Dynamic Conditional Score (DCS) Models: Volatility, EGARCH-M, Changing Correlation and Heavy Tails

Andrew Harvey (ach34@cam.ac.uk)

Faculty of Economics, University of Cambridge

March 2017

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.
5. Extensions to multivariate time series. Correlation or association may change over time. Time-varying copulas.
6. Changing coefficients in ARs and changing signal-noise ratios. Second-order dynamics.
***
Harvey, A.C. **Dynamic models for volatility and heavy tails**. CUP 2013
http://www.econ.cam.ac.uk/DCS
Creal et al (2011, JBES, 2013, JAE)..

Introduction to dynamic conditional score (DCS) models

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 $\xi_t's$ to have a heavy-tailed distribution does not deal with the problem as a large observation becomes incorporated into the level and takes time to work through the system.
An ARMA models with a heavy-tailed distribution is designed to handle *innovations outliers*, as opposed to *additive outliers*. See the **robustness** literature.
But a *model-based approach* is not only simpler than the usual robust methods, but is also more amenable to diagnostic checking and generalization.

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

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

where κ is an unknown parameter.

Dynamic location model

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

where ε_t is serially independent, standard t-variate and

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

where $v_t = y_t - \mu_{t|t-1}$ is the prediction error and $\varphi = \exp(\lambda)$ is the (time-invariant) scale.
$u_t \longrightarrow 0$ as $|y| \longrightarrow \infty$. In the robustness literature this is called a redescending M-estimator. It is a gentle form of *trimming*.

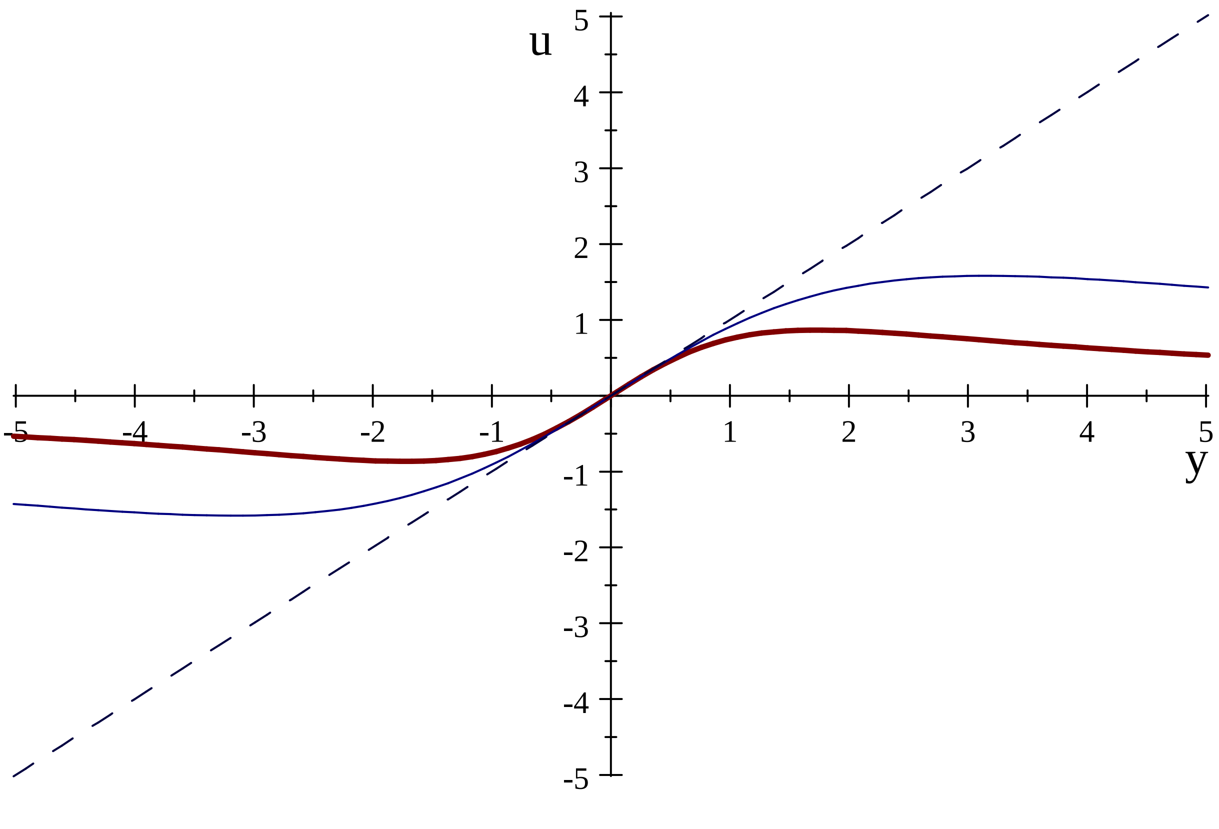


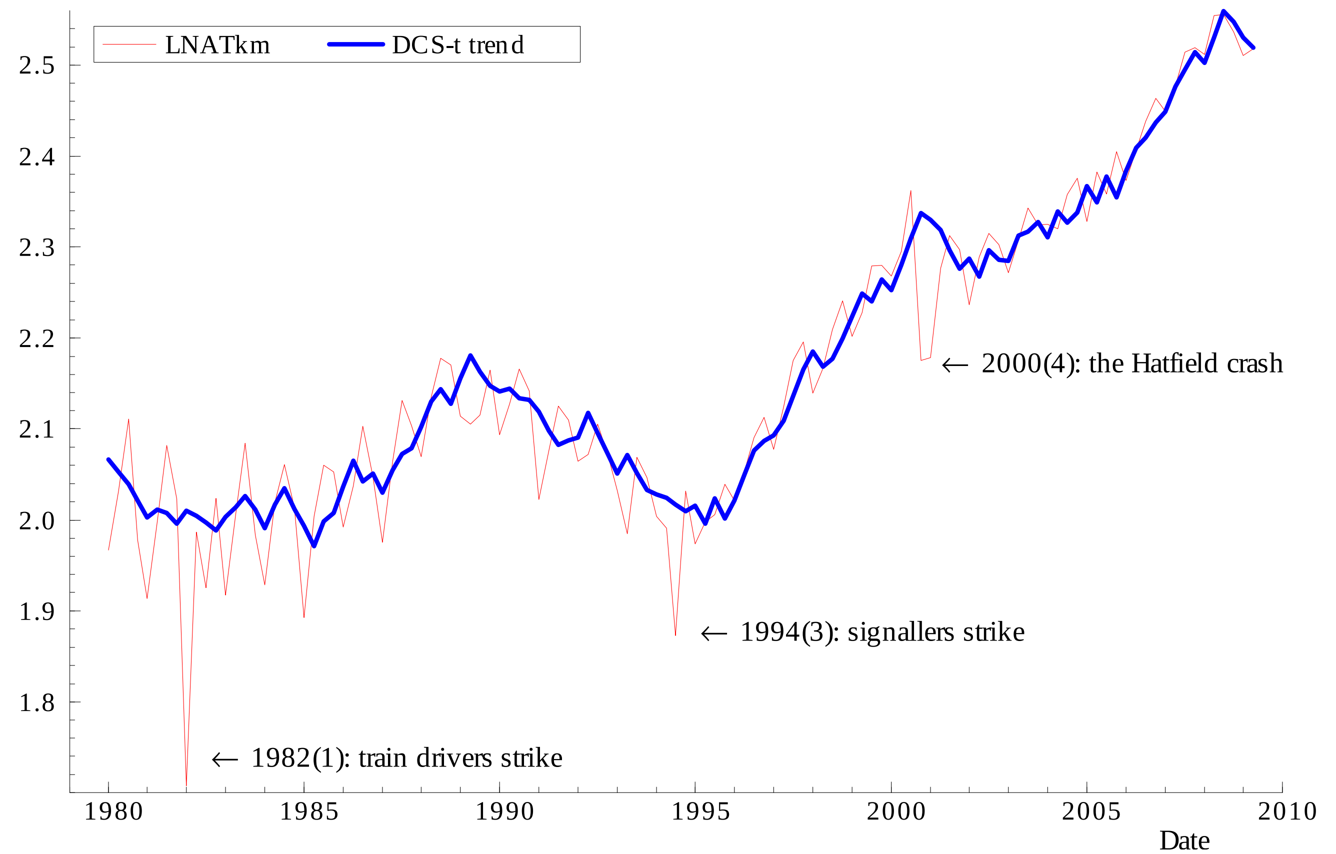
Figure: Impact of u_t for t_ν (with a scale of one) for $\nu = 3$ (thick), $\nu = 10$ (thin) and $\nu = \infty$ (dashed).

