Dynamic distributions and changing copulas

Andrew Harvey,
Faculty of Economics, University of Cambridge

September 3, 2008

Abstract

A copula models the relationships between variables independently of their marginal distributions. When the variables are time series, the copula may change over time. A statistical framework is suggested for tracking these changes over time. When the marginal distributions change, pre-filtering is necessary before constructing the indicator variables on which the tracking of the copula is based. This entails solving an even more basic problem, namely estimating time-varying quantiles. The methods are applied to the Hong Kong and Korean stock market indices. Some interesting movements are detected, particularly after the attack on the Hong Kong dollar in 1997.

KEYWORDS: Concordance, contagion, exponentially weighted moving average; quantiles; signal extraction, tail dependence.

JEL Classification: C14, C22

1 Introduction

Stock returns are known to be non-normal with a distribution that changes over time. The most pervasive form of time variation is changing variance or volatility. However, features other than scale, such as skewness or kurtosis, may also change.

Just as the normal distribution is inadequate for modeling univariate time series, so the bivariate normal distribution is not suitable for modeling the relationship between two assets. As well as the asset returns not being normally distributed, their comovements may not adequately captured by correlation coefficients. For example, marginal distributions tend to be characterized by fat tails and the probability of two markets both exhibiting a relatively high movement (in same direction) may be much higher than can typically be captured with a bivariate normal distribution.

A copula models the relationships between two variables independently of their marginal distributions. It does so by means of a joint distribution function with standard uniform marginals. Hence it gives the probability that the observations in two series are each below certain quantiles. The separation of the dependence structure from the marginals introduces more flexibility into modeling.

There is evidence to suggest that copulas may sometimes change over time; see for example, Van Der Goorbergha, Genest and Werker (2005), Rodriguez (2007) and Patton (2006). The aim of this paper is to suggest a way in which this might be done. The proposed method is based on the filter used in Harvey and Fernandes (1989) to estimate the underlying probability in a binary series. The filter takes the form of an exponentially weighted moving average (EWMA). Although the construction of the filter draws on Bayesian technology in its use of conjugate distributions, it yields a likelihood function that can be maximized to give an estimate of the discount coefficient in the EWMA. Smoothed estimates can also be computed by drawing on the correspondence with the Gaussian local level model. The approach is different from the one employed by Patton (2006). He estimates conditional copula models in which the parameters are assumed to be functions of past observations

Tracking the movements in different parts of the copula may point to a variety of changes in the relationship between the two series. In particular we may wish to focus on movements in lower (upper) tail dependence as characterized by the probability that one series is below (above) a given quantile, given that the other is below (above) a given quantile. If a single measure of dependence is required, it may be appropriate to consider the probability that both observations are below their respective medians. A simple transformation of this measure yields *Blomqvist's beta*, which, because it lies in the range $[-1, 1]$, is comparable with other measures of association; see Kruskal (1958) and Fermanian and Scaillet (2003). At any point in time, Blomqvist's beta and measures of tail dependence are given directly from estimates of the copula. Approximations to Kendall's Tau and Spearman's rank correlation can also be obtained.

If the medians are constant over time, Blomqvist's beta is unaffected by changes in volatility. However, even if the medians are constant, which is not necessarily the case for stock returns, other parts of the copula will certainly be affected by changes in the marginal distributions. Hence some kind of pre-filtering is needed. The most general solution is to try to track the distribution functions of the marginals. This problem is a more fundamental one than tracking the copula, but it can be solved by generalizing the filter for binary observations so as to deal with categorical data. The categories correspond to parts of the distribution and the filters for the proportions in each category are EWMAs, as in the binary case. The discount factor, or factors, can be estimated by maximizing a likelihood function based on the multinomial distribution. Given these proportions, in what may be regarded as a time-varying histogram, quantiles can be estimated by interpolation. Indicator variables can be constructed using these time-varying quantiles and the copula estimated.

The plan of the paper is as follows. Section 2 discusses the method for estimating the changing probability in a binary time series and indicates its relevance to time-varying copulas. Section 3 sets out the proposal for estimating time-varying quantiles and explores the relationship of this method to non-parametric procedures, as in Yu and Jones (1998). Another possibility, that of estimating the density by a discounted kernel, is also discussed. Section 4 returns to the copula and sets out a method, similar to that adopted for computing changing proportions in univariate distributions, for simultaneously estimating all parts of the copula. The estimation of parametric copulas and measures of association, such as Spearman's rank correlation coefficient and Kendall's Tau, is then discussed. Section 5 applies the techniques to exploring the relationship between the Hong Kong (Hang Seng) and Korean (SET) stock price indices, with special emphasis on the issue of contagion stemming from the speculative attack on the Hong Kong dollar in 1997; see Dungey *et al* (2005). Tracking the copula provides a coherent description and provides some new insights. Figures 1 and 2 show daily returns[1] in the two markets from 27/11/79 to 27/11/07. Three events are marked: (i) Black Monday, October, 19th, 1987; (ii) the attack on Hong Kong dollar on 20 Oct 1997; (iii) the 'high tech.' crash of 2nd October 2000. The increase in volatility immediately after 20th October 1997 is clearly discernible.

