Dynamic Models for Volatility and Heavy Tails
2. Location and Scale

Andrew Harvey

Cambridge University

November 2012

.

Econometric Society *Monograph, No. 52*
Cambridge University Press
to appear April 2013.
http://www.cambridge.org/gb/knowledge /isbn/item7091594/?site_locale=en_GB

.

For Table of contents and Chapter 1, see
http://www.econ.cam.ac.uk/faculty/harvey/Pages-from-AHbook.pdf

Dynamic location model

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

where ε_t is serially independent, standard t-variate and the conditional score is

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

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

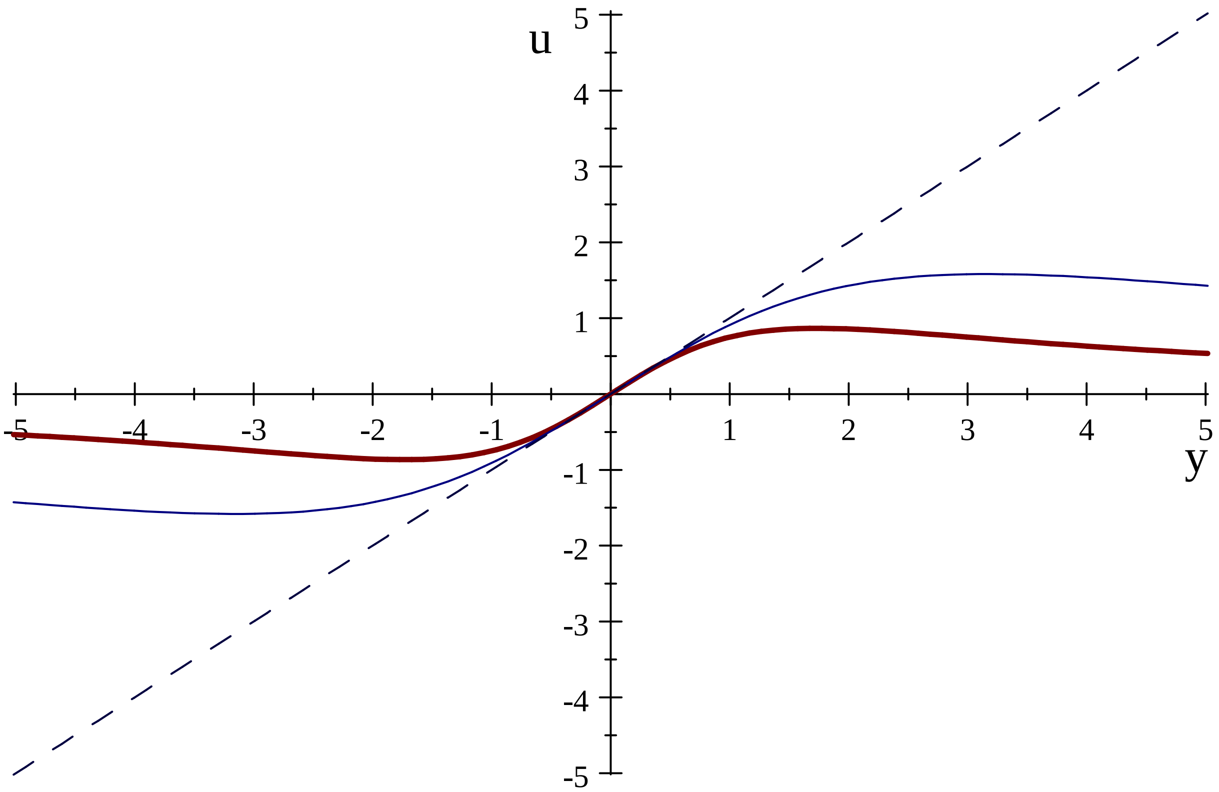


Figure: Impact of u_t for t_ν (with a scale of one) for $\nu = 3$ (thick), $\nu = 10$ (thin) and $\nu = \infty$ (dashed).

Basic properties

$$u_t = (1 - b_t)(y_t - \mu_{t|t-1}), \qquad (1)$$

where

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

is distributed as $beta(1/2, \nu/2)$. The $u_t^{\prime}s$ are $IID(0, \sigma_u^2)$ and symmetrically distributed.
The fact that b_t has a beta distribution follows from the property of the t-distribution

Basic properties

The filter may be generalized to:

$$\mu_{t+1|t} = \phi_1 \mu_{t|t-1} + ... + \phi_p \mu_{t-p+1|t-p} + \kappa_0 u_t + \kappa_1 u_{t-1} + ... + \kappa_r u_{t-r}.$$

Such a filter is denoted as $QARMA(p, r)$. The full model will be called $DCS - t - QARMA(p, r)$. It corresponds to an unobserved component signal plus noise model in which the signal is $ARMA(p, r)$.

Basic properties

In the Gaussian case $u_t = v_t$. If q is defined as $max(p, r+1)$, we may write

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + v_t - (\phi_1 - \kappa_0) v_{t-1} - ... - (\phi_q - \kappa_q) v_{t-q},$$

which is an $ARMA(p, q)$ with MA coefficients $\theta_i = \phi_i - \kappa_{i-1}, i = 1, .., q$. The invertibility conditions apply to $\theta_i = \phi_i - \kappa_{i-1}, i = 1, .., q$ rather than to $\kappa_i, i = 0, ...q$. But more generally, for a t_ν-distribution with $\nu < \infty$,

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \kappa_0 u_{t-1} + ... + \kappa_q u_{t-q} + v_t - \phi_1 v_{t-1} - ... - \phi_p v_{t-p}$$

and the MA disturbances are not identically distributed as each is a different combination of variables, u_t and v_t, which have different (non-normal) distributions. In fact they do not all have the same variances. The process is still $ARMA(p, q)$, but the MA coefficients are not $\phi_i - \kappa_{i-1}, i = 1, .., q$.

Basic properties: moments

When $\mu_{t+1|t}$ is stationary, the location can be written as an infinite moving average,

$$\mu_{t|t-1} = \omega + \sum_{j=1}^{\infty} \psi_j u_{t-j}, \qquad \sum_{j=1}^{\infty} \psi_j^2 < \infty, \tag{3}$$

where $\omega = \delta/(1 - \phi_1 - ... - \phi_p)$, so

$$y_t = \omega + \sum_{j=1}^{\infty} \psi_j u_{t-j} + v_t.$$

The existence of moments of y_t is not affected by the dynamics.

Basic properties: autocovariances

In the first -order model the form of the ACF is that of an ARMA(1,1) since

$$\rho_v(1) = \left[\kappa + \frac{\nu}{3+\nu} \frac{\kappa^2 \phi}{1-\phi^2}\right] / \left[\frac{\nu+1}{\nu-2} + \frac{\nu}{3+\nu} \frac{\kappa^2}{1-\phi^2}\right]$$

depends on κ and ϕ, but thereafter $\rho_v(\tau) = \phi \rho_v(\tau - 1), \quad \tau = 2, 3, ...$

Maximum likelihood estimation

The log-likelihood function for the DCS-t model is

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

Maximization of the log-likelihood function with respect to the unknown dynamic parameters in the vector $\boldsymbol{\psi}$ and the scale and shape parameters, λ and ν, can be carried out by numerical optimization.

Maximum likelihood estimation: information matrix

Let $y_t \mid Y_{t-1}$ have a t_ν-distribution with $\mu_{t|t-1}$ generated by the first-order model. Then, assuming that $|\phi|<1$ and $b<1$,

$$\mathbf{I}\begin{pmatrix}\boldsymbol{\psi}\\ \lambda\\ \nu\end{pmatrix}=\begin{bmatrix}\frac{\nu+1}{\nu+3}\exp(-2\lambda)\mathbf{D}(\boldsymbol{\psi}) & 0 & 0\\ 0 & \frac{2\nu}{\nu+3} & \frac{1}{(\nu+3)(\nu+1)}\\ 0 & \frac{1}{(\nu+3)(\nu+1)} & h(\nu)/2\end{bmatrix}, \qquad (4)$$

where

$$\mathbf{D}\begin{pmatrix}\kappa\\ \phi\\ \omega\end{pmatrix}=\frac{1}{1-b}\begin{bmatrix}\sigma_u^2 & \frac{a\kappa\sigma_u^2}{1-a\phi} & 0\\ \frac{a\kappa\sigma_u^2}{1-a\phi} & \frac{\kappa^2\sigma_u^2(1+a\phi)}{(1-\phi^2)(1-a\phi)} & 0\\ 0 & 0 & \frac{(1-\phi)^2(1+a)}{1-a}\end{bmatrix}$$

Maximum likelihood estimation: information matrix

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

$$b = \phi^2 - 2\phi\kappa\frac{\nu}{\nu+3} + \kappa^2 \frac{\nu\left(\nu^3+10\nu^2+35\nu+38\right)}{(\nu+1)(\nu+3)(\nu+5)(\nu+7)},$$

Figure shows a plot of b against κ for $\phi = 0.9$ and $\nu = 6$. The admissible range is slightly bigger than in the Gaussian case where it is $-0.1 < \kappa < 1.9$.

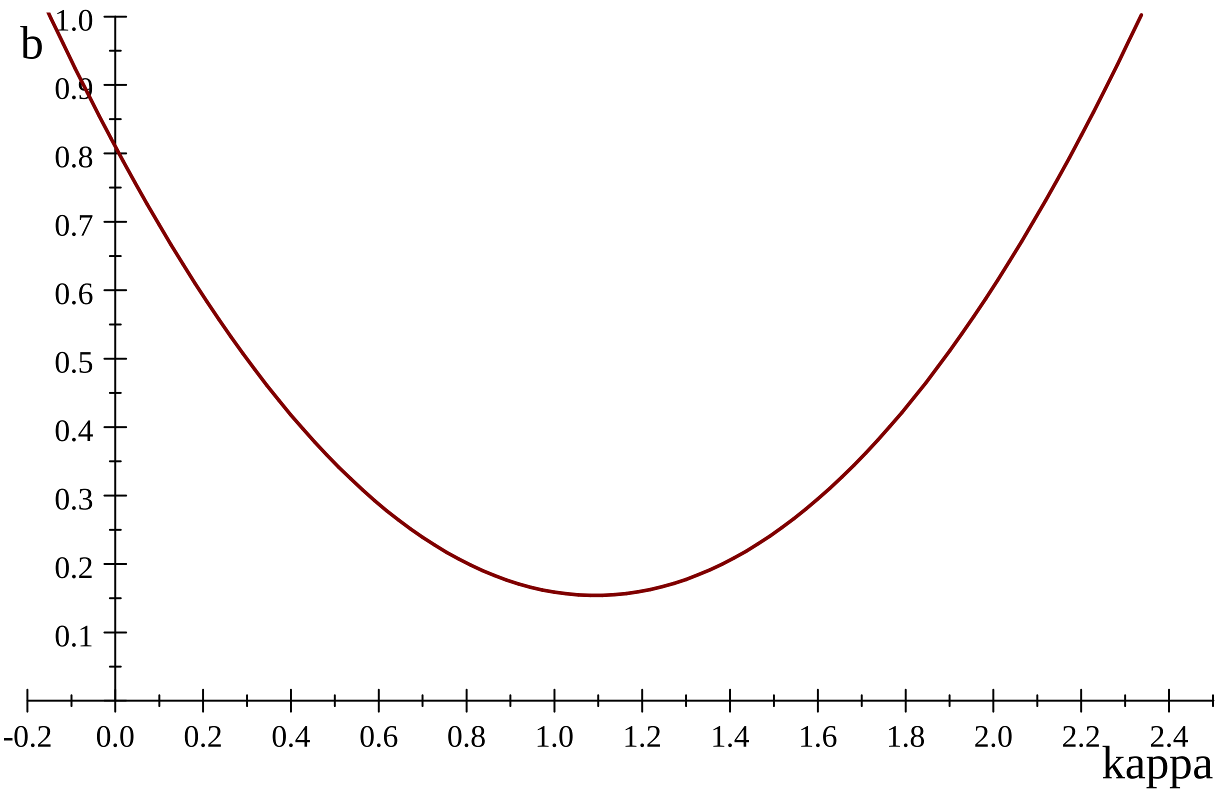


Figure: Plot of b against κ for $\phi = 0.9$ and $\nu = 6$

Maximum likelihood estimation: Gaussian model

For a Gaussian model, $b < 1$ provided that $\phi - 1 < \kappa < \phi + 1$.
The reduced form is the $ARMA(1,1)$ process

$$y_t = \phi y_{t-1} + v_t - \theta v_{t-1}.$$

The condition for strict invertibility in the ARMA(1,1) model is $|\theta| < 1$ and since $\theta = \phi - \kappa$, invertibility ensures that $b < 1$. The condition $\theta \neq \phi$ is needed for identifiability and this condition is equivalent to $\kappa \neq 0$.
When ϕ is known,

$$Var(\widetilde{\kappa}) = 1 - b = 1 - (\phi - \kappa)^2,$$

which is consistent with the standard $MA(1)$ result, $Var(\widetilde{\theta}) = 1 - \theta^2$.

Maximum likelihood estimation: Monte Carlo experiments

Parameter			ML estimates for $T = 1000$				
ϕ	κ		ϕ	κ	λ	ω	ν
0.8	0.5	RMSE	0.037	0.053	0.035	0.093	1.161
		ASE	0.037	0.043	0.029	0.094	0.844
0.8	1.0	RMSE	0.250	0.067	0.031	0.144	0.920
		ASE	0.240	0.045	0.029	0.147	0.844
0.95	0.5	RMSE	0.015	0.048	0.035	0.244	1.100
		ASE	0.012	0.038	0.029	0.269	0.844
0.95	1.0	RMSE	0.012	0.064	0.031	0.387	0.882
		ASE	0.010	0.043	0.029	0.484	0.844

Application to US GDP

A Gaussian AR(1) plus noise model with a constant, was fitted to the growth rate of US Real GDP, defined as the first difference of the logarithm, using the STAMP 8 package. The data were quarterly, from 1947(2) to 2012(1), and the parameter estimates were as follows:

$$\widetilde{\phi} = 0.501, \quad \widetilde{\sigma}_\eta^2 = 7.62 \times 10^{-5}, \quad \widetilde{\sigma}_\varepsilon^2 = 2.30 \times 10^{-5}, \quad \widetilde{\omega} = 0.0078.$$

There was little indication of residual serial correlation, but the Bowman-Shenton statistic is 30.04, which is clearly significant as the distribution under the null hypothesis of Gaussianity is χ_2^2. The non-normality clearly comes from excess kurtosis, which is 1.9, rather than from skewness.