Seasonally adjusted monthly data on U.S. Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing (AWHMAN) from February 1992 to May 2010. DCS-t model

$$y_t = \mu_{t|t-1} + \exp(\lambda)\varepsilon_t$$

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

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



Scale: GARCH

The first-order model, $GARCH\,(1,1)-t$, is

$$y_t = \sigma_{t|t-1}\varepsilon_t, \quad \varepsilon_t \sim t_\nu$$

and

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

The conditions on α and β ensure that the variance remains positive. Can write

$$\sigma^2_{t+1|t} = \delta + \phi\sigma^2_{t|t-1} + \alpha(y_t^2 - \sigma^2_{t|t-1})$$

with $\phi = \alpha + \beta$.
Heavy tails - instantaneous volatility.

Scale: DCS Volatility Models

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

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

by another MD

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

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

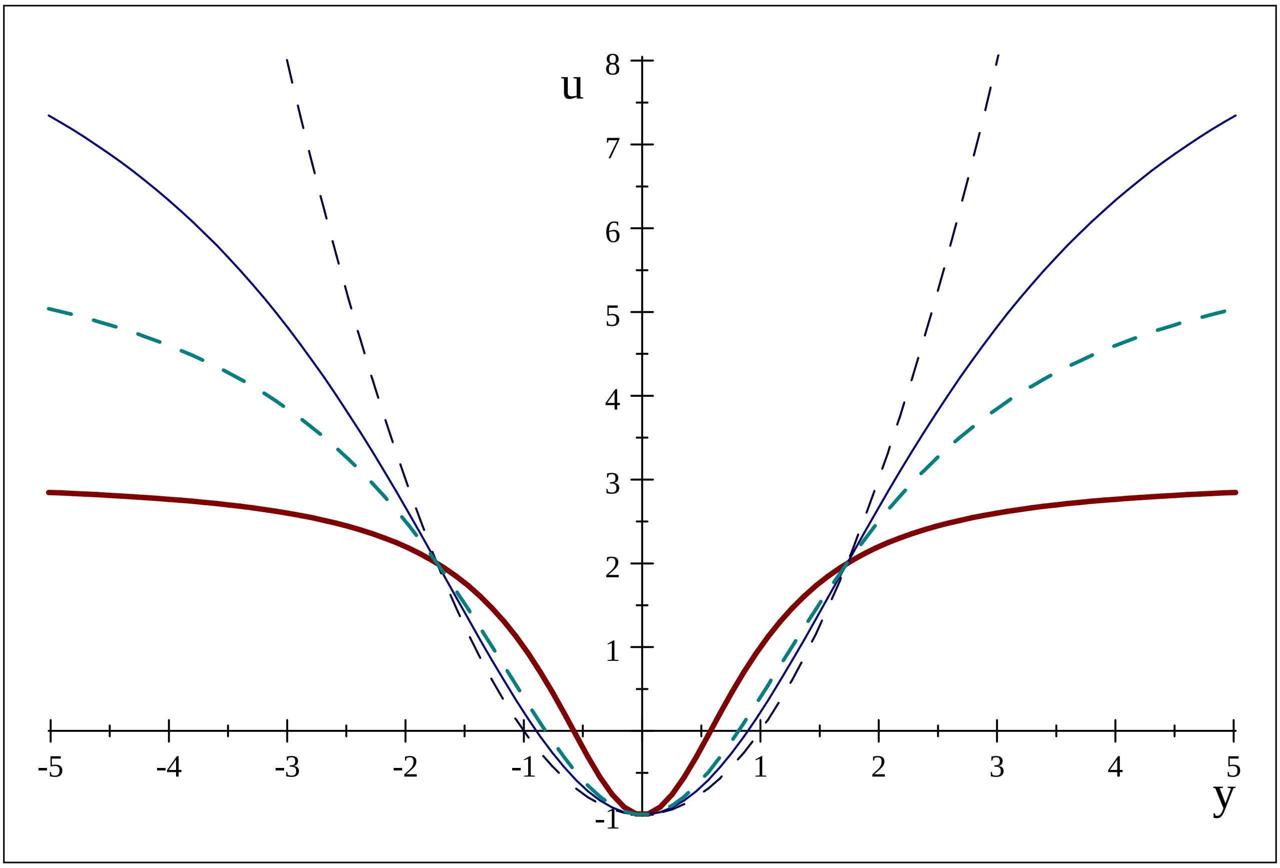


Figure: Impact of u_t for t_ν with $\nu = 3$ (thick), $\nu = 6$ (medium dashed) $\nu = 10$ (thin) and $\nu = \infty$ (dashed).

Exponential DCS Volatility Models: Beta-t-EGARCH

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

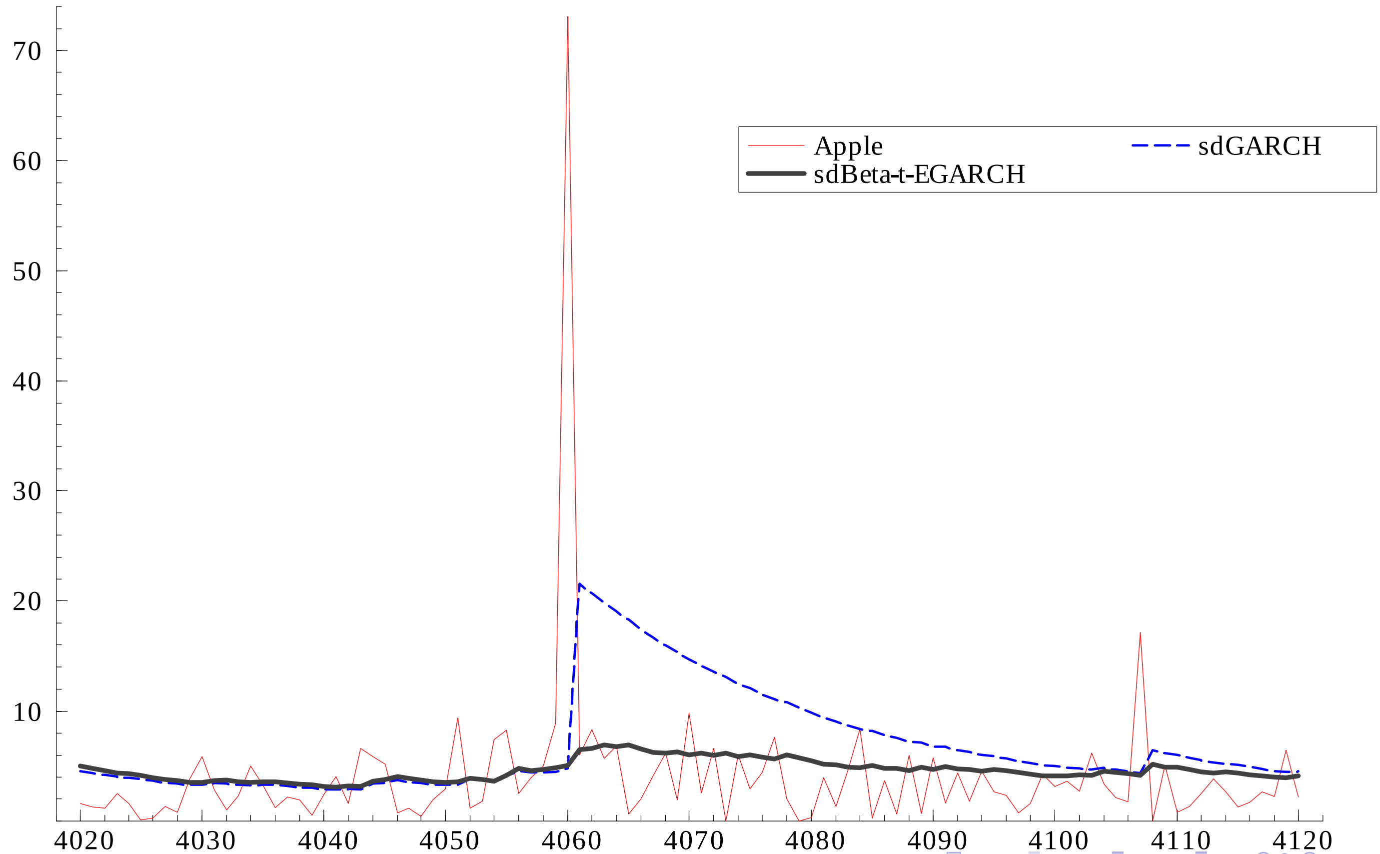
where the serially independent, zero mean variable ε_t has a $t_\nu-$distribution with degrees of freedom, $\nu > 0$, and the dynamic equation for the log of scale is

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

The conditional score is

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

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



Beta-t-EGARCH

The variable u_t may be expressed as

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

where

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

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 $|y_t|^c$.

Gamma-GED-EGARCH

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

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

where υ is a shape parameter.
The score

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

is **gamma** distributed.

