Dynamic Models for Volatility and Heavy Tails

3. Location/scale and multivariate models

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

Non-negative variables: duration, realized volatility and range

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

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

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

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

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

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

Non-negative variables: duration, realized volatility and range

An exponential link function, $\mu_{t|t-1} = \exp(\lambda_{t|t-1})$, not only ensures that $\mu_{t|t-1}$ is positive, but also allows the asymptotic distribution to be derived. The model can be written

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

with dynamics

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

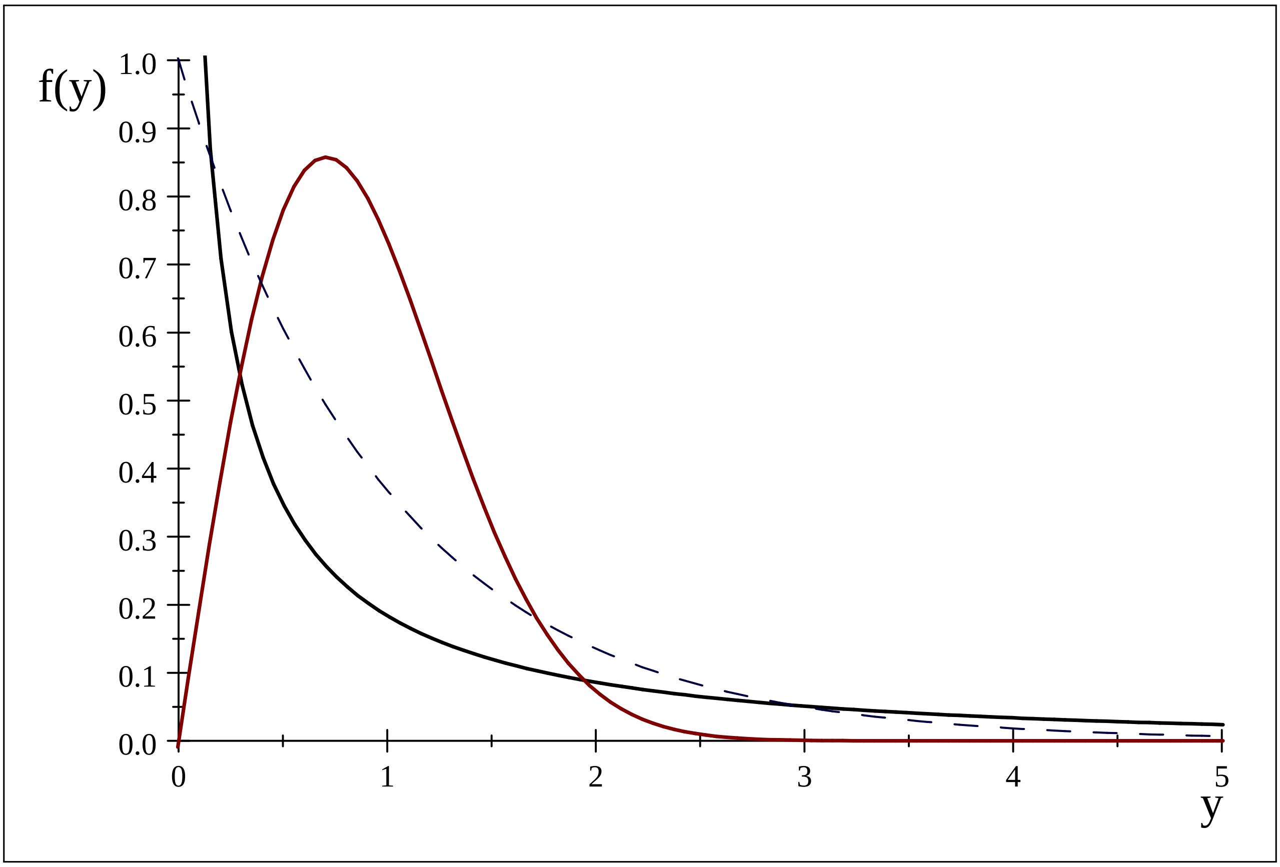
where, for a Gamma distribution

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

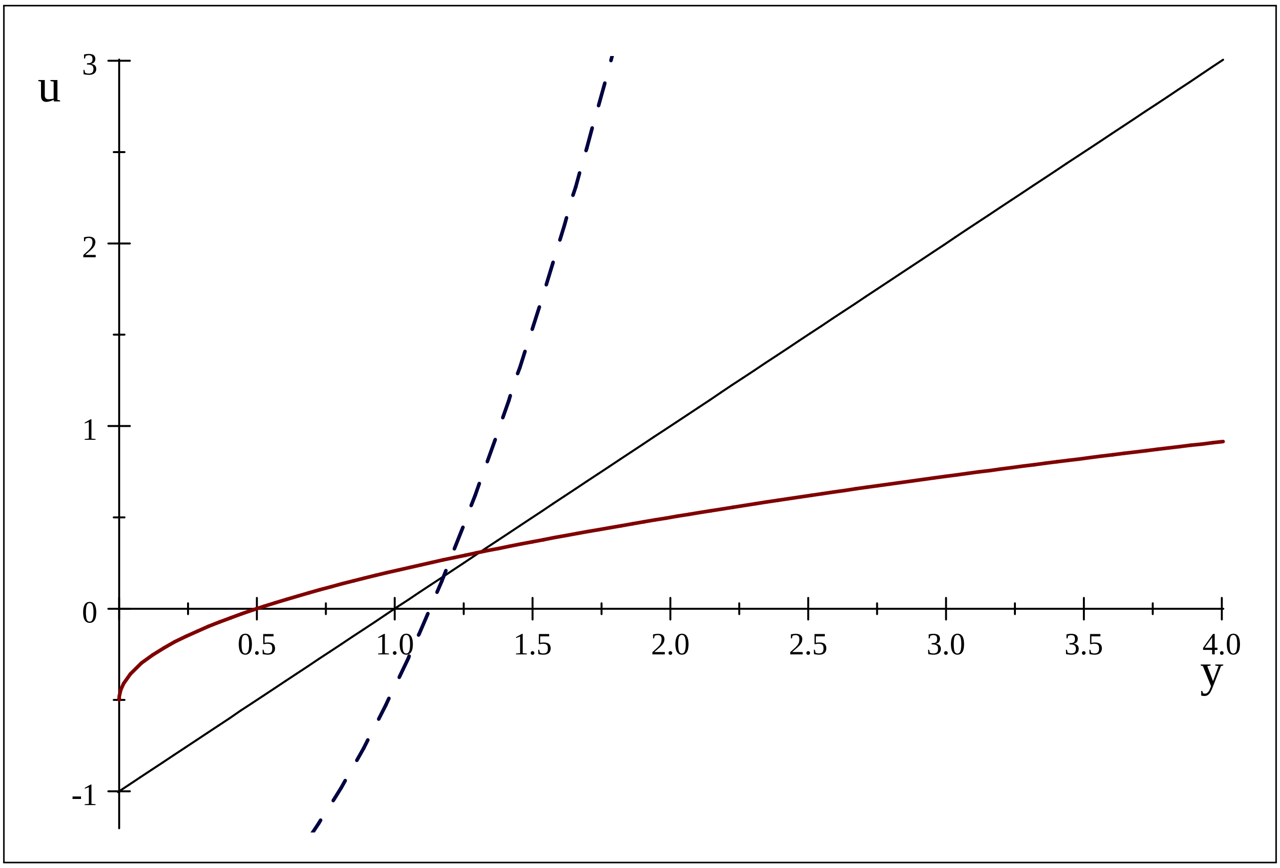
The response is linear but this is not the case for Weibull. The PDFis

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

where α is the scale and υ is the shape parameter. The mean is $\mu = \alpha\Gamma(1+1/\upsilon)$ and the variance is $\alpha^2\Gamma(1+2/\upsilon) - \mu^2$.



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



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

Log-logistic distribution

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

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

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

and so

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

where

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

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

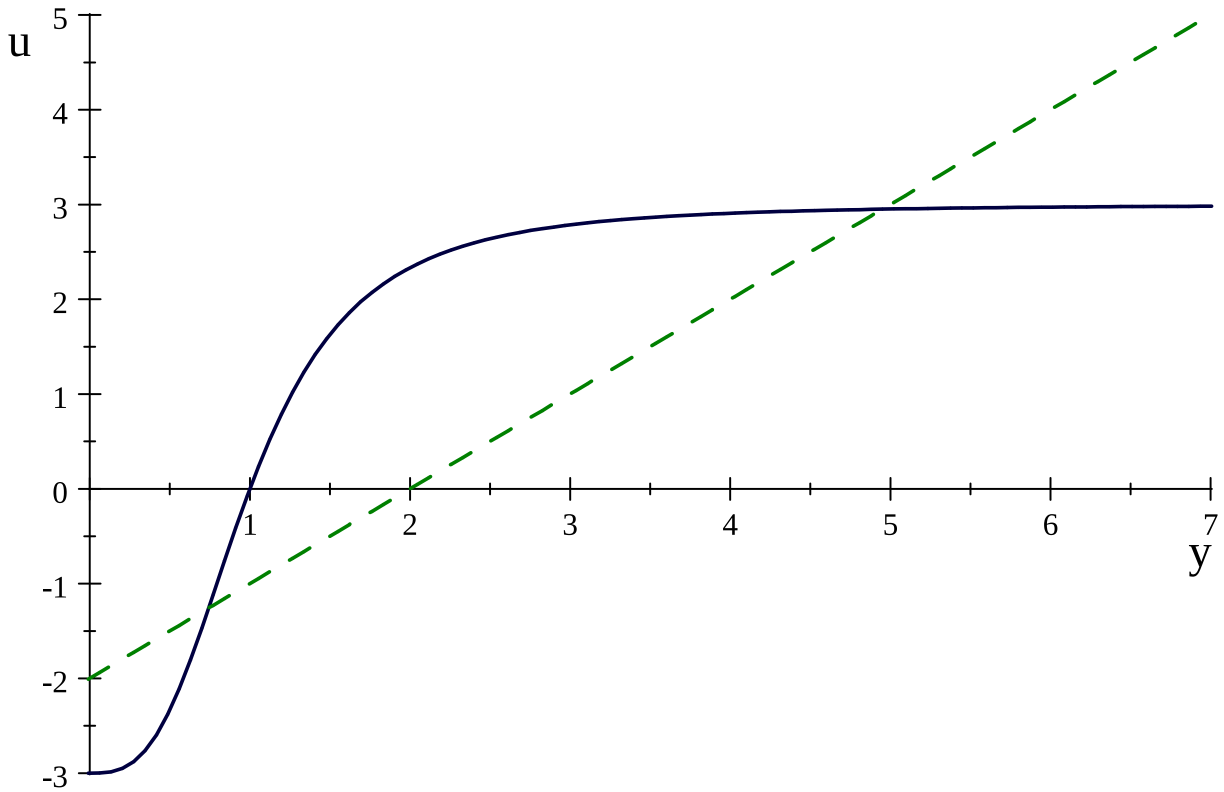


Figure: Impact of u for a log-logistic distribution and a gamma (dashed), with shape parameters $\nu = 3$ and $\gamma = 2$ respectively.