[1] The Hong Kong and Korean stock price indices are in local currency - the Hong Kong dollar and Korean won respectively.

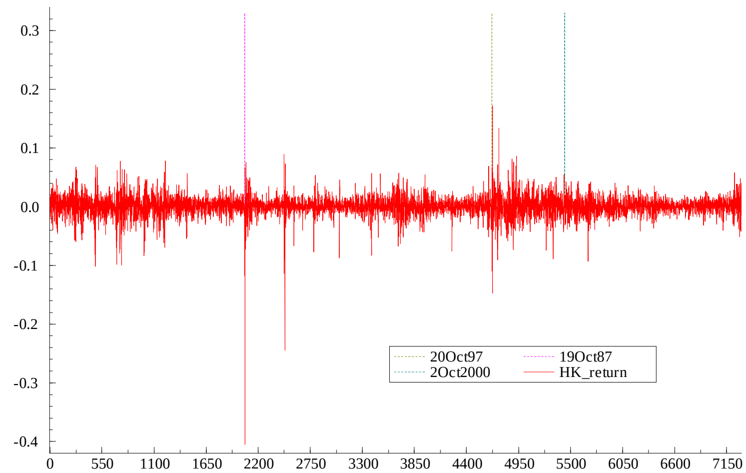

Figure 1: Hong Kong stock returns

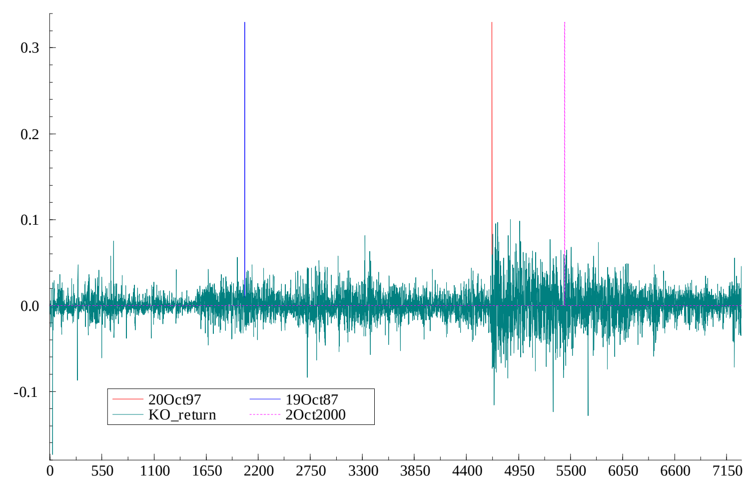

Figure 2: Returns on the Korean SET index

2 Tracking changes in the copula

2.1 Copulas

A copula is a joint distribution function of standard uniform random variables, U_1 and U_2, that is

$$C(u_1, u_2) = \Pr(U_1 \leq u_1, U_2 \leq u_2), \quad 0 \leq u_1, u_2 \leq 1$$

When the variables are independent, $C(u_1, u_2) = \Pr(U_1 \leq u_1). \Pr(U_2 \leq u_2) = u_1 u_2$.

The copula gives the probability that observations on two variables, Y_1 and Y_2 are less than or equal to given quantiles, that is

$$C(\tau_1, \tau_2) = \Pr(Y_1 \leq \xi_1(\tau_1), Y_2 \leq \xi_2(\tau_2)) = F(\xi_1(\tau_1), \xi_2(\tau_2)), \qquad t = 1, ..., T \tag{1}$$

where $\xi_i(\tau_i)$ is the τ_i-quantile for $i = 1, 2$. The probability that both observations lie above their pre-assigned quantiles is known as the survival function and it is equal to

$$\overline{C}(\tau_1, \tau_2) = \Pr(Y_1 > \xi_1(\tau_1), Y_2 > \xi_2(\tau_2)) = 1 - \tau_1 - \tau_2 + C(\tau_1, \tau_2); \tag{2}$$

see, for example, Cherubini et al (2004, p75) or McNeil *et al* (2005, p196). Note that $\overline{C}(0.5, 0.5) = C(0.5, 0.5)$.

The copula provides a flexible way of capturing dependence. The variables Y_1 and Y_2 are said to exhibit *positive quadrant dependence* if $C(\tau_1, \tau_2) \geq \tau_1 \tau_2$. The quadrant association, $\overline{C}(\tau_1, \tau_2) + C(\tau_1, \tau_2)$, gives a measure of dependence in the range $[0, 1]$; see Kruskal (1958, p 818). It can be seen from (2) that quadrant association is a function of $C(\tau_1, \tau_2)$. *Blomqvist's beta*, $2(\overline{C}(0.5, 0.5) + C(0.5, 0.5)) - 1 = 4C(0.5, 0.5) - 1$, lies in the range $[-1, 1]$ and is zero when the series are independent.

