Dynamic Conditional Score (DCS) Models and Realized Variance

Andrew Harvey (ach34@cam.ac.uk)

Faculty of Economics, University of Cambridge

May 2018

Introduction to dynamic conditional score (DCS) models

1. A unified and comprehensive theory for a class of nonlinear time series models in which the dynamics of a changing parameter, such as location or scale, is driven by the score of the conditional distribution.
2. Dynamics are driven by the *score* of the conditional distribution.
3. For EGARCH, analytic expressions may be derived for (unconditional) moments, autocorrelations and moments of multi-step forecasts. An asymptotic distributional theory for ML estimators can be obtained, sometimes with analytic expressions for the asymptotic covariance matrix.
4. Similar results for location/scale models based on a GB2 distribution.
5. Extensions to multivariate time series. Correlation or association may change over time. Time-varying copulas.

Introduction to dynamic conditional score (DCS) models

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

***
GAS package R by David Ardia, Kris Boudt, and Leopoldo Catania.
Computer code: R package GAS.
The 'development' version is available from GitHub at
Development code: Development R package GAS
and will be updated more regularly then the one from CRAN.
Vignette: "Generalized Autoregressive Score Models in R: The GAS Package".

Introduction to dynamic conditional score (DCS) models

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.

Unobserved component models

A simple Gaussian signal plus noise model is

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim NID\left(0, \sigma_\varepsilon^2\right), \quad t = 1, ..., T$$

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

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

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

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

Unobserved component models

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:

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t \\ \mu_{t+1|t} &= \phi\mu_{t|t-1} + k_t v_t \end{aligned}$$

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 $\phi - \kappa = \theta$.

Outliers

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 ξ_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.

Unobserved component models for non-Gaussian noise

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 (2012).

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.

Observation driven model based on the score

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*. 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 innovations form becomes

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \qquad\qquad t = 1, ..., T \\ \mu_{t+1|t} &= \phi\mu_{t|t-1} + \kappa u_t \end{aligned}$$

where κ is an unknown parameter.

Dynamic location model

$$\begin{aligned} y_t &= \omega + \mu_{t|t-1} + \varphi\varepsilon_t, \\ \mu_{t+1|t} &= \phi\mu_{t|t-1} + \kappa u_t, \end{aligned}$$

where ε_t is serially independent, standard t-variate and

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

where $v_t = y_t - \omega - \mu_{t|t-1}$ is the prediction error and φ is the 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*.

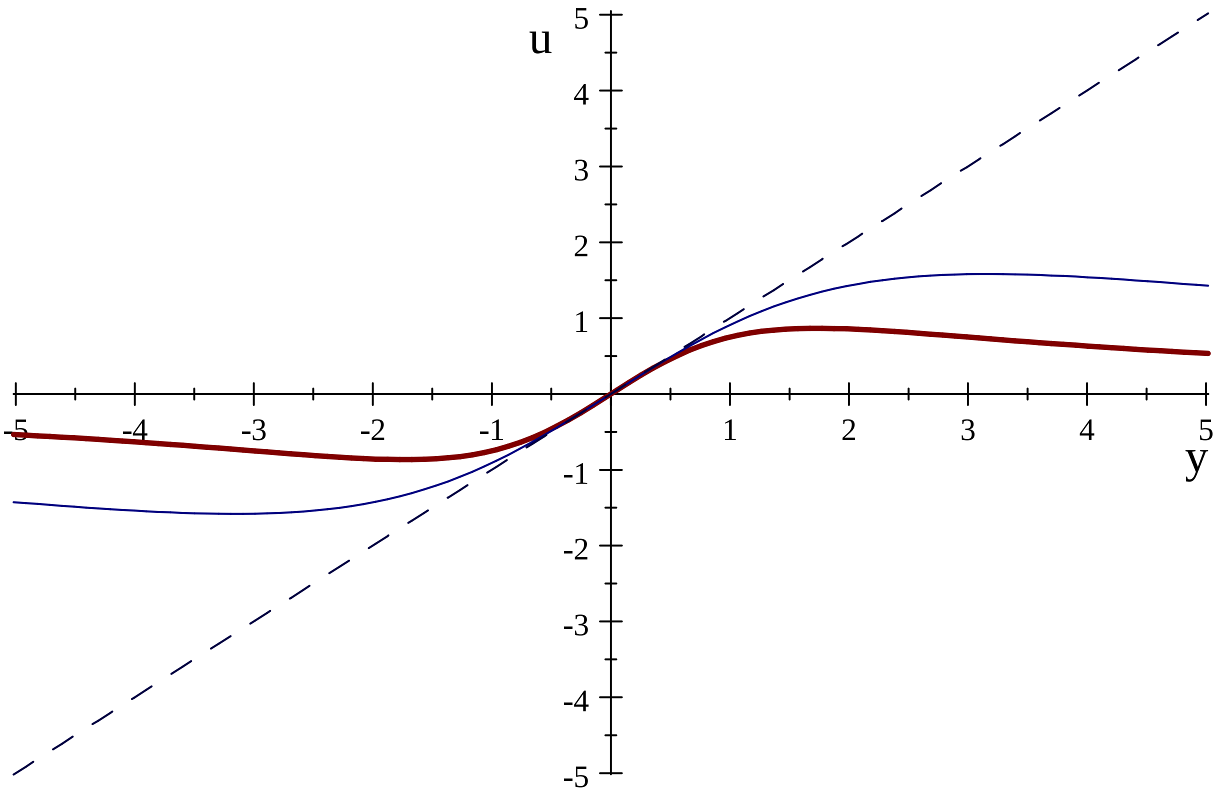


Figure: Impact of u_t for t_ν (with a scale of one) for $\nu = 3$ (thick), $\nu = 10$ (thin) and $\nu = \infty$ (dashed).

Dynamic location model with trend and seasonals

$$y_t = \mu_{t|t-1} + \gamma_{t|t-1} + \varphi\varepsilon_t, \quad t = 1, ..., T,$$

The filter for the trend is

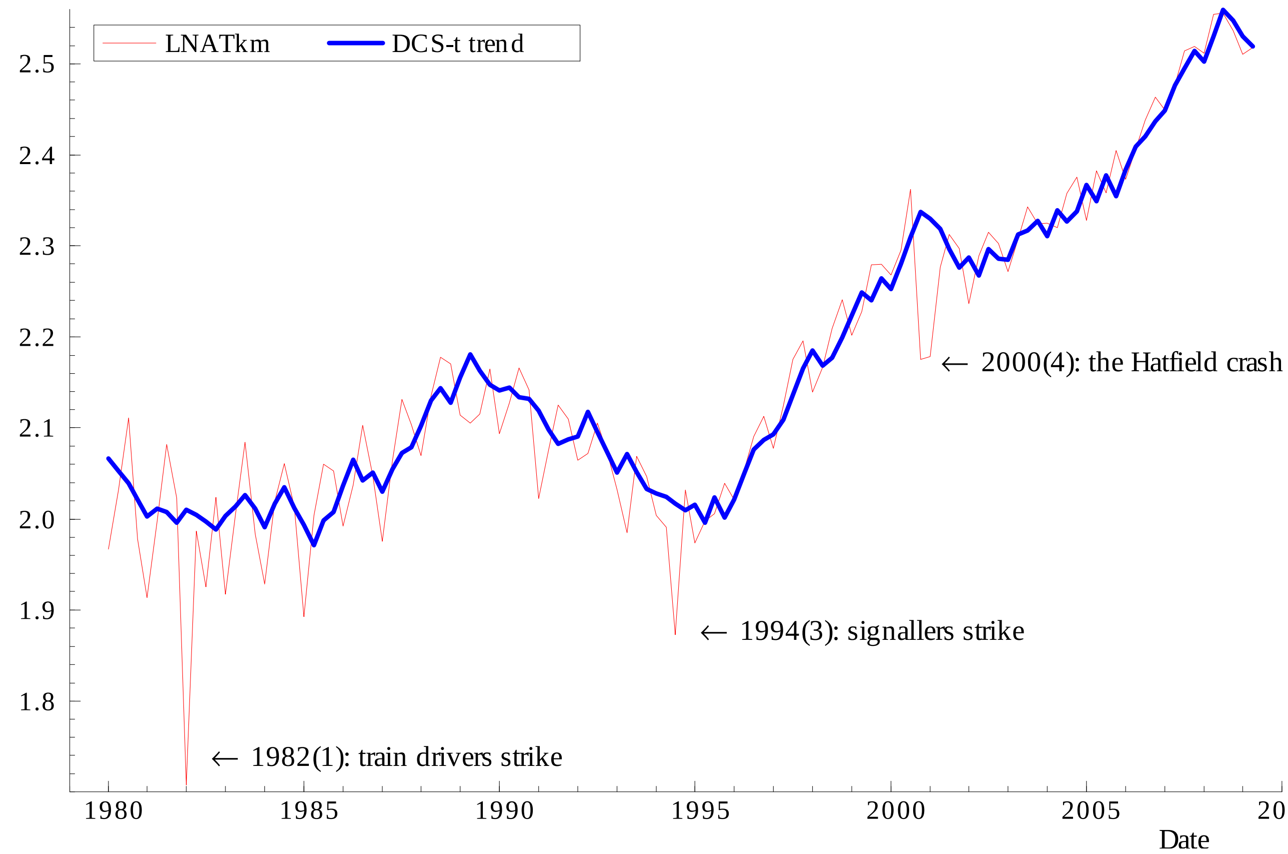
$$\begin{aligned} \mu_{t+1|t} &= \mu_{t|t-1} + \beta_{t|t-1} + \kappa_1 u_t \\ \beta_{t+1|t} &= \beta_{t|t-1} + \kappa_2 u_t. \end{aligned}$$