Log-logistic distribution

The asymptotic theory is not complicated. Differentiating the score gives

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

Proposition

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

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

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

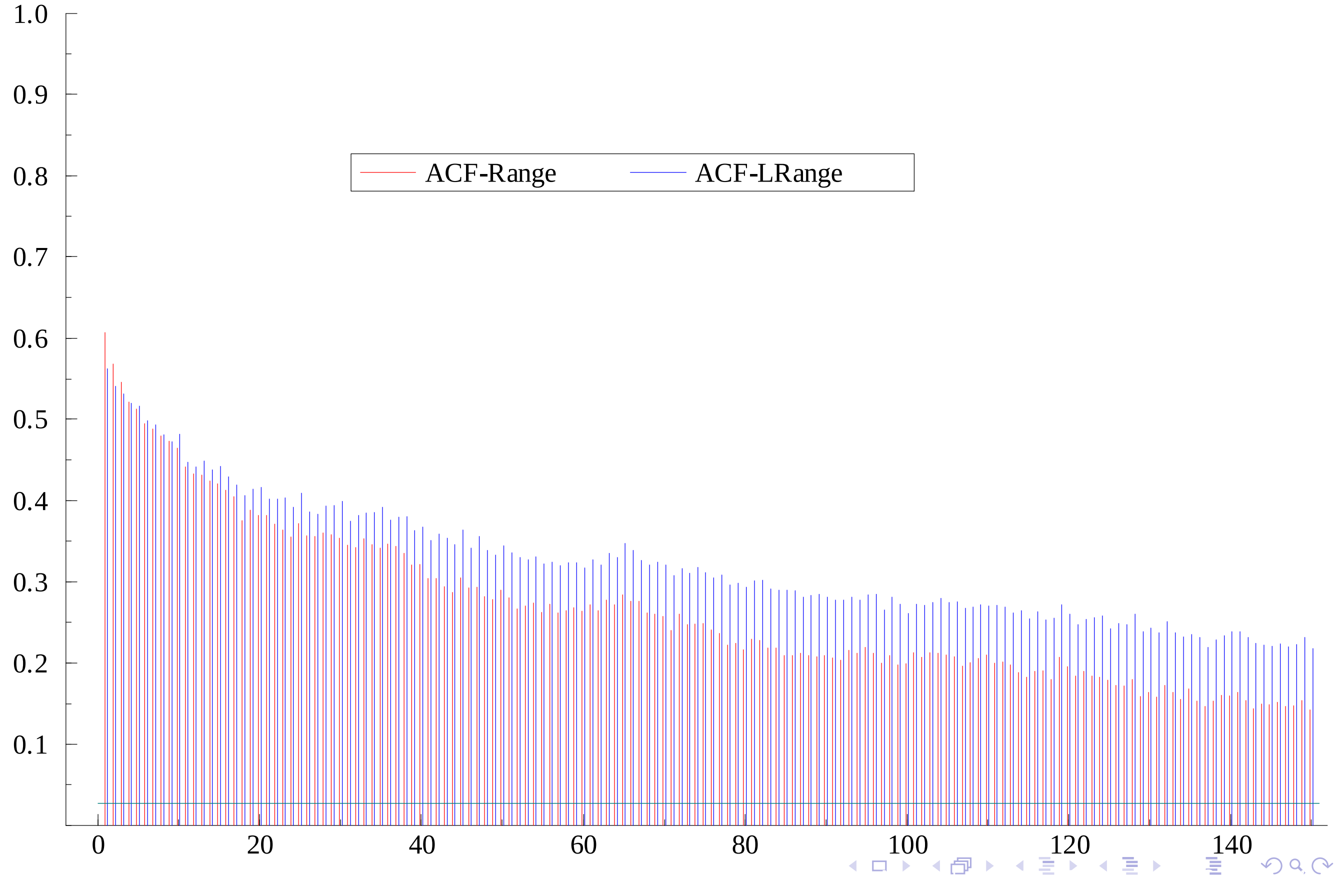
$$\begin{aligned} a &= \phi - \kappa\nu^2/3 \\ b &= \phi^2 - (2/3)\nu^2\phi\kappa + 2\kappa^2\nu^4/15 \quad \text{and} \quad c = 0. \end{aligned}$$

Generalized gamma and beta distributions

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

Tests and model selection

Model selection requires decisions to be made about the distribution and the form of the dynamic equation for the scale. The starting point is testing against serial correlation in the observations. But just as squared observations may be unduly influenced by outliers in returns, so the observations themselves may not be robust here. A square root or logarithmic transformation may be better. The correlograms for the daily range of the CAC index and its logarithm are shown below. The first few sample autocorrelations for the raw observations are bigger than those given by the logarithmic transformation, but at higher lags the autocorrelations of the logarithms are bigger and die away more slowly.



Tests and model selection

Tests for changing scale can be carried out using the Box-Ljung statistic. Under the null hypothesis the observations are independent so any transformation can be used. If a distribution is specified at the outset, Lagrange multiplier tests can be carried out with the $Q_u(P)$ statistic, based on the score.

Example

With the F-distribution,

$$\widetilde{u}_t = \frac{\widetilde{\nu}_1 + \widetilde{\nu}_2}{2} \frac{\widetilde{\nu}_1 y_t \exp(-\widetilde{\lambda})/\widetilde{\nu}_2}{1 + \widetilde{\nu}_1 y_t \exp(-\widetilde{\lambda})/\widetilde{\nu}_2} - \frac{\widetilde{\nu}_1}{2}, \quad t = 1, ..., T,$$

where $\widetilde{\lambda}$, $\widetilde{\nu}_1$ and $\widetilde{\nu}_2$ are the ML estimators of the location/scale parameter and the degrees of freedom parameters in the static distribution.

When the distribution is gamma, the LM test simply uses the observations, the y_t's. For distributions from the generalized gamma family, the LM test will use the autocorrelations for y_t^{υ}, where the shape parameter, υ, is estimated.

Tests and model selection

When a DCS model is fitted, diagnostic tests of serial correlation and distribution can be based on the scores, the residuals, that is $y_t \exp(-\lambda_{t|t-1})$, the PITs of the residuals and the normalized PITs. The Lagrange multiplier test principle suggest that the scores be used to test against serial correlation. However, a test based on the residuals may also be informative. An attraction of making the probability integral transformation to the residuals is that it may yield serial correlation tests which are more robust. Furthermore the PITs are comparable for different conditional distributions and their histograms are very useful for assessing goodness of fit. Figure below shows the PIT residuals from fitting a DCS gamma model to the duration data for Boeing - it is clearly unsatisfactory.

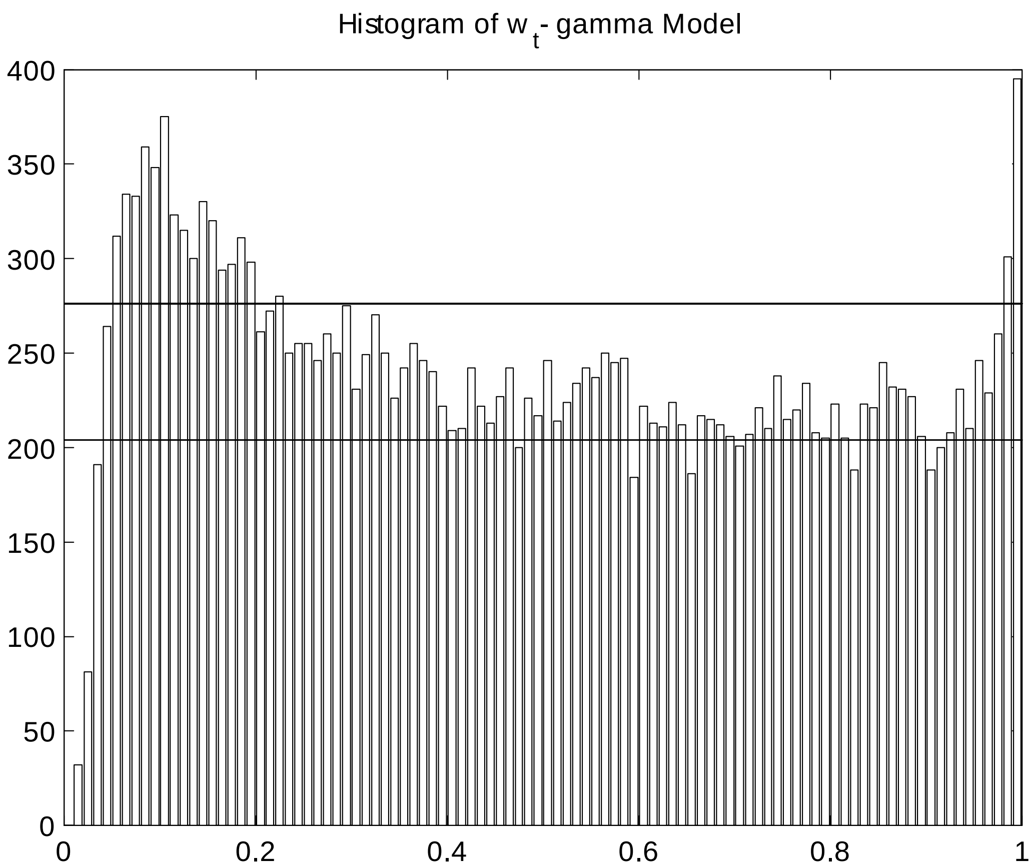


Figure: PITs from fitting a DCS gamma model to data on duration of trades for Boeing