Lower tail dependence, $C(u_1, u_2)/u_2 = \Pr(U_1 \leq u_1 \mid U_2 \leq u_2)$, is the probability that an observation from the first series is below u_1, given that the observation from the second series is below u_2. Upper tail dependence is $\overline{C}(u_1, u_2)/u_2$. As an example, the Clayton copula is defined as

$$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \qquad \theta \in [-1, \infty) \tag{3}$$

Lower tail dependence with respect to the $\tau-$quantile is

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

For $\theta = 1$, $C(\tau, \tau)/\tau$ is 0.526 for $\tau = .10$, but if $\theta = 5$ it goes up to 0.870. The coefficient (index) of lower (left) tail dependence is $\lambda_L = \lim_{\tau \to 0} C(\tau, \tau)/\tau$, which for the Clayton copula with $\theta > 0$ is $\lambda_L = 2^{-1/\theta}$. The coefficient of upper tail dependence is $\lambda_U = \lim_{\tau \to 0} \overline{C}(\tau, \tau)/(1 - \tau)$, which for the Clayton copula is zero.

Figure 3 shows a scatter plot of the ranks, divided by T, of $T = 200$ observations generated from a Clayton copula with $\theta = 5$. The strong lower tail dependence shows up in the concentration of points in the lower left hand corner.

There are other characteristics of copulas apart from association. The *survival copula,* denoted here as $\overline{C}^*(1 - u_1, 1 - u_2)$, is equal to $\overline{C}(u_1, u_2)$ and two variables are said to be (radially) *symmetric* if and only if $\overline{C}^*(1 - u_1, 1 - u_2) = C(u_1, u_2)$. Thus $C(.25, .25) - \overline{C}(.75, .75)$ and $C(.10, .10) - \overline{C}(.90, .90)$ might give informative measures of asymmetry.

2.2 Estimation of a constant copula

The copula can be estimated by counting the number of pairs of observations less than or equal to the relevant quantiles, and dividing by T; the estimator of (1) will be denoted by $\widehat{C}(\tau_1, \tau_2)$. The same estimator can be obtained from the ranks, $r_{1,t}, r_{2,t}, t = 1, .., T$. The scatter plot of the ranks, divided by T, is defined on a lattice, in the unit square, in which each axis is broken into T equal spaces delineated by the points i_1/T and i_2/T,with $i_1, i_2 = 0, .., T$. This forms the domain of the *empirical copula*, defined as

$$\widehat{C}(i_1/T, i_2/T) = \frac{1}{T} \sum_{t=1}^{T} I(r_{1,t} \leq i_1) I(r_{2,t} \leq i_2)$$

where $I(r_{1,t} \leq i_1)$ is the indicator function. The empirical copula frequency, $\widehat{c}(i_1/T, i_2/T)$, is $1/T$ if the ranked observations i_1 and i_2 are elements of the sample, and is zero otherwise; see Nelsen (1999, p 219).

A *grouped empirical copula* can be constructed by first defining τ_{1j} and τ_{2k} so as to partition the unit interval on the u_1 and u_2 axes into n sub-intervals, $0 = \tau_{10} < \tau_{11} < \tau_{12} < ... < \tau_{1,n} = 1$ and similarly for $\tau_{2k}, k = 0, 1, ..., n$. To simplify matters it will be assumed that the sub-intervals are equal. The estimates of the copula for the whole grid are then given by

$$\widehat{C}(\tau_{1j}, \tau_{2k}) = \frac{1}{T} \sum_{t=1}^{T} I(r_{1,t}/T \leq \tau_{1j}) I(r_{2,t}/T \leq \tau_{2k}), \quad j, k = 1, ..., n,$$

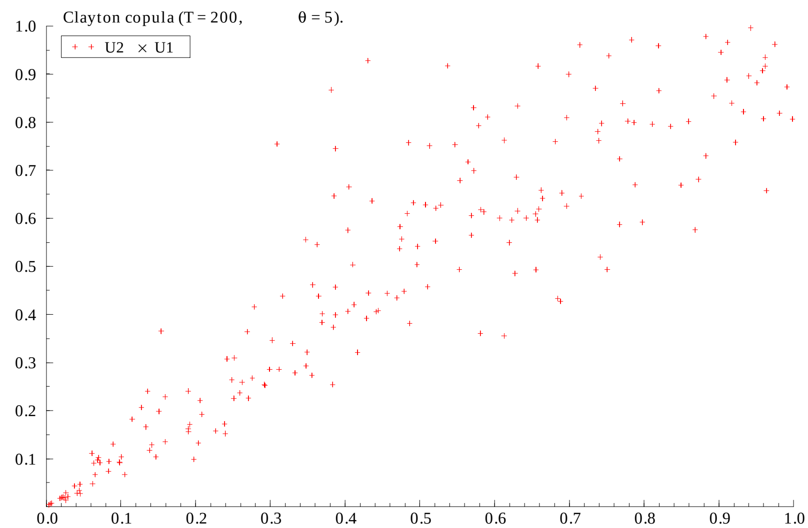

Figure 3: Scatter plot of 200 ranked observations from a Clayton copula with $\theta = 5$.

