psi}) & \mathbf{0} \\ \mathbf{0}' & 1.430\nu^2 \end{bmatrix},$$

where $\mathbf{0}$ is a vector of zeroes and $\mathbf{D}(\boldsymbol{\psi})$ is as given earlier with

$$\begin{aligned} a &= \phi - \kappa\nu^2/3 \\ b &= \phi^2 - (2/3)\nu^2\phi\kappa + 2\kappa^2\nu^4/15 \quad \text{and} \quad c = 0. \end{aligned}$$

Generalized gamma and beta distributions

The statistical theory of DCS models for non-negative variables is simplified by the fact that for the gamma and Weibull distributions the score and its derivatives are dependent on a gamma variate, while for the Burr, log-logistic and F-distributions the dependence is on a beta variate.
Gamma and Weibull distributions are special cases of the generalized gamma distribution.
Burr and log-logistic distributions are special cases of the generalized beta distribution.
The $F-$distribution is related to the generalized beta distribution in that the special case when the degrees of freedom are the same is equivalent to a special case of the generalized beta.
Members of the generalized beta class are particularly useful in situations where there is evidence of heavy tails.

Other Volatility models: Classic EGARCH

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

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

but the dynamics are driven by:

$$g_t = |\varepsilon_t| - E\,|\varepsilon_t| = |\varepsilon_t| - \mu_{|\varepsilon|}.$$

where the leverage term, ε_t, has been dropped.
ML estimator is consistent and asymptotically normal. Analytic information matrix similar to DCS models
For t-distribution - classic EGARCH is of academic interest only - it has no moments.

Log-EGARCH

Nelson (1991, p 91) - mentions Econometric Reviews, 1986. Francq and Zakoian (2013, JE)

$$g_t = \ln|\varepsilon_t| - E\ln|\varepsilon_t|$$

so

$$\lambda_{t+1|t} = \omega(1-\phi) + \phi\lambda_{t|t-1} + \kappa(\ln|\varepsilon_t| - E(\ln|\varepsilon_t|))$$

The unconditional moments depend on the existence of $E(|\varepsilon_t|^{m\psi_{\max}})$ and $E(|\varepsilon_t|^{m\psi_{\min}})$ because $E(\exp m\psi_k g_t) = E(|\varepsilon_t|^{m\psi_k})$.

In Beta-t-EGARCH, the existence of unconditional moments depends solely on ν.

As with classic EGARCH, ML estimator is consistent and asymptotically normal. Analytic information matrix similar to DCS models

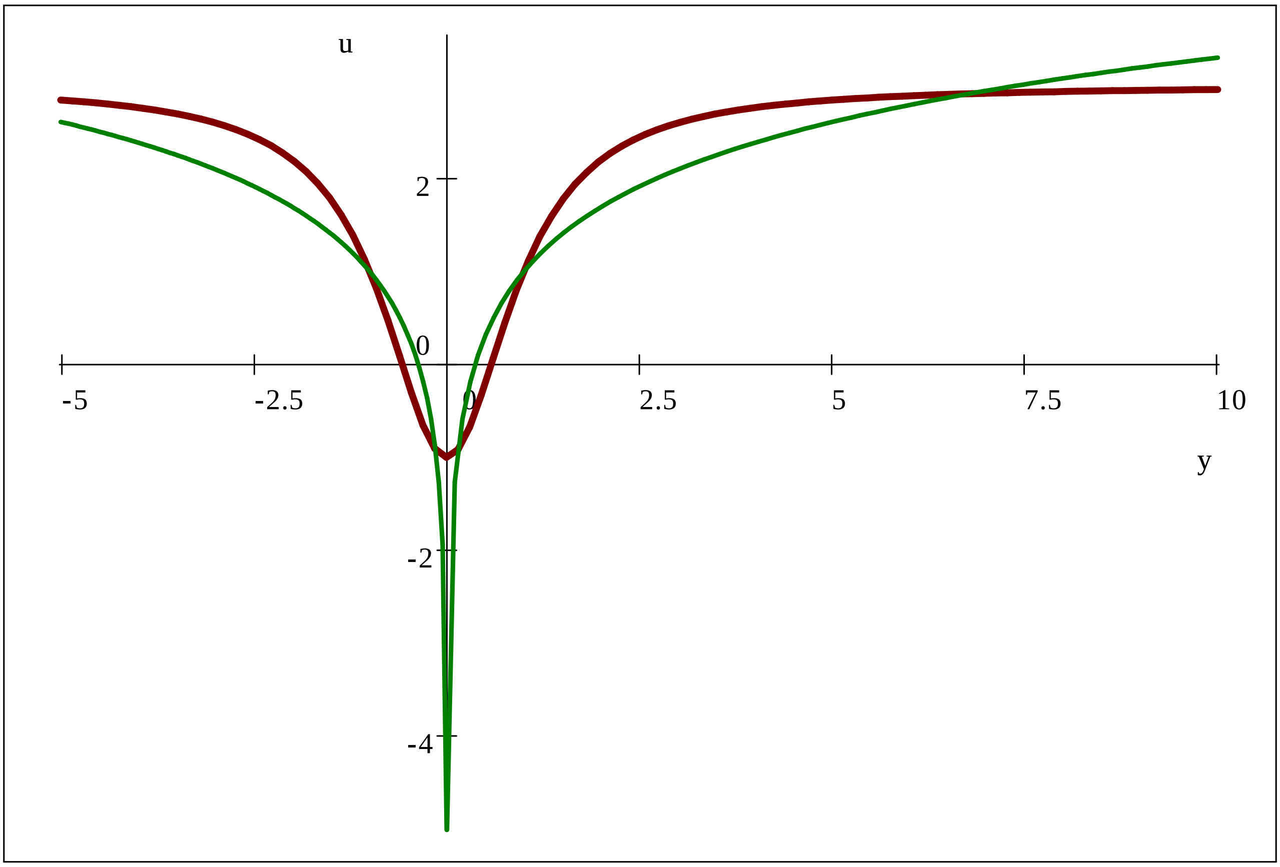


Figure: Impact of u for t_ν with $\nu = 3$ (thick), $\nu = 10$ (thin) and $\nu = \infty$ (dashed).