Application to US GDP

DCS-location-t model. The estimated degrees of freedom of 6.3 means that the DCS filter is less responsive to more extreme observations, such as the fall of 2009(1).

Parameter	κ	ϕ	λ	ω	ν
Estimate	0.520	0.497	-4.878	0.0079	6.303
Num SE	0.098	0.102	0.073	0.0009	2.310
ASE	0.090	0.140	0.057	0.0009	1.807

Higher-order models and the state space form

The observation in the state space form is related to an $m \times 1$ state vector, $\boldsymbol{\alpha}_t$, through a measurement equation,

$$y_t = \omega + \mathbf{z}'\boldsymbol{\alpha}_t + \varepsilon_t, \quad t = 1, ..., T,$$

where $\mathbf{z}$ is an $m \times 1$ vector and ε_t is a serially uncorrelated disturbance with $E(\varepsilon_t) = \mathbf{0}$ and $Var(\varepsilon_t) = \sigma_\varepsilon^2$. The transition equation is

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\delta} + \mathbf{T}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \quad t = 1, ..., T.$$

The Kalman filter can be written as a single set of recursions going directly from $\boldsymbol{\alpha}_{t|t-1}$ to $\boldsymbol{\alpha}_{t+1|t}$, that is

$$\boldsymbol{\alpha}_{t+1|t} = \boldsymbol{\delta} + \mathbf{T}\boldsymbol{\alpha}_{t|t-1} + \mathbf{k}_t v_t, \quad t = 1, ..., T,$$

where $v_t = y_t - \omega - \mathbf{z}_t'\boldsymbol{\alpha}_{t|t-1}$ is the innovation and $f_t = \mathbf{z}'\mathbf{P}_{t|t-1}\mathbf{z} + \sigma_\varepsilon^2$ is its variance. The gain vector, $\mathbf{k}_t$, is

$$\mathbf{k}_t = (1/f_t)\mathbf{T}\mathbf{P}_{t|t-1}\mathbf{z}, \quad t = 1, ..., T.$$

Higher-order models and the state space form

Re-arranging the KF equations gives the innovations form

$$\begin{aligned} y_t &= \omega + \mathbf{z}'\boldsymbol{\alpha}_{t|t-1} + v_t, \quad t = 1, ..., T, \\ \boldsymbol{\alpha}_{t+1|t} &= \boldsymbol{\delta} + \mathbf{T}\boldsymbol{\alpha}_{t|t-1} + \mathbf{k}_t v_t. \end{aligned} \tag{5}$$

A general location DCS model may be set up in the same way as the innovations form of a Gaussian state space model. The model corresponding to the steady-state of (5) is

$$\begin{aligned} y_t &= \omega + \mathbf{z}'\boldsymbol{\alpha}_{t|t-1} + v_t, \quad t = 1, ..., T, \\ \boldsymbol{\alpha}_{t+1|t} &= \boldsymbol{\delta} + \mathbf{T}\boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\kappa} u_t. \end{aligned} \tag{6}$$

Trend and seasonality

Stochastic trend and seasonal components may be introduced into UC models for location. These models, called structural time series models, are implemented in the STAMP package of Koopman *et al* (2009).
The Gaussian random walk plus noise or *local level* model is

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim NID(0, \sigma_\varepsilon^2), \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2), \end{aligned} \tag{7}$$

where $E(\varepsilon_t \eta_s) = 0$ for all t and s. The signal noise ratio is $q = \sigma_\eta^2 / \sigma_\varepsilon^2$.
The range of κ in the steady-state innovations form is $0 < \kappa \leq 1$. In this case $\mu_{t+1|t}$ is an EWMA

$$\mu_{t+1|t} = (1 - \kappa)\mu_{t|t-1} + \kappa y_t. \tag{8}$$

For a semi-infinite sample

$$\mu_{t+1|t} = \kappa \sum_{i=0}^{\infty} (1 - \kappa)^i y_{t-i} \tag{9}$$

and the weights on past observations sum to one.

Trend and seasonality

For the DCS-t filter

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \\ \mu_{t+1|t} &= \mu_{t|t-1} + \kappa u_t. \end{aligned} \tag{10}$$

and the initial value, $\mu_{1|0}$,is treated as an unknown parameter that needs to estimated along with κ and ν.
Since $u_t = (1 - b_t)(y_t - \mu_{t|t-1})$, re-arranging the dynamic equation in (10) gives

$$\mu_{t+1|t} = (1 - \kappa(1 - b_t))\mu_{t|t-1} + \kappa(1 - b_t)y_t, \quad t = 1, ..., T. \tag{11}$$

Trend and seasonality

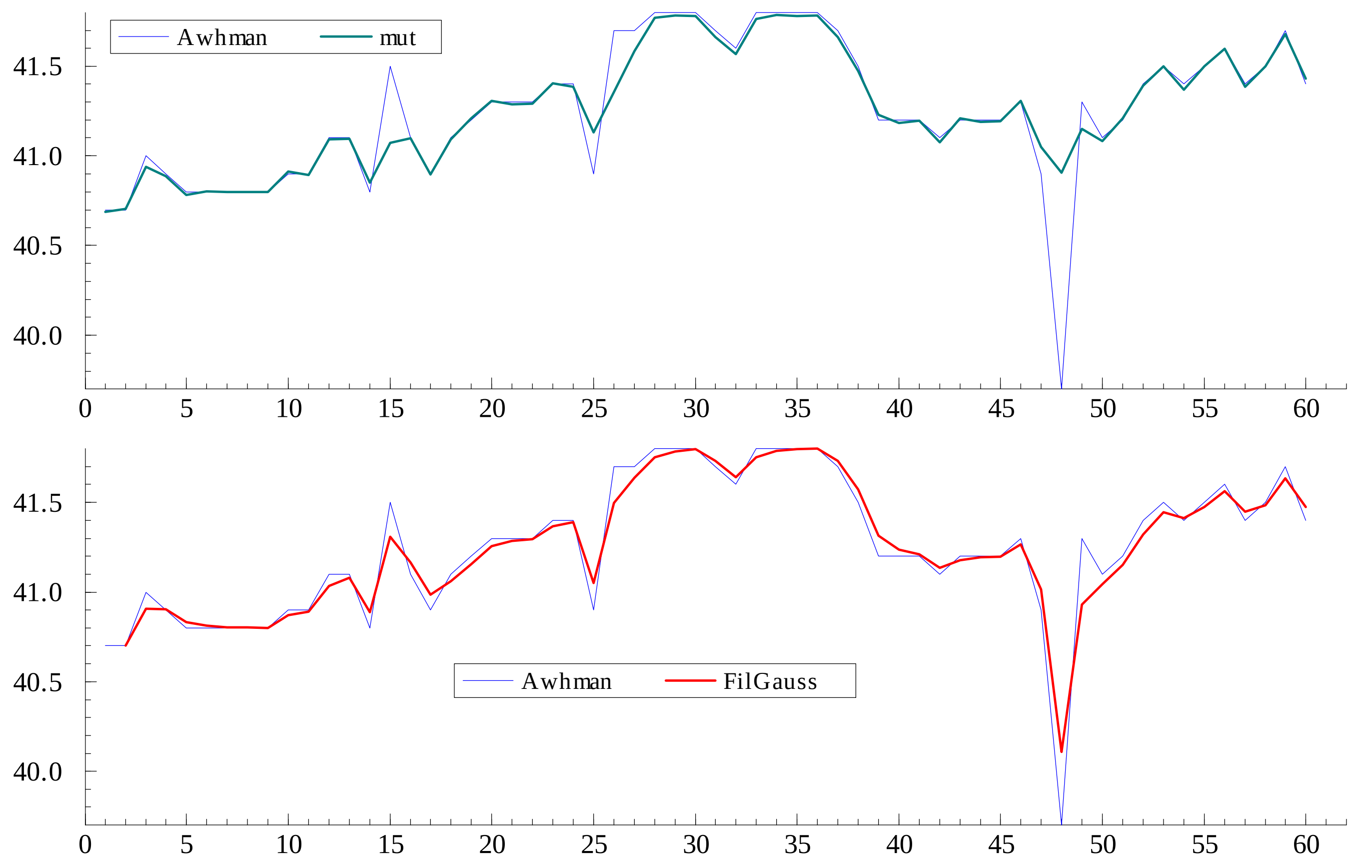
Fitting a local level DCS model (initialized with $\mu_{2|1} = y_1$) to seasonally adjusted monthly data on U.S. Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing (AWHMAN) from February 1992 to May 2010 (220 observations) gave

$$\widetilde{\kappa} = 1.246 \qquad \widetilde{\lambda} = -3.625 \qquad \widetilde{\nu} = 6.35$$

with numerical (asymptotic) standard errors

$$SE(\widetilde{\kappa}) = 0.161(0.090) \qquad SE(\widetilde{\lambda}) = 0.120(0.062) \qquad SE(\widetilde{\nu}) = 1.630(1.991)$$

A drift term was initially included but it was statistically insignificant. The value of b is 0.151. Although $\widetilde{\kappa}$ is greater than one, the resulting filter is perfectly consistent with the properties of the series. Figure shows (part of) the series together with the contemporaneous filter, which for the random walk is $\mu_{t|t} = \mu_{t+1|t}$. Unusually large prediction errors result in a small value of $\kappa(1 - b_t)$ and most of weight in the filter is assigned to $\mu_{t|t-1}$.



Local linear trend

The DCS local linear trend filter is

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \\ \mu_{t+1|t} &= \mu_{t|t-1} + \beta_{t|t-1} + \kappa_1 u_t \\ \beta_{t+1|t} &= \beta_{t|t-1} + \kappa_2 u_t. \end{aligned} \tag{12}$$

The initialization $\beta_{3|2} = y_2 - y_1$ and $\mu_{3|2} = y_2$ can be used, but, as in the local level model, initializing in this way is vulnerable to outliers at the beginning. Estimating the fixed starting values, $\mu_{1|0}$ and $\beta_{1|0}$, may be a better option.
The model may be extended to include a stochastic seasonal.

Seasonal

A fixed seasonal pattern may be modeled as

$$\gamma_t = \sum_{j=1}^{s} \gamma_j z_{jt}$$

where s is the number of seasons and the dummy variable z_{jt} is one in season j and zero otherwise. In order not to confound trend with seasonality, the coefficients, γ_j, $j = 1, ..., s$, are constrained to sum to zero. The seasonal pattern may be allowed to change over time by letting the coefficients evolve as random walks. If γ_{jt} denotes the effect of season j at time t, then

$$\gamma_{jt} = \gamma_{j,t-1} + \omega_{jt}, \quad \omega_t \sim NID(0, \sigma_\omega^2), \quad j = 1, ..., s.$$

Only one seasonal affects the observations at a given time, that is $\gamma_t = \gamma_{jt}$ when season j is prevailing at time t.

Seasonal

The requirement that the seasonal components evolve in such a way that they always sum to zero, that is $\sum_{j=1}^{s} \gamma_{jt} = 0$, is enforced by the restriction that the disturbances sum to zero at each point in time. This restriction is implemented by the correlation structure in

$$Var\left(\boldsymbol{\omega}_t\right) = \sigma_\omega^2 \left(\mathbf{I} - s^{-1}\mathbf{i}\mathbf{i}'\right),$$

where $\boldsymbol{\omega}_t = \left(\omega_{1t}, ..., \omega_{st}\right)'$, coupled with initial conditions requiring that the seasonals sum to zero at $t = 0$.
It can be seen that $Var\left(\mathbf{i}'\boldsymbol{\omega}_t\right) = 0$.

In the state space form, the transition matrix is just the identity matrix, but the $\mathbf{z}$ vector must change over time to accommodate the current season. Apart from replacing $\mathbf{z}$ by $\mathbf{z}_t$, the form of the KF remains unchanged. Adapting the innovations form to the DCS observation driven framework gives

$$\begin{aligned} y_t &= \mathbf{z}_t'\boldsymbol{\alpha}_{t|t-1} + v_t \\ \boldsymbol{\alpha}_{t+1|t} &= \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\kappa}_t u_t, \end{aligned} \tag{13}$$

where $\mathbf{z}_t$ picks out the current season, $\gamma_{t|t-1}$, that is $\gamma_{t|t-1} = \mathbf{z}_t'\boldsymbol{\alpha}_{t|t-1}$.
The only question is how to parameterize $\boldsymbol{\kappa}_t$.

The seasonal dummies in the UC model are constrained to sum to zero and the same is true of their filtered estimates. Thus $\mathbf{i}'\boldsymbol{\kappa}_t = 0$ in the Kalman filter and this property should carry across to the DCS filter.
If κ_{jt}, $j = 1, .., s$, denotes the $j - th$ element of $\boldsymbol{\kappa}_t$ in (13), then in season j we set $\kappa_{jt} = \kappa_s$, where κ_s is a non-negative unknown parameter, while

$$\kappa_{it} = -\kappa_s/(s-1), \qquad i \neq j.$$

The amounts by which the seasonal effects change therefore sum to zero.

Seasonal

The seasonal recursions can be combined with the trend filtering equations of (12) in order to give a structure similar in form to that of the Kalman filter for the stochastic trend plus seasonal plus noise model, sometimes known as the 'basic structural model'. Thus

$$y_t = \mu_{t|t-1} + \gamma_{t|t-1} + v_t, \tag{14}$$

where $\mu_{t|t-1}$ is as in (12).
The filter can be initialized by regressing the first $s+1$ observations on a constant, time trend and seasonal dummies constrained so that the coefficients to sum to zero.
Alternatively, the initial conditions at time $t = 0$ are estimated by treating them as parameters.