with $\widehat{C}_t(\tau_{1,n}, \tau_{2,n}) = 1$ by definition. The *grouped empirical copula frequency,* $\widehat{c}(\tau_{1j}, \tau_{2k})$, is given by the proportion of observations in each of the n^2 squares. Figure 4 shows the grouped empirical copula frequency for the scatter plot in figure 3. The reason for wanting to work with grouped observations will become apparent when we estimate changing copulas.

2.3 Tracking movements in a binary series

Consider a binary series, I_t, taking the value 0 or 1. At any point in time

$$E(I_t) = \pi_t, \qquad t = 1, ..., T \tag{5}$$

where $\pi_t = \Pr(I_t = 1), t = 1, ..., T$. This probability may be estimated by a filter of the form described in Smith (1981) and Harvey and Fernandes (1989); see also Harvey (p350-60). It is assumed that the distribution of π_{t-1}, given information up to and including time $t-1$ is beta with parameters $a_{1,t-1}$ and $a_{2,t-1}$. Then the distribution of π_t, given information up to and including time $t-1$, is beta with parameters

$$a_{1,t|t-1} = \omega a_{1,t-1}, \quad a_{2,t|t-1} = \omega a_{2,t-1}, \qquad t = 1, ..., T \tag{6}$$

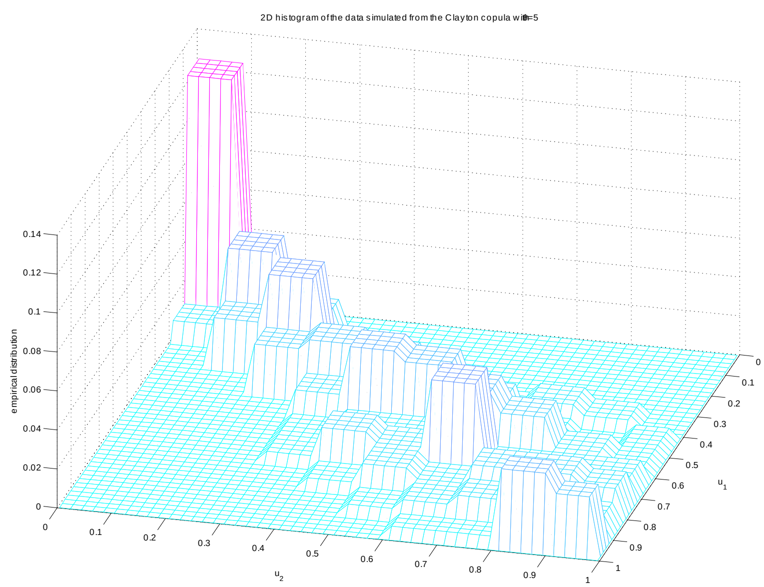

Figure 4: Bivariate histogram of 200 observations simulated from a Clayton copula with $\theta = 5$.

where ω is a discount parameter, $0 < \omega \leq 1$. When the t-th observation becomes available, the conjugacy of the beta and binomial distributions leads to the updating equations

$$a_{1,t} = a_{1,t|t-1} + I_t, \qquad a_{2,t} = a_{2,t|t-1} + 1 - I_t, \quad t = 1, ..., T \tag{7}$$

so providing the parameters for the new beta distribution. The recursions may be initialized with a non-informative uniform prior, which means that $a_{1,0} = 1$ and $a_{2,0} = 1$.

The estimated proportion at time t is the mean of the conditional distribution of π_t, that is

$$E(\pi_t | I_j(\tau), j = 1, ..., t) = \widetilde{\pi}_t = a_{1t}/(a_{1t} + a_{2t}), \qquad t = 1, ..., T.$$

The variance is

$$Var(\pi_t | I_j(\tau), j = 1, ..., t) = \frac{a_{1,t} a_{2,t}}{(a_{1,t} + a_{2,t})^2 (a_{1,t} + a_{2,t} + 1)}. \tag{8}$$

A more general formulation has the series recording the number of hits, y_t, from a binomial distribution with n_t trials. The updating equations are as in (7), but with I_t replaced by y_t and $1 - I_t$ replaced by $n_t - y_t$. The predictive distribution for y_t is beta-binomial. However, with a binary series, the predictive distribution reduces to a Bernouilli distribution and the log-likelihood function is therefore

$$\log L(\omega) = \sum_{t=2}^{T} \{ I_t \ln \widetilde{\pi}_{t|t-1} + (1 - I_t) \ln(1 - \widetilde{\pi}_{t|t-1}) \}, \tag{9}$$

with $\widetilde{\pi}_{t|t-1} = \widetilde{\pi}_{t-1}$.