QML

QML is consistent, *provided that the dynamics are right*.
Only attractive if (i) relatively easy to compute (especially compared with ML eg SV model), (ii) reasonably efficient over a wide range of distributions; (iii) parameters have some meaning eg we want variance.
If (i) holds then might use as starting values.

QML

Can compare asymptotic distribution with that of an ML estimator for a model in which the dynamic equation is the **same**.
Let $\widehat{\boldsymbol{\psi}}$ be the estimator obtained by maximizing a quasi-likelihood function wrt a vector of parameters, $\boldsymbol{\psi}$, that determine the dynamic evolution of a time-varying parameter, $\theta_{t|t-1}$. Then $\widehat{\boldsymbol{\psi}}$ is asy normal with

$$Var(\widehat{\boldsymbol{\psi}}) = \left[E\left(\frac{\partial^2 \ln f_t}{\partial \theta^2} \right) \right]^{-2} E\left[\left(\frac{\partial \ln f_t}{\partial \theta} \right)^2 \right] \mathbf{D}^{-1}(\boldsymbol{\psi})$$

Efficiency is

$$Eff(\widehat{\psi}_i) = \frac{\left[E\left(\frac{\partial^2 \ln f_t}{\partial \theta^2} \right) \right]^2}{I.E\left[\left(\frac{\partial \ln f_t}{\partial \theta} \right)^2 \right]}, \quad i = 1, ..., m,$$

where I is the information quantity for the true distribution.

Conclusions on different dynamics and QML

The shortcomings of (i) $|\varepsilon_t|$ - no moments for heavy-tailed distributions - and (ii) ln $|\varepsilon_t|$ - problem when $y = 0-$ (highlights the benefits of the score !)
Thus $|\varepsilon_t|$ is not viable for t and ln $|\varepsilon_t|$ is not a good way to handle light-tailed. Need to decide whether the tails are sufficiently heavy to justify having the dynamics driven by the log.
Backed up by efficiency. GB2 and t could be very inefficient for QML in levels. Light-tailed inefficient for QML in logs.
Both considerations together suggest :
(1) **with light tailed use classic form for dynamics with QML on the level**
and
(2) **with heavy tailed use log form for dynamics with QML on the log**.

Conclusions on different dynamics and QML

Decision on whether to do QML in logs or levels and whether to specify the dynamics as being driven by $|\varepsilon_t|$ or ln $|\varepsilon_t|$ requires that the researcher takes a stance on the kind of distribution likely to be encountered.
Half-way towards specifying a distribution. Why not therefore fit t and EGB2 and compare the fit ?
Why not by-pass the whole issue of which dynamic equation to use by having a DCS model and estimating it by ML?

NB The coherency principle suggests that classic EGARCH should be coupled with minimization of absolute values, that is

$$\rho(\lambda_{t|t-1}; y_t) = \sum_{t=1}^{T} \left| y_t \exp(-\lambda_{t|t-1}) \right| .$$

QML minimizes a sum of squares function and so would have a GARCH-type equation in squares - this is clearly useless in EGARCH because it lacks moments when there is excess kurtosis.

Conclusions on different dynamics and QML

Same issues for MEM or ACD models.
Bauwens *et al* (2004) propose a 'log-ACD' specification has the conditional mean set to $\mu_{t|t-1} = \exp(\lambda^*_{t|t-1})$, where

$$\lambda^*_{t+1|t} = \delta + \beta\lambda^*_{t|t-1} + \alpha \ln y_t$$

or

$$\lambda^*_{t+1|t} = \delta + \beta\lambda^*_{t|t-1} + \alpha y_t \exp(-\lambda^*_{t|t-1}).$$

Better to use a DCS model with a GG or a GB2.
Such DCS models fit better - see Andres and Harvey (2012). Andres (2013) provides more details.

Robust estimators

M-estimators. Time series models coherent when dynamics driven by $\rho'(.;.)$.
Analytic information matrix difficult to derive, though asymptotics will surely go through.
But nothing to be gained compared with a parametric approach.
Just fit a t or EGB2.
The fundamental case for DCS comes from the fact that it is a coherent, likelihood based approach.

Dynamic shape parameters

Generalized Pareto distribution reparameterized so that

$$p(y) = \alpha^{-1}\varsigma\left[1 + (y - y_0)/\alpha\right]^{-(\varsigma+1)}, \quad \alpha > 0, \quad 0 < \varsigma < \infty,$$

Special case of the Burr distribution in which $\nu = 1$. (Exponential when $\varsigma \to \infty$.) The scale parameter may therefore be allowed to evolve over time, according to $\alpha_{t|t-1} = \exp(\lambda_{t|t-1})$, and the moments, forecasts and asymptotic distribution can be found from the general results for the GB2 distribution.
The tail index is ς. If it is time-varying, we might set $\varsigma_{t|t-1} = \exp(\theta_{t|t-1})$ so no constraints are needed. With this parameterization

$$\mathbf{I}\begin{pmatrix}\lambda \\ \theta\end{pmatrix} = \begin{bmatrix} \frac{e^{\theta}}{2+e^{\theta}} & \frac{-1}{1+e^{\theta}} \\ \frac{-1}{1+e^{\theta}} & 1 \end{bmatrix}$$