The filter for the seasonal is

$$\gamma_{t|t-1} = \mathbf{z}_t' \gamma_{t|t-1}, \qquad \gamma_{t+1|t} = \gamma_{t|t-1} + \boldsymbol{\kappa}_t u_t,$$

where the $s \times 1$ vector $\mathbf{z}_t$ picks out the current season from the vector $\gamma_{t|t-1}$. If κ_{jt}, $j = 1, .., s$, denotes the $j - th$ element of $\boldsymbol{\kappa}_t$, then in season j we set $\kappa_{jt} = \kappa_s$, where κ_s is a non-negative unknown parameter, whereas $\kappa_{it} = -\kappa_s/(s-1)$, $i \neq j$, $\quad i = 1, .., s$. The amounts by which the seasonal effects change therefore sum to zero.

The initial conditions at time $t = 0$ are estimated by treating them as parameters.



EGARCH: El Classico

In the classic EGARCH model of Nelson (Econometrica, 1991)

$$y_t = \sigma_{t|t-1}\varepsilon_t, \quad \varepsilon_t \backsim IID\,(0,1)\,.$$

$$\ln \sigma^2_{t+1|t} = \gamma + \phi \ln \sigma^2_{t|t-1} + \alpha\varepsilon_t + \beta\left[|\varepsilon_t| - E\,|\varepsilon_t|\right]$$

The term $\alpha\varepsilon_t + \beta\left[|\varepsilon_t| - E\,|\varepsilon_t|\right]$ is a zero mean, *IID* sequence, which is able to respond asymmetrically (when $\alpha \neq 0$) to rises and falls in stock price (leverage).
Stationary if $|\phi| < 1$.
*****
Nelson notes that if ε_t is t_ν with ν, moments of $\sigma^2_{t|t-1}$ and y_t rarely exist even though they are strictly stationary.
Also - is it invertible? Straumann and Mikosch (2006) show there exist invertible models when $\phi = 0$.

Exponential DCS Volatility Models: Beta-t-EGARCH

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

where the serially independent, zero mean variable ε_t has a t_ν−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

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

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

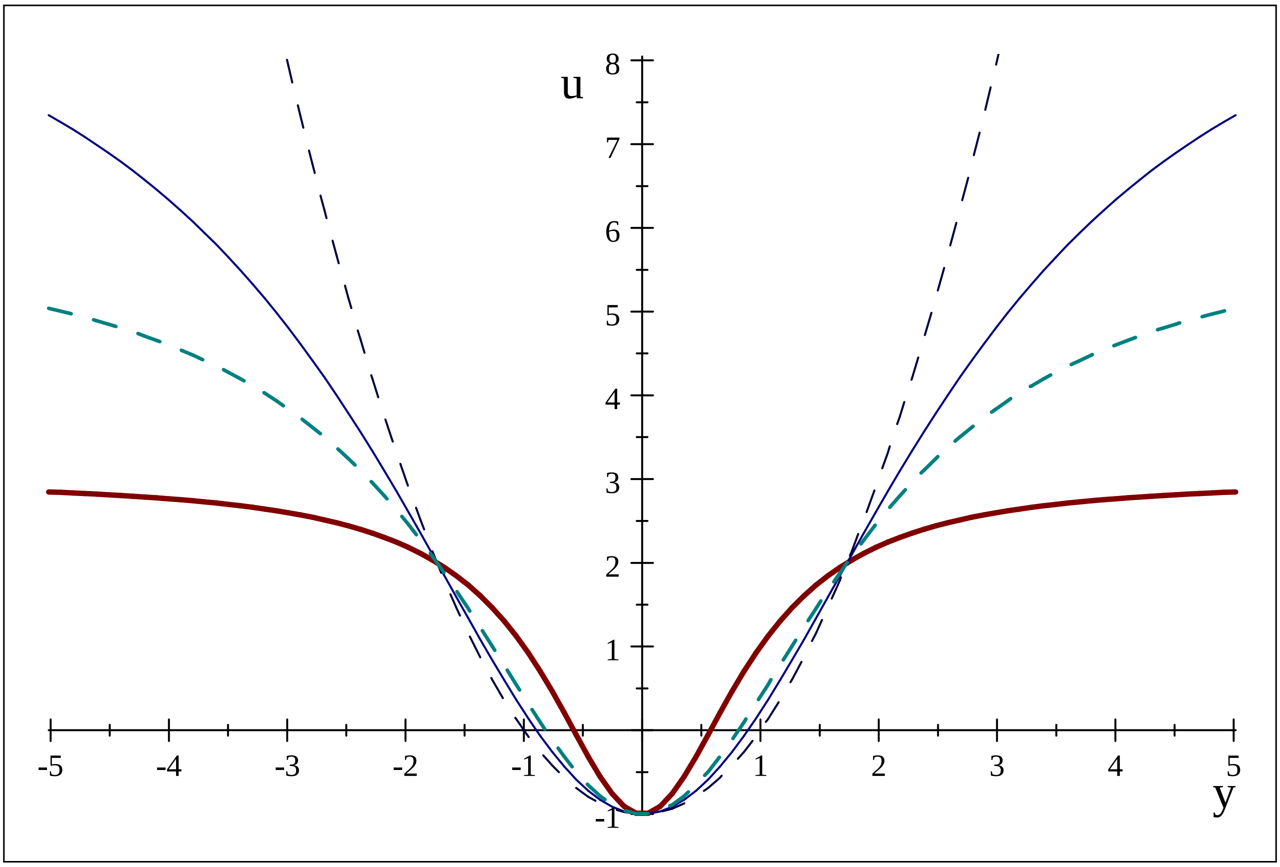
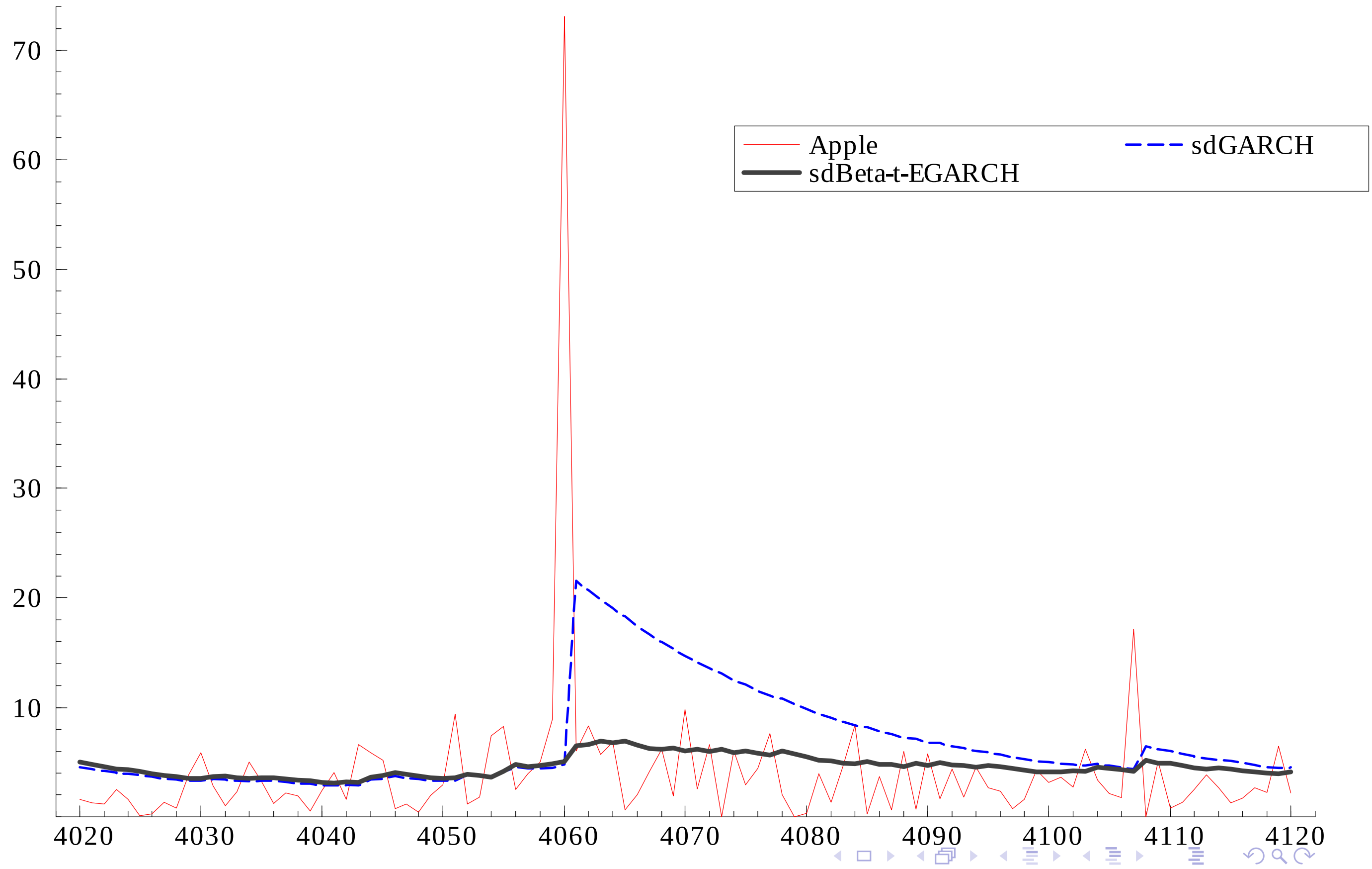