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

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

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

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

where α is the scale parameter, ν, ξ and ς are shape parameters and $B(\xi, \varsigma)$ is the beta function; see Kleiber and Kotz (2003, ch6).
The GB2 distribution contains many important distributions as special cases, including the Burr ($\xi = 1$) and log-logistic ($\xi = 1, \varsigma = 1$).
Furthermore Generalized Gamma (includes Weibull as well as gamma) is a special limiting case.
GB2 distributions are **fat tailed** for finite ξ and ς with upper and lower tail indices of $\eta = \varsigma\nu$ and $\overline{\eta} = \xi\nu$ respectively.
The absolute value of a t_f variate is GB2$(\varphi, 2, 1/2, f/2)$ with tail index is $\eta = \overline{\eta} = f$.
Score is *bounded*.

A unified theory for volatility: EGARCH and Generalized Student-t distribution

The flexibility of the generalized-t enables it to capture a variety of shapes at the peak of the distribution as well as in the tails. Student-t and GED are special cases. (NB. Absolute value of gen-t is GB2)
The flexibility goes a long way towards meeting the objection that parametric models are too restrictive and hence vulnerable to misspecification. McDonald and Newey (Econometrica, 1987) argued that the flexibility of the generalized-t model made it 'partially adaptive'. They highlighted the robustness of the generalized-t assumption for models of location (or more generally static regression).
Extends to handle skewness and/or asymmetry.
Harvey and Lange (2016, JTSA). Volatility Modeling with a Generalized t-distribution.

Generalized Student-t distribution*

Our preferred parameterization sets one of the shape parameters equal to the tail index and so

$$f(y) = \frac{1}{\varphi} K(\eta, \upsilon) \left(1 + \frac{1}{\eta}\left|\frac{y-\mu}{\varphi}\right|^{\upsilon}\right)^{-(\eta+1)/\upsilon}, \quad -\infty < y < \infty,$$

where φ is a scale parameter, υ and η are positive shape parameters and

$$K(\eta, \upsilon) = \frac{\upsilon}{2\eta^{1/\upsilon}} \frac{1}{B(\eta/\upsilon, 1/\upsilon)}$$

with $B(.,.)$ denoting the beta function.
The range of η is $0 < \eta \leq \infty$, so if we define $\overline{\eta} = 1/\eta$, then $0 \leq \overline{\eta} \leq 1$ for $1 \leq \eta \leq \infty$. The range of υ is $0 < \upsilon < \infty$ but in practice $\upsilon < 1$ is unlikely.
The distribution has fat tails for finite η and so moments only exist up to, but not including, η.

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.