Dynamic shape parameters

The Pareto distribution is obtained with $\alpha = 1$ and $y_0 = 1$, because $p(y) = \varsigma\left[1 + (y-1)\right]^{-(\varsigma+1)} = \varsigma y^{-\varsigma-1}$, $y \geq 1$. With $\varsigma_{t|t-1} = \exp(\theta_{t|t-1})$, the score is

$$u = 1 - e^{\theta}\ln y, \quad y \geq 1.$$

Because $\ln y$ is exponentially distributed with scale $1/\varsigma$, all its moments exist and it can be seen immediately that the mean of the score is zero. Unlike the score for scale in the Burr, this is not bounded. However, $I(\theta) = 1$ and the information matrix for the first-order DCS model has

$$a = \phi - \kappa, \; b = \phi^2 - 2\phi\kappa + 2\kappa^2, \quad c = 1$$

Furthermore $u' = u'' = -e^{\theta}\ln y$ and it is not difficult to see that the conditions on p 40-45 of Harvey (2013) are satisfied.
Not surprising as the shape parameter becomes a scale parameter when logs are taken. (cf GB2 and EGB2)

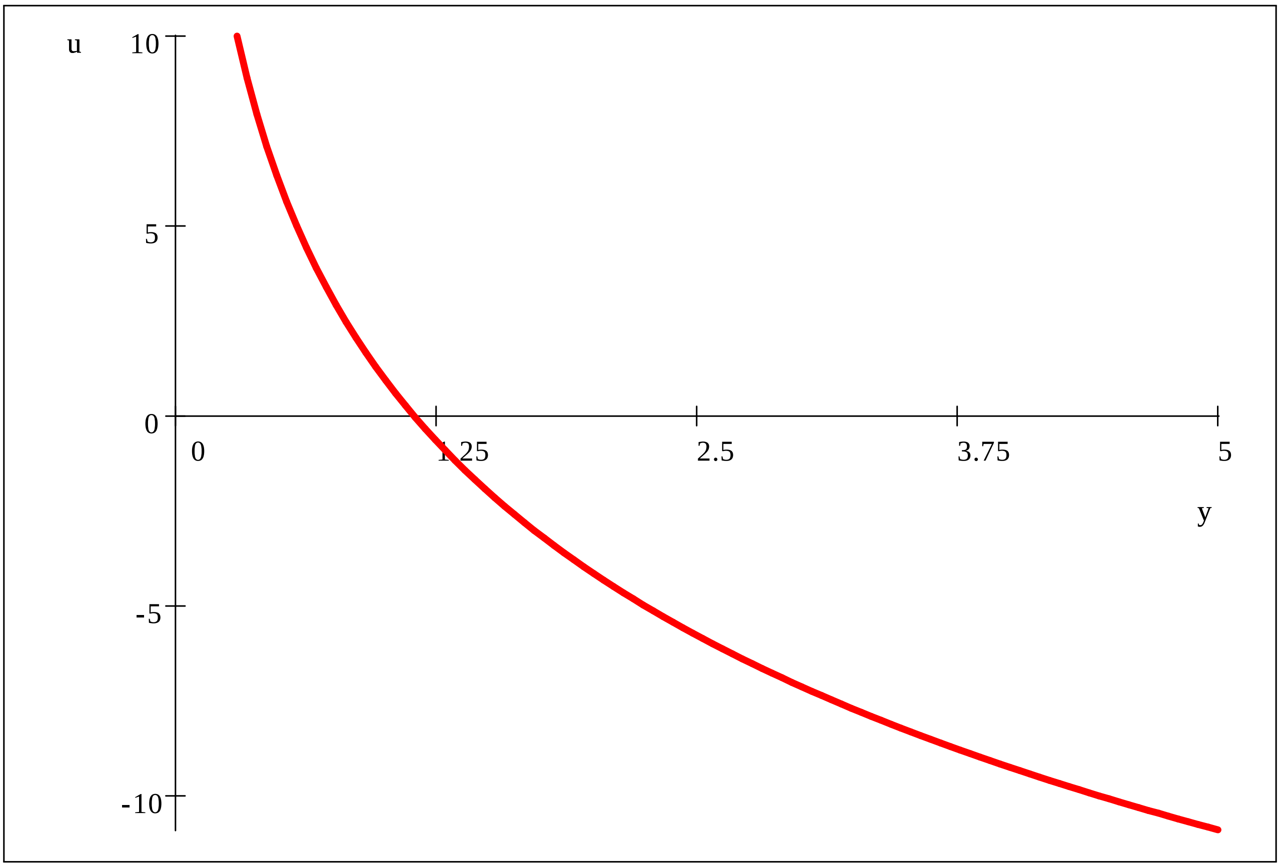


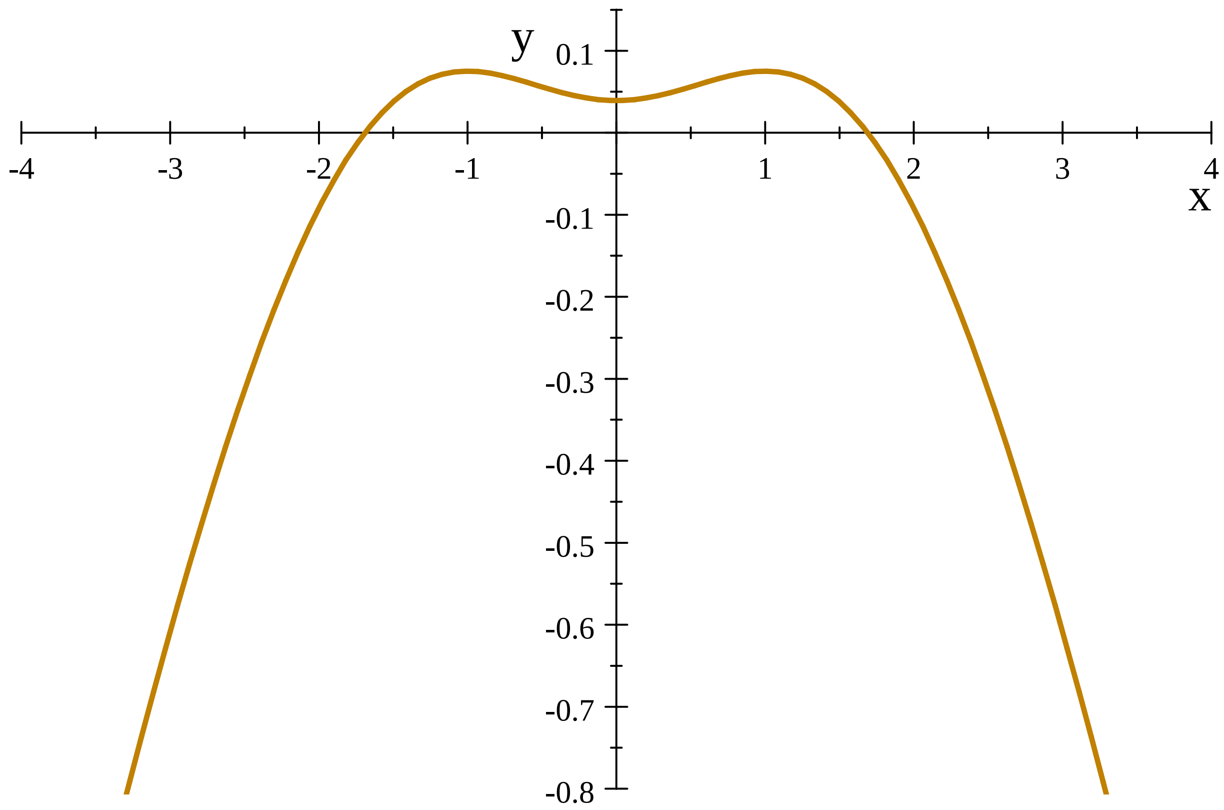
Figure: Score

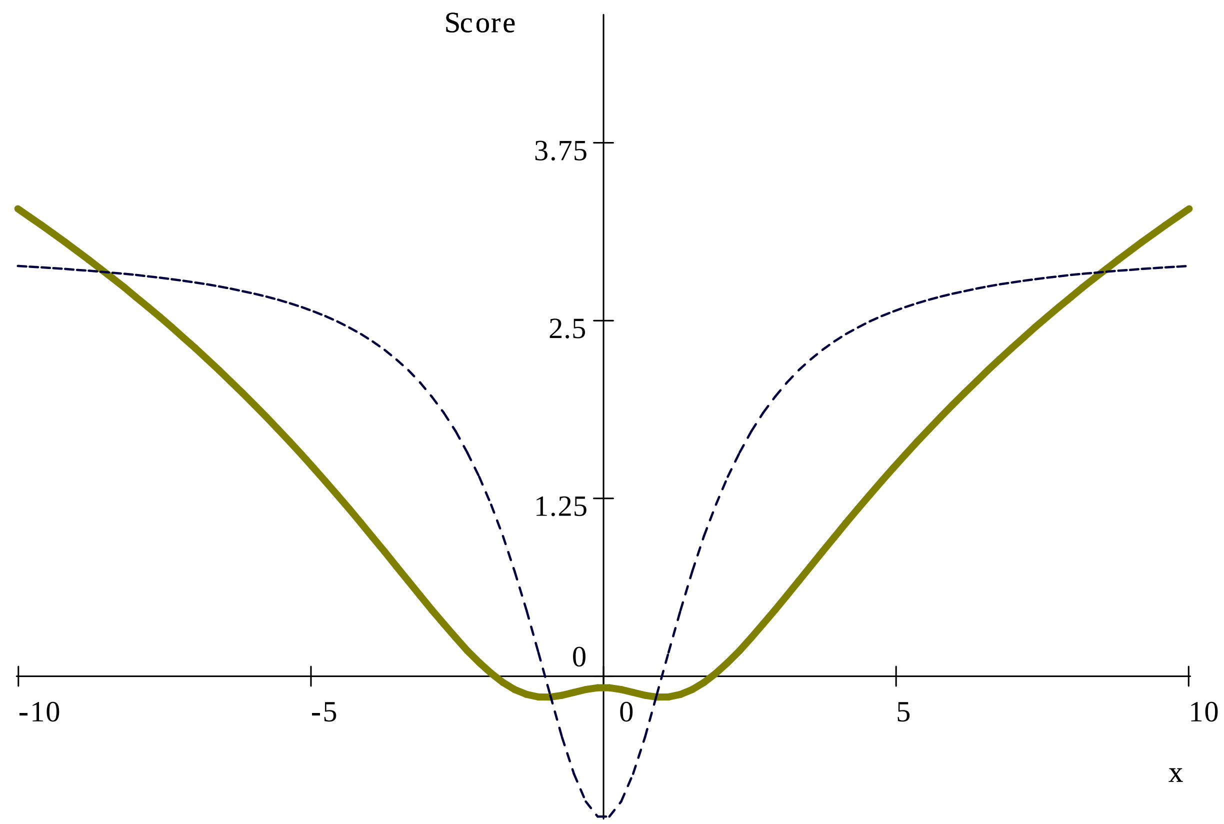
Dynamic shape parameters

Dynamic degrees of freedom,ν, in t-distribution. Score is

$$u_\nu = -\frac{1}{2}\,\mathrm{Psi}\,(\nu/2) + \frac{1}{2}\,\mathrm{Psi}\,((\nu+1)/2) - \frac{1}{2\nu} + \frac{\nu+1}{2\nu}b + \frac{1}{2}\ln(1-b)$$

where b is beta$(1/2, \nu/2)$.
Figure shows u_ν and then compares score for $-\ln\nu$ with the score for scale $(\nu = 3)$


y
0.1
-4
-3
-2
-1
1
2
3
4
x
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8


Score
3.75
2.5
1.25
0
-10
-5
0
5
10
x

Correlation and copulas: Estimating changing correlation

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 $x_2 = 4$.