The filtered estimates are an EWMA in the indicators, that is

$$\widetilde{\pi}_t = \frac{\sum\limits_{j=0}^{t-1} \omega^j I_{t-j} + a_{1,0} \omega^t}{\sum\limits_{j=0}^{t-1} \omega^j + (a_{1,0} + a_{2,0}) \omega^t}, \quad t = 1, 2, ... \tag{10}$$

with $a_{1,0} = a_{2,0} = 1$ for a non-informative prior. The median lag in this EWMA is $\ln(0.5)/\ln \omega - 1 = -0.693/\ln \omega - 1$.

The fact that the predictive distribution depends only on $\widetilde{\pi}_{t+1|t}$ suggests calculating an approximation to (10), and hence $\widetilde{\pi}_{t+1|t}$, from the following EWMA[2]

$$\widetilde{\pi}_{t+1|t}=(1-\omega)I_t+\omega\widetilde{\pi}_{t|t-1},\quad t=1,..,T, \tag{11}$$

with $\widetilde{\pi}_{1|0}=1/2$. Hence, the recursions for $a_{1,t}$ and $a_{2,t}$ are unnecessary. From (8), the MSE of $\widetilde{\pi}_t$ for large t is

$$MSE(\widetilde{\pi}_t)\simeq\widetilde{\pi}_t(1-\widetilde{\pi}_t)(1-\omega)/(2-\omega). \tag{12}$$

If π_t were fixed the same MSE would be obtained with a sample size of approximately $(2-\omega)/(1-\omega)$. When ω is close to one, $MSE(\widetilde{\pi}_t)\simeq\widetilde{\pi}_t(1-\widetilde{\pi}_t)(1-\omega)$. Thus for $\omega=.99$, the RMSEs for $\pi=0.5, 0.25$ and 0.1 are approximately, .050, .043 and .030 respectively.

A two-sided smoothed estimator of π_t corresponding to the EWMA filter may be constructed by analogy with a Gaussian random walk plus noise model in which the signal-noise ratio, that is the variance of the disturbance driving the random walk to the variance of the noise, is

$$q=(1-\omega)^2/\omega.$$

An efficient algorithm saves the filtered estimates, $\widetilde{\pi}_{t|t-1}$, and then calculates

$$r_{t-1}=\omega(r_t+I_t-\widetilde{\pi}_{t|t-1}),\qquad t=T,...,2,$$

where $r_T=0$. The smoothed estimates are then given by the forward recursion

$$\widetilde{\pi}_{t|T}=\omega\widetilde{\pi}_{t|t-1}+(1-\omega)(r_t+I_t),\qquad t=1,...,T,$$

with $\widetilde{\pi}_{1|T}=r_1+I_1$; see appendix. In the middle of a large sample

$$\widetilde{\pi}_{t|T}\simeq\frac{1-\omega}{1+\omega}\sum_j\omega^{|j|}I_{t+j} \tag{13}$$

The adoption of the recursion in (11) suggests the possibility of a change in interpretation whereby the model is *defined* by the predictive distribution. Such a model is said to be 'observation driven'. The role of $\widetilde{\pi}_{t|t-1}$ is analogous

[2] Although the estimates of π_t obtained from (10) and the preceding recursion are identical for small t, the notation does not distinguish between the two.

to that of the variance in a GARCH(1,1) model.[3] An advantage of this revised interpretation is that observations may be simulated. In the earlier setup the transition equation leading to (6) is only defined implicitly; see the discussion in Smith and Miller (1986). Another advantage of letting the predictive distribution define the model is that (11) may be modified to yield different dynamics. In particular, we might consider the filter

$$\pi_{t+1|t} = (1 - \omega^* - \omega)\pi^* + \omega^* I_t + \omega \pi_{t|t-1}, \quad t = 1, .., T, \tag{14}$$

where the notation $\pi_{t+1|t}$ accords with that used by Andersen et al (2006) for the variance in a GARCH model. This filter is stable if $\omega^* + \omega < 1$, but reverts to the EWMA if $\omega^* + \omega = 1$. Estimation may be simplified by setting π^* equal to the (unconditional) proportion in the sample; this is similar to the use of 'variance targeting' in GARCH estimation, as in Laurent (2007, p25).