Figure: Impact of u_t for t_ν with $\nu = 3$ (thick), $\nu = 6$ (medium dashed) $\nu = 10$ (thin) and $\nu = \infty$ (dashed).



Beta-t-EGARCH

The variable u_t may be expressed as

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

where

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

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

Beta-t-EGARCH: moments

The score is bounded.
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 $|y_t|^c$.

Generalized t-distribution

General Error distribution (GED) leads to Gamma-GED-EGARCH model.
The score is **gamma** distributed.
Student-t and GED are special cases of generalized-t . (NB. Absolute value of gen-t is GB2)
The flexibility of Gen-t goes a long way towards meeting the objection that parametric models are too restrictive and hence vulnerable to misspecification -see McDonald and Newey (Econometrica, 1987).
Harvey and Lange (2016, JTSA). Volatility Modeling with a Generalized t-distribution.

Invertibility

$$\Lambda_t(\boldsymbol{\psi}) := \sup |x_t|$$

where $x_t = d\lambda_{t+1|t}/d\lambda_{t|t-1} = \phi + \kappa(\partial u_t/\partial\lambda_{t|t-1})$. Blasques et al (2018) show that a sufficient condition for invertibility is $E \ln \Lambda_0(\boldsymbol{\psi}) < 0$ over all admissible $\boldsymbol{\psi}$.
A sufficient condition for invertibility of Beta-t-EGARCH is

$$|\phi - \kappa(\nu+1)/2| < 1.$$

Obtained as $\partial u_t/\partial\lambda_{t|t-1} = -(\nu+1)b_t(1-b_t) < 0$ and the maximum value of $b_t(1-b_t)$ is $1/4$. Then $|x_t| < 1$ for all t and so $E \ln \Lambda_0(\boldsymbol{\psi}) < 0$.
When $\kappa > 0$

$$\kappa < \frac{2(1+\phi)}{\nu+1}$$

Invertibility

The condition is probably much stronger than necessary. However, so long as tail index is not too large, it seems to include parameter values that arise in practice.

Example

Beta-t-EGARCH with $\nu = 9$ and $\phi = 0.999$, requires $\kappa < 0.40$. Even with $\nu = 39$, $\kappa < 0.10$. These bounds are halved for $\phi = 0.5$, but are still comfortably high and ϕ is normally above 0.9.

***

Empirical

$$\frac{1}{T}\sum \ln |x_t| < 0$$

Asymmetric impact curve (leverage)

Returns may have a different effect on volatility depending on whether they are positive or negative:

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

where $u_t^* = sgn\,(\mu - y_t)(u_t + 1)$ and κ^* is a parameter.
The effect of the extra term is to add or subtract, depending on $sgn(y_t - \mu)$, a fraction of the impact curve plus one.