Application to rail travel

Figure shows the logarithm of National Rail Travel, defined as the number of kilometres traveled by passengers.
An unobserved components model was fitted to this series using the STAMP 8 package of Koopman et al (2009).
Trend, seasonal and irregular components were included but the model was augmented with intervention variables to take out the effects of observations that are known to be unrepresentative.
The intervention dummies were: (i) the train driver strikes in 1982(1,3); (ii) the Hatfield crash and its aftermath, 2000(4) and 2001(1); and (iii) the signallers strike in 1994(3).

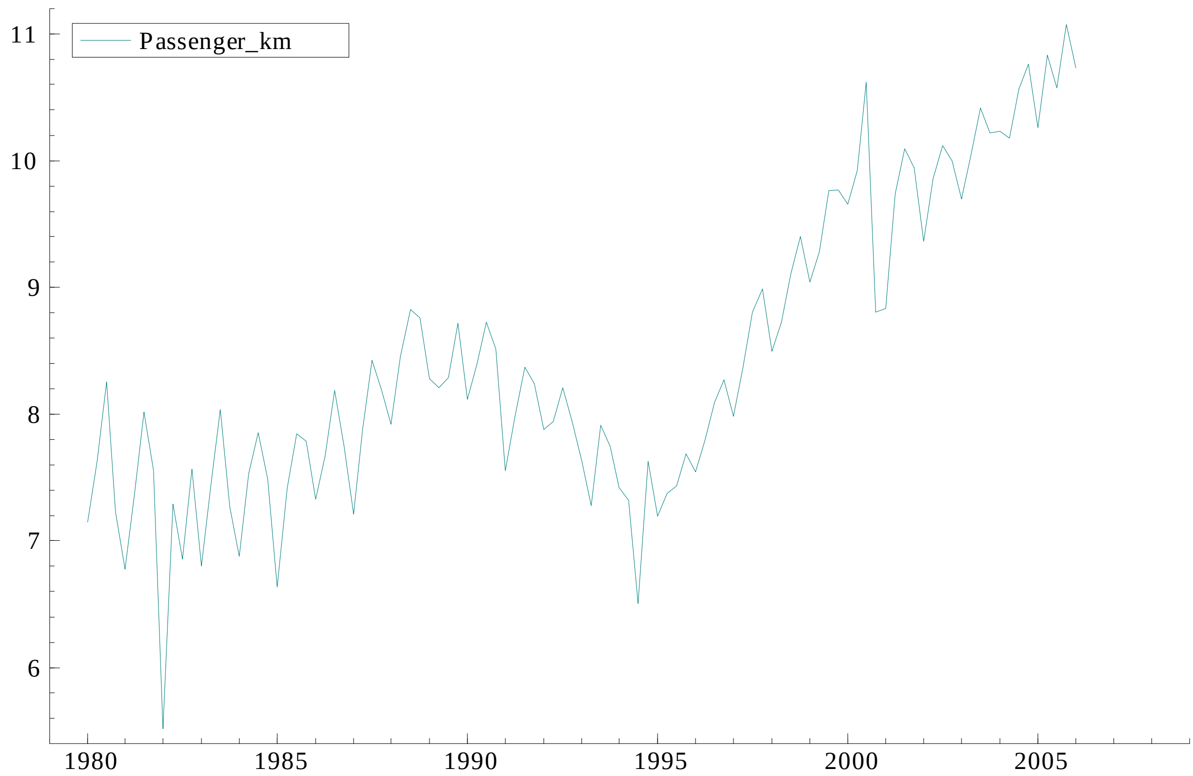


Figure: Logarithm of National Rail Travel in the UK (number of kilometres)

Fitting a DCS model with trend and seasonal, that is (14), avoids the need to deal explicitly with the outliers. The ML estimates for the parameters in a model with a random walk plus drift trend, that is

$$\mu_{t+1|t} = \mu_{t|t-1} + \beta + \kappa_1 u_t,$$

are

$$\begin{array}{rclll} \widetilde{\kappa}_1 & = & 1.421(0.161) & \widetilde{\kappa}_s = 0.539\ (0.070) & \widetilde{\lambda} = -3.787\ (0.053) \\ \widetilde{\nu} & = & 2.564\ (0.319) & \widetilde{\beta} = 0.003\ (0.001) & \end{array}$$

with initial values

$$\begin{array}{rclll} \widetilde{\mu} & = & 2.066(0.009) & & \\ \widetilde{\gamma}_1 & = & -0.094(0.007) & \widetilde{\gamma}_2 = -0.010(0.006) & \widetilde{\gamma}_3 = 0.086(0.006) \end{array}$$

The last seasonal is $\widetilde{\gamma}_4 = 0.018$; it has no standard error (SE) as it was constructed from the others.

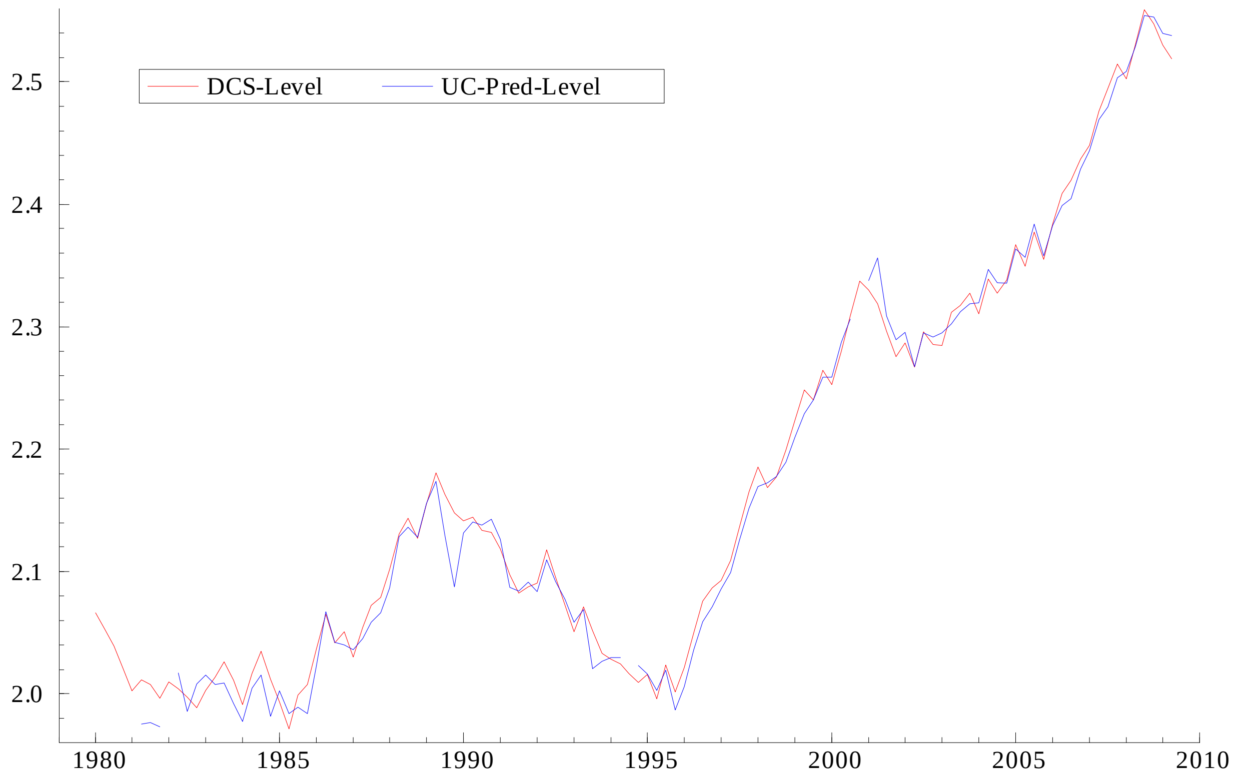


Figure: Trends in National Rail Travel from UC and DCS models

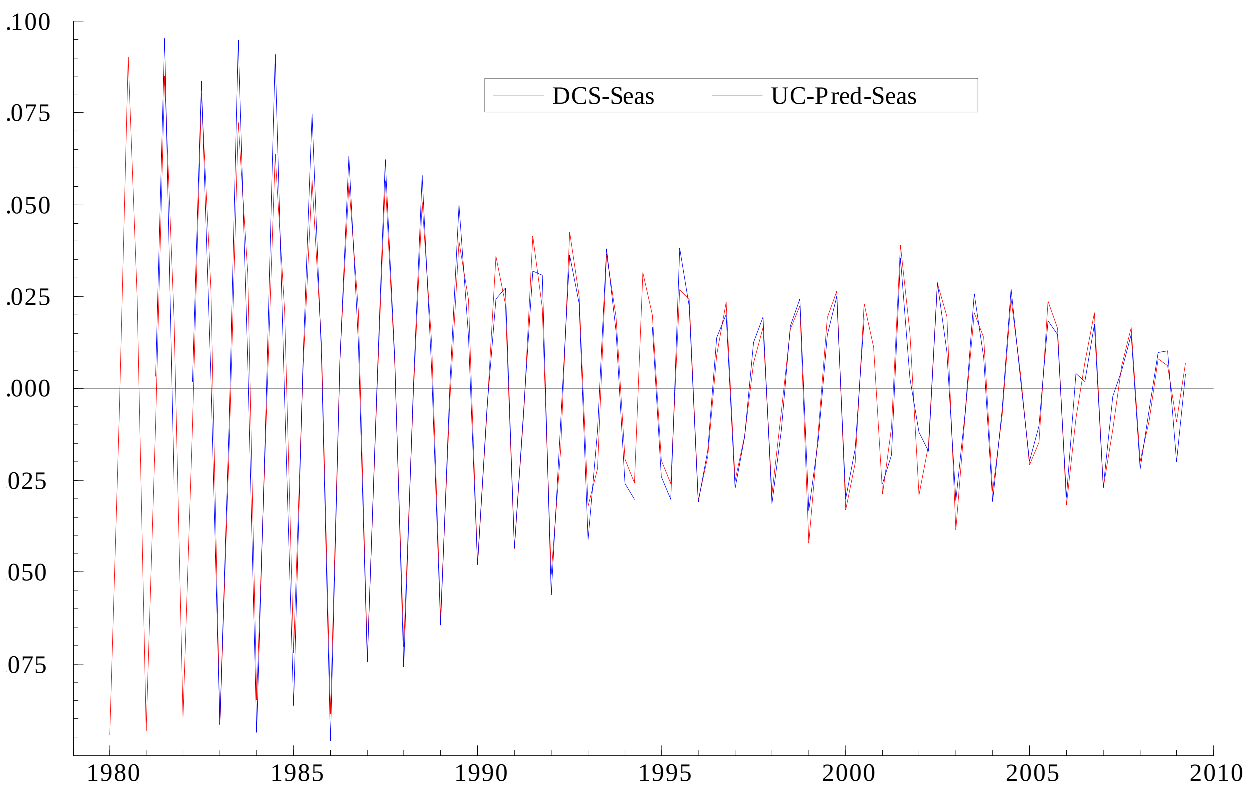


Figure: Seasonals in National Rail Travel from UC and DCS models

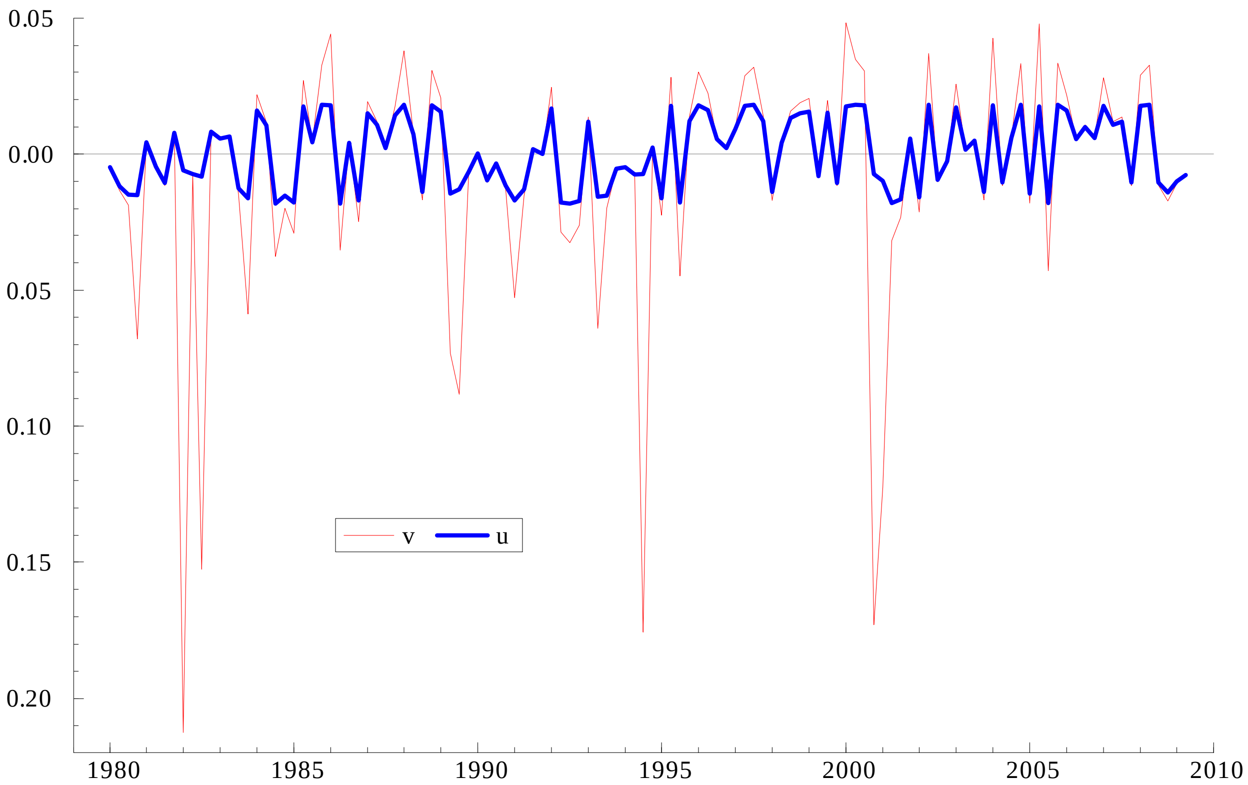


Figure: Residuals and scores from DCS-t model

The outliers, removed by dummies in the UC model, show up clearly in the irregular. (The Bowman-Shenton statistic is 137.82 indicating a massive rejection of the null of Gaussianity.) In the score series the outliers are downweighted. As a result, the autocorrelations for the score are slightly bigger than those of the residuals as they are not weakened by aberrant values; see Figure The Box-Ljung Q(12) statistic is 19.78 for the score and 12.40 for the residuals. If it can be assumed that only the number of fitted dynamic parameters affects the distribution of the Box-Ljung statistic, its distribution under the null hypothesis of correct model specification is χ^2_{10}, which had a 5% critical value of 18.3. Thus the scores reject the null hypothesis, albeit only marginally, while the residuals do not. Having said that, the score autocorrelations do not exhibit any clear pattern and it is not clear how the dynamic specification might be improved.