Tests and model selection

While an inspection of the histogram of PITs or normalized PITs is often sufficient to eliminate a distribution from further consideration, the choice between competing candidates is best made by goodness of fit criteria. The AIC or BIC may be used within the sample, while outside the sample, the predictive likelihood (sometimes called the log-score) is simple and effective. Looking at the post sample residuals, scores and PITs may also provide valuable information. Mitchell and Wallis (2011) provide a recent discussion of the issues involved.

Range

The **log-logistic** model fitted to Paris CAC gave the following estimates, with numerical SEs in parentheses:

$$\begin{array}{ccccc} \widetilde{\omega} = & -4.828 & \widetilde{\phi}_1 = 0.998 & \widetilde{\phi}_2 = 0.962 & \\ & (0.054) & (0.002) & (0.009) & \\ \widetilde{\kappa}_1 = & 0.025 & \widetilde{\kappa}_2 = 0.066 & \widetilde{\kappa}_L = 0.040 & \widetilde{\nu} = 4.546 \\ & (0.010) & (0.009) & (0.031) & (0.034) \end{array}$$

The first component is highly persistent, and little would be lost by simply setting ϕ_1 to unity.
As regards the short-term component, the effect of leverage is that the overall response, $\widetilde{\kappa}_2 + \widetilde{\kappa}_L$, is 0.026 for positive returns and 0.106 for negative returns.

Duration

Duration models are widely used in financial econometrics to capture the changing intensity governing the time between events. Thus they may used, for example, to model the times between trades of an asset. In this context here is a relationship with volatility in that higher volatility tends to be associated with more trades. Duration models are also used in other areas.

Bauwens *et al* (2004) investigate a wide range of autoregressive conditional duration models for price, volume and trade duration data.

A trade duration is given by the time interval between two consecutive trade events.

A price duration is measured by the time interval between two bid-ask quotes during which a cumulative change in the mid-price of at least \$0.125 is observed.

A volume duration denotes the time interval between two bid-ask quotes during which the cumulative traded volume amounts to at least 25,000 shares.

Duration

Bauwens *et al* (2004) argue that price durations are perhaps the most interesting duration processes due to their close links to market microstructure and options pricing. They find that employing the basic MEM specifications with the exponential and Weibull distributions is not advisable. An exponential link function gives much better results for the Weibull distribution. However, their preference is for the generalized gamma and Burr distributions, again with exponential link functions. Their 'log-ACD' specification has the conditional mean set to $\mu_{t|t-1} = \exp(\lambda^*_{t|t-1})$, where

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

The first of the above dynamic equations corresponds to the DCS model for a lognormal distribution, while the second is the DCS model for a gamma distribution. Neither resembles the DCS equation for any member of the generalized beta family.

Duration

Bauwens *et al* (2004) reach similar conclusions regarding the best models when volume duration data is used.

Table, adapted from Andres and Harvey (2012), shows the results of fitting various DCS models to the Boeing volume duration data used by Bauwens *et al* (2004). The (asymptotic) standard errors were computed analytically. The first 1200 observations were used for estimation with the remaining reserved for post-sample evaluation.

The Burr distribution gives the best fit, followed closely by Weibull.

The Weibull shape parameter is greater than one, meaning that the distribution has the humped shape.

The log-logistic distribution does not give a good fit and the hypothesis that the second shape parameter in the Burr, ς, is unity is easily rejected using a LR test.

The gamma and F-distributions are only marginally worse than the Weibull, but the lognormal fit is very bad.

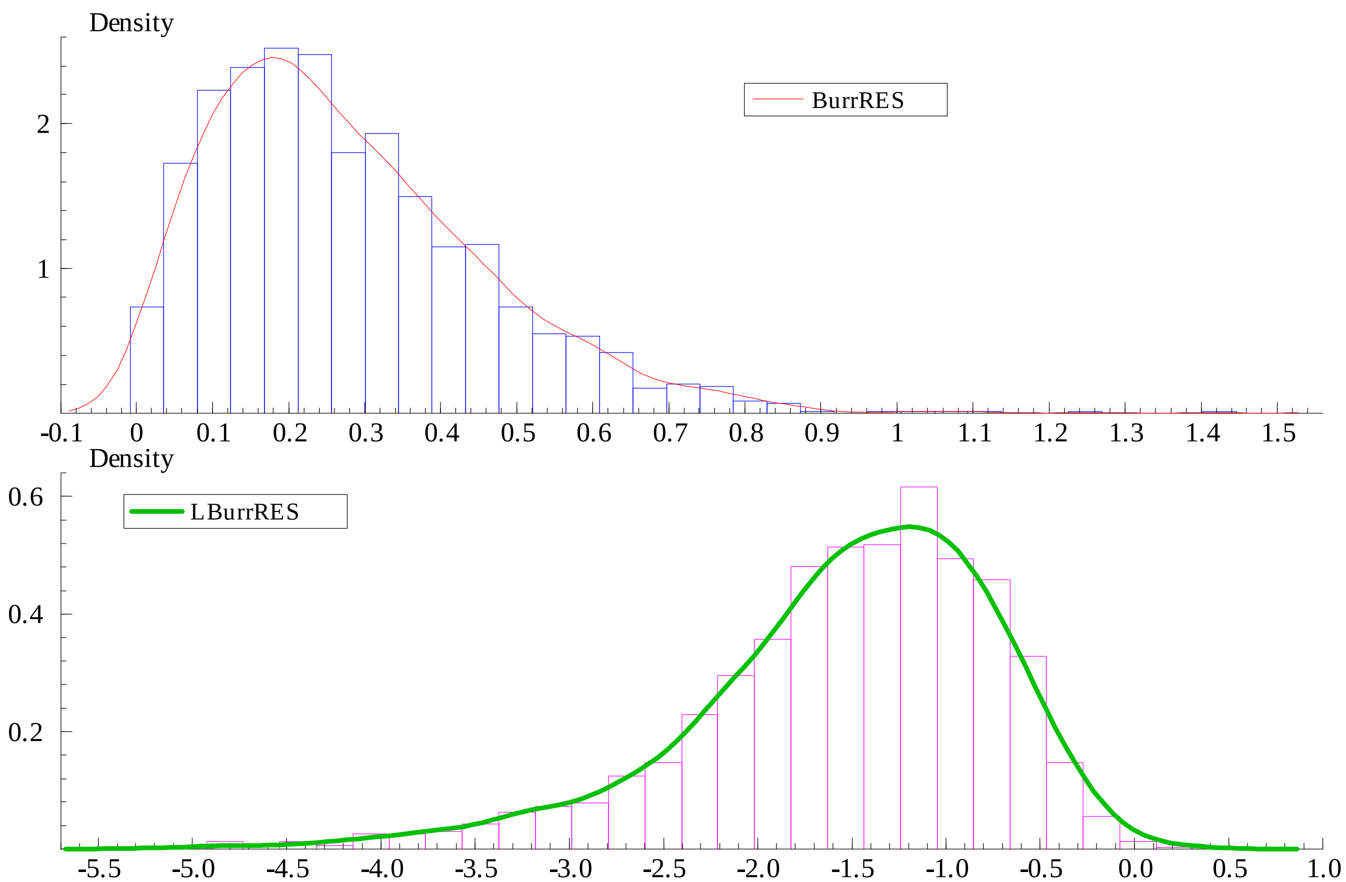
Duration

One particularly interesting feature of the results is that although the maximized likelihood function for the Weibull distribution is only marginally worse than that of the Burr distribution, its shape parameter of 1.57 means that, in contrast to the Burr distribution, it does not have a heavy tail. The QQ plots indicates that there are six or seven observations that are outliers for the Weibull, but not for the Burr. The corresponding graphs for the scores tell the same story, but the outlying Weibull observations do not show up in the histogram of the PITs.

Although all Burr distributions have a heavy tail, a value of less than one for the ς scale parameter means that the distribution of the logarithm of the variable is skewed to the left. Figure shows the histogram of the residuals from the fitted Burr model, together with the histogram of their logarithms.

The diagnostics give little indication of residual serial correlation. In contrast to the Q-statistics for the Dow-Jones range data, the Q-statistics shown are all rather similar for scores, residuals and PITs. The same is true in the post-sample period.