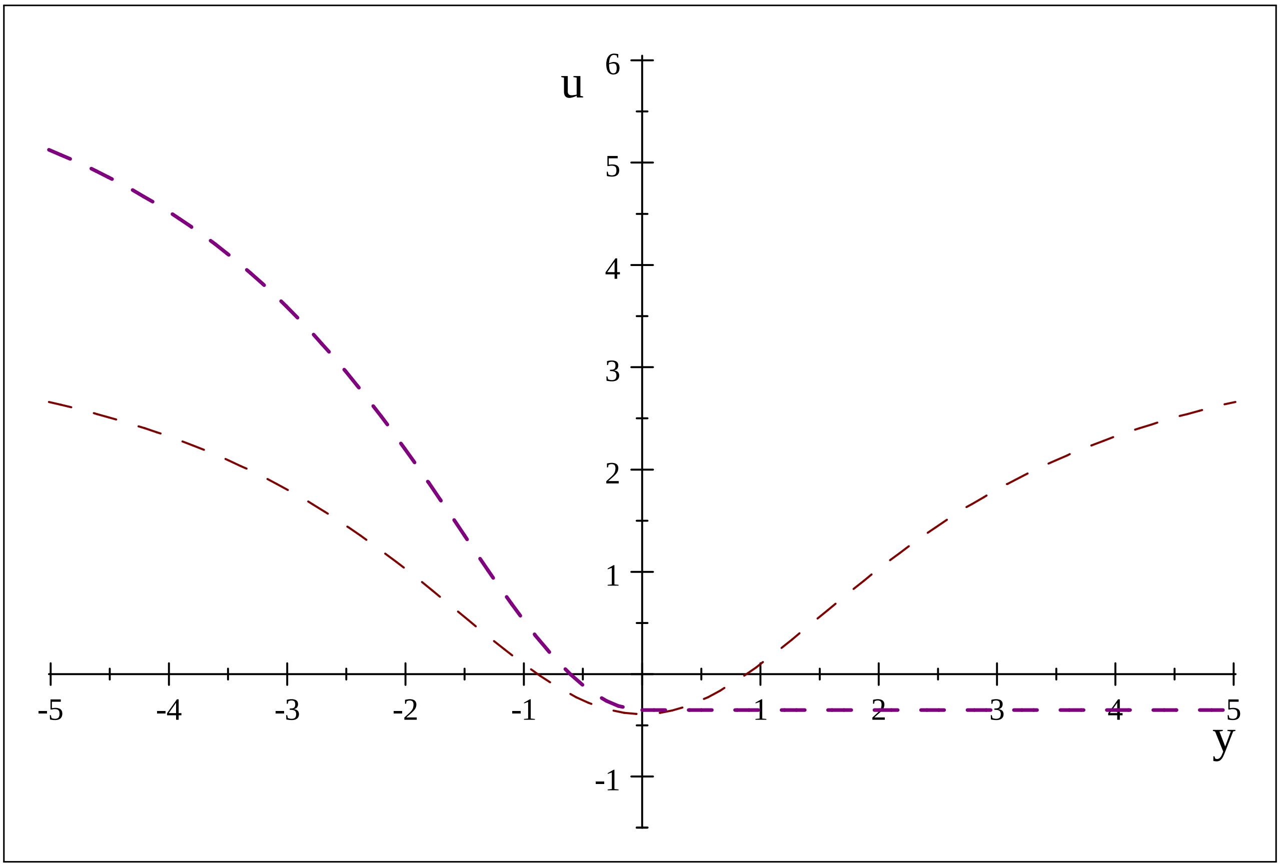


Figure: Impact of u for t_9. Purple is $\kappa = \kappa^*$

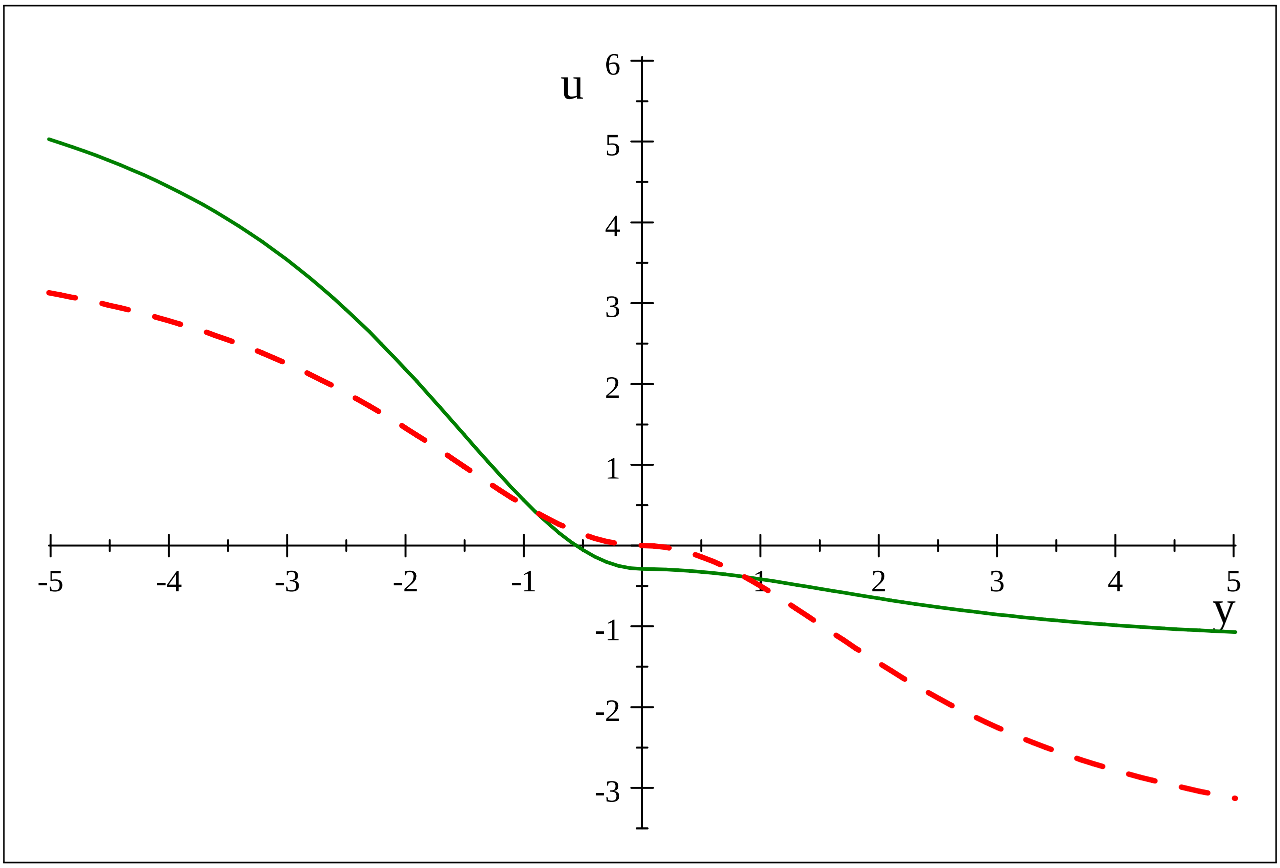


Figure: Red has $\kappa = 0$.

Two components

Instead of capturing long memory by a fractionally integrated process, two components may be used.

$$\lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1},$$
$$\lambda_{i,t+1|t} = \phi_i \lambda_{i,t|t-1} + \kappa_i u_t + \kappa_i^* sgn(-\varepsilon_t)(u_t + 1), \qquad i = 1, 2,$$

It is often found that the leverage effect is confined to the short-term component. In this case, the evolution of the long-run component will be less susceptible to the influence of strongly negative returns and so may be more suitable for capturing the ARCH-M effect.

Location/scale models for positive 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,

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

Positive 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

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

Generalized gamma and beta distributions for positive variables with changing location/scale

The statistical theory of DCS models for positive 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** (GG) distribution.
Burr and log-logistic distributions are special cases of the **generalized beta of the second kind** (GB2) distribution.
GB2 has fat tails except in a limiting case when it goes to GG.

Generalized gamma and beta distributions for positive variables with changing location/scale

The PDF of a GB2 is

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

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 ($\xi = 1, \varsigma = 1$).
Furthermore Generalized Gamma (includes Weibull as well as gamma) is a special limiting case.
The F-distribution is related to GB2 in that for an F-distribution with (ν_1, ν_2) degrees of freedom, $(\nu_1, \nu_2)F$ is $GB2(\nu_1/2, \nu_2/2)$.
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.

Generalized gamma and beta distributions for positive variables with changing location/scale

Score is *bounded*.
Invertibility

$$\kappa < \frac{4(1+\phi)}{\upsilon^2(\xi+\varsigma)}$$

The condition appears to place a constraint on the magnitude of ξ and ς. However, this difficulty can be avoided by analysing what happens when we work with the logarithm of y_t.

Realized volatility

If p_t as the log closing price of the index at time t, $\log P_t$, the return r_t is defined as $r_t = p_t - p_{t-1}$. If each daily interval is divided into m subintervals, the Realised Variance (RV) daily estimator can be constructed as

$$RV_t = \sum_{i=1}^{m} r_{t,i}^2$$

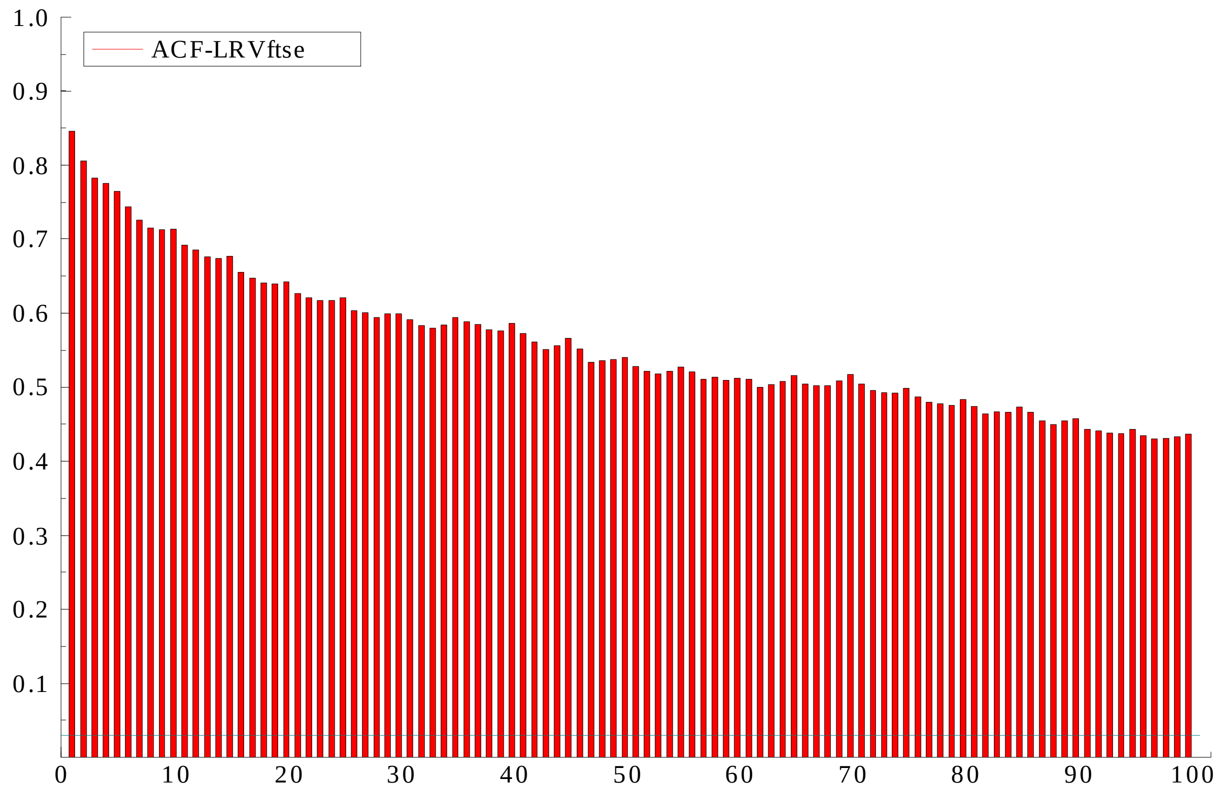
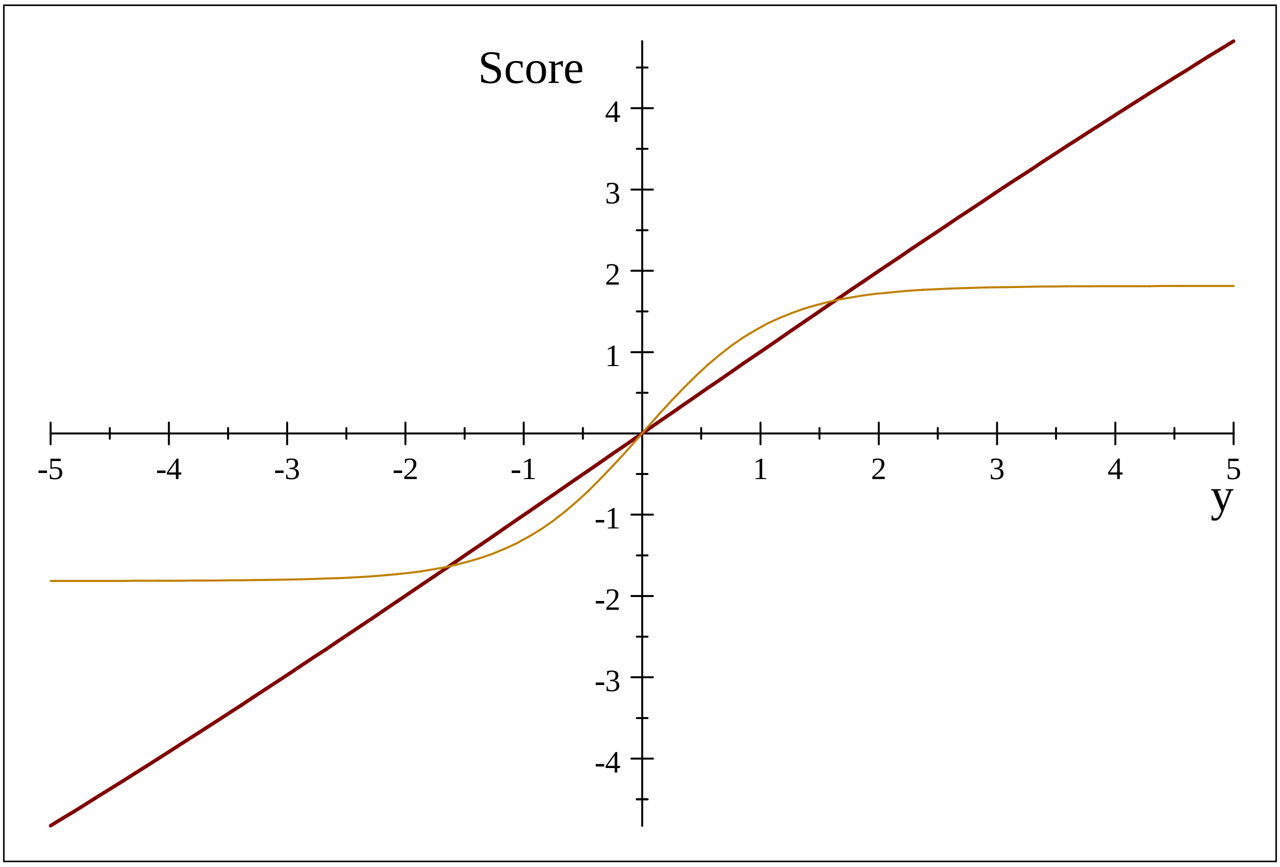