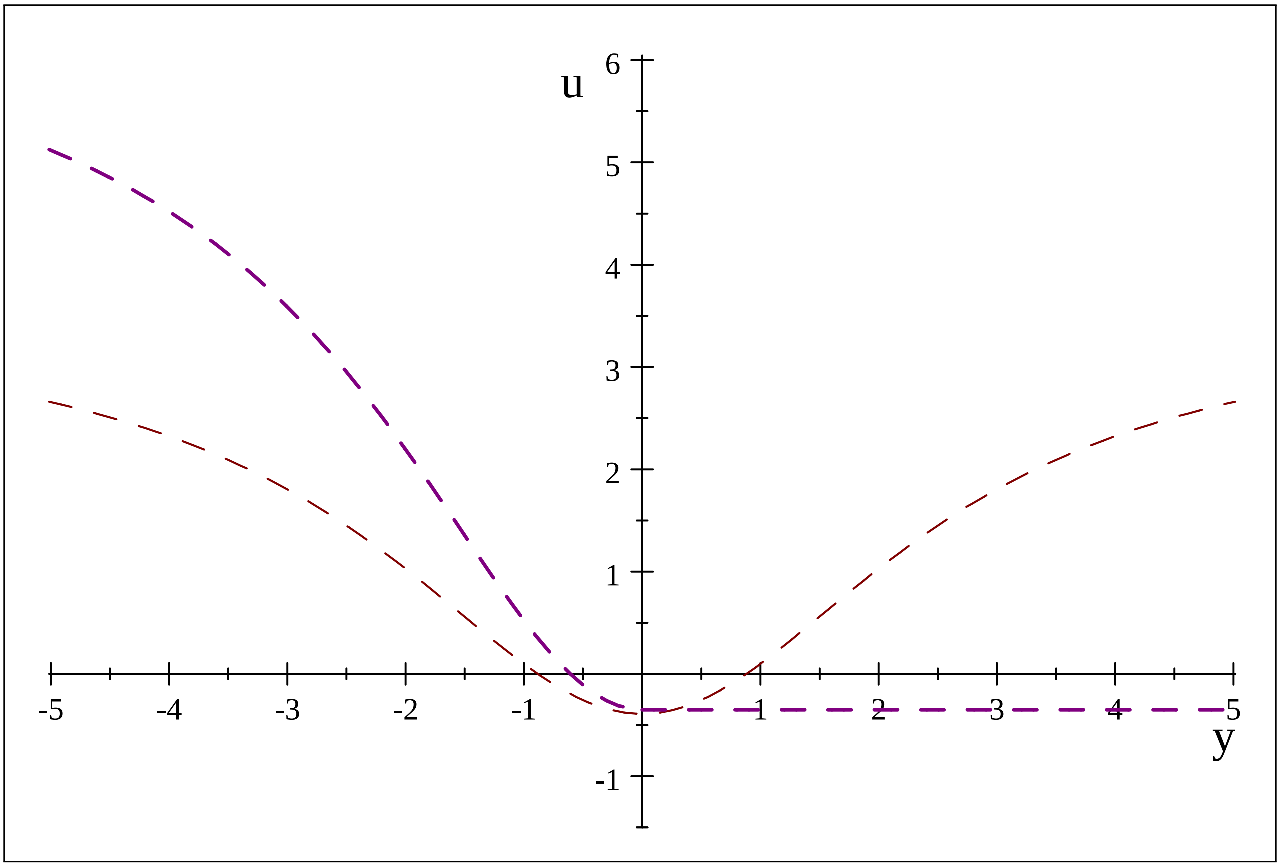


Figure: Impact of u for t_9. Purple is $\kappa = \kappa^*$

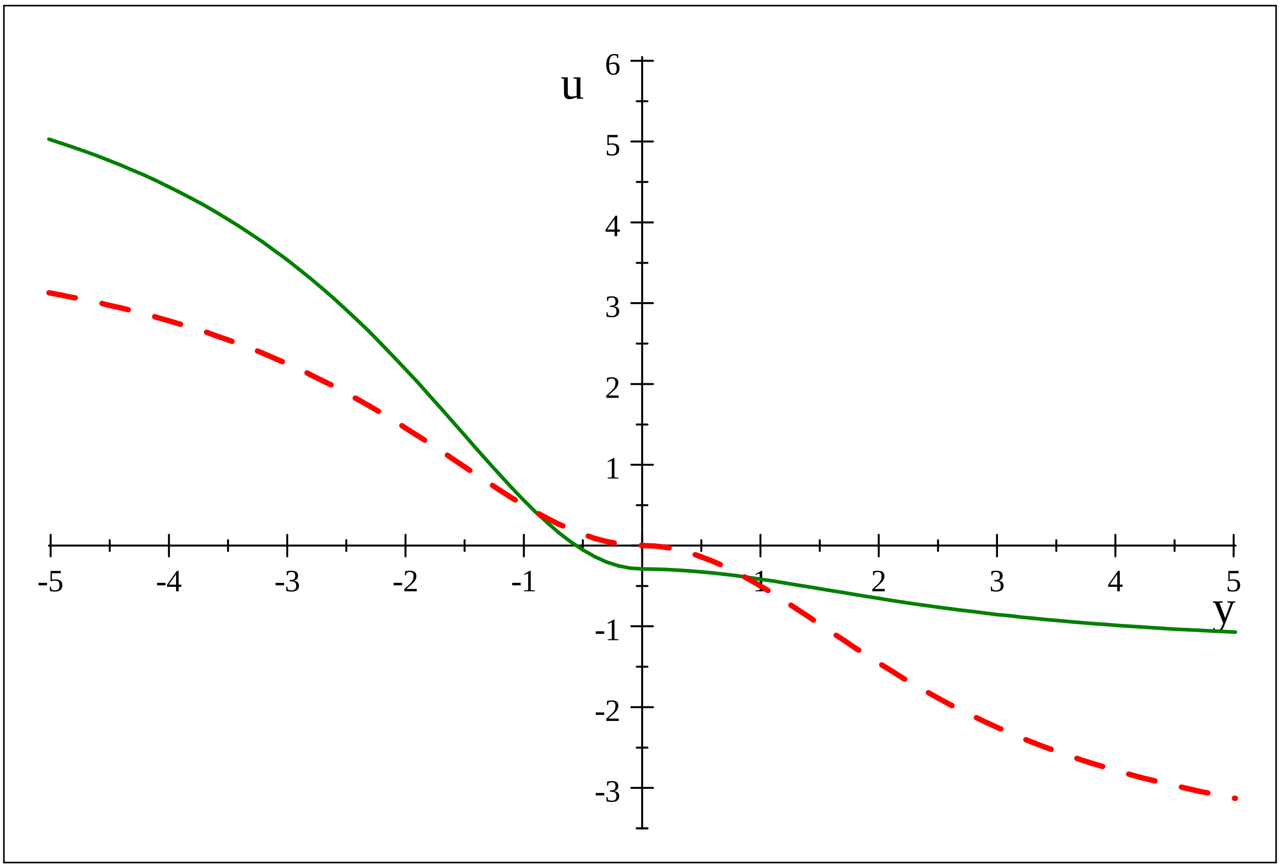
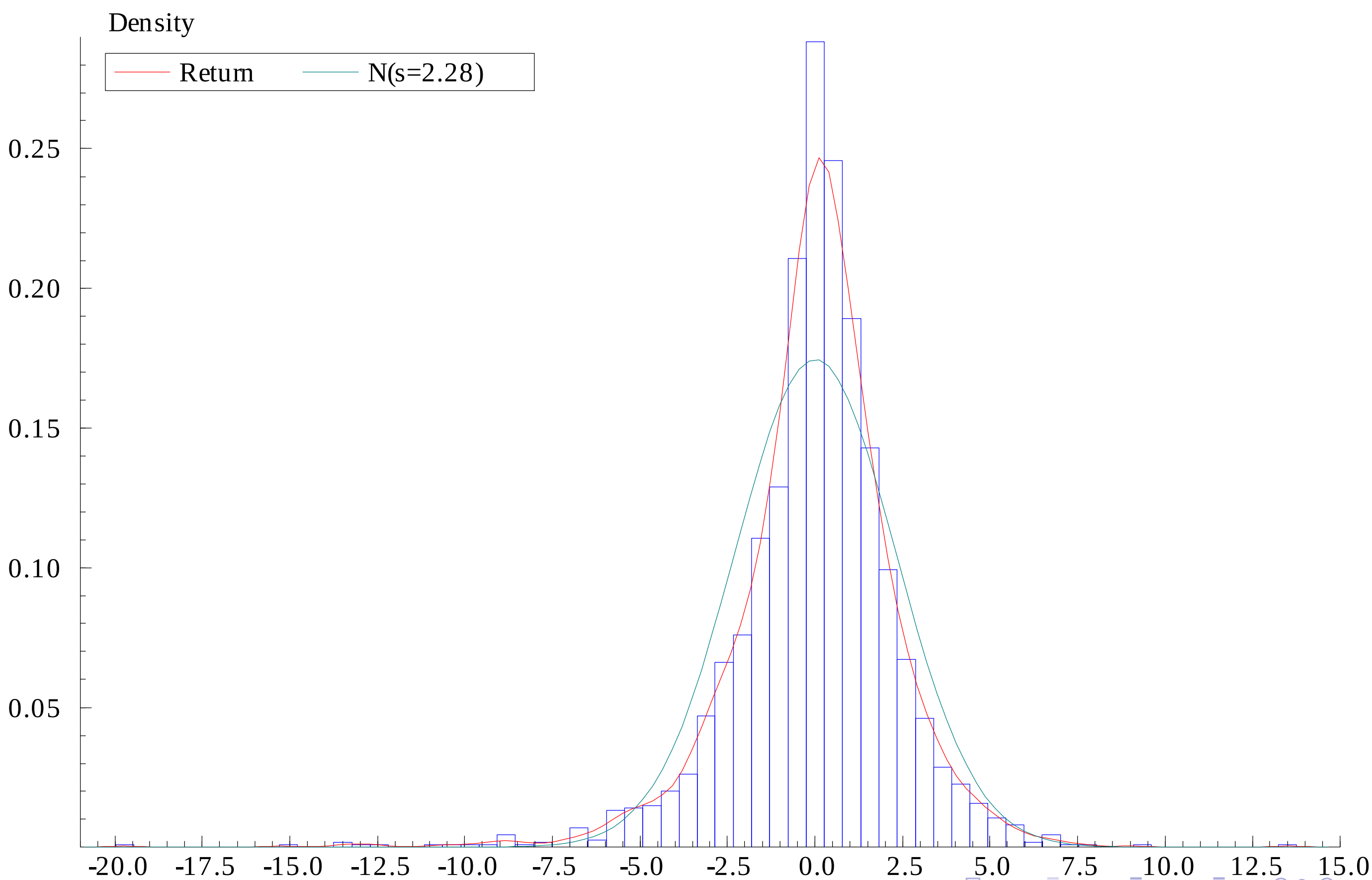


Figure: Red has $\kappa = 0$.



Silver

$\upsilon = 1.34$ $\overline{\eta}$ =0.075
$\upsilon = 2$ $\overline{\eta} = 0.220$ η goes from 13.31 to 4.54.
The SE of $\overline{\eta}$ is reduced from 0.043 to 0.023.
All three models fit well according to the Q(20) statistics. However Beta-t-EGARCH is worst on the AIC and BIC. The hypothesis that $\upsilon = 2$ is rejected by Wald and LR tests.
The LR statistic of the null hypothesis that $\overline{\eta} = 0$ is 3.60 which is less than 3.84, the 5% significance value for a χ_1^2 distribution. However, the correct 5% significance value is only 2.70 because of the one-sided alternative.
Thus there may be a small gain from using Gen-t rather than GED.

Extensions: Two components

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

$$\begin{aligned}\lambda_{t|t-1} &= \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}, \\ \lambda_{i,t+1|t} &= \phi_i \lambda_{i,t|t-1} + \kappa_i u_t, \quad i = 1, 2,\end{aligned}$$

where $\phi_1 > \phi_2$ if $\lambda_{1,t|t-1}$ denotes the long-run component.

Leverage and Two Components

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

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.

Extensions: ARCH in mean

A DCS EGARCH-M model may be set up as

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

where α is a parameter. It it straightforward to extend the theory for the conditional t-distribution, used in Harvey and Lange (2015), to generalized-t.

The ARCH-M effect could be simply captured by total volatility. However, because volatility has two components, it is possible to investigate the impact of both long-run and short-run movements. Thus

$$y_t = \mu + \alpha_1 \exp(\omega + \lambda_{1,t|t-1}) + \alpha_2[\exp(\lambda_{2,t|t-1}) - 1] + \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, ...,$$

When volatility is at its equilibrium level, the risk premium is $\mu + \alpha_1 \exp(\omega)$.

Asymmetry and skewness

Skewness can be introduced into the generalized Student-t distribution by means of the Fernandez and Steel (1998) method, as used by Harvey and Sucarrat (2014) for the Student-t distribution.
Zhu and Galbraith (2010) make a further extension to asymmetry by allowing different degrees of freedom above and below μ.
Extending the generalized-t distribution to handle skewness and asymmetry is, in principle, straightforward: we now have v_1 and v_2 as well as η_1 and η_2.

Asymmetry and skewness: silver

	κ^*	κ	ϕ	ω	μ	$\bar{\eta}_1$	$\bar{\eta}_2$	υ	LogL
Estimate	0.004	0.040	0.990	0.499	0.098	0.136	0.000	1.349	-4628.4
SE	0.004	0.007	0.002	0.104	0.034	0.052	0.048	0.144	

Left tail index $\eta_1 = 7.35$. Light right tail $\eta_2 = \infty$.
Asymmetric score (not leverage)

κ^*	κ	ϕ	ω	μ	α	$\bar{\eta}_1$	$\bar{\eta}_2$	υ	LogL
0.045	0.026	0.988	-0.535	0.040		0.041		1.497	-3361.5
0.005	0.005	0.001	0.102	0.014		0.024		0.114	
0.049	0.036	0.986	-0.376	0.023		0.121	0.004	1.617	-3347.5
0.005	0.005	0.001	0.100	0.013		0.039	0.044	0.149	
0.050	0.030	0.986	-0.233	0.015	0.573	0.031		1.478	-3342.0
0.005	0.005	0.001	0.115	0.014	0.011	0.029		0.122	
0.047	0.032	0.987	-0.323	0.021	0.542	0.068	0.004	1.561	-3341.5
0.004	0.005	0.001	0.116	0.014	0.023	0.032	0.056	0.153	

Table 3 SP500 daily excess returns from 2 Jan 2004 to 31 Dec 2013 ($T = 2,517$).