Boeing Volume Duration						
	Gamma		Weibull		Lognormal	
	Estimate	ASE	Estimate	ASE	Estimate	ASE
ω	-0.001	0.003	0.002	0.003	-0.011	0.005
ϕ	0.966	0.013	0.971	0.010	0.961	0.015
κ	0.118	0.019	0.067	0.011	0.102	0.016
ν, υ or σ^2	2.133	0.082	1.551	0.034	0.590	0.024
LogL	-1012.78		-1014.34		-1092.65	
AIC/BIC	2033.6	2028.3	2036.7	2031.46	2193.30	2188.07
	Log-logistic		Burr		F	
	Estimate	ASE	Estimate	ASE	Estimate	ASE
ω	-0.010	0.013	0.036	0.013	-0.001	0.004
ϕ	0.952	0.017	0.971	0.010	0.967	0.005
κ	0.162	0.028	0.082	0.012	0.055	0.013
ν, ν or ν_1	2.316	0.057	1.680	0.044	4.256	0.016
ς or ν_2	-	-	8.006	0.751	1000.0	7.607
LogL	-1064.35		-1010.35		-1012.93	
AIC/BIC	2136.7	2131.46	2030.7	2023.9	2035.86	2029.08



Andrew Harvey (Cambridge University) Volatility and Heavy Tails / 66

Figure: Histograms of residuals and their logarithms from fitting a conditional

Duration

Distribution	Scores		PITs		Residuals	
	Q(10)	Q(50)	Q(10)	Q(50)	Q(10)	Q(50)
Gamma	12.74	39.77	15.47	44.96	12.74	39.77
Weibull	8.20	32.78	17.32	47.97	10.59	37.52
Lognormal	16.13	46.00	17.05	47.28	14.94	42.92
Log-logistic	15.13	46.52	15.26	46.76	14.07	43.23
Burr	11.58	38.08	15.72	44.53	12.67	39.12
F	12.81	39.97	15.46	44.97	12.77	39.81

Conclusions

Letting the dynamics for the scale in a time series model for a non-negative variables be driven by the score yields a class of models that can be applied to a wide range of distributions. The generalized beta and generalized gamma distributions play a unifying role. The statistical properties of the models can be found because the scores are either beta or gamma distributed. For a first-order model, an analytic expression can be derived for the information matrix and Monte Carlo evidence shows that resulting asymptotic standard errors provide a good approximation in moderate size samples. Indeed they often appear to be more reliable than numerical standard errors.

The Burr distribution featured prominently for both range and duration. This has important implications for model performance since the response of dynamic conditional score models to large observations is bounded for generalized beta distributions.

Conclusions

Dynamic conditional score models can be used to model realized volatility, the case for their use being the same as for the range. Measures of realized volatility can be biased by market microstructure and so their logarithms may not be normally distributed. For example, Taylor (2005, pp 327-42) notes there appears to be significant skewness and kurtosis.
The structure of dynamic conditional score models is such that they can be extended to include time-varying trend and seasonal effects. For intra-day data, the seasonality translates into a diurnal effect; see Brownlees et al (2010, p 11). The usual approach in the literature is to remove such effects prior to any estimation. However, there is evidence to suggest that the diurnal effect is time-varying and future work will attempt to capture such effects within the model by using a limited number of trigonometric terms or by a time-varying periodic spline as in Harvey and Koopman (1993).

Multivariate UC models

Multivariate structural time series models are described in some detail in Harvey (1989, Ch. 8) and implemented in STAMP 8. The prototypical model is

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\omega} + \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim NID\left(\mathbf{0}, \Sigma_\varepsilon\right), \quad t = 1, ..., T \\ \boldsymbol{\mu}_{t+1} &= \Phi\boldsymbol{\mu}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim NID(\mathbf{0}, \Sigma_\eta). \end{aligned} \tag{1}$$

The statistical treatment of all such models is based on the Kalman filter.

Multivariate UC models

The breakdown into signal and noise provides the basis for a rich description of multivariate time series. Setting $\Phi = \mathbf{I}$ gives the multivariate **local level** model and letting $rank(\Sigma_\eta) < N$ gives rise to common trends and hence co-integration. If $rank(\Sigma_\eta) = J < N$, an appropriate ordering of the series enables the model to be written with J common levels, or trends, $\boldsymbol{\mu}_t^\dagger$:

$$\begin{aligned} \mathbf{y}_{1t} &= \boldsymbol{\mu}_t^\dagger + \boldsymbol{\varepsilon}_{1t} \\ \mathbf{y}_{2t} &= \Pi\boldsymbol{\mu}_t^\dagger + \overline{\boldsymbol{\mu}} + \boldsymbol{\varepsilon}_{2t}, \end{aligned} \tag{2}$$

where Π is an $(N-J) \times J$ matrix of coefficients, $\overline{\boldsymbol{\mu}}$ is an $(N-J) \times 1$ vector of constants, and

$$\boldsymbol{\mu}_t^\dagger = \boldsymbol{\mu}_t^\dagger + \boldsymbol{\eta}_t^\dagger, \quad \boldsymbol{\eta}_t^\dagger \sim NID(\mathbf{0}, \Sigma_\eta^\dagger), \qquad \Sigma_\eta^\dagger \text{ is pd.}$$

The presence of common trends implies **co-integration**. In the above model there exist $R = N - J$ co-integrating vectors, such that pre-multiplication of (2) by these vectors yields R stationary time series. If the matrix of co-integrating vector is $\mathbf{A} = (-\Pi, \mathbf{I}_R)$,

$$\mathbf{y}_{2t} = \Pi\mathbf{y}_{1t} + \overline{\boldsymbol{\mu}} + \boldsymbol{\varepsilon}_t, \quad \text{where} \quad \boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}_{2t} - \Pi\boldsymbol{\varepsilon}_{1t}.$$

Multivariate UC models

When there is only one common trend, $\mu_t^\dagger$, the matrix Π is an $(N-1) \times 1$ vector, $\boldsymbol{\pi}$. There are $N-1$ co-integrating vectors, but these can be chosen in different ways. Restricting $\boldsymbol{\pi}$ to be a vector of ones gives *balanced growth*.

Example

National income, consumption and investment exhibit balanced growth when there is a single common trend for their logarithms. If the common trend is associated with income and the two co-integrating equations correspond to the 'great ratios' of consumption and investment to income then

$$\mathbf{A} = \begin{bmatrix} -\boldsymbol{\pi} & \mathbf{I}_{N-1} \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

The Gaussian noise in (1) could be replaced by heavy-tailed noise with multivariate $t-$distribution, that is $\boldsymbol{\varepsilon}_t \sim t_\nu(\mathbf{0}, \Omega_\varepsilon)$. Once the Gaussianity assumption is dropped, computer intensive techniques are needed. A DCS model offers an alternative.

Multivariate DCS models

The DCS location model is

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

$$\boldsymbol{\mu}_{t+1|t} = \Phi\boldsymbol{\mu}_{t|t-1} + \mathbf{K}\mathbf{u}_t.$$

where the vector $\mathbf{u}_t$ depends on the score.
The log-density for the t-th observation is

$$\ln f_t(\boldsymbol{\omega},\Phi, \mathbf{K},\Omega,\nu) = \ln(\Gamma(\nu+N)/2) - \ln\Gamma(\nu/2) - \frac{N}{2}\ln\pi\nu - \frac{1}{2}\ln|\Omega| - \frac{\nu+N}{2}\ln$$

where

$$w_t = 1 + (1/\nu)(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1})'\Omega^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1})$$

Multivariate models

The score vector with respect to $\boldsymbol{\mu}_{t|t-1}$ is

$$\frac{\partial \ln f_t}{\partial \boldsymbol{\mu}_{t|t-1}} = \frac{1}{w_t}\frac{\nu+N}{\nu}\Omega^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1}). \tag{3}$$

If $\mathbf{u}_t$ is set equal to the score vector in the multivariate dynamic equation, an outlier in any one series will affect the others because they are connected by Ω, but the effect will be mitigated because of the consequent downweighting from w_t. This weight is the same for all series - even if they are uncorrelated with each other.
The intuition lies in the fact that the multivariate t is constructed from normal variates with a common chi-squared denominator.

Multivariate models

The $\mathbf{u}_t$ vector may be modified by premultiplying by the inverse of the information matrix. Dropping constant terms then gives a generalization of the variable used in the univariate location model, namely

$$\mathbf{u}_t = w_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1})$$

An outlier in one series will not now affect the other elements of $\mathbf{u}_t$ directly, although there is still the downweighting from w_t. How the different models will finally work out depends on the constraints put on $\mathbf{K}$. If $\mathbf{K}$ is diagonal the specification of $\mathbf{u}_t$ will be crucial but this will not be the case if $\mathbf{K}$ is merely constrained to be symmetric.

Multivariate models

A multivariate local level DCS model can be specified by setting $\Phi = \mathbf{I}$. The specification can be adapted to allow for common trends. Thus

$$\begin{aligned}
\mathbf{y}_{1t} &= \boldsymbol{\mu}^{\dagger}_{t|t-1} + \boldsymbol{\nu}_{1t} \\
\mathbf{y}_{2t} &= \Pi\boldsymbol{\mu}^{\dagger}_{t|t-1} + \overline{\boldsymbol{\omega}} + \boldsymbol{\nu}_{2t}, \\
\boldsymbol{\mu}^{\dagger}_{t+1|t} &= \Phi\boldsymbol{\mu}^{\dagger}_{t|t-1} + \mathbf{K}^{\dagger}\mathbf{u}_t,
\end{aligned}$$

where $\boldsymbol{\mu}^{\dagger}_{t|t-1}$ is now a $J \times 1$ vector and $\mathbf{K}^{\dagger}$ is a $J \times N$ matrix.
For a single common trend, $\mathbf{K}^{\dagger\prime} = \boldsymbol{\kappa}$ is an $N \times 1$ vector. In a balanced growth model it may be appropriate to let $\boldsymbol{\kappa}$ be proportional to a vector of ones times the inverse of the information matrix.

Figure shows scatter plots of 3397 daily returns for IBM and General Motors from April 7th, 1986, to April 7th, 1999. The top graph is for the first 500 observations while the bottom graph is for the remainder. The correlation of 0.74 and slope of 0.76 for the earlier period contrasts with a correlation of 0.26 and a slope of 0.27 for the later one. The pair of observations for the crash of 1987, seen in the bottom left hand corner of the top graph, has had a considerable influence. Making allowance for conditional heteroscedasticity will not change the basic message.