Figure: ACF of $\ln RV$

Realized volatility - EGB2

$$\ln y_t = x_t = \lambda_{t|t-1} + \ln \varepsilon_t$$

Location/scale, $\lambda_{t|t-1}$, now becomes location and the error is additive. Whereas y_t is intrinsically heteroscedastic, $\ln \varepsilon_t$ has constant (conditional) variance.
The distribution is exponential GB2 (EGB2). The EGB2 is symmetric when $\xi = \varsigma$. It includes both normal, when $\xi = \varsigma \to \infty$, and Laplace, when $\xi = \varsigma \to 0$, as special cases; see McDonald and Xu (1995) and Caivano and Harvey (2014). Hence the EGB2, like the GED, is a light tailed distributions covering the space between normal and Laplace. The logistic distribution sets $\xi = \varsigma = 1$.
The model can be estimated in levels or logs - same result.



Standardized score for logistic and normal

Realized volatility

Replace υ, which is now a scale parameter, by h/σ where σ is the sd of y_t and $h = \sqrt{\psi'(\xi) + \psi'(\varsigma)}$. The forcing variable in the dynamic equation is then

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

As $\varsigma = \xi \to \infty$, the distribution becomes normal and so for large ς and ξ, $u_t \simeq y_t - \mu_{t|t-1}$ and $\sigma_u^2 \to \sigma^2$.
The invertibility condition is

$$\left|\phi - \kappa h^2 \varsigma / 2\right| < 1$$

For a logistic distribution, when $\xi = \varsigma = 1$, the invertibility condition is $|\phi - 1.645\kappa| < 1$. But the coefficient of κ goes to one rapidly as ς increases. Since $\varsigma h^2 \to 2$ as $\varsigma \to \infty$, it follows that $-\sigma^2 u_t' \to 1$ and the standard invertibility condition for a Gaussian model is obtained, that is $|\phi - \kappa| < 1$.

Long memory, Fractional integration and HAR

Long memory in RV can be modeled by a FI model for $\lambda_{t|t-1}$, as in FIEGARCH; see Bollerslev and Mikkelsen (1996). Estimation is by ML.
The Heterogeneous Autoregression (HAR) model for RV is a parsimonious approximation to the high-order autoregressions implied by long memory; see Corsi (2009). The standard HAR regresses RV on the past 1-day, 5-day, and 22-day average realized variances. Thus

$$\overline{y}_{h,t+h} = \mu + \beta_d y_t + \beta_w \overline{y}_{w,t} + \beta_m \overline{y}_{m,t} + \zeta_{h,t+h},$$

where y_t is the RV series, $\overline{y}_{h,t} = \left(\sum_{i=1}^{h} y_{t-i}/h\right)$ for $h = w = 5$ and $h = m = 22$, and $\widehat{y}_{h,t+h}$ is the h-day cumulative average for $h = 1, 2, ..$
Usually $h = 1$. The disturbance term is $\zeta_{h,t+h}$.
Better to work in logs, particularly when $h = 1$. The fact that the log model is additive makes aggregation more appealing because the (conditional) variance is constant.

Realized volatility

A preliminary investigation of the properties of RV is best carried out by taking logarithms. Taking logarithms is particularly attractive when $\ln \varepsilon_t$ not too far from normality. Taking logarithms also facilitates a comparison with HAR.
If the disturbance is treated as Gaussian, linear unobserved components models may be fitted, using a package such as STAMP.
Fitting one AR(1) component gives a very high $Q(67)$ with a high $r(1) = 0.078$. We therefore fit two AR(1) components. One of these has a coefficient close to unity and so can be replaced by a RW. Including a second AR(1) reduces $r(1)$ to 0.033 but correlations at lags of multiples of 5 are apparent. The spectrum clearly shows peaks at 2/5 and its harmonic 4/5. The model is therefore augmented by including a weekly seasonal component.