Estimating changing correlation

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

Estimating changing correlation

The dynamic equation for correlation is defined as

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

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

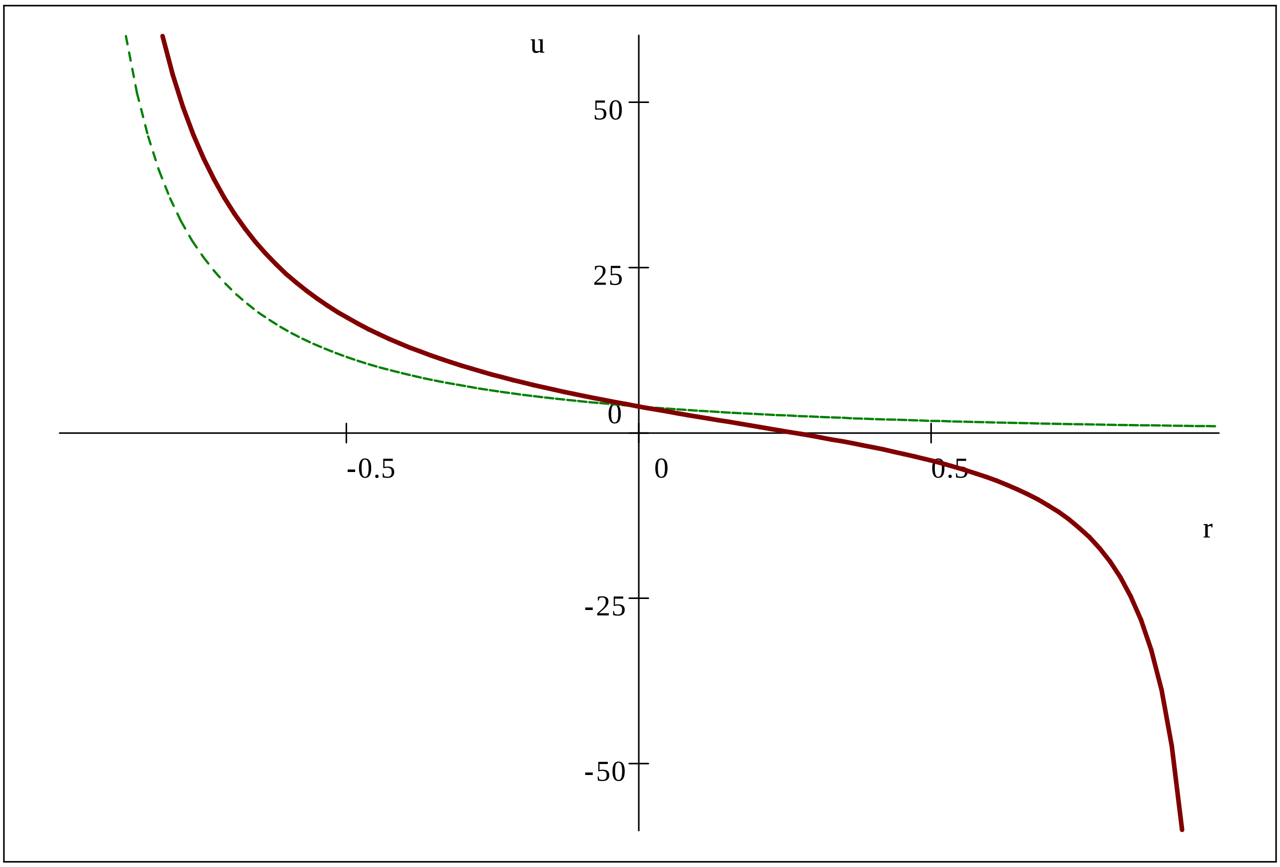
$$\begin{aligned}\frac{\partial \ln f_t}{\partial \gamma_{t|t-1}} &= u_{\gamma t} = \frac{(x_{1t}+x_{2t})^2}{4}\exp(-2\gamma_{t|t-1}) \\ &\quad -\frac{(x_{1t}-x_{2t})^2}{4}\exp(2\gamma_{t|t-1}) + \frac{\exp(2\gamma_{t|t-1})-1}{\exp(2\gamma_{t|t-1})+1},\end{aligned}$$

where the first two terms are mutually uncorrelated.

Written in terms of $\rho_{t|t-1}$,

$$u_{\gamma t} = \frac{(x_{1t}+x_{2t})^2}{4}\frac{1-\rho_{t|t-1}}{1+\rho_{t|t-1}} - \frac{(x_{1t}-x_{2t})^2}{4}\frac{1+\rho_{t|t-1}}{1-\rho_{t|t-1}} + \rho_{t|t-1}. \quad (7.17)$$

The score reduces to $x_{1t}x_{2t}$ when $\rho_{t|t-1} = 0$, but when $\rho_{t|t-1}$ is close to one, the weight given to $(x_{1t}+x_{2t})^2$ is small and the second term dominates unless x_{1t} and x_{2t} are very close; in this case x_{1t} and x_{2t} provide evidence of strong correlation and there is little reason to reduce $\rho_{t|t-1}$. As $\rho_{t|t-1}$ becomes smaller, the first term in (7.17) becomes relatively more important, leading to an increase in correlation when x_{1t} and x_{2t} are close. The dashed line in Figure ? shows what happens when $x_{1t} = x_{2t}$. The solid line tells a different story: here, x_{1t} and x_{2t} are not very close and so when $\rho_{t|t-1}$ is near one, it is reduced in the next time period, the reduction being bigger the closer it is to one.



Plot of standardized score, u, against correlation, r, for $x_1 = x_2$ (dash) and $x_1 = 4, x_2 = 1$.