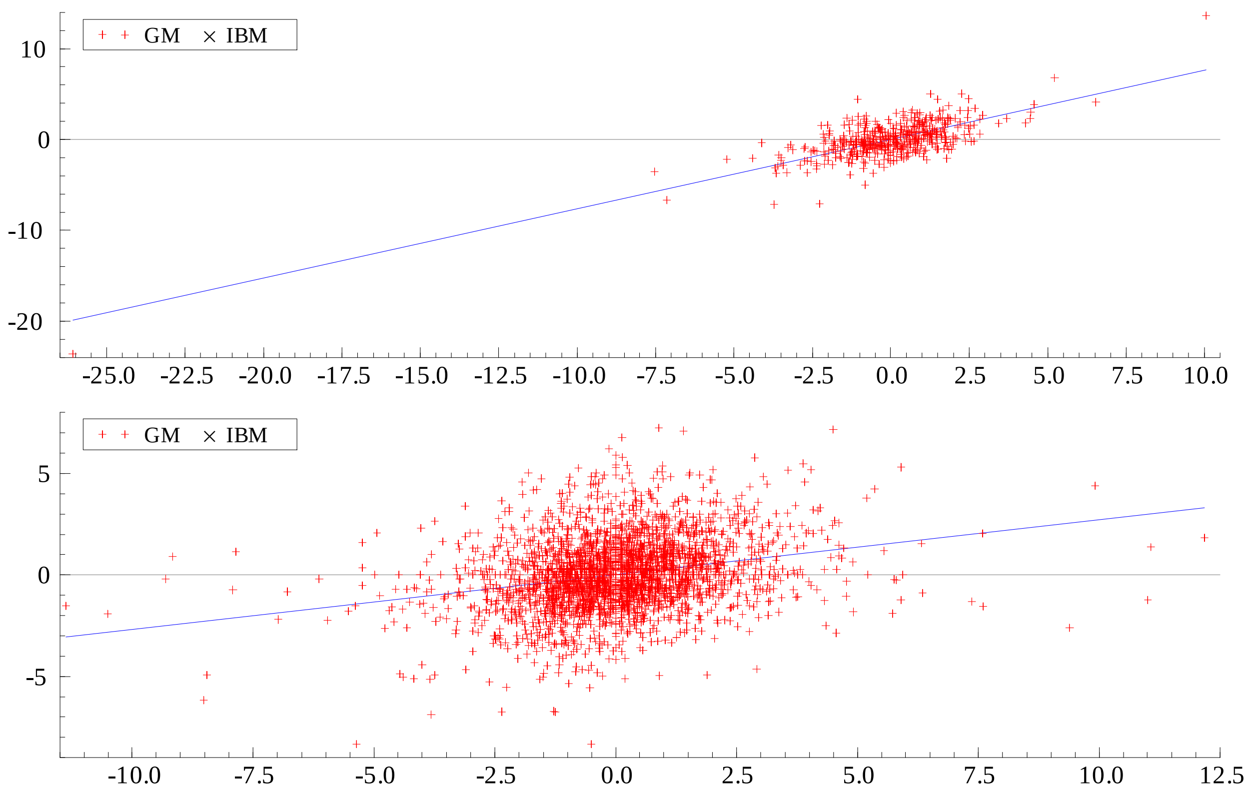


Figure: Daily returns for IBM against GM returns

Andrew Harvey (Cambridge University) Volatility and Heavy Tails / 66

Multivariate models for changing scale

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

Issues of interpretation aside, differentiation of the matrix exponential is not straightforward.

Better to follow the approach in Creal et al (2011, JBES) and let

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

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

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

Changes in correlation in the GM-IBM dataset used earlier are not just associated with the 1987 crash observations. When the first 3000 observations are divided into six group of 500, and the remaining 397 observations are iassigned to a seventh group, the correlation coefficients and Kendall's tau are as in the tablebelow. The correlations are higher at the beginning than at the end. The same is true of Kandall's tau, but because Kandall's tau is a robust measure of association it is less influenced by outliers and the differences are not so marked.

Group	1	2	3	4	5	6	7
r	.74	.51	.42	.16	.18	.10	.36
tau	.37	.34	.26	.10	.15	.08	.22

Table - Correlations and Kendall's tau for GM-IBM daily returns

Estimating changing correlation

Assume a bivariate model with a conditional Gaussian distribution. Zero means and variances time-invariant.
How should we drive the dynamics of the filter for changing correlation, $\rho_{t|t-1}$, *and with what link function ?*
Specify the standard deviations with an exponential link function so $Var(y_i) = \exp(2\lambda_i)$, $i = 1, 2$.
A simple moment approach would use

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

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

Estimating changing correlation

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

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

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

Estimating changing correlation

The dynamic equation for correlation is defined as

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

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

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

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

Copulas

When two series are non-Gaussian, correlation may not be the best way of capturing the association between them. The multivariate t-distribution is a partial attempt to break out of the Gaussian mould in that it accommodates heavy tails, but association is still measured by correlation. Copulas offer a more radical and flexible way of modeling association between variables that is independent of their marginal distributions. Modeling the relationshop between two variables in this way exploits the fact that the PIT of any random variable has a uniform distribution.

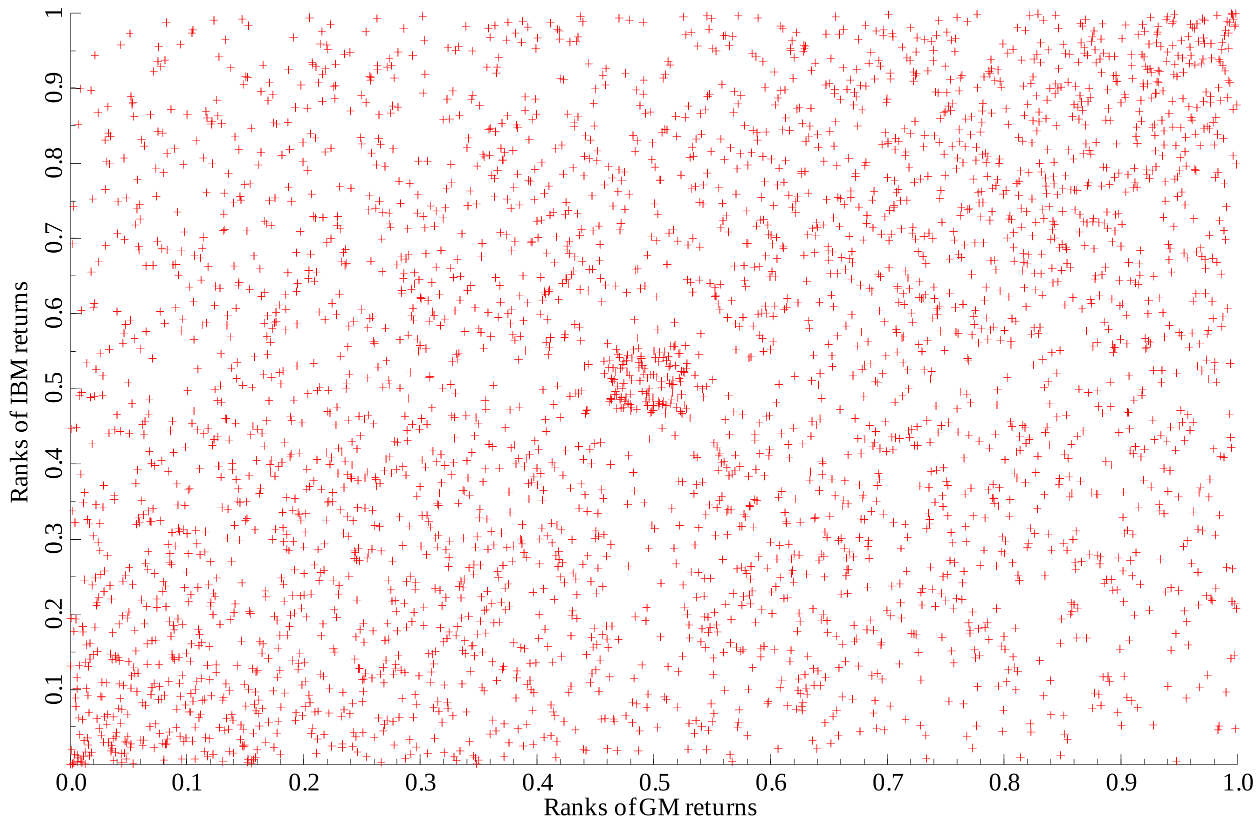


Figure: Scatter plot of ranked GM and IBM returns

Copulas and quantiles

Figure shows a scatter plot of the ranked GM and IBM returns. There is more association in the upper and lower tails than can be captured by a bivariate Gaussian distribution.
In population terms, the probability that an observation from the first series is less than the τ_1-quantile, $\xi(\tau_1)$, at the same time as the corresponding observation from the second series is below the τ_2-quantile, $\xi(\tau_2)$, is

$$\Pr(Y_1 \leq \xi(\tau_1), Y_2 \leq \xi(\tau_2)) = F(\xi(\tau_1), \xi(\tau_2)), \quad 0 \leq \tau_1, \tau_2 \leq 1.$$

Such probabilities are given by the *copula*, $C(\tau_1, \tau_2)$.

The copula is a joint distribution function of standard uniform random variables, that is

$$C(\tau_1, \tau_2) = \Pr(U_1 \leq \tau_1, U_2 \leq \tau_2), \quad 0 \leq \tau_1, \tau_2 \leq 1.$$

The **Clayton copula** is defined as

$$C(\tau_1, \tau_2) = \begin{cases} (\tau_1^{-\theta} + \tau_2^{-\theta} - 1)^{-1/\theta}, & \theta \in [-1, \infty),\ \theta \neq 0 \\ \tau_1 \tau_2, & \theta = 0. \end{cases}$$

Figure shows a scatter plot of two hundred observations generated with $\theta = 5$. The concentration of points in the lower left hand corner indicates *tail dependence*.