Realized volatility

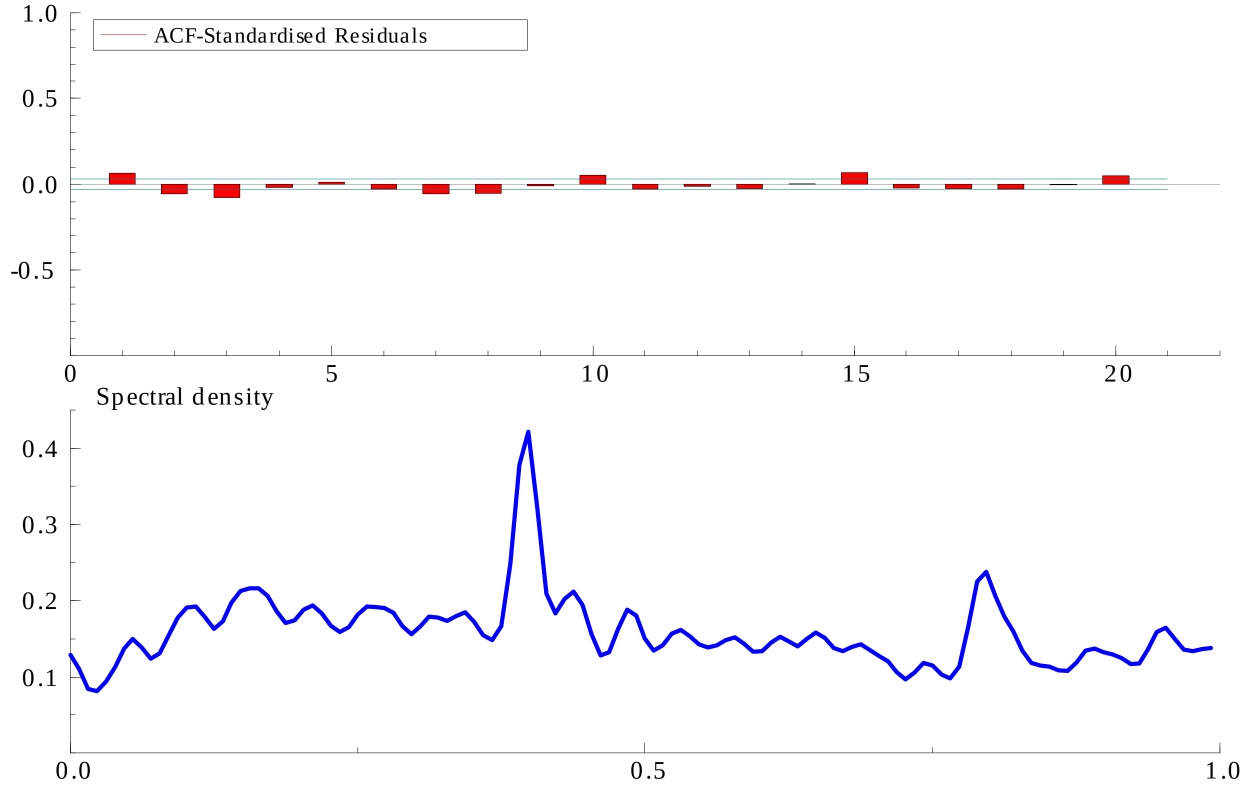


Figure: ACF and spectrum of residuals from fitting a RW and AR1 to ln RV

Realized volatility

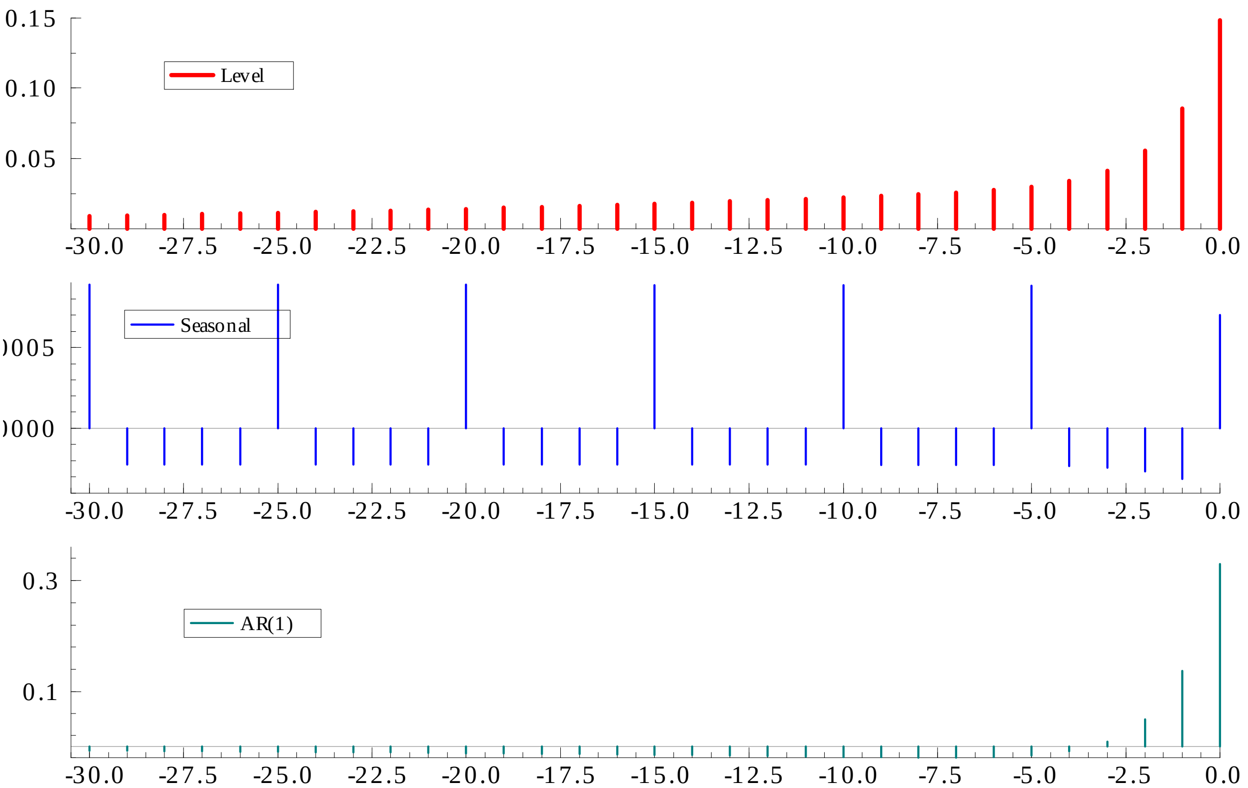


Figure: Filter weights from DCS model

Realized volatility

For range or RV, the leverage term is governed by $sgn(-r_t)$, where r_t denotes mean-adjusted returns. Thus

$$\lambda_{i,t+1|t} = \phi_i \lambda_{i,t|t-1} + \kappa_i u_{i,t} + \kappa_i^* sgn(-r_t)(u_t + 1), \quad i = 1, 2,$$

Other components may be added, eg

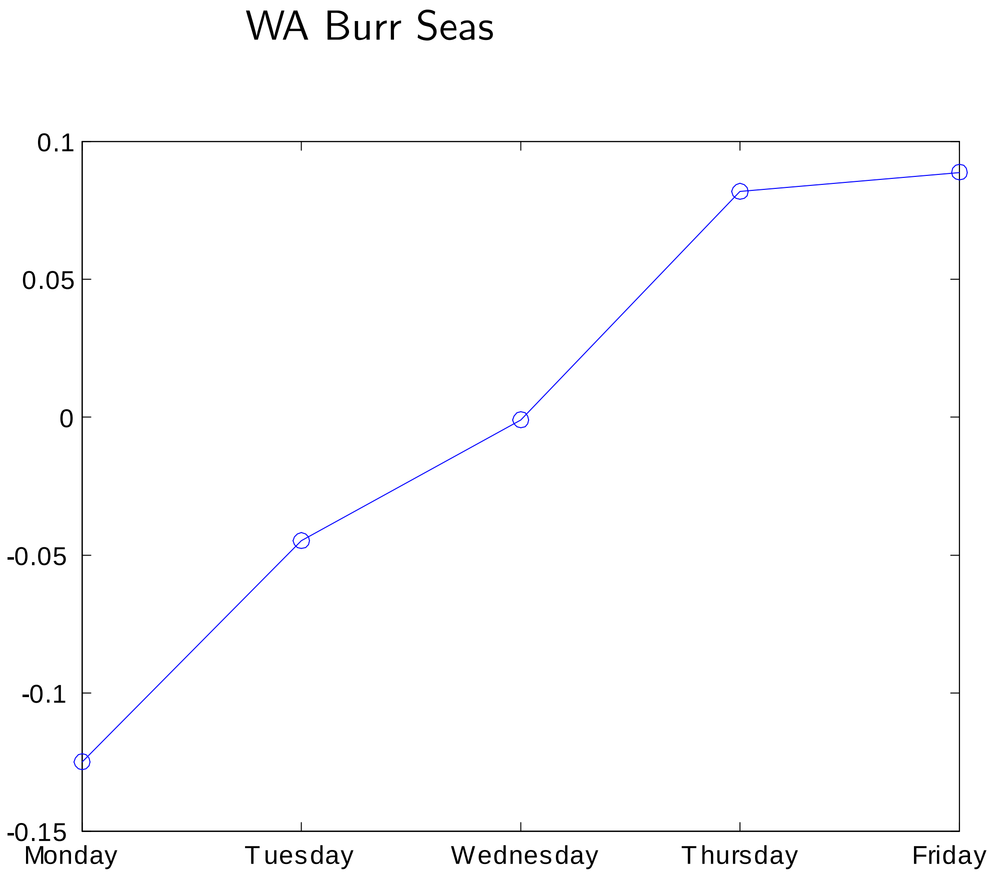
$$\lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1} + \gamma_{t|t-1},$$

where $\gamma_{t|t-1}$ is a seasonal component.

Realized volatility

A GB2 (or EGB2) can be fitted to RV with two components, leverage and a seasonal.
Fit with a Burr distribution or balanced GB2, that is $\xi = \varsigma$, is best.
The HAR model does rather well given its simplicity but Q is still high and ln L is much smaller. The sum of the coefficients is 0.961.
The HAR model eliminates serial correlation at lag 5 - but not at multiples of 5 - *so it is not dealing with the weekly effect*. There is no detectable monthly effect - but there wasn't before the model was fitted. (Of course if the dependent variable is the average over the week, the weekly effect will be eliminated.)