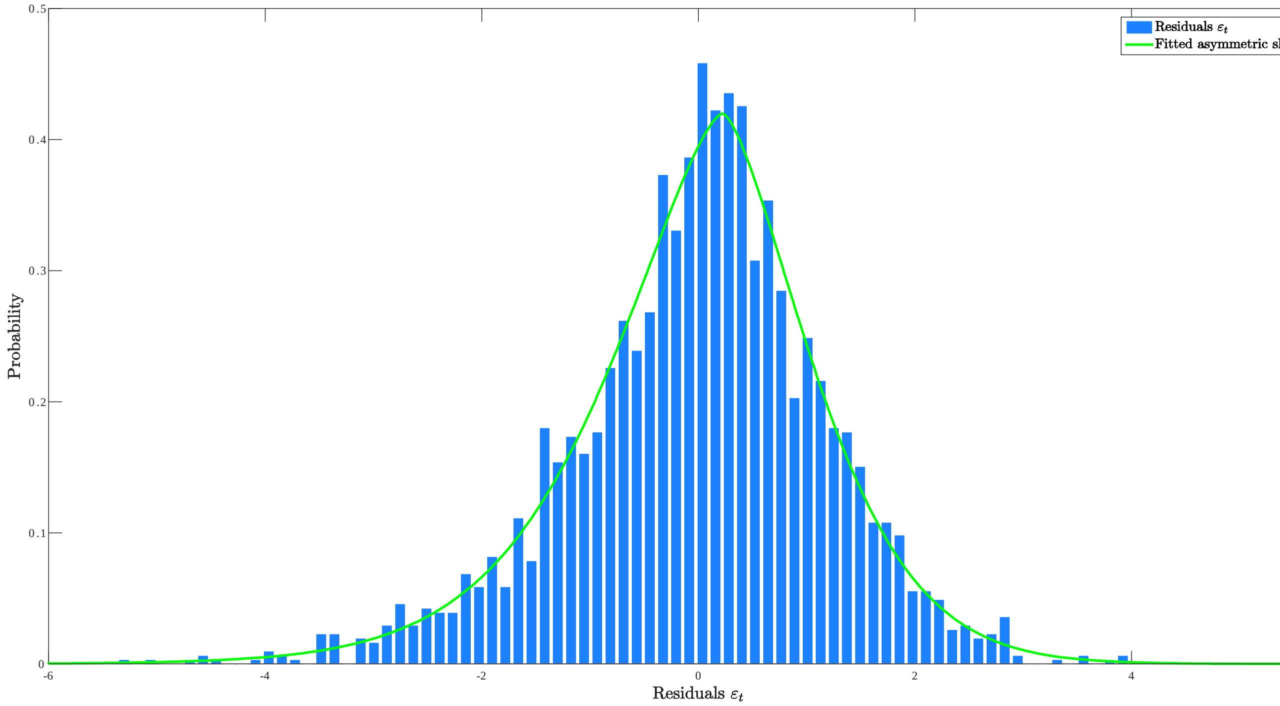


Figure: Standardized residuals from a skewed asymmetric generalized t EGARCH model fitted to daily SP500 returns

Stochastic location and stochastic scale

The Student t model for time-varying location may be combined with Beta-t-EGARCH. In other words $y_t \mid Y_{t-1}$ has a t_ν distribution with mean $\mu_{t|t-1}$ and scale $\exp(\lambda_{t|t-1})$, that is

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

where $\lambda_{t|t-1}$ depends on

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

Estimation by ML is straightforward.

Stochastic location and stochastic scale: inflation

Seasonally adjusted rate of inflation (CPI) in the United States from 1947(1) to 2007(2)

The rate of inflation is often taken to follow a random walk plus noise. Thus for the DCS-t model

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa^{\dagger} u_t,$$

where the dagger serves to differentiate $\kappa^{\dagger}$ from the κ parameter in the dynamic scale equation. Fitting a Gaussian model using the STAMP package gave an estimate of 0.579 for $\kappa^{\dagger}$.

The plot of the filtered level, $\mu_{t+1|t}$, shows it to be sensitive to extreme values and the ACF of the absolute values of the residuals provides strong evidence of serial correlation in variance.

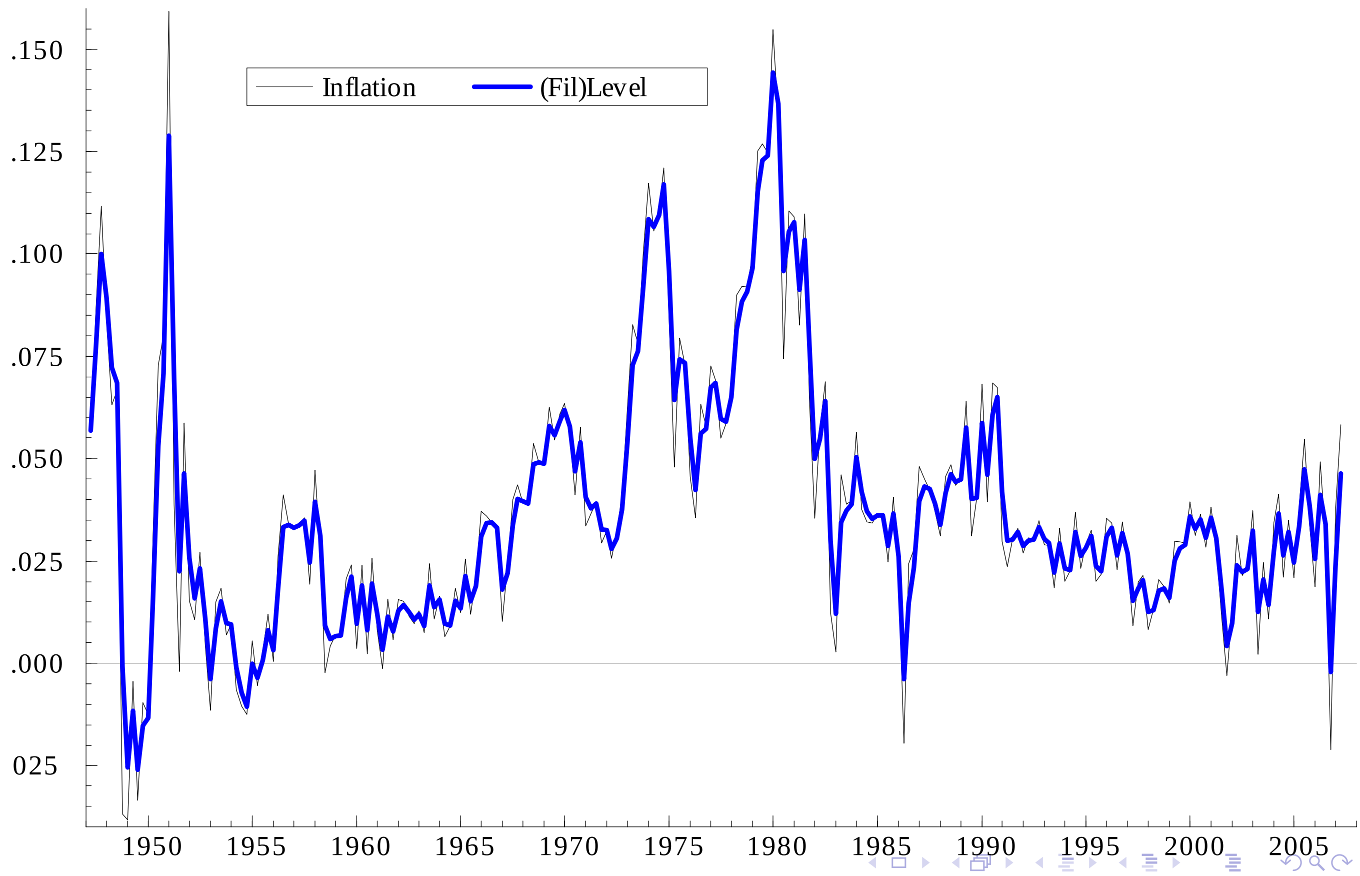


Figure: Filtered estimates of level from a Gaussian random walk plus noise fitted

Estimating a model in which filtered location is a random walk and scale evolves as a first-order Beta-t-EGARCH process gives the following ML estimates (with standard errors in parentheses): for location, $\widetilde{\kappa}^{\dagger} = 0.699(0.097)$, and for scale, $\widetilde{\delta} = -0.370(0.214)$, $\widetilde{\phi} = 0.912(0.051)$ and $\widetilde{\kappa} = 0.\ 118(0.041)$, with $\widetilde{\nu} = 11.71(4.58)$.
The filtered estimates respond less to extreme values than those from the homoscedastic Gaussian model.