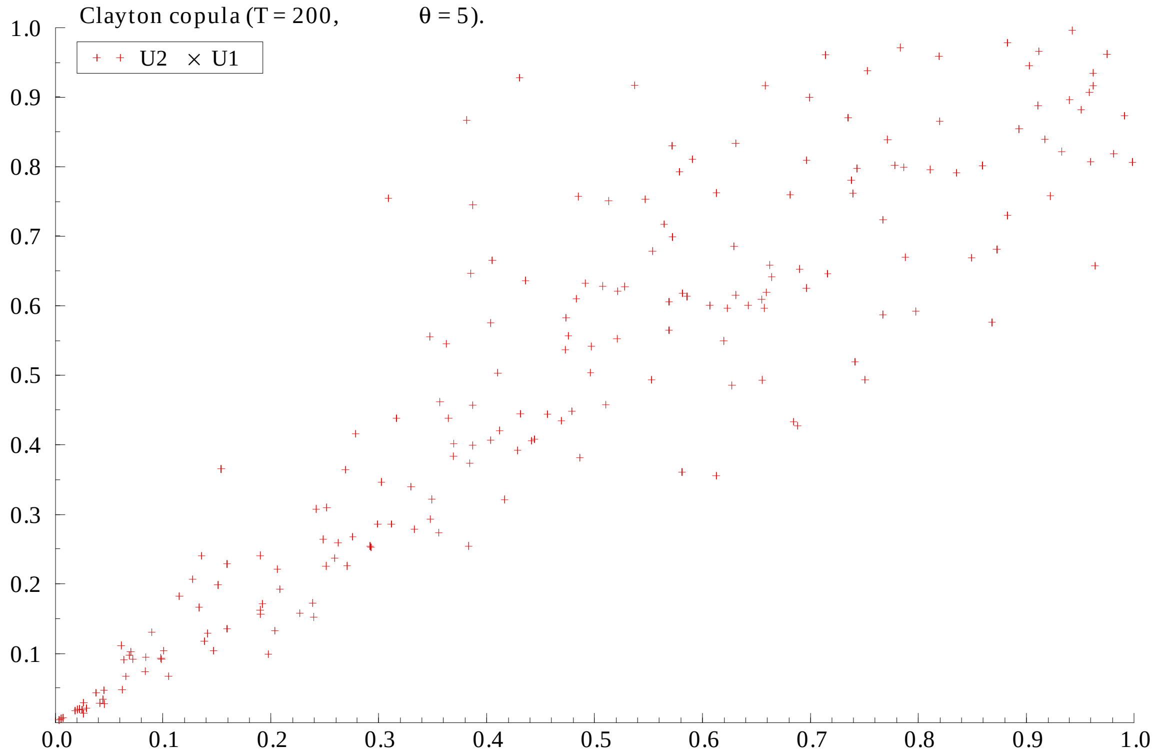


Figure: Scatter plot of 200 ranked observations from a Clayton copula with $\theta = 5$.

Since the PIT, $F(Y)$, of a random variable has a uniform distribution, the copula may be combined with the marginal distribution functions to give the full joint distribution function. Specifically, a copula computed at $\tau_1 = F_1(y_1)$, $\tau_2 = F_2(y_2)$ gives $F(y_1, y_2)$ because

$$\begin{aligned} C(F_1(y_1), F_2(y_2)) &= \Pr(U_1 \leq F_1(y_1), U_2 \leq F_2(y_2)) \\ &= \Pr(F_1^{-1}(U_1) \leq y_1, F_2^{-1}(U_2) \leq y_2) \\ &= \Pr(Y_1 \leq y_1, Y_2 \leq y_2) = F(y_1, y_2). \end{aligned}$$

When y_1 and y_2 are the quantiles $\xi(\tau_1)$ and $\xi(\tau_2)$, the copula is $C(\tau_1, \tau_2)$.

Sklar's theorem states that if $F(y_1, y_2)$ is a joint distribution function with continuous marginals $F_1(y_1)$ and $F_2(y_2)$, then there exist a unique copula. Marginal distributions do not need to be of the same form, nor is the choice of copula constrained by the choice of marginals. Hence, given the joint distribution function, the univariate marginals and the dependence structure can be separated, with the dependence structure represented by the copula.

When a Clayton copula is combined with marginal distributions of Y_1 and Y_2, both of which are exponential and so have CDF's $F_i(y) = 1 - \exp(-y_i/\alpha_i)$, $i = 1, 2$, the joint distribution function of Y_1 and Y_2 is

$$F(y_1, y_2) = ((1 - \exp(-y_1/\alpha_1)^{-\theta} + (1 - \exp(-y_2/\alpha_2)^{-\theta} - 1)^{-1/\theta}.$$

The copula density is

$$c(\tau_1, \tau_2) = \frac{\partial^2 C(\tau_1, \tau_2)}{\partial \tau_1 \partial \tau_2}, \quad 0 \leq \tau_1, \tau_2 \leq 1.$$

For a Clayton copula

$$c(\tau_1, \tau_2) = (1+\theta)\tau_1^{-\theta-1}\tau_2^{-\theta-1}\left(\tau_1^{-\theta} + \tau_2^{-\theta} - 1\right)^{-(1+2\theta)/\theta}, \quad 0 \leq \tau_1, \tau_2 \leq 1.$$

Figure shows the conditional distribution of τ_2 given that $\tau_1 = 0.1$ for $\theta = 1$, $\theta = 5$ and $\theta = 0$. When $\theta = 5$, the probability that τ_2 is close to the value taken by τ_1 is quite high. In contrast, τ_2 and τ_1 are independent when $\theta = 0$ and $c(\tau_2 \mid \tau_1) = 1$ for $0 \leq \tau_2 \leq 1$.

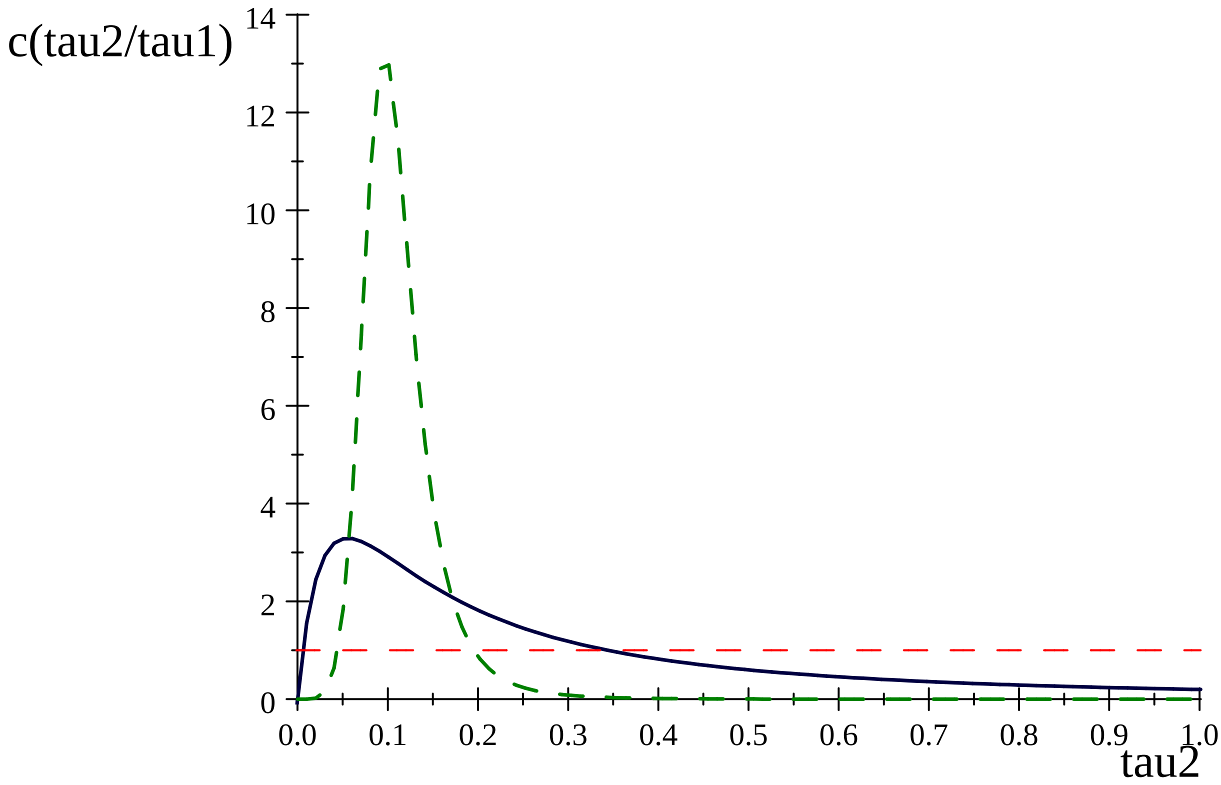


Figure: Conditional distribution of τ_2 given that $\tau_1 = 0.1$ for a Clayton copula with $\theta = 1$ (solid line), $\theta = 5$ (thick dashed) and $\theta = 0$ (thin dashed).

The joint probability density function of y_1 and y_2 is

$$f(y_1, y_2) = c(F(y_1), F(y_2)).f_1(y_1).f_2(y_2).$$

If the marginal densities are uniform, the joint density function is the copula density. If not, its shape is stretched and contracted by the form of the probability density functions.
When the variables are independent

$$C(\tau_1, \tau_2) = \Pr(U_1 \leq \tau_1).\Pr(U_2 \leq \tau_2) = \tau_1\tau_2, \quad 0 \leq \tau_1, \tau_2 \leq 1.$$

This is the product copula. The *copula density*, $c(\tau_1, \tau_2)$, is unity and so $f(y_1, y_2) = f_1(y_1).f_2(y_2)$.
As noted earlier, the variables in a bivariate normal or bivariate t distribution are linearly related. However, linear relationships are the exception rather than the rule. Bouyé and Salmon (2009) show how to derive the distribution of one variable conditional on another when they are related by a given parametric copula.