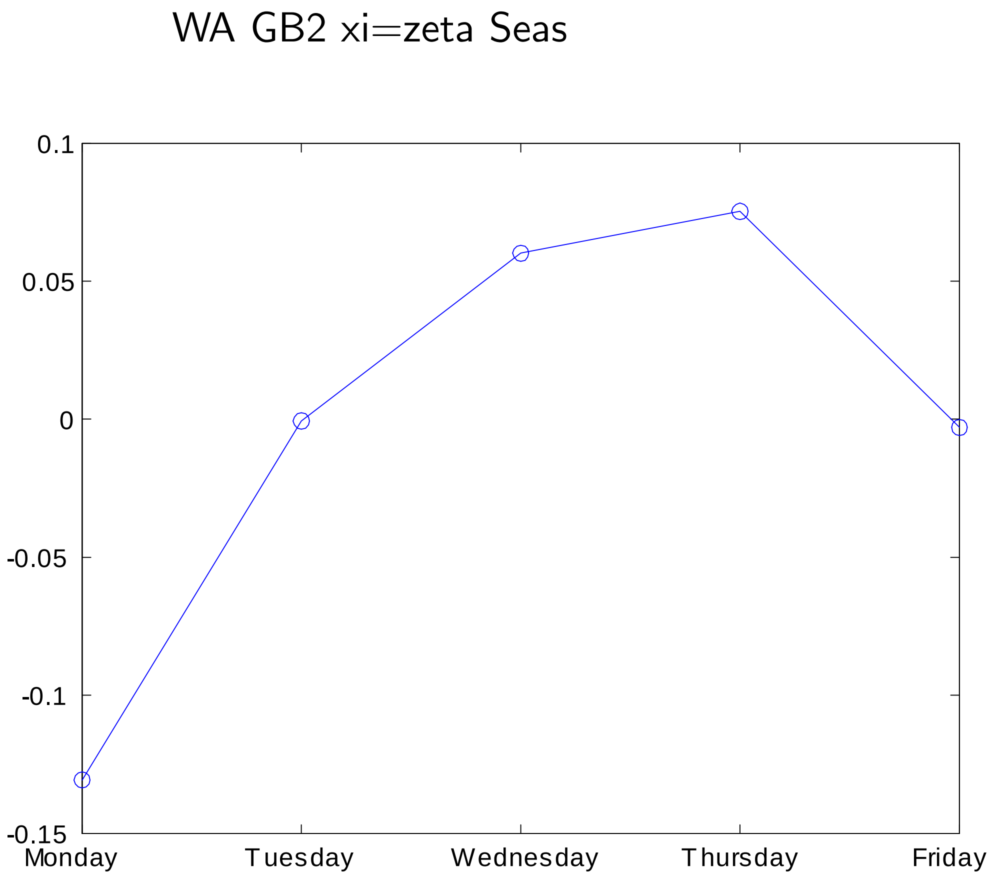
Pattern



$_1 1.pdf$

Figure: FTSE

Pattern



$_1 2.pdf$

Figure: S and P

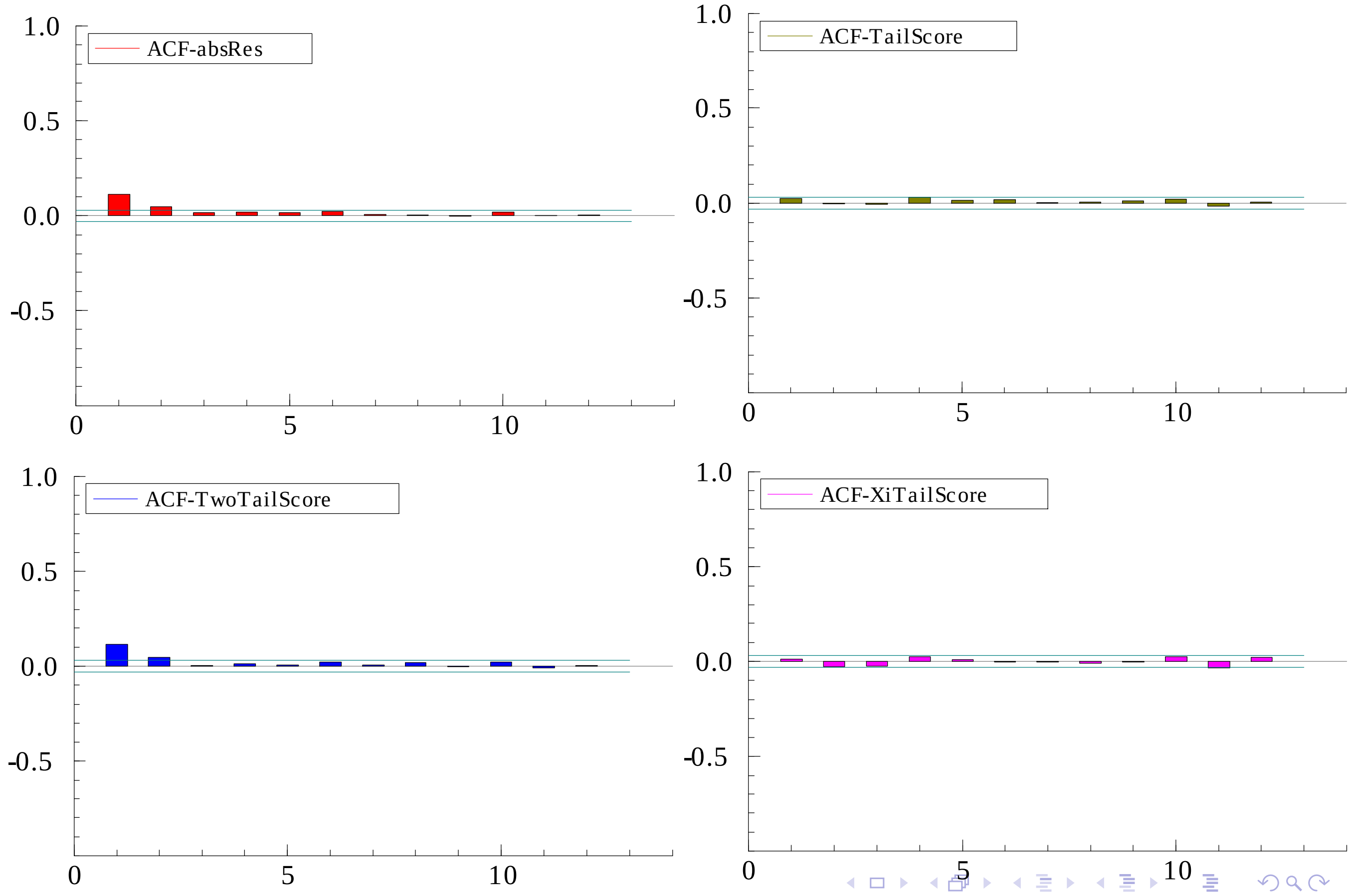


Figure: ACFs of scores for absolute values of residuals from preferred ln *RV*

Heteroscedasticity and a Changing Tail Index

Although the residuals from the preferred Gaussian RW+AR1+daily model have relatively little serial correlation, this is not true for their squares or absolute values. For fixed ξ and ς, the scale of the EGB2 is $1/\upsilon$ and the score wrt $\overline{\upsilon} = \exp(-\upsilon)$ is

$$\partial \ln f_t / \partial \overline{\upsilon} = (\xi + \varsigma)\varepsilon_t b_t - \xi \varepsilon_t - 1,$$

where $\varepsilon_t = (x_t - \lambda_{t|t-1})/\overline{\upsilon}$ and

$$b_t = \frac{\exp\{(x_t - \lambda_{t|t-1})/\overline{\upsilon}\}}{1 + \exp\{(x_t - \lambda_{t|t-1})/\overline{\upsilon}\}}.$$

The score is symmetric but unbounded. Caivano and Harvey (2014, p 566) show that in the limit as $\xi = \varsigma \to \infty$,
$\partial \ln f_t / \partial \overline{\upsilon} = (x_t - \lambda_{t|t-1})^2 / Var(x_t) - 1$, which is the score for a Gaussian EGARCH model. At the other extreme, when $\xi = \varsigma = 0$, the score is $\sqrt{2}\left|x_t - \lambda_{t|t-1}\right| / SD(x_t) - 1$.

Heteroscedasticity and a Changing Tail Index

How should heteroscedasticity in the logarithm of a variable that is already subject to changing variance be interpreted? As was noted earlier, the (upper) tail index for a GB2 is $\eta = \upsilon\varsigma$. Thus for a fixed value of ς, a dynamic υ implies that η is dynamic. It also might imply that the lower tail index, $\upsilon\xi$, is dynamic.
We could let the tail index be dynamic directly or let ς be dynamic and keep υ fixed. In the latter case, setting $\xi = \varsigma$ gives

$$\partial \ln f_t / \partial \varsigma = -2\psi(\varsigma) + 2\psi(2\varsigma) + \varepsilon_t + 2\ln(1 - b_t)$$