Figure: Residual correlograms for irregular and score residuals from DCS-t model fitted to National Rail Travel (lines are $\pm 1/\sqrt{T}$)

Andrew Harvey (Cambridge University) Volatility and Heavy Tails / 107

Explanatory variables

The location parameter may depend on a set of observable explanatory variables, denoted by the $k \times 1$ vector $\mathbf{w}_t$, as well as on its own past values and the score. The model with a stochastic trend can be set up as

$$y_t = \mu^{\dagger}_{t|t-1} + \mathbf{w}_t'\gamma + \nu_t,$$

where

$$\mu^{\dagger}_{t+1|t} = \phi\mu^{\dagger}_{t|t-1} + \kappa u_t$$

Corollary

Assume that the explanatory variables are weakly stationary with mean $\boldsymbol{\mu}_w$ and second moment Λ_w and are strictly exogenous in the sense that they are independent of the u_t's in all time periods, that is $E(\mathbf{w}_t u_s) = \mathbf{0}$ for all $t, s = 1, .., T$. The information matrix is

$$\mathbf{I}\begin{pmatrix} \kappa \\ \phi \\ \gamma \end{pmatrix} = \frac{\sigma_u^2}{k^2(1-b)}\begin{bmatrix} A & D & \mathbf{0} \\ D & B & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_w \end{bmatrix},$$

where A, B and D, E are as before while

$$\mathbf{C}_w = (1+\phi^2)\Lambda_w - 2\phi\Lambda_w(1) + \frac{2a(1-\phi)^2}{1-a}\boldsymbol{\mu}_w\boldsymbol{\mu}_w',$$

with $\Lambda_w(1) = E(\mathbf{w}_t\mathbf{w}_{t-1}') = E(\mathbf{w}_{t-1}\mathbf{w}_t')$.

Corollary

*When $\mu^{\dagger}_{t|t-1}$ is known to be a random walk with drift, β, as in (**??**), and $\mu^{\dagger}_{1|0}$ is fixed and known, the information matrix is as in (4) but with*

$$\mathbf{D}\begin{pmatrix}\kappa\\ \gamma\\ \beta\end{pmatrix}=\frac{1}{1-b}\begin{bmatrix}\sigma_u^2 & \mathbf{0}' & \mathbf{0}'\\ \mathbf{0} & \mathbf{C}_{\Delta w} & \boldsymbol{\mu}_{\Delta w}\\ \mathbf{0} & \boldsymbol{\mu}'_{\Delta w} & 1\end{bmatrix},$$

where $\boldsymbol{\mu}_{\Delta w}=E(\Delta \mathbf{w}_t)$ and $\mathbf{C}_{\Delta w}=E(\Delta \mathbf{w}_t\Delta \mathbf{w}'_t)$. It is assumed that $b<1$ and $\mathbf{C}_{\Delta w}$ is positive definite.

The first differences of the explanatory variables must be weakly stationary but their levels may be nonstationary. It follows from the above result that the covariance matrix of the limiting distribution of $\sqrt{T}\widetilde{\gamma}$ is

$$Var(\widetilde{\gamma})=\left(2\kappa\frac{\nu}{\nu+1}-\kappa^2\frac{\nu\left(\nu^3+10\nu^2+35\nu+38\right)}{(\nu+1)^2\left(\nu+5\right)\left(\nu+7\right)}\right)e^{2\lambda}(\mathbf{C}_{\Delta w}-\boldsymbol{\mu}_{\Delta w}\boldsymbol{\mu}'_{\Delta w})^{-1}$$

In principle, the above Corollary may be extended to models where seasonals are included.

Application to rail travel

Potential explanatory variables for the rail travel series are: (i) Real GDP (in £2003 prices), (ii) Real Fares, obtained by dividing total revenue by the number of kilometres travelled and the retail price index (RPI), and (iii) Petrol and Oil index (POI), divided by RPI. The fares series was smoothed by fitting a univariate UC model,
Fitting an unobserved components time series model using STAMP gave the following estimates for the coefficients of the logarithms of the explanatory variables:

$$LGDP = 0.716(0.267) \qquad Lfare = -0.416(0.245) \qquad LPOI = 0.050(0.065)$$

All the estimates are all plausible. The coefficient of the petrol index is not statistically significant at any conventional level, but at least it has the right sign. The appendix shows the print-out of the full set of results.
Failure to deal with outliers in a time series regression can lead to **serious distortions** and this is well-illustrated by the rail series when the intervention variables are not included. In particular the fare estimate is *plus* 0.36.

Fitting a DCS-t model gave the following results:

$$\begin{aligned} \widetilde{\kappa}_1 &= 2.212 \qquad \widetilde{\kappa}_s = 0.771 \quad \widetilde{\lambda} = -4.059 \\ \widetilde{\nu} &= 2.070 \qquad \widetilde{\beta} = 0.0004 \end{aligned}$$

with initial values $\widetilde{\mu} = -6.162$, $\widetilde{\gamma}_1 = -0.084$, $\widetilde{\gamma}_2 = -0.007$ and $\widetilde{\gamma}_3 = 0.070$.
The coefficients of the explanatory variables were:

$$LGDP = 0.734 \qquad Lfare = -0.427 \quad LPOI = 0.056$$

The Box-Ljung $Q(12)$ statistic is 5.30 for the score and 16.12 for the residuals. This result is surprising because in the univariate model the $Q - statistic$ for the score was bigger than that of the residuals.

Figure shows the stochastic trend with a constant factor added so that it is at a level comparable with that of the series. A good deal, but by no means all, of the growth from the mid-nineties is due to the increase in GDP. The continued fall in rail travel after the economy had moved out of the recession of the early nineties is explained by the fact that fares increased sharply in 1993 in anticipation of rail privatisation and continued to increase till 1995.

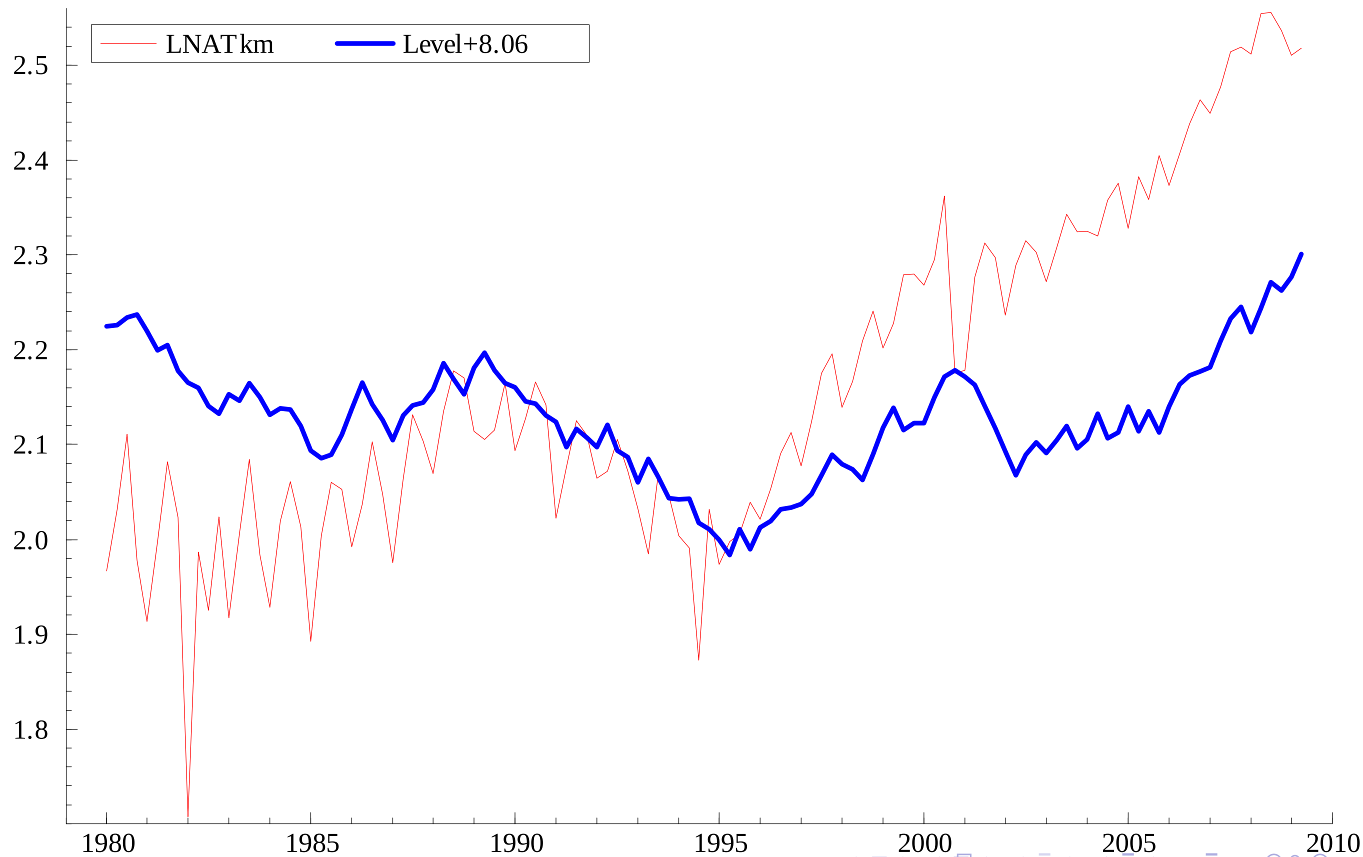


Figure: Adjusted level (trend) in rail travel when explanatory variabes are taken

Smoothing

In a linear Gaussian model, the smoothed estimate is $\boldsymbol{\alpha}_{t|T}$, the conditional expectation of $\boldsymbol{\alpha}_t$, and hence its MMSE, based on all the observations in the sample. For a DCS model the smoothing filter is defined by a symmetry argument rather than being derived as an optimal estimate. However it can also be rationalized by an argument based on the conditional mode of the posterior distribution of the state. Indeed this signal extraction interpretation probably provides a more solid foundation for DCS models than the case set out earlier.

The filtering equations in a DCS model have the same form as the Kalman filter in a corresponding linear Gaussian UC model. The KF defines an implicit set of weights for current and past observations. Similarly the backward recursive equations in a fixed interval smoother for a linear Gaussian UC model implicitly define a set of weights for all observations. Smoothing in a DCS model amounts to using a set of smoothing weights that match the weights for the DCS filter.

Smoothing

In a simple linear Gaussian UC model there are explicit formulae for these weights; see Whittle (1983, Chapters 6 and 7). For the AR(1) plus noise model, the weights are a function of the parameter θ in the ARMA(1,1) reduced form. To be specific, the weights for the filter in a semi-infinite sample,

$$\mu_{t+1|t} = \sum_{j=0}^{\infty} w_j y_{t-j},$$

are

$$w_j = (\phi - \theta)\theta^j, \qquad j = 0, 1, 2, \ldots, \tag{15}$$

For the smoother in a doubly infinite sample,

$$\mu_{t|T} = \sum_{j=-\infty}^{\infty} w_j y_{t-j}, \qquad \textit{where} \tag{16}$$

$$w_j = \frac{(1+\theta^2)\phi - (1+\phi^2)}{(1-\theta^2)\phi}\theta^{|j|}, \qquad j = 0, \pm 1, \pm 2, \ldots$$

The expressions for the weights may be modified to take account of a finite sample. More generally the weights for finite samples can be computed numerically from the state space form of any linear model using the algorithm in Koopman and Harvey (2003). There will be a different set of weights for each value of t from $t = 1$ to T, although those in the middle will typically be very close. Figure shows the weights four periods from the end for a random walk plus noise model with signal-noise ratio, q, equal to 0.1.