Measures of association

The *survival function*

$$\overline{C}(u_1, u_2) = \Pr(U_1 > \tau_1, U_2 > \tau_2),$$

gives the probability that two variables both lie above pre-assigned quantiles, that is $\overline{C}(\tau_1, \tau_2) = \Pr(y_{1t} > \xi_1(\tau_1), y_{2t} > \xi_2(\tau_2))$. The quadrant association,

$$\overline{C}(\tau_1, \tau_2) + C(\tau_1, \tau_2), \quad 0 \leq \tau_1, \tau_2 \leq 1,$$

gives a measure of dependence in the range $[0, 1]$. However, it can be shown that quadrant association depends only on $C(\tau_1, \tau_2)$ and is equal to $1 - \tau_1 - \tau_2 + 2C(\tau_1, \tau_2)$; see Cherubini et al (2004, p. 75) or McNeil *et al* (2005, p. 196).
There is *positive quadrant dependency* if $C(\tau_1, \tau_2) \geq \tau_1\tau_2$. *Blomqvist's beta*, $4C(0.5, 0.5) - 1$, is the quadrant association at $\tau_1 = \tau_2 = 0.5$, standardized so as to lie in the range $[-1, 1]$ and to be zero when the series are independent.

Measures of association

Conditional probabilities for measuring dependence depend on the copula. The probability that an observation from the first series is less than a given quantile, $\xi(\tau_1)$, given that the corresponding observation from the second series is below a given quantile, $\xi(\tau_2)$, is

$$F(\xi(\tau_1), \xi(\tau_2))/F(\xi(\tau_2)) = C(\tau_1, \tau_2)/\tau_2.$$

For the Clayton copula with $\tau_1 = \tau_2 = \tau$

$$C(\tau, \tau)/\tau = \left(2 - \tau^{\theta}\right)^{-1/\theta} \tag{4}$$

Figure plots $C(\tau, \tau)/\tau$ for three values of θ. When $\theta = 1$ the tail dependence for $\tau = 0.10$ is 0.526, but if $\theta = 5$, it goes up to 0.870. The conditional density of gives a complementary picture of the association.

Measures of association

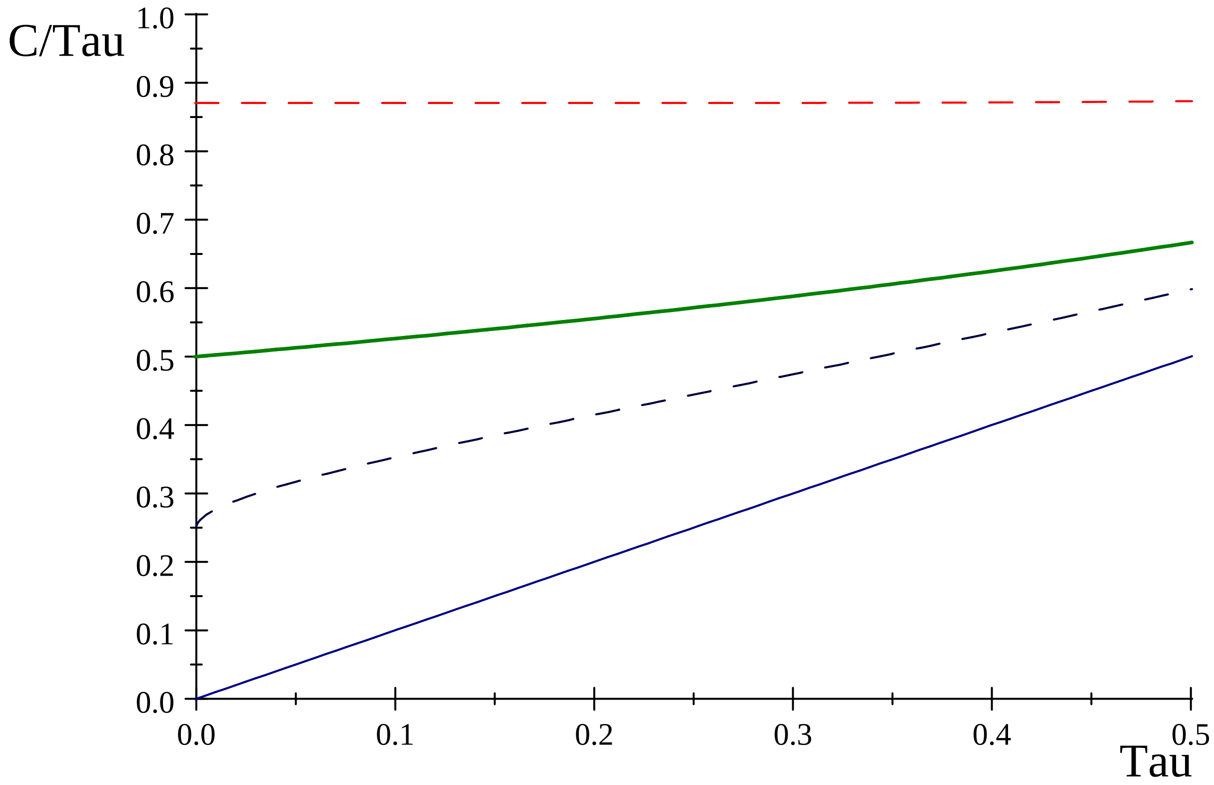


Figure: Lower tail dependence for Clayton copula

Measures of association

The coefficients of tail dependence provide measures that depend only on copula shape parameters; see McNeil *et al* (2005, p. 208). The *coefficient of lower (left) tail dependence*, or *lower tail index,* is

$$\lambda_L = \lim_{\tau \to 0} C(\tau, \tau)/\tau,$$

while the *coefficient of upper (right) tail dependence* is

$$\lambda_U = \lim_{\tau \to 1} \overline{C}(\tau, \tau)/(1 - \tau).$$

If two variables have a bivariate normal distribution, with $|\rho| < 1$, they are asymptotically independent in the tails as the coefficients of tail dependence are both zero.

Measures of association

For $\theta > 0$, the Clayton copula exhibits lower tail dependence, with $\lambda_L = 2^{-1/\theta}$, as is easily seen from (4). For $\theta = 1$, $C(\tau, \tau)/\tau \simeq 1/2$ for small τ and $\lambda_L = 0.5$. For $\theta = 5$, $\lambda_L = 0.870$, the same as was calculated for $\tau = 0.1$. As $\theta \to \infty$, $C(\tau, \tau)/\tau \to 1$. The practical implications are that with a small τ, such as 0.05 or 0.01, $C(\tau, \tau)$ may be close to τ and the probability of one variable being below its $\tau-$quantile given that the other is below its $\tau-$quantile is close to unity.

Maximum likelihood estimation

The log-likelihood function of the observations y_{1t}, y_{2t}, $t = 1, ..., T$, is

$$\ln L(\boldsymbol{\psi}) = \sum_{t=1}^{T} \ln c(F(y_{1t}), F(y_{2t})) + \sum_{t=1}^{T} \ln f_1(y_{1t}) + \sum_{t=1}^{T} \ln f_2(y_{2t}),$$

where $\boldsymbol{\psi}$ includes the parameters of both copula and marginals.
The calculations may be simplified by first estimating the parameters in the marginal distributions and then the copula. This is called the inference for the margins method. According to Cherubini *et al* (2004), it entails very little loss in efficiency.

Dynamic copulas: estimating changing association

Time-varying copulas can be modeled using the conditional score to drive a dynamic equation for the shape parameter. Since the conditional score takes account of the specification of the copula, it would seem to be a better way of proceeding than the essentially *ad hoc* approach of Patton (2006). Creal *et al* (2012) illustrate the viability and relevance of the DCS approach in an application of dynamic Gaussian copulas to exchange rate data.

First-order dynamic equation - possible RW. The form of the score for the joint density function depends only on the copula, but the marginal distributions affect its value through the probability integral transforms applied to the raw data. Unfortunately, the information quantity is not usually easy to derive, and so there is little hope of developing an asymptotic theory for ML estimation as in Lecture 1. Nevertheless, the simulation results for the Clayton copula reported by Creal *et al* (2012) show that ML estimation works well.

Dynamic copulas: estimating changing association

The conditional score for the Clayton copula is

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

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

Dynamic copulas: estimating changing association

First 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.