Dynamic copulas: estimating changing association

The conditional score for the Clayton copula is

$$\frac{\partial \ln f(y_{1t}, y_{2t}, \theta_{t|t-1})}{\partial \theta_{t|t-1}} = -\ln(\tau_{1t}\tau_{2t}) + (1+\theta_{t|t-1})^{-1} + \theta^{-2}\ln(\tau_{1t}^{-\theta_{t|t-1}} + \tau_{2t}^{-\theta_{t|}}$$

$$+ \left(\frac{1+2\theta_{t|t-1}}{\theta_{t|t-1}}\right)\frac{(\tau_{1t}^{-\theta_{t|t-1}}\ln\tau_{1t} + \tau_{2t}^{-\theta_{t|t-1}}\ln\tau_{2t})}{\tau_{1t}^{-\theta_{t|t-1}} + \tau_{2t}^{-\theta_{t|t-1}} - 1},$$

where $\tau_{it} = F(y_{it})$, $i = 1, 2$. The response to a pair of observations is not as readily interpretable as it is for the bivariate normal distribution. However, the basic point to note is that the first term involves the product $\tau_{1t}\tau_{2t}$, and so is a little like the product $x_{1t}x_{2t}$. In the Gaussian model the score modifies the impact of $x_{1t}x_{2t}$ by taking account of how the product was formed and the current parameter value. The same is true here.

Figure shows the response of the score when τ_2 varies, but τ_1 is fixed. Two points are worth noting.
1) As expected, the response is asymmetric in the sense that the behaviour when τ_1 fixed at 0.9 is not a mirror image of the behaviour for τ_1 fixed at 0.1.
2) When $\tau_1 = 0.1$, the score is only positive for vaues of τ_2 close to 0.1, the effect being more pronounced when $\theta = 5$, as opposed to $\theta = 1$. This behaviour is entirely consistent with the conditional density shown earlier : if τ_2 is not close to 0.1, it suggests that $\theta_{t|t-1}$ is too big and the role of the negative score in the dynamic equation is to make $\theta_{t+1|t}$ smaller.

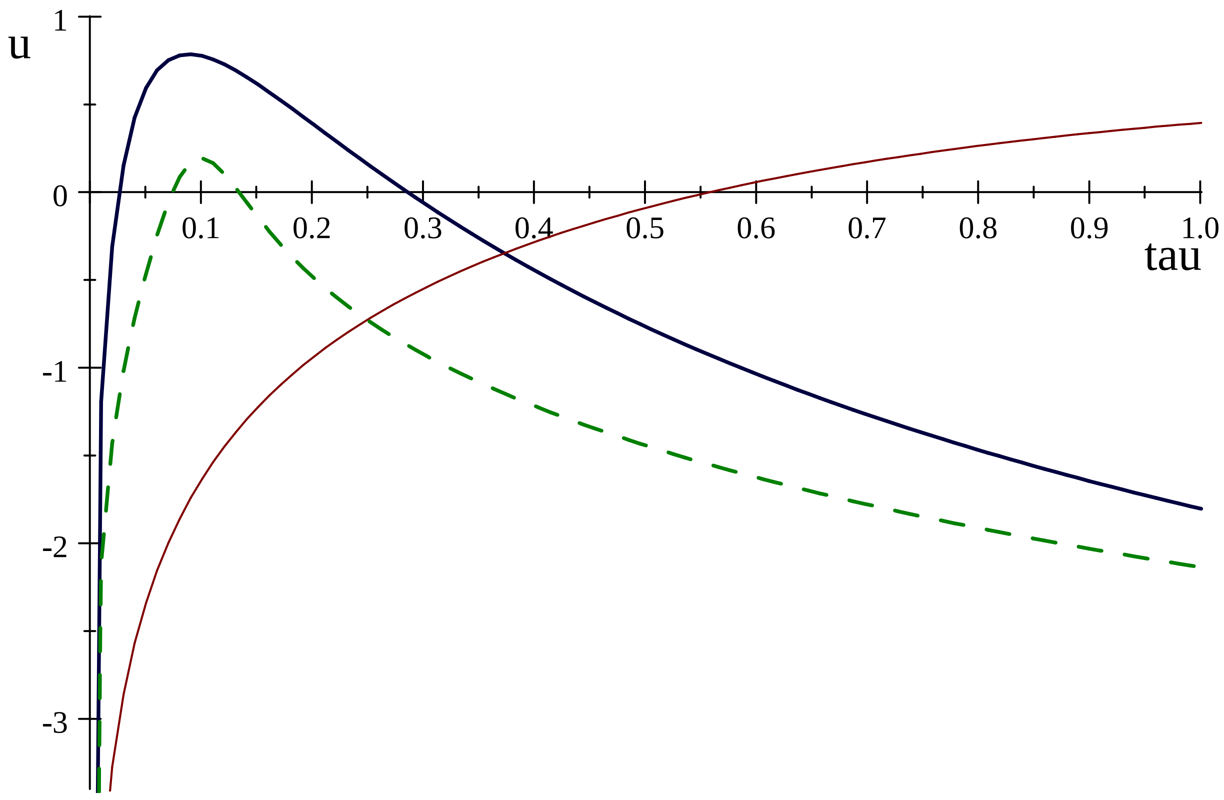


Figure: Response of u for fixed τ_1

Nonparametric and kernels

Using a sample of T observations drawn from a distribution $F(y)$ with a corresponding probability density function $f(y)$, a kernel estimator of $f(y)$ at point y is given by

$$\overline{f}_T(y) = \frac{1}{Th}\sum_{i=1}^{T} K\left(\frac{y - y_i}{h}\right), \tag{9}$$

where $K(.)$ is the kernel and h is the bandwidth. The kernel, $K(\cdot)$, is symmetric about the origin and everywhere non-negative. It integrates to one when divided by h.
The Epanechnikov kernel is

$$K(z) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{z^2}{25}), & |z| < \sqrt{5} \\ 0, & |z| \geq \sqrt{5} \end{cases}. \tag{10}$$

Nonparametric and kernels- filtering

A weighting scheme may be introduced into the kernel estimator so as to make predictions of the density at time $t+1$, based on information at time t. Thus

$$f_{t+1|t}(y) = \frac{1}{h}\sum_{i=1}^{t} K\left(\frac{y - y_i}{h}\right) w_{t,i}, \quad t = 1, ..., T, \tag{11}$$

while, for the distribution function,

$$F_{t+1|t}(y) = \sum_{i=1}^{t} H\left(\frac{y - y_i}{h}\right) w_{t,i}. \tag{12}$$

The weights, $w_{t,i}$, $i = 1, ..., t$, $t = 1, ..., T$, may change over time, although in the steady-state, $w_{t,i} = w_{t-i}$. For an EWMA scheme, the weights sum to unity. The expressions for smoothing are similar except that the summations run from $t = 1$ to T.