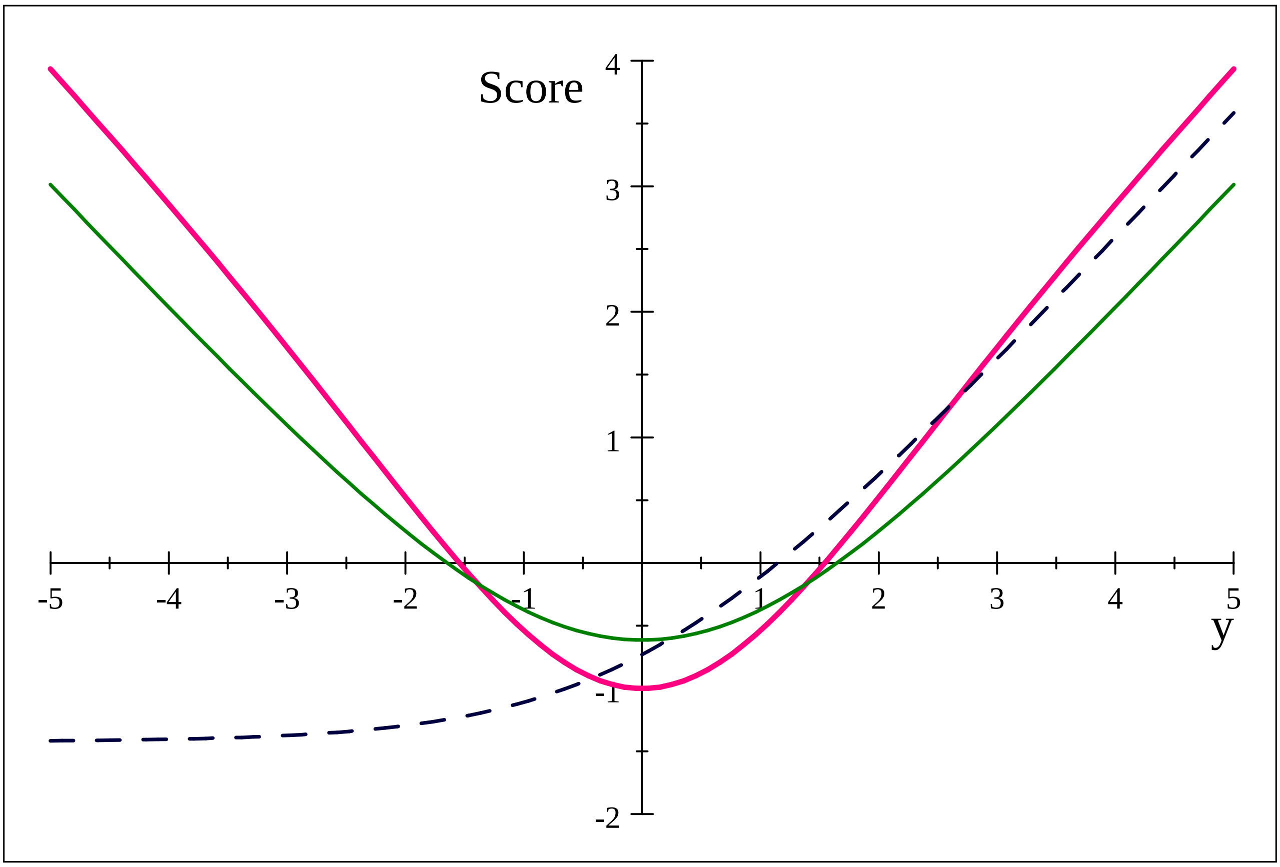
This is symmetric and for large x_t,
$\partial \ln f_t / \partial \varsigma \simeq -2\psi(\varsigma) + 2\psi(2\varsigma) - |\varepsilon_t| / \overline{\upsilon}$. The ACF for this score is similar to the one for $\overline{\upsilon}$. This is not surprising because both scores are given approximately by the absolute values of the residuals.

Heteroscedasticity and a Changing Tail Index

If $\xi = \varsigma$ is not assumed,

$$\partial \ln f_t / \partial \varsigma = -\psi(\varsigma) + \psi(\varsigma + \xi) + \ln(1 - b_t)$$

and, as can be seen from the graph, the (dashed) curve is asymmetric with large positive (negative) observations lowering (raising) the value of ς.



Scores wrt $\overline{\upsilon}$ and $\overline{\varsigma}$ ($=\xi$). Both parameters set to 1. The dash is for the score wrt $\overline{\varsigma}$ alone.

Multivariate models

The F-distribution gives a good fit to RV, although it is not the best. However, it does generalize to the modelling of an $N \times N$ realized volatility covariance matrix, $\mathbf{Y}_t$; see Opschoor et al (2016). The multivariate F has pdf

$$f(\mathbf{Y}_t \mid \Omega_{t|t-1}, \nu_1, \nu_2) = K(\nu_1, \nu_2) \frac{\left|\Omega_{t|t-1}\right|^{-\nu_1/2} \left|\mathbf{Y}_t\right|^{(\nu_1-N-1)/2}}{\left|\mathbf{I}+\Omega_{t|t-1}^{-1}\mathbf{Y}_t\right|^{(\nu_1+\nu_2)/2}},$$

where $\nu_1, \nu_2 > N-1$, $\Omega_{t|t-1} = (\nu_2 - N - 1)/\nu_1)\mathbf{V}_{t|t-1}$ is a scale matrix, such that $\mathbf{V}_{t|t-1} = E(\mathbf{Y}_t)$ for $\nu_1, \nu_2 > N-1$, and

$$K(\nu_1, \nu_2) = \frac{\Gamma_N((\nu_1 + \nu_2)/2)}{\Gamma_N(\nu_1/2)\Gamma_N(\nu_2/2)},$$

where $\Gamma_N(.)$ is the multivariate gamma function. When $\nu_2 \longrightarrow \infty$, the distribution becomes a Wishart distribution, which is the multivariate generalization of χ^2. A single entry on the diagonal of $\mathbf{Y}_t$, $y_{ii,t}$, $i = 1, ..., N$, is distributed as $F(\nu_1, \nu_2 - N - 2)$.

Multivariate models

A parsimonious model adopted by Opschoor et al (2016) is

$$\mathbf{V}_{t+1|t} = \mathbf{V} + \phi \mathbf{V}_{t|t-1} + \kappa \mathbf{U}_t, \tag{1}$$

where $\mathbf{U}_t$ is a scaled score matrix for $\mathbf{V}_{t|t-1}$. The assumptions $0 < \kappa < \phi < 1$ and $\mathbf{V}$ is a pd matrix ensure the stationarity of $\mathbf{Y}_t$ with the unconditional expectation of $\mathbf{Y}_t$ being $(1-\phi)^{-1}\mathbf{V}$. A simple extension is

$$\mathbf{V}_{t+1|t} = \mathbf{V} + \Phi \mathbf{V}_{t|t-1}\Phi' + \mathbf{K}\mathbf{U}_t\mathbf{K}'$$

where Φ and $\mathbf{K}$ are lower triangular matrices (and $\mathbf{V}$ is also decomposed as such). In principle $\mathbf{V}_{t|t-1}$ or, equivalently ˙$_{t|t-1}$, could have been modeled as ˙$_{t|t-1} = \mathbf{D}_{t|t-1}\mathbf{R}_{t|t-1}\mathbf{D}_{t|t-1}$.

Multivariate models

The matrix of scores is taken wrt a general non-symmetric $\mathbf{V}_{t|t-1}$ matrix and after scaling by a (scalar) multiple of $\mathbf{V}_{t|t-1} \otimes \mathbf{V}_{t|t-1}$ is

$$\mathbf{U}_t = \frac{\nu_1}{\nu_1+1}\left[\frac{\nu_1+\nu_2}{\nu_2-N-1}\mathbf{Y}_t\left(\mathbf{I}_N + \frac{\nu_1}{\nu_2-N-1}\mathbf{V}_{t|t-1}^{-1}\mathbf{Y}_t\right)^{-1} - \mathbf{V}_{t|t-1}\right]$$

The theoretical properties are obtained by using the fact that a transformation of the scaled score has a multivariate beta distribution; compare the result for a univariate F - a special case of GB2 with $\upsilon = 1$ and $\varsigma = \xi = 2\nu_1 = 2\nu_2$ as in Harvey (2013, pp 167-8).
When $\nu_2 \to \infty$,

$$\mathbf{U}_t \to \frac{\nu_1}{\nu_1+1}\left[\mathbf{Y}_t - \mathbf{V}_{t|t-1}\right]$$

which is the scaled score for the Wishart distribution. The results in Opschoor et al (2016) indicate a much better fit for the multivariate F.

Conclusion

DCS is statistically coherent and is able to model daily effects and asymmetric response (leverage) in a straighforward and transparent manner using ML. Structure and asymptotics similar to EGARCH (eg Harvey/Lange). The boundedness of the forcing variable in the dynamic equation makes the filtered volatility resistant to extreme values (outliers). It also ensures invertibility for models with parameters that tend to arise in practice.

Working in logarithms is useful in that the normal distribution is a limiting case and preliminary investigation is easily carried out using linear SSMs. Other models, although nonlinear, are relatively simple in that they only have one extra parameter, eg LL and bEGB2 and F. Simple and practical.

The analysis reveals a day of the week effect, which has not been observed in returns (or range).

HAR a benchmark. Best in logs. Similar forecasting performance but in general unable to beat our model. Asymmetry less easy to deal with and daily effect ruled out by construction.

Generalizes to multivariate case.