2.4 The changing copula

Consider two serially independent time series, with time invariant marginal distributions, observed as y_{1t} and y_{2t}, $t = 1, ..., T$. Let $C_t(\tau_1, \tau_2)$ denote the copula at time t for $t = 1, ..., T$. The indicator variable taking the value one if both observations are less than or equal to pre-assigned quantiles, that is $I(y_{1t} \leq \xi_1(\tau_1)).I(y_{2t} \leq \xi_2(\tau_2)), t = 1, ..., T$, contains information on changes in the copula since its expected value is $C_t(\tau_1, \tau_2)$. If the quantiles are unknown then the indicator is replaced by the sample indicator

$$I_t(\tau_1, \tau_2) = I(y_{1t} \leq \widetilde{\xi}_1(\tau)).I(y_{2t} \leq \widetilde{\xi}_2(\tau)), \qquad t = 1, ..., T \tag{15}$$

The filter of the previous sub-section may be applied with $C_t(\tau_1, \tau_2)$ playing the role of π_t. Thus

$$C_{t+1|t} = (1 - \omega) I_t(\tau_1, \tau_2) + \omega C_{t|t-1}, \quad t = 1, .., T$$

A suitable initialization is obtained by noting that independence implies that $C_{1|0}(\tau_1, \tau_2) = \tau_1 \tau_2$.

If the copula is constant, the estimates of $C(\tau_1, \tau_2)$ and $\overline{C}(\tau_1, \tau_2)$ satisfy an identity of the form (2). This is no longer the case when filtering. This suggests the use of an estimator of $\overline{C}(\tau_1, \tau_2)$ to help to estimate the movements in $C_t(\tau_1, \tau_2)$ more accurately. The indicator defined by

[3] The notational convention adopted by Andersen et al (2006) in their review of GARCH models is $\sigma^2_{t|t-1}$, rather than simply σ^2_t,stressing that $\sigma^2_{t|t-1}$ is a filter.

$I(y_{1t} > \widetilde{\xi}_1(\tau)).I(y_{2t} > \widetilde{\xi}_2(\tau)) = \overline{I}_t(\tau_1, \tau_2)$ may be used to estimate $\overline{C}_t(\tau_1, \tau_2)$, initialized with $(1 - \tau_1)(1 - \tau_2)$. The *modified estimator* of $C_t(\tau_1, \tau_2)$ is

$$\widehat{C}_t(\tau_1, \tau_2) = \frac{\widetilde{C}_t(\tau_1, \tau_2) + \widetilde{\overline{C}}_t(\tau_1, \tau_2) - 1 + \tau_1 + \tau_2}{2}. \tag{16}$$

We might proceed by estimating the quadrant association, $\widetilde{C}_{QA,t}(\tau_1, \tau_2) = \widetilde{C}_t(\tau_1, \tau_2) + \widetilde{\overline{C}}_t(\tau_1, \tau_2)$, by adding the indicators $I_t(\tau_1, \tau_2)$ and $\overline{I}_t(\tau_1, \tau_2)$ and initializing with $1 - \tau_1 - \tau_2 + 2\tau_1\tau_2$.

Lower tail dependence is estimated from (16) as $\widehat{C}_t(\tau_1, \tau_2)/\tau_2$, $\tau_2 \leq 0.5$. The formula for upper tail dependence is

$$\frac{\widehat{\overline{C}}_t(\tau_1, \tau_2)}{1 - \tau_2} = \frac{\widehat{C}_t(\tau_1, \tau_2) + 1 - \tau_1 - \tau_2}{1 - \tau_2}, \quad \tau_2 > 0.5$$

The emphasis will usually be on the movements in these measures when $\tau_1 = \tau_2$. For $\tau_1 = \tau_2 = 0.5$ the lower and upper tail dependence measures are both equal to the quadrant association. Note that Blomqvist's beta is $2(\widetilde{C}_t(\tau_1, \tau_2) + \widetilde{\overline{C}}_t(\tau_1, \tau_2)) - 1 = 2\widetilde{C}_{QA,t}(\tau_1, \tau_2) - 1$.

Figure 6 shows estimates of the quadrant association, $\widetilde{C}_{QA}(\tau_1, \tau_2)$, for $\tau = 0.25, 0.5$ and 0.75 from 2000 observations simulated from a bivariate normal distribution in which the marginals are constant but the correlation coefficient changes from zero to 0.75 half way through the sample. The discount factor is set at 0.995. Figure 5 contrasts $\widetilde{C}_t(\tau, \tau)$ with $\widehat{C}_t(\tau, \tau)$. It can be seen that $\widetilde{C}_t(\tau, \tau)$ is far more variable.

In figure 6, $\widetilde{C}_{QA}(\tau, \tau)$ hovers around 0.5 for the first 1000 observations, and then with the introduction of correlation it rises to a new level of between 0.7 and 0.8, reaching 0.7 after approximately 200 observations.[4] For a bivariate Gaussian distribution, the correlation is related to the quadrant association at $\tau = 0.5$ by the formula $C_{QA}(0.5, 0.5) = 0.5 + (1/\pi) \arcsin \rho$. Thus $\rho = 0.75$, corresponds to $C_{QA}(0.5, 0.5) = 0.77$.