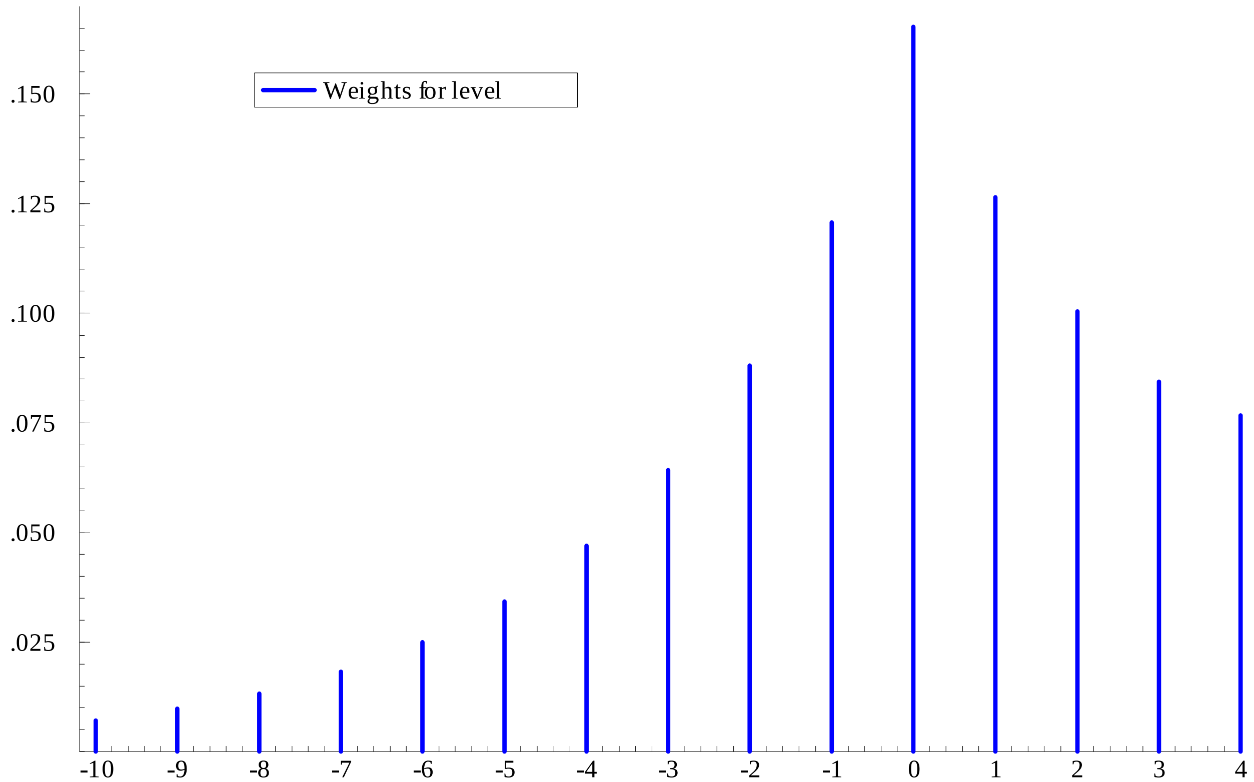


Figure: Smoothing weights at $t = T - 4$ for a random walk plus noise model

Smoothing

Smoothing DCS models is complicated by the fact that the filter is driven by a nonlinear function of the observations that is itself dependent on the output of the filter. The DCS filter for the first-order model can be written as

$$\mu_{t+1|t} = (\phi - \kappa)\mu_{t|t-1} + \kappa y_t(\mu_{t|t-1}), \qquad (17)$$

showing that the weighting of the pseudo-observations,

$$y_t(\mu_{t|t-1}) = u_t + \mu_{t|t-1}, \quad t = 1, ..., T, \qquad (18)$$

is the same as in the KF. For the conditional $t-$distribution,

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

Note that b_t depends on $\mu_{t|t-1}$: hence the notation $y_t(\mu_{t|t-1})$.

Smoothing

Finite sample two-sided smoothing weights can be obtained by computing them for the corresponding UC Gaussian model, that is one with

$$q = \phi(1 + (\phi - \kappa)^2)/(\phi - \kappa) - 1 - (\phi - \kappa)^2, \qquad (19)$$

using the algorithm in Koopman and Harvey (2003). These weights can be applied to *pseudo-observations*

$$y_t(\mu_{t|T}) = (1 - b_{t|T})y_t + b_{t|T}\mu_{t|T}, \quad t = 1, ..., T, \qquad (20)$$

where

$$b_{t|T} = \frac{(y_t - \mu_{t|T})^2/\nu \exp(2\lambda)}{1 + (y_t - \mu_{t|T})^2/\nu \exp(2\lambda)}. \qquad (21)$$

Since the $y_t(\mu_{t|T})'s$ depend on the $\mu'_{t|T}s$, the weights would need to be applied repeatedly until there is convergence.

Smoothing

The state space smoothing recursions, as given in Durbin and Koopman (2012). may be adapted to give a set smoothed values, $\mu_{t|T}$, $t = 1, ..., T$, which are the same as those given by applying the weights from the equivalent UC model to the pseudo-observations $y_t(\mu_{t|T})$. These recursions are also based on repeatedly revising the values of the $b_t's$ to reflect the latest $\mu'_{t|T}s$.

The DCS filter is first run and then followed by a backward smoother (which will usually have time-invariant system matrices) in which the innovations are replaced by the scores, u_t, $t = 1, ..., T$. The forward recursion gives smoothed values of the signal, that is a set of $\mu'_{t|T}s$. The DCS filter is then run again with the score variable evaluated at $b_{t|T}(\mu_{t|T})$, as in (21), that is

$$u_t(\mu_{t|T}) = (1 - b_{t|T})(y_t - \mu_{t|t-1}), \quad t = 1, ..., T. \tag{22}$$

The smoother is then run with the $u_t(\mu_{t|T})'s$ and the whole process of filtering and smoothing repeated until the $\mu'_{t|T}s$ converge.

In the first-order DCS$-t$ model the backward filter becomes

$$r_{t-1} = (\phi - \kappa)r_t + (1 - \kappa/\phi)u_t, \qquad t = T, ..., 2,$$

where u_t is repeatedly updated as $u_t(\mu_{t|T})$ after the first iteration. The forward recursion is either

$$\mu_{t|T} = \mu_{t|t-1} + \kappa(r_t + u_t(\mu_{t|T})/\phi), \qquad t = 1, \ldots T,$$

where the smoothed signal in $u_t(\mu_{t|T})$ is from the previous round, or

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

with q given by (19). Since the estimate of the constant, ω, does not change, neither does the initial value, $\mu_{1|0}$. The same is true in the local level when $\mu_{1|0}$ is treated as a fixed parameter.

Generalization of the above recursive method, or indeed the method based on weighting pseudo-observations, appears not to be straightforward. The difficulty lies in finding a UC model which yields the same filter as the DCS model.

Conditional mode estimation and the score

In an unobserved components model, the distribution of the signal conditional on the observations can be written as the joint PDF of the observations and signal divided the PDF of the observations. Taking logarithms gives

$$\ln p(\boldsymbol{\mu} \mid \mathbf{y}) = \ln p(\boldsymbol{\mu}, \mathbf{y}) - \ln p(\mathbf{y}), \tag{23}$$

and maximizing $\ln p(\boldsymbol{\mu} \mid \mathbf{y})$ with respect to $\boldsymbol{\mu}$ gives the conditional modes of the series of signals. For a linear Gaussian state space model, these modes are the same as the conditional expectations of the signals. Hence they are the smoothed estimates. Note that the second term in (23), that is $\ln p(\mathbf{y})$, can be ignored, and so $\ln p(\boldsymbol{\mu} \mid \mathbf{y})$ may be replaced by the more straightforward expression $\ln p(\mathbf{y}, \boldsymbol{\mu}) = \ln p(\mathbf{y} \mid \boldsymbol{\mu}) + \ln p(\boldsymbol{\mu})$.

Conditional mode estimation and the score

Consider a model in which the PDF of y_t given μ_t is $p(y_t \mid \mu_t)$ and dynamic equation for μ_t is a Gaussian AR(1). Then

$$\ln p(\mathbf{y}, \boldsymbol{\mu}) = -\sum_{t=1}^{T} p(y_t \mid \mu_t) - \frac{1}{2\sigma_\eta^2}\sum_{t=2}^{T}\left(\mu_t - \phi\mu_{t-1}\right)^2 - \frac{1}{2p_{1|0}}\left(\mu_1 - \mu_{1|0}\right)^2.$$

When μ_t is stationary, $\mu_{1|0} = 0$ and $p_{1|0} = \sigma_\eta^2/(1-\phi^2)$. When μ_t is a random walk initialized with a diffuse prior, $p_{1|0} \to \infty$ and the last term disappears.
For a linear Gaussian model $p(y_t \mid \mu_t) = (y_t - \mu_t)^2/\ 2\sigma_\varepsilon^2$. Differentiating $\ln p(\mathbf{y}, \boldsymbol{\mu})$ with respect to each element of $\boldsymbol{\mu}$ then gives a set of equations, which, when set to zero and solved, yield the minimum mean square error estimates of the $\mu_t's$. These smoothed estimates may be computed efficiently by the smoothing recursions given in the last sub-section but one.

Conditional mode estimation and the score

More generally, for any conditional distribution, $p(y_t \mid \mu_t)$, with a continuous first derivative,

$$\frac{\partial \ln p(\boldsymbol{\mu} \mid \mathbf{y})}{\partial \mu_1} = -\frac{\partial p(y_1 \mid \mu_1)}{\partial \mu_1} + \frac{\phi}{\sigma_\eta^2}(\mu_2 - \phi\mu_1) - \frac{\mu_1}{p_{1|0}} \qquad (24)$$

$$\frac{\partial \ln p(\boldsymbol{\mu} \mid \mathbf{y})}{\partial \mu_t} = -\frac{\partial p(y_t \mid \mu_t)}{\partial \mu_t} - \frac{1}{\sigma_\eta^2}(\mu_t - \phi\mu_{t-1}) + \frac{\phi}{\sigma_\eta^2}(\mu_{t+1} - \phi\mu_t), \qquad t=2,\ldots$$

$$\frac{\partial \ln p(\boldsymbol{\mu} \mid \mathbf{y})}{\partial \mu_T} = -\frac{\partial p(y_t \mid \mu_T)}{\partial \mu_T} - \frac{1}{\sigma_\eta^2}(\mu_T - \phi\mu_{T-1}).$$

The conditional modes satisfy the equations obtained by setting these derivatives equal to zero. Durbin and Koopman (2012, pp. 252-3) discuss optimality properties of the conditional modes as estimates of the $\mu_t' s$. The first terms on the right hand side of the equations in (24) are the scores at time t. When evaluated at the conditional modes they are proportional to the $u_t(\mu_{t|T})'s$. Working back to the filter, there is now a rationale for the conditional score, $u_t = u(\mu_{t|t-1})$.

Conditional mode estimation and the score

Summing the equations in (24) when they are evaluated at the conditional modes gives the following result when $\phi = 1$, because the terms involving the first differences of the $\mu_t' s$ cancel each other out.

Proposition

When μ_t is a random walk initialized with a diffuse prior, the scores sum to zero when evaluated at the conditional modes, that is

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

De Rossi and Harvey (JE, 2009) show that this result holds generally for stochastic trends, such as the integrated random walk. The formal requirement is that for the transition matrix in (**??**), the first column of $\mathbf{T} - \mathbf{I}$ consists solely of zeroes.

Forecasting

The one-step ahead predictive distribution is given directly by the model. The concern here is with multi-step prediction.

When $\mu_{t+1|t}$ is of the $QARMA(p, r)$ form the predictor, $\mu_{T+\ell|T}$, is usually best computed recursively, as in an ARMA model. Thus for $\ell = 2, 3, ..,$

$$\mu_{T+\ell|T} = \phi_1 \mu_{T+\ell-1|T} + ... + \phi_p \mu_{T+\ell-p|T} + \kappa_0 u_{T+\ell-1|T} + ... + \kappa_r u_{T+\ell-r|T},$$

where $\mu_{T+j|T}$ is known for $j \leq 1$, $u_{T+j|T}$ is known for $j \leq 0$ and $u_{T+j|T} = 0$ for $j > 0$. A recursion of this form can be used even if $\mu_{t+1|T}$ is nonstationary.

$MSE(\mu_{T+\ell|T})$ is computed in a similar way to an ARMA model

The predictor of the observation at time $T + \ell$ is

$$y_{T+\ell|T} = \mu_{T+\ell|T}, \quad \ell = 1, 2, 3, ...,$$

Provided that $\nu > 2$, $y_{T+\ell|T}$ is the MMSE $\ell -$ *step* ahead forecast of $y_{T+\ell}$ with

$$MSE(y_{T+\ell|T}) = MSE(\mu_{T+\ell|T}) + Var(v_{T+\ell}), \quad \ell = 1, 2, .$$

Forecasting

Multi-step predictions and associated MSEs can be similarly computed using the state space form.

**

A formula for the multi-step conditional *distribution* cannot be found unless the model is Gaussian. However, simulation is a viable option. The error associated with the predictor $y_{T+\ell|t}$ is

$$\sum_{j=1}^{\ell-1} \psi_j u_{T+\ell-j} + v_{T+\ell}, \quad \ell = 2, 3, ...$$

and u_{T+j}, $j = 1, .., \ell - 1$, and $v_{T+\ell}$ can be generated from independent $t-$distributions.

Components and long memory

Fractionally integrated white noise is

$$(1-L)^d y_t = \varepsilon_t, \quad t = 1, ..., T, \tag{25}$$

where d need not be an integer, as it would be for a simple differencing operation. The model may be expressed as an infinite autoregression by expanding the operator as

$$(1-L)^d = 1 - dL - \frac{1}{2}d(1-d)L^2 - ..., \qquad d > -1.$$

Conversely a moving average is obtained from $(1-L)^{-d}$.
The model is stationary if $d < 1/2$, in which case the autocorrelations are

$$\rho(\tau) = \frac{\Gamma(1-d)\Gamma(\tau+d)}{\Gamma(d)\Gamma(\tau+1-d)}, \quad \tau = 0, 1, 2, ...$$

When $d > 0$, the observations exhibit long memory. The ACF decays hyperbolically the ACF of the $AR(1)$ decays exponentially.

Components and long memory

One interpretation of long memory is as an approximation to a mixture of components. Suppose that the location is the sum of two unobserved first-order autoregressions, that is

$$\begin{aligned} y_t &= \omega + \mu_{1,t} + \mu_{2,t} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_i^2) \\ \mu_{i,t} &= \phi_i \mu_{i,t-1} + \eta_{it}, \quad \eta_{it} \sim NID(0, \sigma_i^2), \quad i = 1, 2. \end{aligned} \tag{26}$$

Figure 10 shows ACF of a model in which $\phi_1 = 0.5$ and $\sigma_1^2 = 37.5$ while $\phi_2 = 0.99$ with $\sigma_2^2 = 1$. As can be seen, the ACF of this model is close to that of the long memory model.