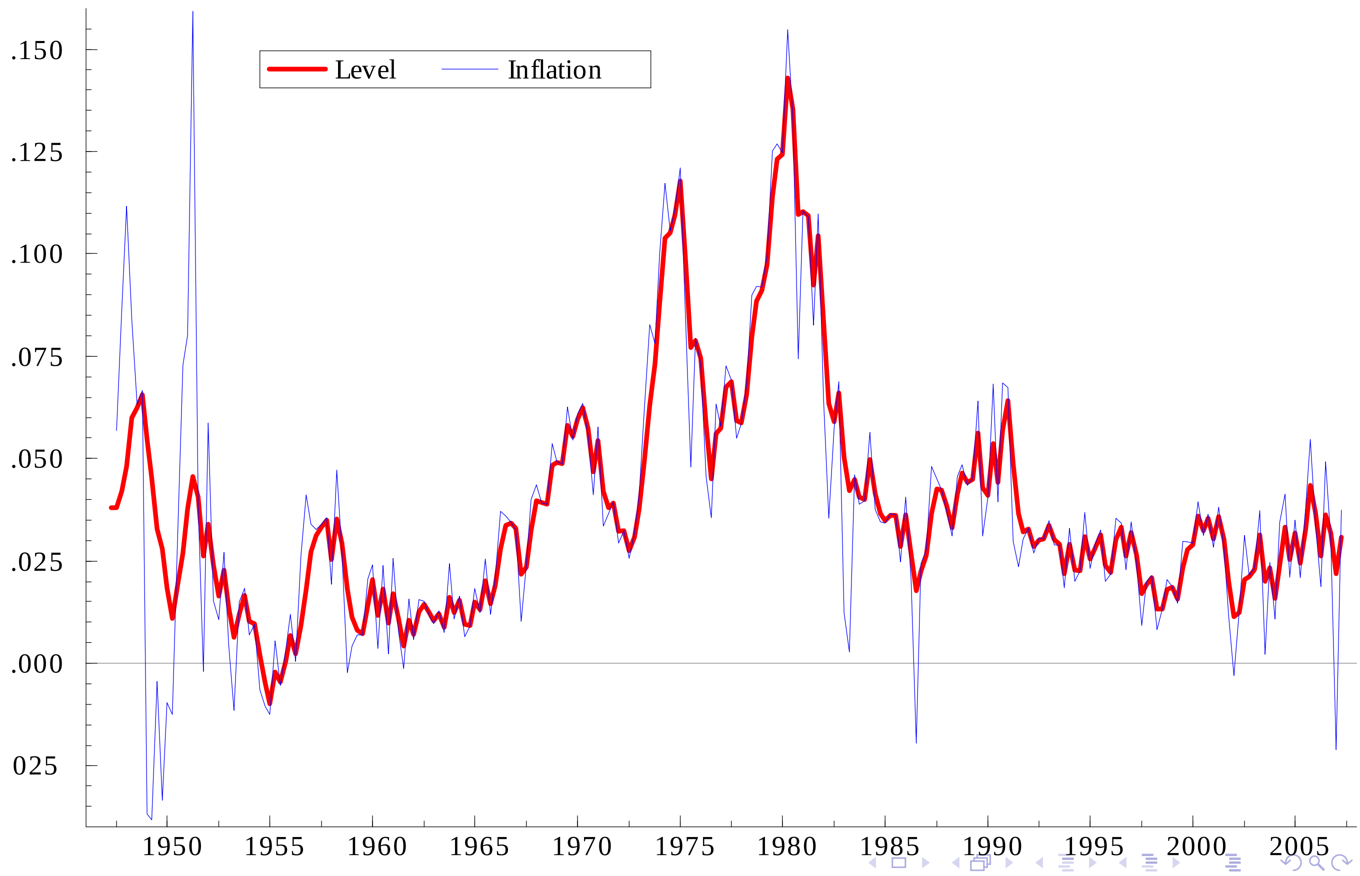


Figure: Estimated level of US inflation from a random walk plus noise model with

Stochastic location and stochastic scale: inflation

Stock and Watson UC-SV (2007)

$$y_t = \mu_t + \varepsilon_t \exp(\lambda_t^{\varepsilon}), \qquad \varepsilon_t \sim NID\left(0, \sigma_{\varepsilon}^2\right)$$

with stochastic level

$$\mu_{t+1} = \mu_t + \eta_t \exp(\lambda_t^{\eta}), \qquad \eta_t \sim NID\left(0, \sigma_{\eta}^2\right)$$

and SV

$$\lambda_{t+1}^{\varepsilon} = \lambda_t^{\varepsilon} + \zeta_t^{\varepsilon} \quad \text{and} \quad \lambda_{t+1}^{\eta} = \lambda_t^{\eta} + \zeta_t^{\eta}$$

where ζ_t^{ε} and ζ_t^{η} are mutually independent $NID(0, \sigma_{\zeta}^2)$, ie $Var(\zeta_t^{\varepsilon}) = Var(\zeta_t^{\eta})$.
Shephard (2016) proposes a similar model but parameterized so that the signal-noise ratio changes.

Stochastic location and stochastic scale: inflation

Stock and Watson (2015, R E Stat) allow for outliers. The model is

$$y_t = \mu_t + s_t \varepsilon_t \exp(\lambda_t^{\varepsilon})$$

with dynamics

$$\mu_{t+1} = \mu_t + \eta_t \exp(\lambda_t^{\eta}).$$

where $s_t = 1$ with prob p and $U[2, 10]$ with prob $1 - p$.

Stochastic location and stochastic scale: inflation

In an **adaptive filtering** model, the parameters in the dynamic equation are themselves allowed to change over time. Here

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

If $\kappa_{t+1|t}$ is evolves according to a first-order equation, then

$$\kappa_{t+1|t} = \kappa(1-\alpha) + \alpha\kappa_{t|t-1} + \beta w_t,$$

where w_t is the conditional score or an approximation to it. The conditional score is

$$\frac{\partial \ln f_t}{\partial \kappa_{t|t-1}} = \frac{\partial \ln f_t}{\partial \mu_{t|t-1}} \frac{\partial \mu_{t|t-1}}{\partial \kappa_{t|t-1}} = u_t \frac{\partial \mu_{t|t-1}}{\partial \kappa_{t|t-1}},$$

with $\partial\mu_{t|t-1}/\partial\kappa_{t|t-1}$ obtained recursively.

Stochastic location and stochastic scale: inflation

In a Gaussian model, w_t is (approximately) a nonlinear function of the observations, constructed from cross-products of current and lagged residuals.

$$w_t \simeq v_t v_{t-1} + \sum_{j=1}^{t-1} \left[\prod_{i=0}^{j-1} (1 - \kappa_{t-i|t-i-1}) \right] v_t v_{t-j-1}.$$

Variables involving such cross-products have appeared in nonlinear models before, an example being the bilinear model proposed by Granger and Andersen (1978), but the key point here is that w_t has a specific form that derives from a model set up to address a particular problem.
Test based on autocorrelations of the $w_t's$.

Gaussian model: test

LM tests against time-variation are based on scores. Here

$$w_t = v_t v_{t-1} + \sum_{j=1}^{t-1} \left[(1-\kappa)^j \right] v_t v_{t-j-1}.$$

Weights decline exponentially when $0 < \kappa < 2$ (in practice $\kappa \leq 1$).
May respond more quickly to jumps.

Multivariate Volatility

A multivariate normal distribution for an $N \times 1$ vector $\mathbf{y}$ is parameterized in terms of an $N \times 1$ mean vector, $\boldsymbol{\mu}$, and an $N \times N$ covariance matrix.
The most common multivariate t-distribution has PDF

$$f(\mathbf{y}_t; \boldsymbol{\mu}, \Omega, \boldsymbol{\nu}) = \frac{\Gamma(\nu + N)}{2(\pi\nu)^{N/2}\Gamma(\nu/2)\, |\Omega|^{1/2}\, w_t^{(\nu+N)/2}},$$

where $w_t = 1 + (1/\nu)(\mathbf{y}_t - \boldsymbol{\mu})'\Omega^{-1}(\mathbf{y}_t - \boldsymbol{\mu})$ and Ω is the scale matrix. It may be decomposed as $\Omega = \mathbf{DRD}$ where $\mathbf{R}$ is a correlation matrix and $\mathbf{D}$ is a diagonal matrix with the $i - th$ diagonal element equal to a scale parameter, φ_i.
In the dynamic case

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

where the $i - th$ diagonal element is $\varphi_i = \exp(\lambda_{i,t|t-1})$. See Creal et al (2011) and Harvey (2013, ch 7)

Multivariate Volatility

Consider a bivariate with zero means

$$\mathbf{R}_{t|t-1} = \begin{bmatrix} 1 & \rho_{t|t-1} \\ \rho_{t|t-1} & 1 \end{bmatrix}$$

where $\rho_{t|t-1}$ denotes the (changing) correlation, based on information at time $t-1$.