Note that lower tail dependence is $(\widetilde{C}_{QA}(\tau, \tau) - 1 + 2\tau)/2\tau$, while upper tail dependence is $(\widetilde{C}_{QA}(\tau, \tau) + 1 - 2\tau)/2(1 - \tau)$. A quadrant association of 0.8 yields tail dependence coefficients of 0.6 for $\tau = 0.25$ and 0.75. Indeed this will always be the case for complementary $\tau' s$, that is τ and $1 - \tau$. Recall that for $\tau = 0.5$, tail dependence is equal to quadrant association.

[4] The quadrant association test statistics of Busetti and Harvey (2008) for $\tau = .1, .25, .5, .75$ and $.9$ are 3.87, 5.21, 7.40, 4.20 and 2.07 respectively. All are highly significant.

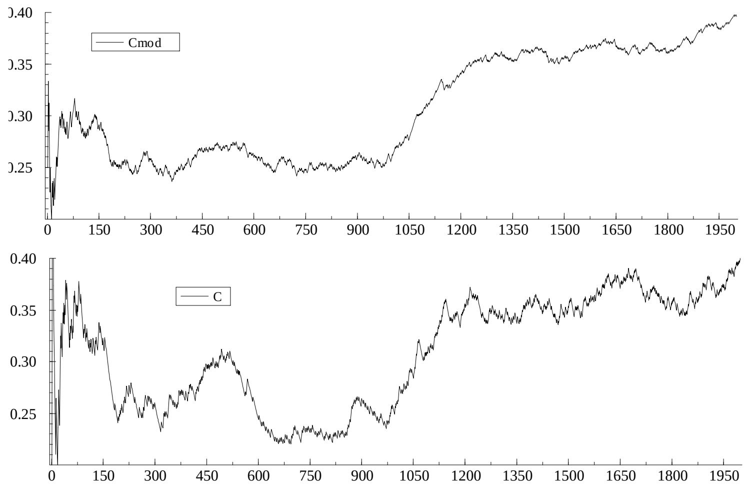

Figure 5: Modified estimator, $\widehat{C}_t(\tau_1, \tau_2)$, in upper graph, compared with direct estimator, $\widetilde{C}_t(\tau_1, \tau_2)$.

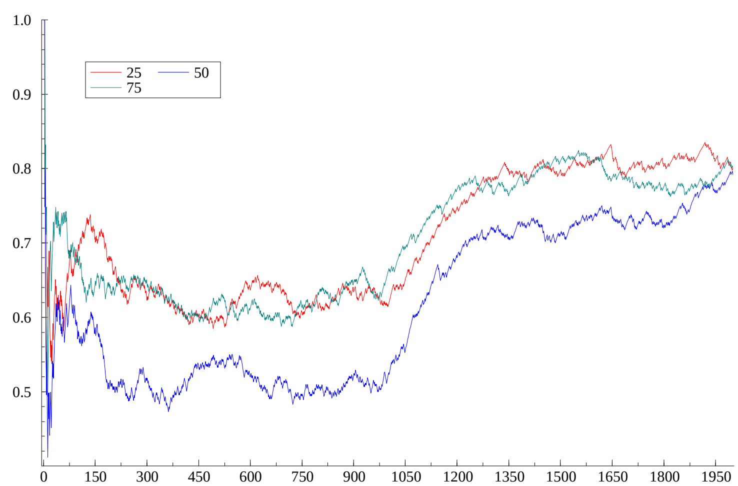

Figure 6: Quadrant association for $\tau = 0.25, .5$ and 0.75.

Changing asymmetry may be tracked with estimators of $(C(\tau,\tau)-\overline{C}^*(1-\tau,1-\tau))/\tau$ for, say, $\tau = .25$ and $.10$. The modified estimator is

$$\begin{aligned}
& \frac{\widetilde{C}_t(\tau,\tau)+\widetilde{\widetilde{C}}_t(\tau,\tau)-\widetilde{C}_t(1-\tau,1-\tau)-\widetilde{\widetilde{C}}_t(1-\tau,1-\tau)}{2\tau} \qquad (17) \\
= \quad & \widehat{C}_t(\tau,\tau)\tau-\widehat{C}_t(1-\tau,1-\tau))/\tau
\end{aligned}$$

Dividing by τ means that maximum asymmetry gives a measure equal to one (perfect lower TD and no upper TD) or minus one (perfect upper TD and no lower TD).

2.5 Changing marginals

When the marginal distributions are symmetric about a constant mean, the median is constant, so $C(0.5, 0.5)$ can be tracked with no pre-filtering. The fact that $C(0.5, 0.5)$ may be estimated without estimating the quantiles of the marginals or correcting for stochastic volatility is an important advantage over other measures of association.

If there are only changes in scale, the marginals can be standardized by dividing by a measure of volatility. The time-varying copula can then be fitted.

More generally, if the marginals change over time, the indicator variables need to be defined in terms of changing quantiles, $\xi_{1t}(\tau_1)$ and $\xi_{2t}(\tau_2)$.