Nonparametric and kernels- filtering

Filters may be constructed for the predictive density for a given value of y. The first-order filter is

$$f_{t+1|t}(y) = \delta_y + \beta f_{t|t-1}(y) + \alpha \frac{1}{h} K\left(\frac{y - y_t}{h}\right), \quad t = 1, .., T,$$

where $\alpha \geq 0$, $\beta \geq 0$ and $\delta_y > 0$. Alternatively,

$$f_{t+1|t}(y) = \delta_y + \phi f_{t|t-1}(y) + \kappa u_t, \quad t = 1, .., T, \qquad (13)$$

where $\phi = \alpha + \beta$, $\kappa = \alpha$ and the innovation for the density is

$$u_t(y) = \frac{1}{h} K\left(\frac{y - y_t}{h}\right) - f_{t|t-1}(y), \qquad (14)$$

which is similar in form to the conditional score in a DCS model.

Nonparametric and kernels- filtering

The filter updates the estimate of the PDF at y. The maximum impact is when there is a direct hit, that is the observation coincides with y. In this case $h^{-1}K(0) - f_{t|t-1}(y)$ must be positive. On the other hand an observation far from y will have little or no effect and so $u_t(y)$ is close to $-f_{t|t-1}(y)$. Figure 7 shows the impact of an Epanechnikov kernel (10) with $f_{t|t-1}(y)$ (arbitrarily) set to 0.1.

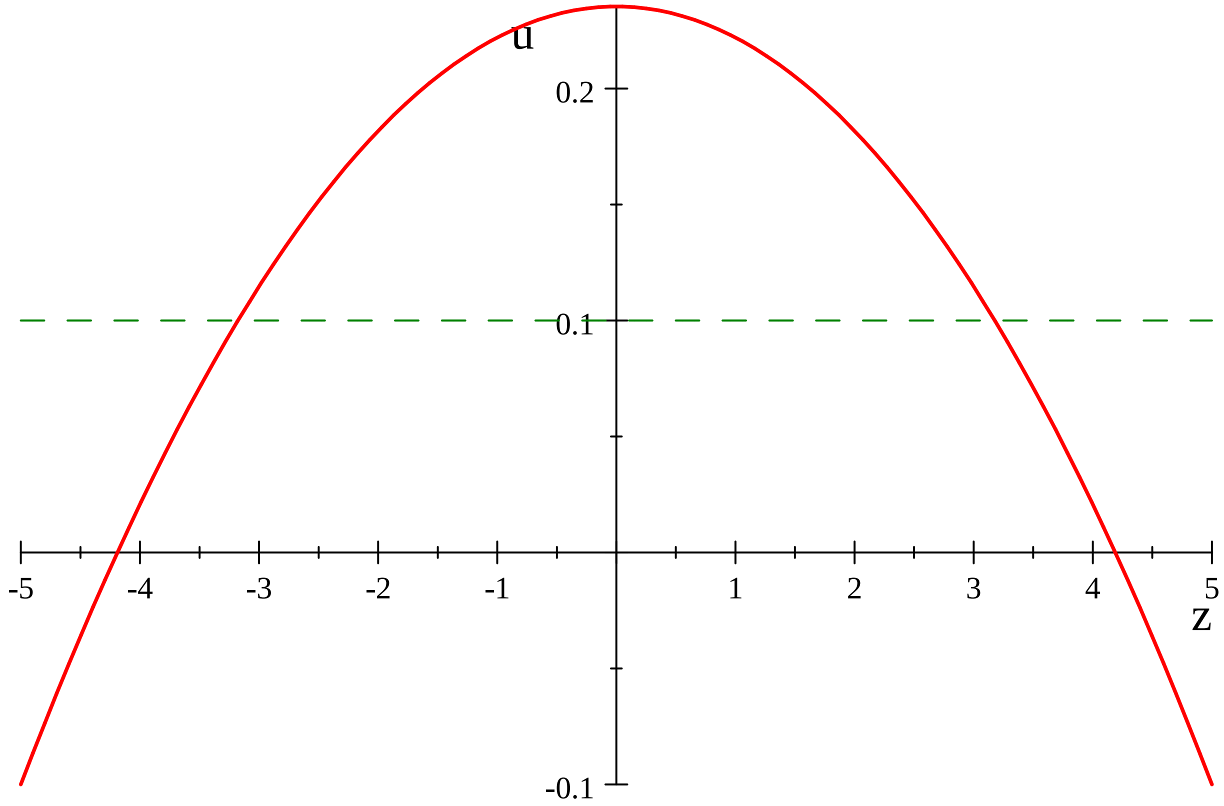


Figure: Impact of Epanechnikov kernel when $f_{t|t-1}(y) = 0.1$

Nonparametric and kernels- estimation

The recursive nature of the filter leads naturally, but perhaps surprisingly, to a maximum likelihood procedure for estimating unknown parameters, as contained in a vector denoted $\boldsymbol{\psi}$. These parameters include the bandwidth, h, as well as any parameters governing the dynamics, such as κ and ϕ. The log-likelihood function is

$$\begin{aligned} \ln L(\boldsymbol{\psi}) &= \sum_{t=m}^{T-1} \ln f_{t+1|t}(y_{t+1}) \\ &= \sum_{t=m}^{T-1} \ln \left[\frac{1}{h} \sum_{i=1}^{t} K\left(\frac{y_{t+1} - y_i}{h} \right) w_{t,i} \right], \end{aligned} \tag{15}$$

where, for a nonstationary filter m is some preset number of observations used to initialize