How should we drive the dynamics of the filter for changing correlation, $\rho_{t|t-1}$, and with what link function ?

A simple moment approach would use

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

to drive the covariance, but the effect of $x_1 = x_2 = 1$ is the same as $x_1 = 0.5$ and $x_2 = 2$.

Multivariate Volatility

Rather than work directly with $\rho_{t|t-1}$, a transformation is applied so as to keep it in the range, $-1 < \rho_{t|t-1} < 1$. The link function

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

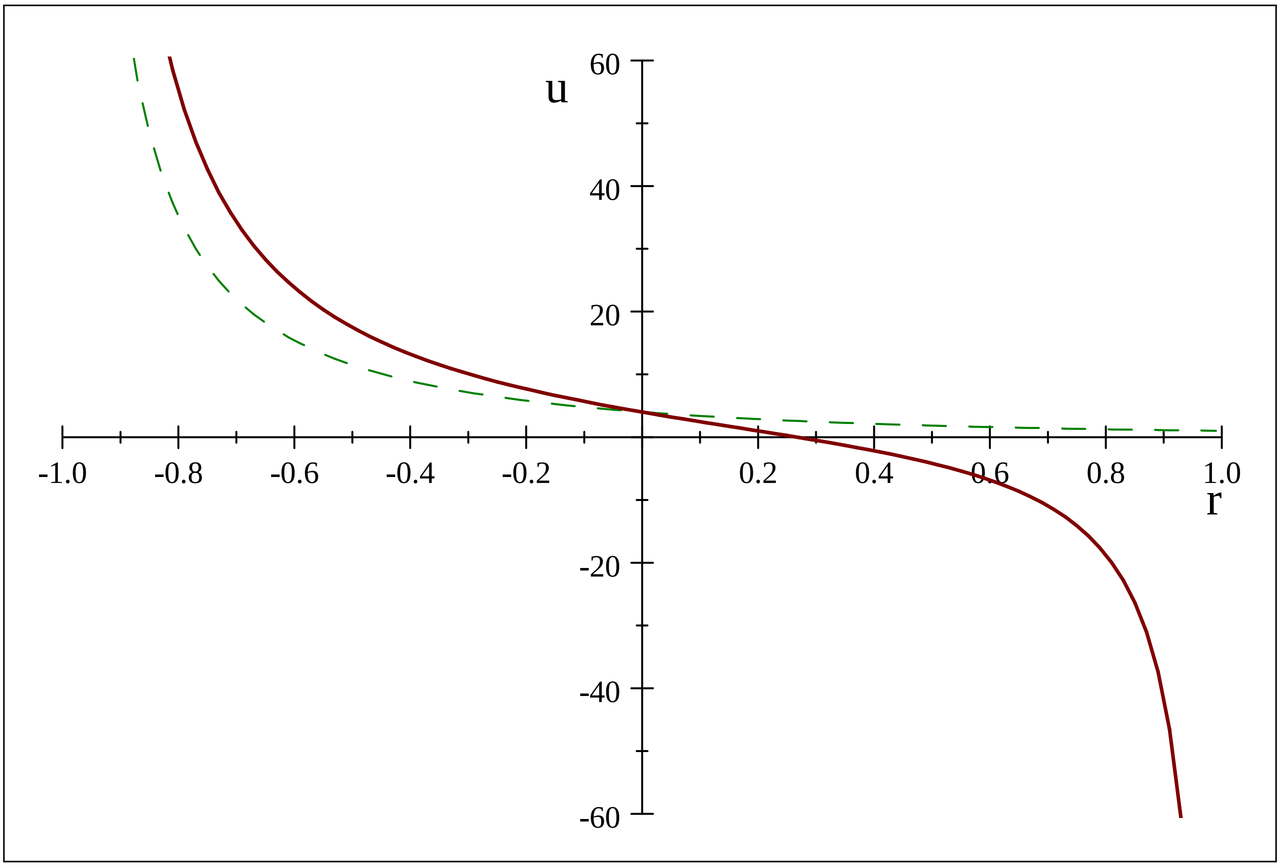
the new variable, $\gamma_{t|t-1}$, to be unconstrained. The inverse is the **arctanh** transformation.

The dynamic equation for correlation is

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

The score, written in terms of $\rho_{t|t-1}$, is

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



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

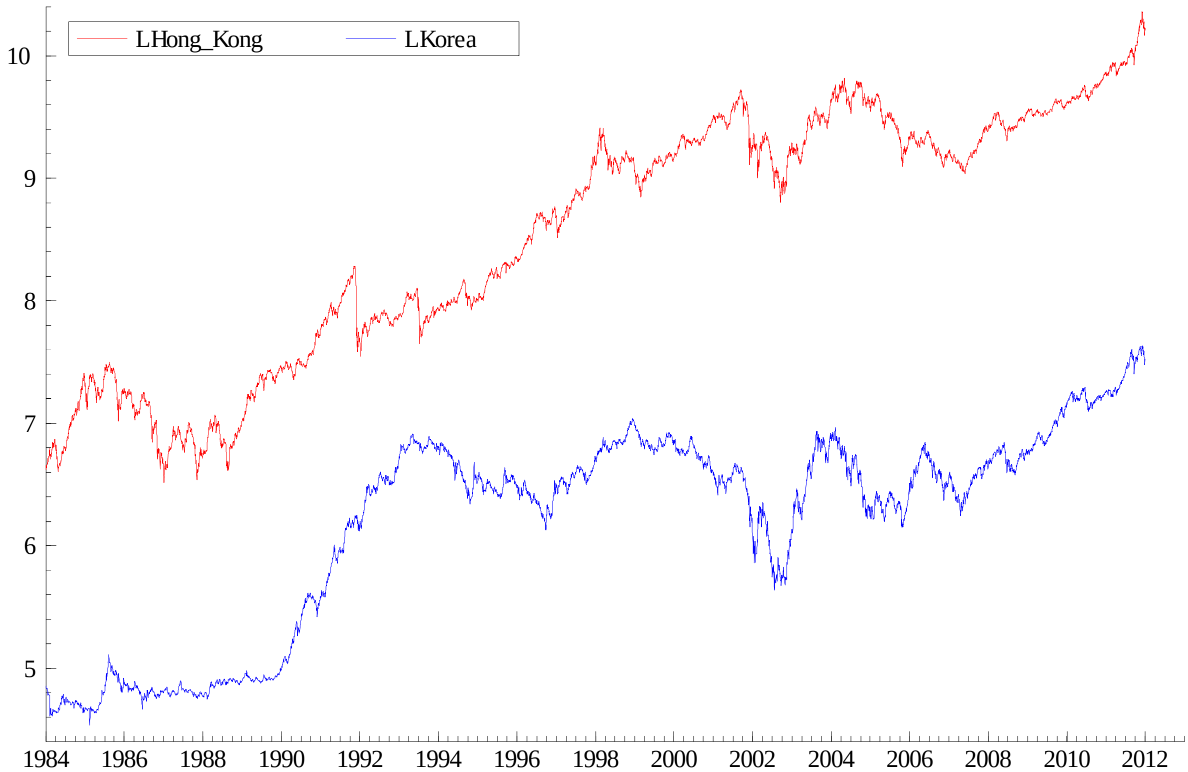
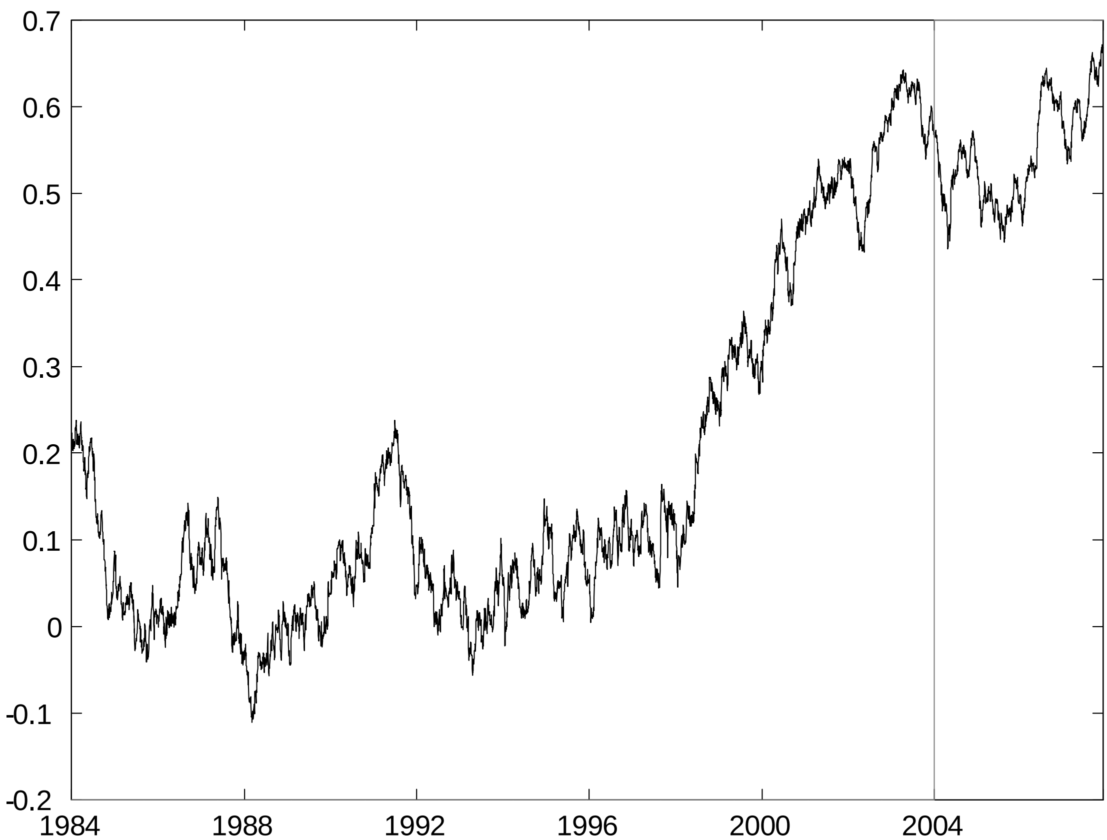