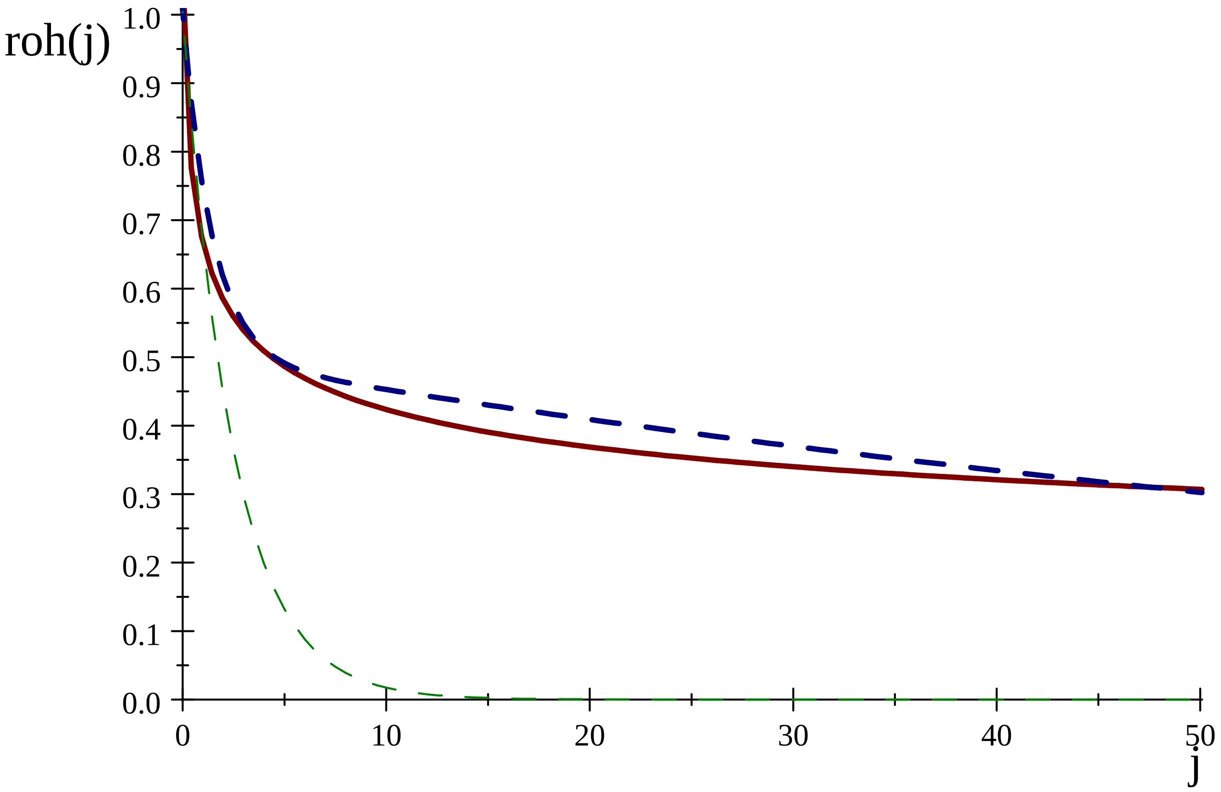


Figure: Long memory (solid line) and two AR(1)'s. Thin dashed line is an AR(1) with same r(1) as long memory.

Components and long memory

The corresponding DCS model is

$$\begin{aligned} y_t &= \omega + \mu_{1,t|t-1} + \mu_{2,t|t-1} + v_t \\ \mu_{i,t+1|t} &= \phi_i \mu_{i,t|t-1} + \kappa_i u_t, \quad i = 1, 2. \end{aligned} \tag{27}$$

where $\phi_1 \neq \phi_2$. Note that u_t appears in both dynamic equations, just as the prediction error does in the innovations form.

Skew distributions

Skewness can be introduced into a t-distribution in such a way that most of the theory set out in this chapter for the DCS location model is unchanged.

In the method of Fernandez and Steel (1998) a standardized probability density function, $f(.)$, which is unimodal and symmetric about zero, is used to construct a skewed probability density function

$$f(\varepsilon_t|\gamma) = \frac{2}{\gamma+\gamma^{-1}}\left[f\left(\frac{\varepsilon_t}{\gamma}\right)I_{[0,\infty)}(\varepsilon_t) + f(\varepsilon_t\gamma)I_{(-\infty,0)}(\varepsilon_t)\right],$$

where $I_{[0,\infty)}(\varepsilon_t)$ is an indicator variable, taking the value one when $\varepsilon_t \geq 0$ and zero otherwise, and γ is a parameter in the range $0 < \gamma < \infty$.

Skew distributions

Symmetry is attained when $\gamma = 1$, whereas $\gamma < 1$ and $\gamma > 1$ produce left and right skewness respectively.

The uncentered moments of ε_t are

$$E(\varepsilon_t^c) = M_c\frac{\gamma^{c+1} + (-1)^c/\gamma^{c+1}}{\gamma+\gamma^{-1}},$$

where

$$M_c = 2\int_0^\infty z^c f(z)dz = E(|z|^c).$$

Hence

$$E(\varepsilon_t) = \mu_\varepsilon = M_1(\gamma - 1/\gamma),$$

which is not zero unless $\gamma = 1$

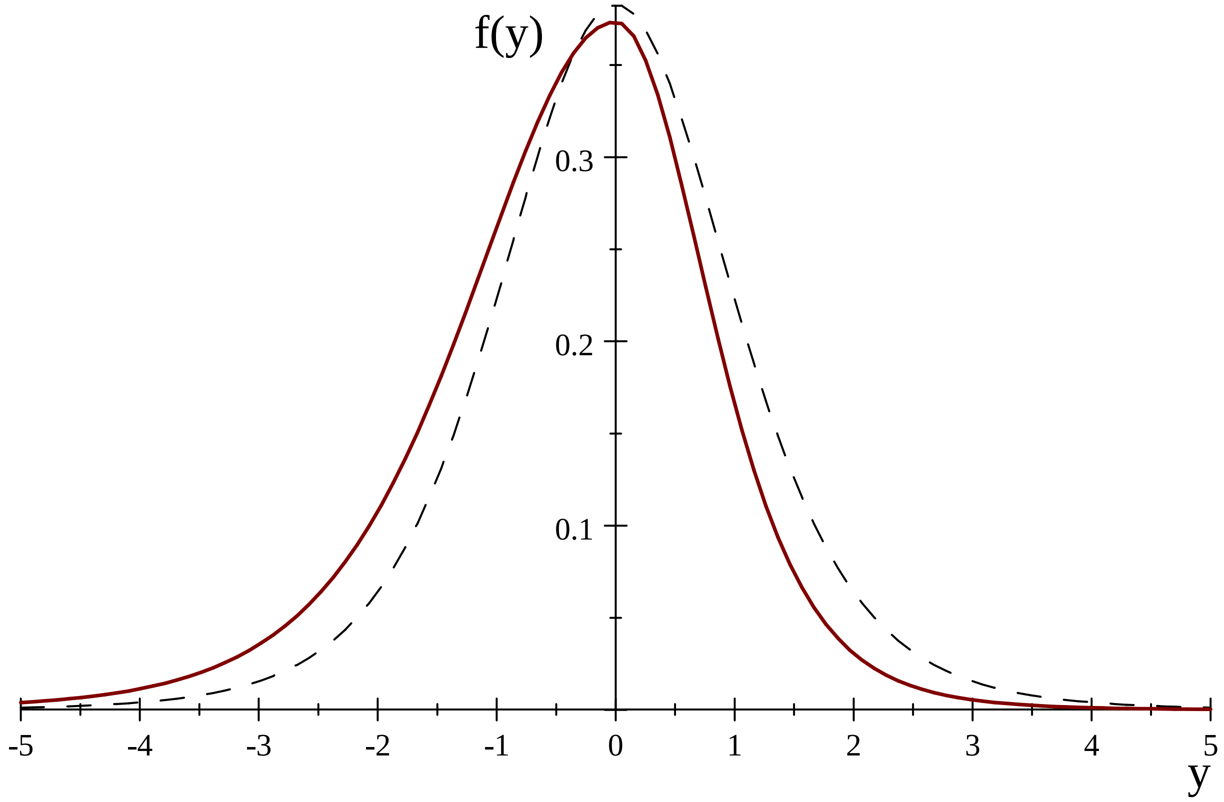


Figure: Skew-t with $\gamma = 0.8$ and normal distribution (dashed)

Skew-t dynamic location model

When the location changes over time, the score is proportional to

$$u_t = u_t^+ I_{[0,\infty)}(y_t - \mu_{t|t-1}) + u_t^- I_{(-\infty,0)}(y_t - \mu_{t|t-1}), \quad t = 1, ..., T, \quad (28)$$

where $u_t = u_t^+$ and $u_t = u_t^-$ are as in $u_t = (1 - b_t)(y_t - \mu_{t|t-1})$, but with b_t defined as

$$b_t^+ = \frac{(y_t - \mu_{t|t-1})^2/\nu \exp(2\lambda)}{1 + (y_t - \mu_{t|t-1})^2/\nu\gamma^2 \exp(2\lambda)} \quad or \quad b_t^- = \frac{(y_t - \mu_{t|t-1})^2/\nu \exp(2\lambda)}{1 + (y_t - \mu_{t|t-1})^2/\nu\gamma^{-2} \exp(2\lambda)}$$

depending on whether $y_t - \mu_{t|t-1}$ is non-negative (b_t^+) or negative (b_t^-). The following result follows because the properties of u_t^+ and u_t^- do not depend on the sign of $y_t - \mu_{t|t-1}$, since they are both linear functions of beta variables with the same distribution.

Skew-t dynamic location model

Corollary

The variables b_t^+ and b_t^- are both distributed as beta$(1/2, \nu/2)$ and u_t, defined in (28), is IID$(0, \sigma_u^2)$.

When γ is known, the information matrix for the skew-t model is exactly as in the symmetric case. The reason is simple: the distribution of the score and its first derivative depend on IID beta variates in exactly the same way as in the symmetric case. When γ is estimated, the asymptotic covariance matrix of the ML estimators of ψ, λ and ν is unaffected as these estimators are independent of the ML estimator of γ.

DCS models for scale: Beta-t-EGARCH

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

where, for example,

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

More generally $\lambda_{t+1|t}$ may be a higher order process of the $QARMA(p, r)$ form. First-order is $(1, 0)$.

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 $m < \nu$. Furthermore, for $\nu > 0$, $\lambda_{t|t-1}$ and $\exp(\lambda_{t|t-1}/2)$ are strictly stationary and ergodic, as is y_t.

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

Beta-t-EGARCH: moments

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

$$
\begin{aligned}
E\left(y_t^m\right) &= E(|\varepsilon_t|^c)e^{m\gamma/2}\prod_{j=1}^{\infty} e^{-\psi_j m/2}\beta_\nu(\psi_j m/2), \qquad m<\nu. \\
&= \frac{\nu^{m/2}\Gamma(\frac{m}{2}+\frac{1}{2})\Gamma(\frac{-m}{2}+\frac{\nu}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})}e^{m\gamma/2}\prod_{j=1}^{\infty} e^{-\psi_j m/2}\beta_\nu(\psi_j m/2)
\end{aligned}
$$

where

$$
\beta_\nu(a) = 1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{1+2r}{\nu+1+2r}\right)\frac{a^k(\nu+1)^k}{k!}, \qquad 0<\nu<\infty.
$$

is Kummer´s (confluent hypergeometric) function.

Beta-t-EGARCH: moments

Proof: The first term in

$$
E\left(|y_t|^c\right) = E\left(|\varepsilon_t|^c\right)E\left(e^{\lambda_{t|t-1}c}\right), \tag{29}
$$

is a property of the t-distribution. The last term depends on a linear combination of independent beta variates so

$$
E\left(e^{\lambda_{t|t-1}m}\right) = e^{m\omega}\prod_{j=1}^{\infty} e^{-\psi_j m}E\left(e^{\psi_j(\nu+1)b_{t-j}m}\right),
$$

and the expectations evaluated by setting $a=\psi_j m$ in $\beta_\nu(a) = E\left(e^{a(\nu+1)b}\right)$, which is a special case of the MGF of a beta variable. *

Beta-t-EGARCH: moments

It follows from Jensen's inequality that

$$E\left(\exp(\lambda_{t|t-1}m)\right) \geq \exp(E(\lambda_{t|t-1}m) = \exp(m\omega).$$

Thus the expected value of a time-varying scale, $\exp(\lambda_{t|t-1})$, is greater than $\exp(\omega)$. On comparing each term in the power series expansion of e^{-a} with the corresponding term in $\beta_\nu(a)$, it can be seen that $e^{-\psi_j m}\beta_\nu(\psi_j m) \geq 1$ for finite ψ_j, with the equality holding when $\psi_j = 0$. For values of ψ_j likely to arise in practice, $e^{-\psi_j m}\beta_\nu(\psi_j m)$ is close to one. Note also that $\beta_\nu(-a) < \beta_\nu(a)$ for $a \neq 0$.

Beta-t-EGARCH: moments

The serial correlation in scale means that, by Jensen's inequality, the kurtosis in y_t exceeds that of ε_t. For a $t_\nu-$ distribution, the kurtosis is $\kappa_\nu = 3(\nu-2)/(\nu-4)$, $\nu > 4$, and the kurtosis of y_t is given by $\kappa_\nu K_\nu$, where K_ν is obtained as follows.
The factor by which the kurtosis increases in the stationary Beta-t-EGARCH model is

$$K_\nu = \frac{E\left(e^{4\lambda_{t|t-1}}\right)}{\left[E\left(e^{2\lambda_{t|t-1}}\right)\right]^2} = \left(\prod_{j=1}^{\infty}\beta_\nu(2\psi_j)\right)^{-2}\prod_{j=1}^{\infty}\beta_\nu(4\psi_j), \qquad \nu > 4.$$

Beta-t-EGARCH: Autocorrelation functions of powers of absolute values

The autocorrelations of the squared observations are given by analytic expressions for GARCH models.
But the ACFs can of Beta-t-EGARCH *be computed for the absolute observations raised to any positive power;* see Harvey and Chakravary (2009).
Heavy-tails tend to weaken the autocorrelations.