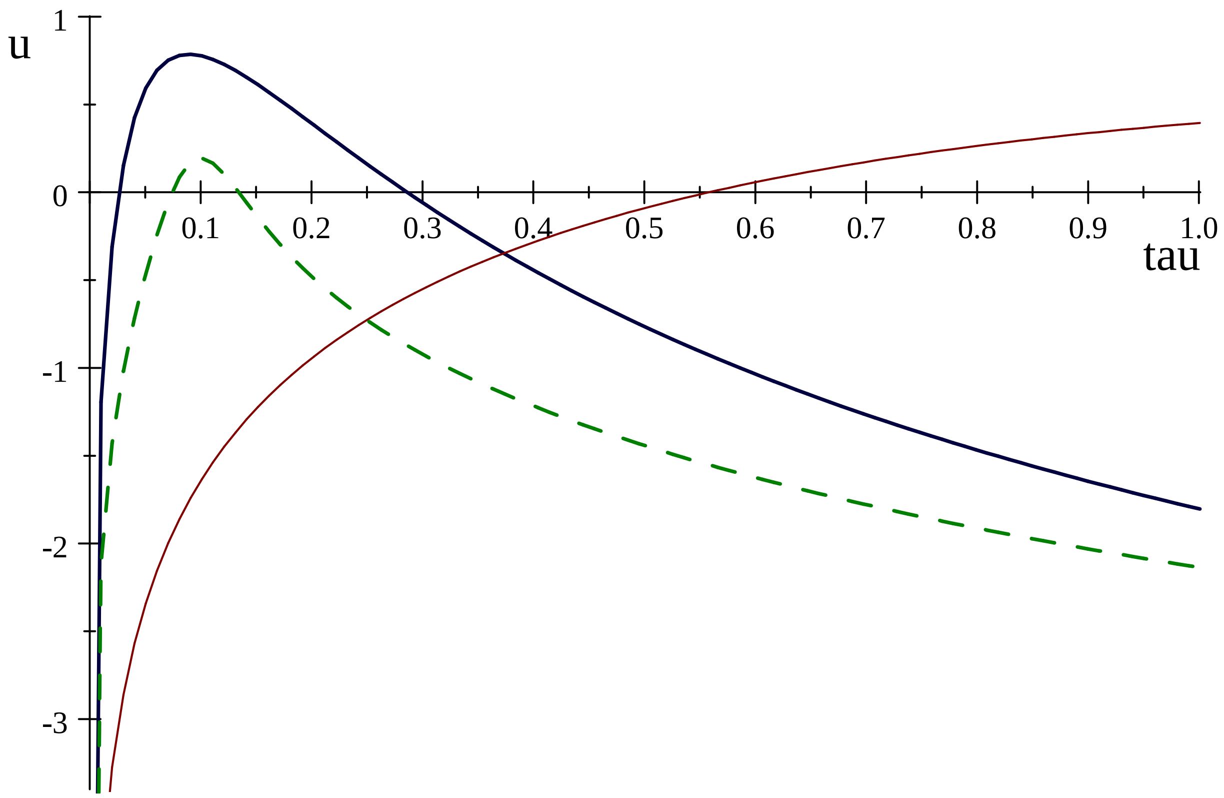


Figure: Response of u for fixed τ_1

Andrew Harvey (Cambridge University) Volatility and Heavy Tails / 66

The response shown in second figure is very different. Here $\theta = 0.0001$, so τ_1 and τ_2 are almost independent. When $\tau_1 = 0.1$, the score increases as τ_2 gets closer to zero, and decreases as it goes towards one. The opposite is true for $\tau_1 = 0.9$.

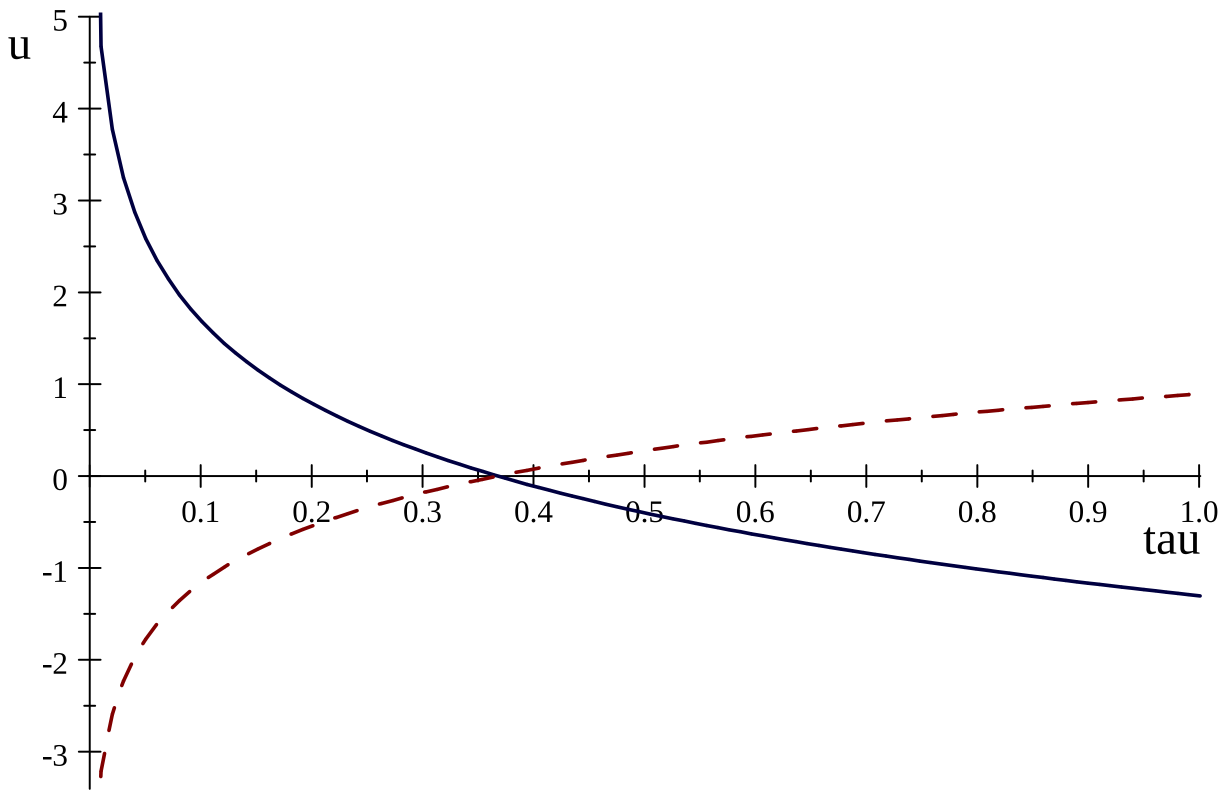


Figure: Response of score, u, for $\theta = 0.0001$ when τ_2 varies, but τ_1 is fixed at 0.1 (solid line) or 0.9 (dashed line).

Dynamic copulas: estimating changing association

A full maximum likelihood approach can, in principle, be used to jointly estimate dynamic volatility and copula parameters by defining $\ln L$ in terms of distributions conditional on past observations. However, a two-step procedure may be more appealing in practice. If a univariate Beta-t-EGARCH is fitted to each series, the PITs can be computed using a subroutine for a regularized incomplete beta function.
Probabilities associated with a constant copula can be estimated nonparametrically simply by counting the number of pairs of observations with the required property, for example both being below a certain quantile. A changing copula can be tracked by using a time series filter, such as an EWMA, to estimate the copula probabilities. Filtering to allow for changing volatility can be done nonparametrically, as described in Chapter 6, or parametrically, by using an EGARCH model. The application described in Harvey (2010) shows how the association between the Hong Kong (Hang Seng) and Korean stock market indices increased in the late 1990's.

Dynamic copulas: Tests against changing association

General tests against time-varying copulas are investigated in Busetti and Harvey (2011). The idea is to look at how the proportion of observations in particular quadrants changes over time. The default is to use a test based on medians, thereby essentially detecting movements in Blomqvist's beta. Such a test will be more robust than one based on the Gaussian score. However, scores derived from the multivariate–t will be less affected by outliers.
A test against time-variation in a given parametric copula could be based on the scores constructed by estimating a static copula from the PITs obtained from fitting univariate Beta-t-EGARCH models.

Conclusions and further directions

Many nonlinear unobserved components models can be approximated by analogous observation driven models in which the dynamics are driven by the score of the conditional distribution. Given a judicious choice of link function everything falls into place. When the dynamic conditional score model is taken to be the true model, the asymptotic theory for maximum likelihood estimation can be developed and in many cases an analytic expression can be derived for the information matrix.

Furthermore ML estimation seems to work well in practice. In all the applications reported here, convergence of the likelihood function for the models estimated was fast and reliable.

Conclusions and further directions

In much of the literature, the way in which observation-driven models are constructed is essentially arbitrarily. The use of the conditional score in models associated with unobserved components parameter driven models provides **guidance** and **discipline**, as well as a unified approach to nonlinear time series modeling.

Even when the asymptotic theory cannot be developed along the lines set out here, as is the case with any parameter associated with the distributions typically employed for count data and qualitative observations, the evidence suggest that the conditional score is still the best way forward.

Conclusions and further directions

A theory of testing, both before and after model estimation, is also developed and the evidence suggests that it is appropriate and effective. For example, the Lagrange multiplier tests for serial correlation are robust to heavy tails and the tests based on the probability integral transform are useful for assessing the validity of distributional assumptions.
The numerical standard errors computed for the parameters in models estimated from real data were sometimes found to be unreliable, particularly for shape parameters. On the other hand, analytic standard errors require moderate size samples to be close to the values indicated by Monte Carlo experiments.
Nevertheless it should be borne in mind that even when the standard errors are reliable, care has to be exercised in testing certain hypotheses, particularly those that pertain to a parameter being zero, since identifiability issues may render Wald tests invalid.

Conclusions and further directions

Extending dynamic conditional score modeling to multivariate time series appears to be relatively straightforward for time-varying location models based on the multivariate t-distribution. Generalizing Beta-t-EGARCH to multivariate series is more difficult. The best way forward seems to be to decompose the covariance, or scale, matrix into two parts, one of which is for the correlations. If only the volatilites change over time, the statistical treatment is relatively straightforward.
It is evolving correlations which pose the challenge. At the same time, the modeling of dynamic correlations, with or without heavy tailed distributions, offers a rich opportunity for extending the scope of time series econometrics. Rather than approaching the modeling of changing relationships from the standpoint of time-varying regression parameters, the signal extraction framework adopted in this monograph suggests that modeling changing correlations is likely to provide a more fruitful line of attack.

Conclusions and further directions

More generally, copulas offer a flexible way of modeling association between two variables that is independent of their marginal distributions. When a time-varying copula is to be modeled, letting the dynamic equations be driven by the conditional score seems to offer the most promising approach. As with dynamic correlations, the asymptotic theory developed here cannot be directly employed. Much remains to be done.