Figure: Logarithms of Hong Kong and Korea stock markets indices from 2/1/1984 to 27/11/2007.



LM test against time-varying correlation

Harvey and Thiele (2016, JEF) show the effectiveness of the score-based test against time-varying correlation.
Much better than standard moment-based test.

Time-varying parameters in regression

Time-varying parameters are usually modeled in regression by letting the coefficient of an explanatory variable change:

$$y_t = \beta_t x_t + \varepsilon_t, \quad t = 1, ..., T,$$

where β_t follows a stochastic process, such as an AR(1) or random walk.
If the explanatory variable is stochastic then it needs to be independent of β_t and ε_t in all time periods. Not only is this assumption restrictive, but if x_t is a linear process, then y_t will, in general, be nonlinear. Furthermore there are concerns about the path followed by the dependent variable when the explanatory variable is integrated; see the discussion in Harvey (1989, p. 409-10).

Time-varying parameters in regression

A bivariate model provides a better starting point. Suppose that y_{1t} and y_{2t} are jointly normal with zero means and constant variances, but a correlation that changes over time. The regression equation then becomes

$$y_{1t} \mid y_{2t} = \rho_t(\sigma_1/\sigma_2)y_{2t} + \varepsilon_t, \quad t = 1, ..., T,$$

where ε_t is distributed independently of y_{2t} with variance $\sigma_1^2 - \rho_t^2\sigma_2^2$. Similarly for the DCS model

$$y_{1t} \mid y_{2t} = \rho_{t|t-1}(\sigma_1/\sigma_2)y_{2t} + \varepsilon_t, \quad t = 1, ..., T,$$

The time-varying parameter is $\beta_{t|t-1} = \rho_{t|t-1}(\sigma_1/\sigma_2)$ and if $\rho_{t|t-1}$ is constrained to lie in the interval $(-1, 1)$, perhaps by using the arctan transformation, the range of $\beta_{t|t-1}$ is $-\sigma_1/\sigma_2$ to σ_1/σ_2.
Note - $\beta_{t|t-1}$ will also change if variances change.

Dynamic copulas: estimating changing association

The conditional score for the Clayton copula with parameter $\theta_{t|t-1}$ is

$$\begin{aligned} u_{\theta t} &= -\ln(\tau_{1t}\tau_{2t}) + (1+\theta_{t|t-1})^{-1} + \theta^{-2}\ln(\tau_{1t}^{-\theta_{t|t-1}} + \tau_{2t}^{-\theta_{t|t-1}} - 1) \\ &+ \left(\frac{1+2\theta_{t|t-1}}{\theta_{t|t-1}}\right)\frac{(\tau_{1t}^{-\theta_{t|t-1}}\ln\tau_{1t} + \tau_{2t}^{-\theta_{t|t-1}}\ln\tau_{2t})}{\tau_{1t}^{-\theta_{t|t-1}} + \tau_{2t}^{-\theta_{t|t-1}} - 1}, \end{aligned}$$

where $\tau_{it} = F(y_{it})$, $i = 1, 2$.
The first term involves the product $\tau_{1t}\tau_{2t}$, and so is a little like the product $x_{1t}x_{2t}$. In the Gaussian model the score modifies the impact of $x_{1t}x_{2t}$ by taking account of how the product was formed and the current parameter value. The same is true here.

Dynamic copulas: estimating changing association

Figure shows the response of the score when τ_2 varies, but τ_1 is fixed. Two points are worth noting.

1) As expected, the response is asymmetric in the sense that the behaviour when τ_1 fixed at 0.9 is not a mirror image of the behaviour for τ_1 fixed at 0.1.

2) When $\tau_1 = 0.1$, the score is only positive for vaues of τ_2 close to 0.1, the effect being more pronounced when $\theta = 5$, as opposed to $\theta = 1$. This behaviour is entirely consistent with the conditional density shown earlier : if τ_2 is not close to 0.1, it suggests that $\theta_{t|t-1}$ is too big and the role of the negative score in the dynamic equation is to make $\theta_{t+1|t}$ smaller.

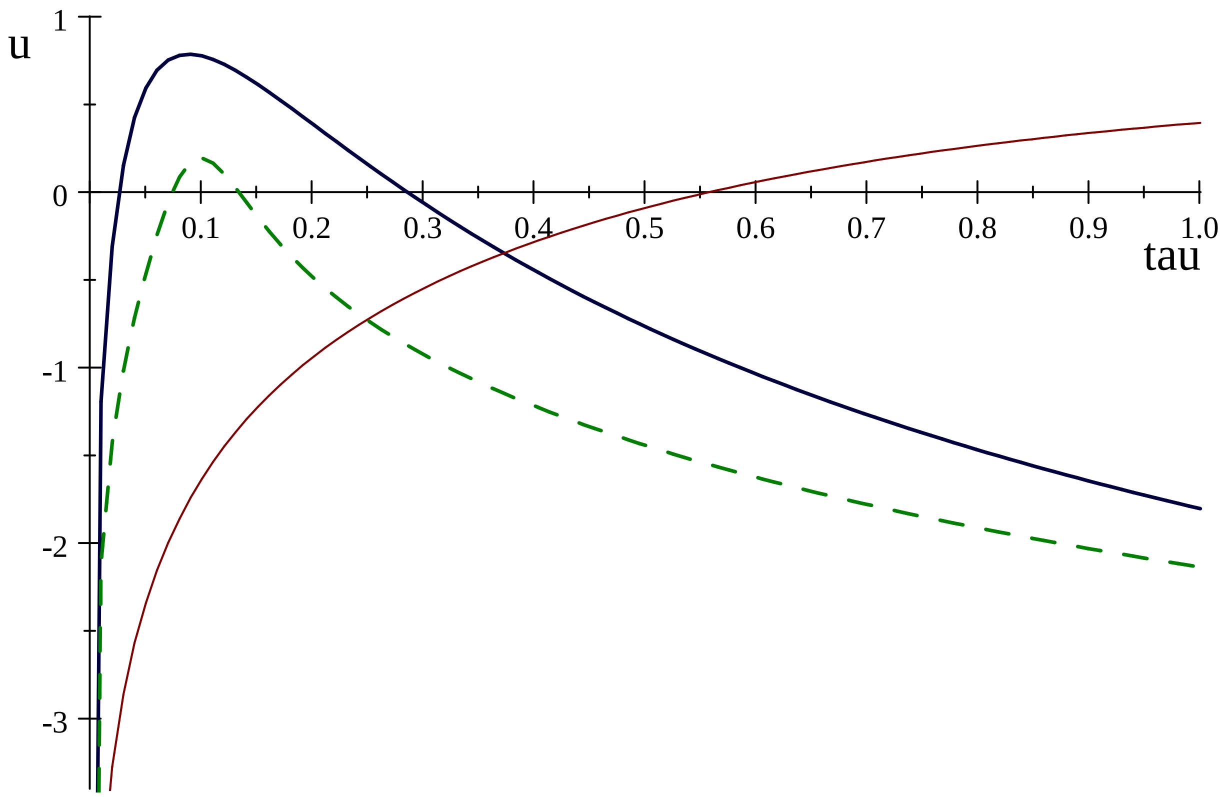


Figure: Response of u for fixed τ_1