Forecasts

The standard EGARCH model readily delivers the optimal $\ell-$step ahead forecast - in the sense of minimizing the mean square error - of future logarithmic conditional variance. Unfortunately, as Andersen *et al* (2006, p 804-5, p 810-11) observe, the optimal forecast of the conditional variance, that is $E_T(\sigma^2_{T+\ell|T+\ell-1})$, where E_T denotes the expectation based on information at time T, generally depends on the entire $\ell-$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 analytic expressions for the conditional scale and variance can be obtained in the same way as expressions were obtained for higher order moments.
Expressions for $\ell - step$ ahead volatility and volatility of volatility.

Forecasts

Full predictive distribution needs to be simulated - Needed for VaR and expected shortfall.In the Beta-t-EGARCH model, the distribution of $y_{T+\ell}$, $\ell = 2, 3, \ldots.$, conditional on the information at time T, is the distribution of

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

An analogous expression can be written down for Gamma-GED-EGARCH. Hence it is not be difficult to simulate the multi-step predictive distribution of the scale and observations; see the discussion in Andersen *et al* (2006, pp. 810-811). The term in square brackets is made up of $\ell - 1$ independent beta variates and this variable can be combined with a draw from a t-distribution, $\varepsilon_{T+\ell}$. The composite variable so obtained depends only on the parameters that determine the ψ_j's. Multiplying by $\exp(\lambda_{T+\ell \mid T})$ gives $y_{T+\ell}$.
The tails of the predictive distribution of $y_{T+\ell}$ become heavier as ℓ increases.

Asymptotic theory for Beta-t-EGARCH

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.

Asymptotic theory for Beta-t-EGARCH

Proposition

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

$$\begin{aligned} a &= \phi - \kappa\frac{2\nu}{\nu+3} \\ b &= \phi^2 - 4\phi\kappa\frac{\nu}{\nu+3} + \kappa^2\frac{12\nu(\nu+1)}{(\nu+5)(\nu+3)} \\ c &= \kappa\frac{4\nu(1-\nu)}{(\nu+5)(\nu+3)}. \end{aligned}$$

Asymptotic theory for Beta-t-EGARCH

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

where $\mathbf{D}(\boldsymbol{\psi})$ was defined in Lecture 1, as (), and where

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

with $\psi'(.)$ being the trigamma function; see, for example, Taylor and Verblya (2004) and Lin and Wang (2009).

Parameter		(a) ML estimates for T=1000					
ϕ	κ	RMSE(ϕ)	ASE(ϕ)	RMSE(κ)	ASE(κ)	RMSE(ω)	ASE(ω)
0.90	0.05	0.075	0.052	0.016	0.016	0.053	0.049
	0.10	0.038	0.032	0.018	0.017	0.065	0.069
0.95	0.05	0.058	0.024	0.014	0.013	0.069	0.069
	0.10	0.019	0.017	0.016	0.015	0.098	0.109
0.99	0.05	0.010	0.006	0.010	0.010	0.198	0.226
	0.10	0.008	0.005	0.013	0.013	0.312	0.428

Table 4.1 Monte Carlo results based on 1000 replications for first-order Beta-t-EGARCH with $\nu = 6$.

Asymptotic theory for Beta-t-EGARCH: unit root

When ϕ is taken to be unity in the first-order model,

$$\lambda_{t+1|t} = \delta + \lambda_{t|t-1} + \kappa u_t, \quad t = 1, ..., T.$$

When $\lambda_{1|0}$ is fixed and known, it follows from the general Proposition that, provided that $b < 1$, the limiting distribution of $\sqrt{T}(\widetilde{\kappa} - \kappa, \widetilde{\delta} - \delta)'$ is multivariate normal with mean zero and covariance matrix $\mathbf{I}^{-1}(\widetilde{\kappa}, \widetilde{\delta})$
When δ is set to zero,

$$I(\widetilde{\kappa}) = \frac{\sigma_u^4}{2\kappa\sigma_u^2 - \kappa^2 E\left[(\partial u_t/\partial\lambda)^2\right]}.$$

For small κ, $I(\widetilde{\kappa}) \simeq \sigma_u^2/2\kappa$. Thus for a $t_\nu-$distribution

$$SE(\widetilde{\kappa}) \simeq \sqrt{\kappa(\nu+3)/\nu T}$$

When the initial value, $\lambda_{1|0}$, is treated as parameter, ω, to be estimated, it appears from the simulation evidence in Table 4.2 that the asymptotic distribution of the ML estimator of κ is unchanged. The approximate asymptotic standard errors for $\kappa = 0.05$ and 0.10 are 0.00274 and 0.00387 respectively and these are almost exactly the same as the figures in Table 4.2.

Parameter		Mean and SD for T=10,000			
ω	κ	Mean(ω)	SD(ω)	Mean(κ)	SD(κ)
0	0.05	0.014	0.309	0.050	0.0027
0	0.10	0.011	0.435	0.100	0.0038

Table 4.2 Monte Carlo results for Beta-t-EGARCH with ϕ known to be 1 and ν known to be 6.

If ϕ is estimated unrestrictedly, it will have a non-standard distribution. (A reasonable conjecture is that the limiting distribution of $T\widetilde{\phi}$ can be expressed in terms of functionals of Brownian motion, as is the case when a series is a random walk and observations are regressed on their lagged values.) The simulations reported in Table 4.3, where ω, ϕ and κ are all unknown parameters, indicate that the distribution of $\widetilde{\kappa}$ is unchanged, which is to be expected since, unlike $\widetilde{\phi}$, $\widetilde{\kappa}$ is not superconsistent. (The parameter ω is not estimated consistently but this should not affect the asymptotic distribution of $\widetilde{\phi}$ and $\widetilde{\kappa}$.)

Parameter		Mean and SD for T=10,000					
ω	κ	Mean(ω)	SD(ω)	Mean(κ)	SD(κ)	Mean(ϕ)	SD(ϕ)
0	0.05	0.012	0.313	0.050	0.0027	1.000	0.00033
0	0.10	0.020	0.435	0.100	0.0038	1.000	0.00031

Table 4.3 Monte Carlo results for Beta-t-EGARCH with $\phi = 1$, but estimated, and ν known to be 6.

Gamma-GED-EGARCH

The log-likelihood function for a Gamma-GED-EGARCH model is

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

The distribution of the score

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

is gamma, as is that of

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

Gamma-GED-EGARCH

For the stationary first-order Gamma-GED-EGARCH model define

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

For a given value of υ and provided that $b < 1$, $\widetilde{\boldsymbol{\psi}}$ is consistent and asymptotically normal.

Gamma-GED-EGARCH

For a Gaussian conditional distribution, $\upsilon = 2$ and so $b = \phi^2 - 4\phi\kappa + 12\kappa^2$ and $a = \phi - 2\kappa$. Hence $b > a^2$, whereas for the Gaussian location model $b = a^2$. These expressions for a and b are also given by letting $\nu \to \infty$ in for the t distibution.
For the Laplace , $\upsilon = 1$ and $b = \phi^2 - 2\phi\kappa + 2\kappa^2$, which, perhaps surprisingly, permits a wider range for κ than does the normal, even though (or perhaps because) the Laplace distribution has heavier tails.

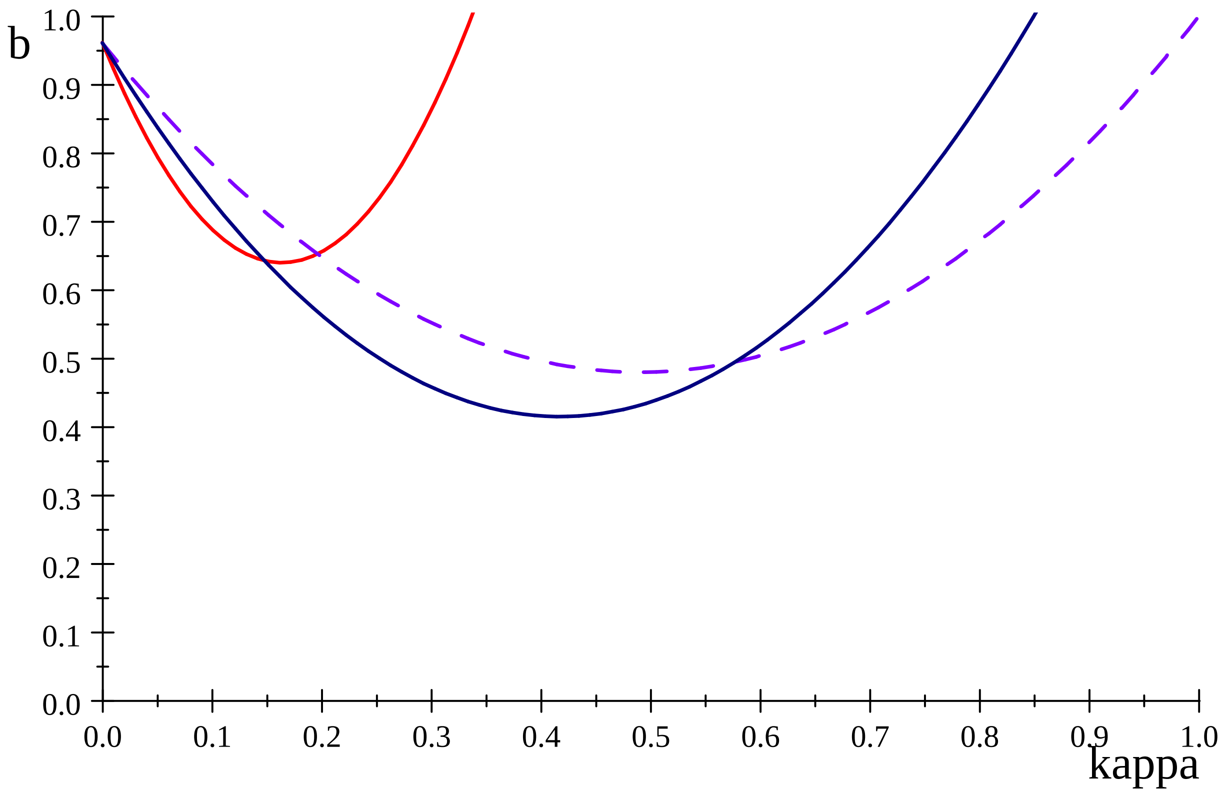


Figure: b against κ for $\phi = 0.98$ and (i) $t-distribution$ with $\nu = 6$ (solid), (ii) normal (upper line), (iii) Laplace (thick dash).

Leverage effects

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 $I(y_t < 0)$.

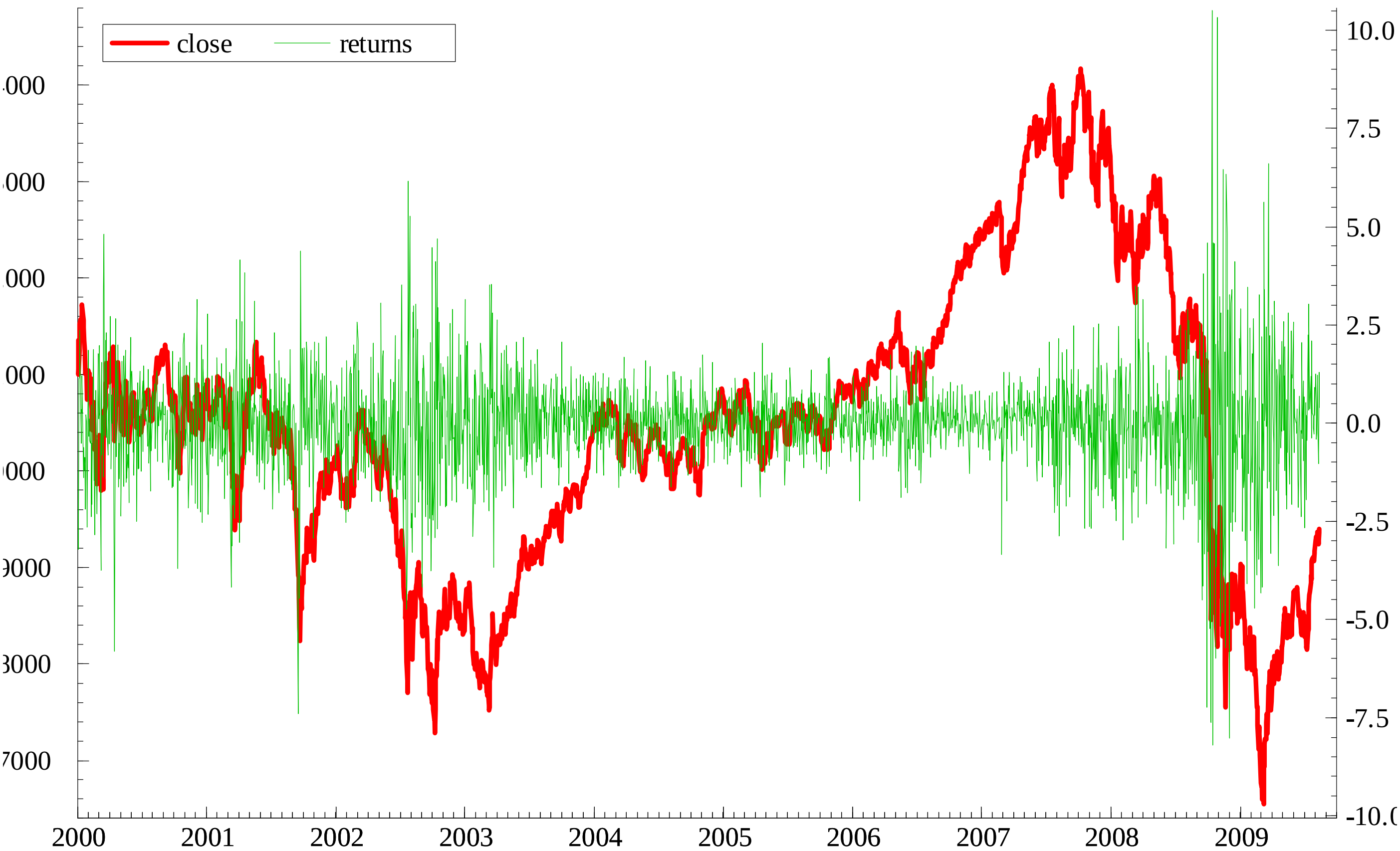
Alternatively, the **sign** of the observation may be used, so

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

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.