Direct estimation of individual quantiles

The $\tau - th$ quantile for a set of T observations, $\widetilde{\xi}(\tau)$, can be obtained as the solution to the minimization of

$$\rho_\tau(\xi) = \sum_{t=1}^{T} \rho_\tau(y_t - \xi) \tag{16}$$

with respect to $\xi = \xi(\tau)$, where $\rho_\tau(.)$ is the *check function* for quantiles, that is

$$\rho_\tau(y_t - \xi) = (\tau - I(y_t - \xi < 0))\,(y_t - \xi)\,, \tag{17}$$

where $I(.)$ is one when $y_t < 0$ and zero otherwise.

Direct estimation of individual quantiles

Differentiating ρ where it is continuous gives the *quantile indicator function:*

$$IQ(y_t - \xi_t(\tau)) = \begin{cases} \tau - 1, & \text{if} \quad y_t < \xi_t(\tau) \\ \tau, & \text{if} \quad y_t > \xi_t(\tau) \end{cases}, \qquad t = 1, ..., T,$$

$IQ(0)$ is not determined, but can be set to zero.
Estimating generating equation because

$$\sum_{t=1}^{T} IQ(y_t - \xi_t(\tau)) = 0$$

Direct estimation of individual quantiles

The smoothing approach to estimating time-varying quantiles results in their having the following important property when the dynamic equation is a stochastic level or trend.

The number of observations which are less than the corresponding quantile, that is the number of occasions on which $y_t < \tilde{\xi}_{t|T}$ for $t = 1, ..., T$, *is no more than* $[T\tau]$, *while the number greater is no more than* $[T(1-\tau)]$.

When the trend reverts to a constant, the usual defining feature of quantiles is satisfied.

Direct estimation of individual quantiles - and expectiles*

Expectiles, denoted $\mu(\omega), 0 < \omega < 1$, are similar to quantiles but they are determined by tail expectations rather than tail probabilities. For a given value of ω, the sample expectile, $\widetilde{\mu}(\omega)$, is obtained by minimizing the asymmetric least squares function,

$$S_\omega(\mu) = \sum \rho_\omega(y_t - \mu) = \sum |\omega - I(y_t - \mu < 0)| (y_t - \mu)^2,$$

with respect to μ. See De Rossi and Harvey (2009)

Direct estimation of individual quantiles

The smoothed estimate of a quantile at the end of the sample is the filtered estimate. For the EWMA scheme derived from the local level model, the filtered estimator must satisfy

$$\xi_{t+1|t} = \kappa \sum_{j=0}^{\infty} (1-\kappa)^j [\xi_{t-j|t} + IQ(y_{t-j} - \xi_{t-j|t})];$$

Thus $\xi_{t+1|t}$ is an EWMA of the pseudo-observations, $\xi_{t-j|t} + IQ(y_{t-j} - \xi_{t-j|t})$. As new observations become available, the smoothed estimates need to be revised.
However, filtered estimates could be used instead, so

$$\xi_{t+1|t}(\tau) = \xi_{t|t-1}(\tau) + \kappa u_t(\tau),$$

where $u_t(\tau) = IQ(y_t - \xi_{t|t-1}(\tau))$ is an indicator which plays a similar role to that of the conditional score in a DCS model.
Asymptotics ?

Direct estimation of individual quantiles

The filter belongs to the class of CAViaR models proposed by Engle and Manganelli (2004) in the context of tracking value at risk. In CAViaR, the first-order conditional quantile is a GARCH-type filter of the form

$$\widehat{\xi}_{t+1|t}(\tau) = \delta + \beta \widehat{\xi}_{t|t-1}(\tau) + \alpha q(y_t),$$

where $q(y_t)$ is a function of y_t. Suggested specifications include an adaptive model, which in a limiting case has the same form as the filter for $\xi_{t+1|t}(\tau)$ with $\alpha = \kappa$ and, in the general first-order model, $\beta = \phi - \kappa$.
Other CAViaR specifications, which are based on actual values, rather than indicators, may suffer from a lack of robustness to additive outliers. See Figure 1 in Engle and Manganelli (2004, p. 373). The evidence on predictive performance in Kuester et al (2006, p. 80-1) indicates a preference for the adaptive specification.

Direct estimation of individual quantiles

The advantage of fitting individual quantiles is that different parameters may be estimated for different quantiles. The disadvantage is that the quantiles may cross; see Gourieroux and Jasiek (2008).

If the parameters across quantiles have to be the same to prevent them crossing, the ability to have different models for different quantiles loses much of its appeal.

Direct estimation of individual quantiles

The conditional mode signal extraction argument used to derive the quantiles may be adapted to the dynamic kernel by defining

$$\rho(y_t|f_t(y)) = -\frac{1}{2}\left[\frac{1}{h}K\left(\frac{y_t - y}{h}\right) - f_t(y)\right]^2$$

for all admissible values of y. For a given value of y, differentiating this new criterion function with respect to $f_t(y)$ for all t, setting the derivatives to zero and solving, gives the smoothed estimate, $f_{t|T}(y)$. The residuals are

$$u_t(f_{t|T}(y)) = \frac{1}{h}K\left(\frac{y_t - y}{h}\right) - f_{t|T}(y), \qquad t = 1, ..., T.$$

For all y, $\sum_{t=1}^{T} u_t(f_{t|T}(y)) = 0$. The variable driving the filter, that is $u_t(y)$, is of the same form as $u_t(f_{t|T}(y))$.