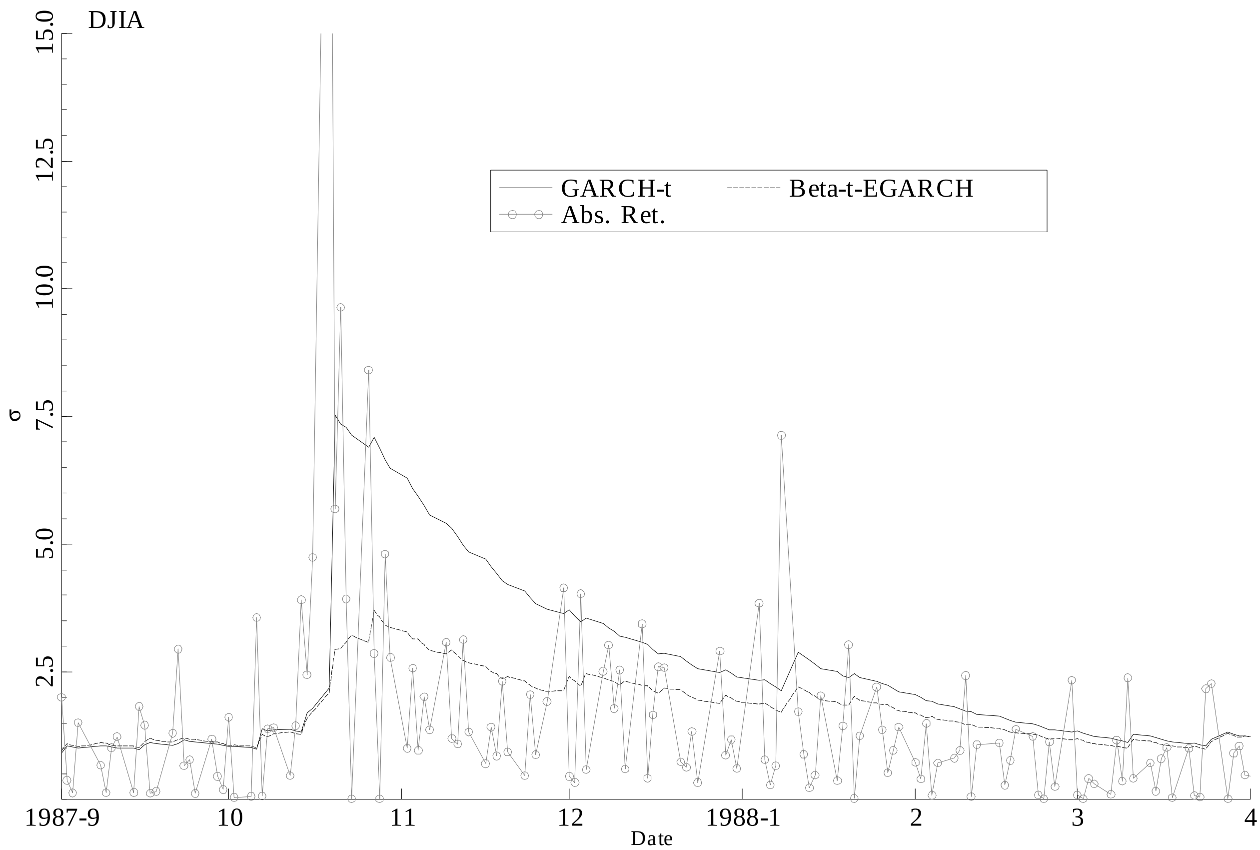
Application of Beta-t-EGARCH to Hang Seng and Dow-Jones

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.

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

Estimates with numerical and asymptotic standard errors

Numerical and ASEs are the similar. $b < 1$.
Graph- Dow-Jones absolute (de-meaned) returns around the great crash of October 1987, together with estimated conditional standard deviations for Beta-t-EGARCH and GARCH-t, both with leverage.



Explanatory variables for volatility

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

Asymptotic theory with explanatory variables

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 $\boldsymbol{\beta}$ are asymptotically independent of the estimators of ψ and ν.

Components

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

$$\begin{aligned} \lambda_{1,t|t-1} &= \phi_1 \lambda_{t-1|t-2} + \kappa_1 u_{t-1} \\ \lambda_{2,t|t-1} &= \phi_2 \lambda_{t-1|t-2} + \kappa_2 u_{t-1} \end{aligned}$$

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.

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

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

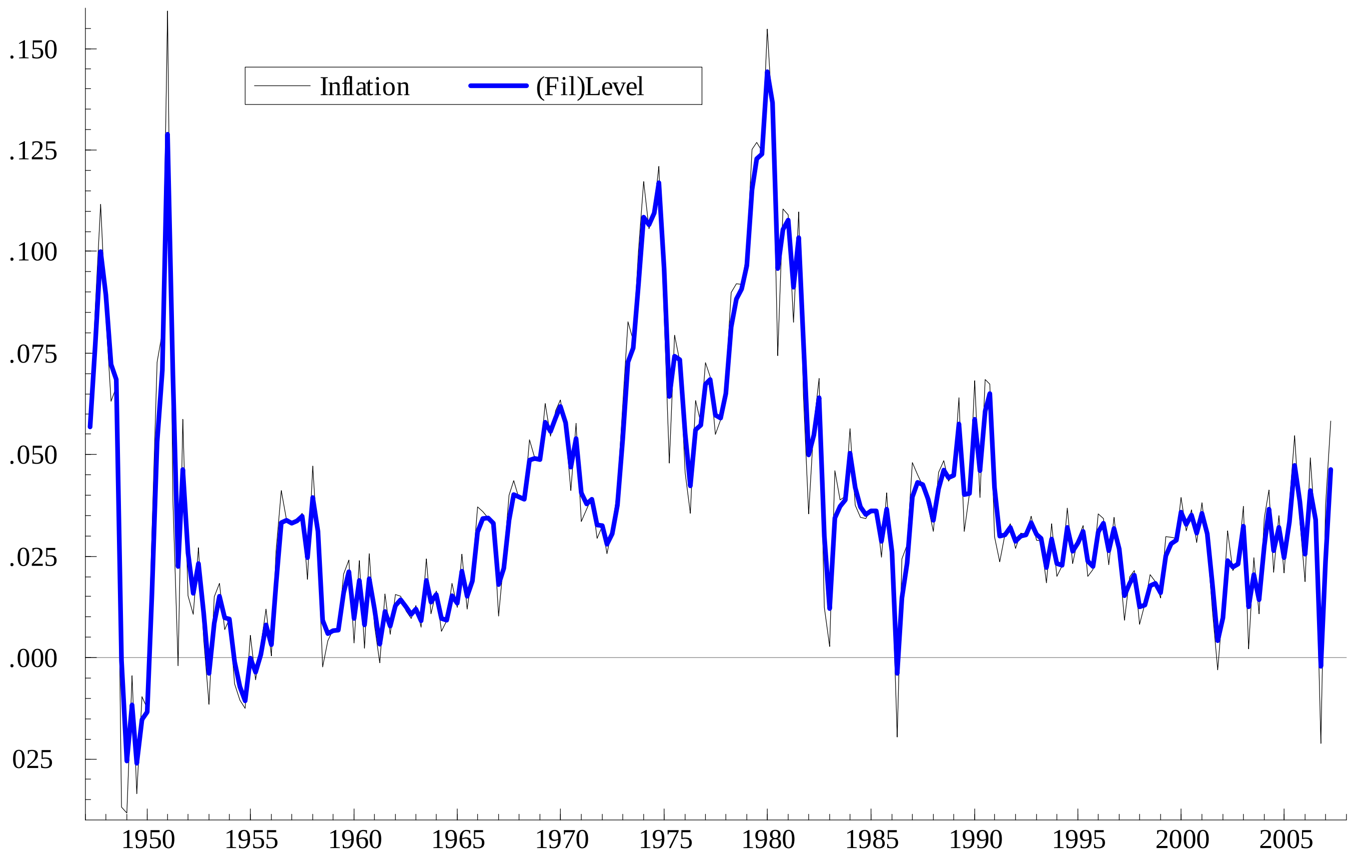
$$u_{\lambda t} = \frac{(\nu + 1)(y_t - \mu_{t|t-1})^2}{\nu \exp(2\lambda_{t|t-1}) + (y_t - \mu_{t|t-1})^2} - 1$$

Stochastic location and stochastic scale: US inflation

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.



Stochastic location and stochastic scale: US inflation

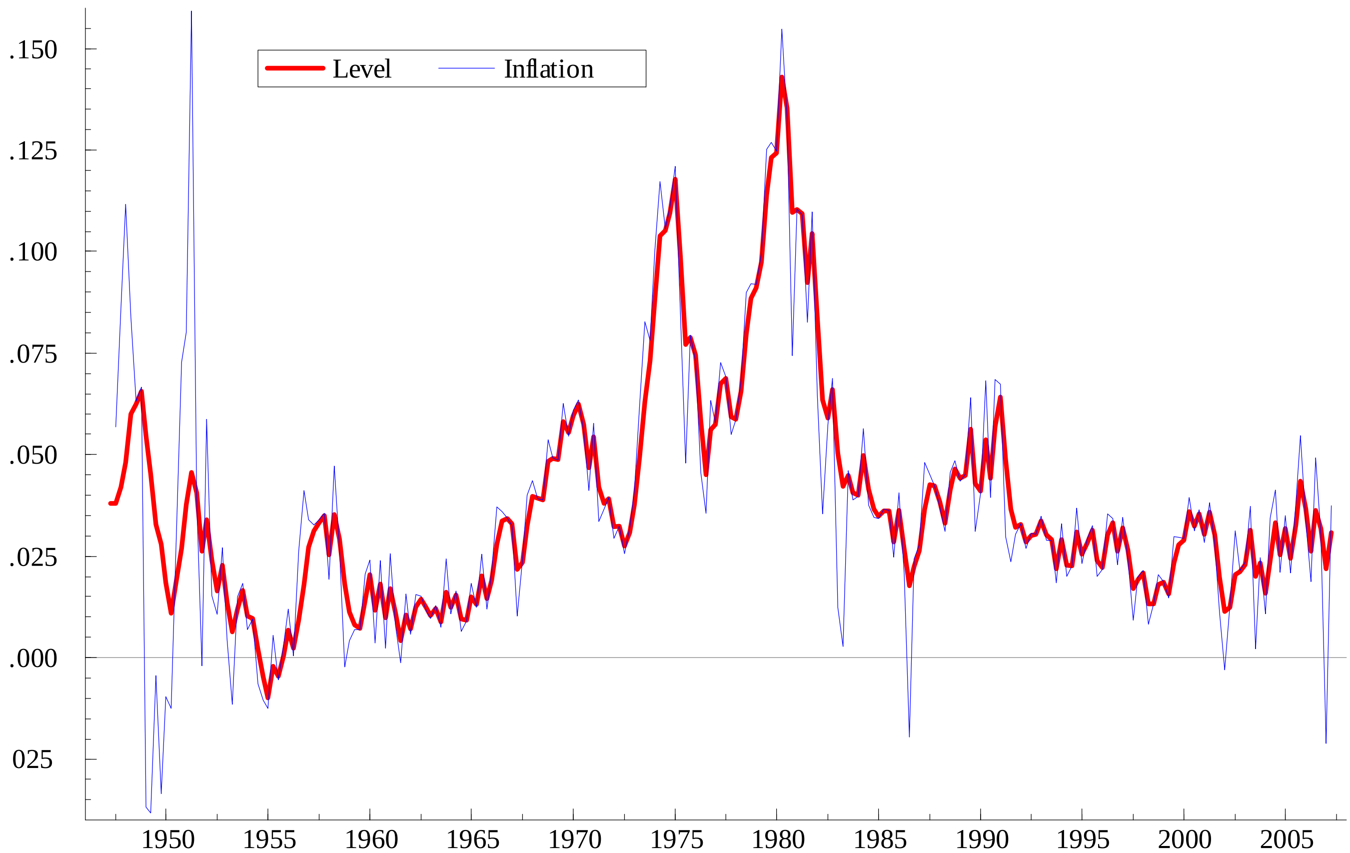
ML estimates (with standard errors in parentheses):
for location

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

for scale

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

and $\quad \widetilde{\nu} = 11.71(4.58).$
The filtered estimates respond less to extreme values than those from the homoscedastic Gaussian model.



Andrew Harvey (Cambridge University) Volatility and Heavy Tails / 107

Tests

A Lagrange multiplier test for changing volatility in a Beta-t-EGARCH model can be carried out with the $Q_u(P)$ statistic. The variables used to construct the sample autocorrelations, $r_u(j)$, $j = 1, 2, ...$, will be written as

$$\widetilde{y}_t(\lambda) = \frac{(\widetilde{\nu}+1)(y_t - \widetilde{\mu})^2}{\widetilde{\nu}\exp(2\widetilde{\lambda}) + (y_t - \widetilde{\mu})^2} - 1, \qquad t = 1, ..., T,$$

where $\widetilde{\mu}$, $\widetilde{\lambda}$ and $\widetilde{\nu}$ are the ML estimators of the location, scale parameter and the degrees of freedom in the $t-$distribution. If the location is modeled as a function of exogenous explanatory variables, the distribution of the test statistic is not affected. The ML estimators, $\widetilde{\mu}$, $\widetilde{\lambda}$ and $\widetilde{\nu}$, are nonlinear, but they can be computed by an iterative procedure, such as the method of scoring. A test statistic constructed from the score variables will be more resistant to outliers than the conventional statistic constructed from the squares (obtained as $\nu \to \infty$).
Martingale difference test. Modify as in Lobato *et al* (2001).

Skew-t

Harvey and Sucarrat (CWPE, 2012) report results of fitting a $Beta - skew - t - EGARCH$ model with two components and leverage.

Analytic results on moments. preditions and asymptotic theory carry over to this case.

Conclusions on volatility models

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.*

Conclusions on volatility models

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.

Conclusions on volatility models

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.

Conclusions on volatility models

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.

Conclusions on volatility models

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 *DCS* models includes changing location and changing scale/location in models for non-negatitve variables.