2.6 Changing joint distributions

For some purposes estimating the joint distribution rather than the copula may be what is required; see the discussion on coexceedances in Bae, Karolyi and Stulz (2003). Thus the estimated probabilities, $\widehat{F}_t(\xi_{1t}(\tau_1), \xi_{2t}(\tau_2)) = \widehat{C}_t(\tau_1, \tau_2)$, are plotted against the (time-varying) quantiles, $\xi_1(\tau_1)$ and $\xi_2(\tau_2)$, rather than against τ_1 and τ_2.

When the observations are no longer IID, the joint distribution may still be estimated with respect to time-invariant (unconditional) quantiles. Tracking movements in $\widehat{F}_t(\xi_1(\tau_1), \xi_2(\tau_2))/\widehat{F}_t(\xi_2(\tau_2))$ may still be useful but are no longer the same as tail dependence in the copula. Changes in the marginals, for example in volatility, will be reflected in movements in $\widehat{F}_t(\xi_1, \xi_2)$.

The joint distribution may also be plotted against the (time-varying) quantiles, $\xi_{1t}(\tau_1)$ and $\xi_{2t}(\tau_2)$, that is $\widehat{F}_t(\xi_{1t}(\tau_1), \xi_{2t}(\tau_2)) = \widehat{C}_t(\tau_1, \tau_2)$.

3 Tracking a univariate distribution

Changing quantiles may be estimated[5] non-parametrically, as in Yu and Jones (1998), by minimizing a local check function. De Rossi and Harvey (2006, 2008) construct a similar estimator by assuming that $\xi(\tau)$ follows a time series model and show how this is related to the non-parametric estimator. An advantage of the more model-based approach is that it automatically determines a weighting pattern at the end of the sample that is consistent with the one in the middle. De Rossi and Harvey (2006) suggest estimating parameters by cross-validation. Unfortunately this turns out to be very time-consuming, particularly if the quantile changes relatively slowly over time.

The approach proposed here tackles the problem indirectly. However, it is somewhat more limited than the method in De Rossi and Harvey (2008) insofar as it is only suitable for series, such as returns, that are stationary or close to being stationary. The statistical method generalizes the one for binary series to categorical data; see Harvey (1989, p 355-6). Indicator variables are defined according to whether each observation lies within a particular pre-assigned range. In the basic model, the discount factor is the same for all groups and it may be estimated by maximizing a likelihood function that comes from a multinomial distribution. The filter tracks the proportions in each category and, from these proportions, time-varying quantiles are extracted. Smoothed estimates can also be computed. It follows from the method of construction that the quantiles cannot cross. The method can be generalized to have discount factors that vary over groups and over time.

3.1 Time-varying histograms

A changing distribution function can be estimated and tracked by first dividing the support into N categories, defined by the boundaries $-\infty, \xi_1, ..., \xi_{N-1}, \infty$. Let $I_{t,j} = 1$ if the observation is in category j, that is $\xi_{j-1} < y_t \leq \xi_j$, $j = 1, ..., N$, and zero otherwise. Then, introducing a discount factor, ω, that is common to all categories, and assuming that the proportions in each category at time $t-1$ are from a Dirichlet distribution with parameters $a_{j,t-1}$,

[5] The simplest way of estimating time-varying quantiles is by 'historical simulation', whereby quantiles are computed for a moving block of data ; see Andersen et al (2006).

$j = 1, ..., N$, the prediction and updating recursions are

$$a_{j,t|t-1} = \omega a_{j,t-1}, \qquad j = 1, ..., N, \quad t = 1, ..., T \tag{18}$$

and

$$a_{j,t} = a_{j,t|t-1} + I_{j,t}, \qquad j = 1, ..., N, \quad t = 1, ..., T \tag{19}$$

with the uniform prior giving $a_{0j} = 1, j = 1, ..., N$.

Solving the recursions yields

$$\widetilde{\pi}_{j,t} = \frac{a_{j,t}}{\Sigma_{h=0}^{N} a_{h,t|t-1}} = \frac{\sum\limits_{i=0}^{t-1} \omega^i I_{j,t-i} + \omega^t}{\sum\limits_{i=0}^{t-1} \omega^i + (N+1)\omega^t}, \quad j = 1, ..., N, \quad t = 1, 2, .., T, \tag{20}$$

and each $\widetilde{\pi}_{j,t}, j = 1, ..., N$ is an EWMA. Note that $\Sigma_{i=1}^{N} \widetilde{\pi}_{i,t} = 1$ for all $t = 1, .., T$ as $\Sigma_{i=1}^{N} I_{i,t} = 1$. Since the initial conditions are strictly positive, $\widetilde{\pi}_{j,t} > 0$ for all $j = 1, ..., N$. Smoothed estimates can be constructed as in the binary case.

The predictive distribution is the Dirichlet-multinomial, or Polya, when there are several draws in each time period, but with a single draw it reduces to the multinomial. Hence the log-likelihood function is just

$$\ln L(\omega) = \sum_{t=2}^{T} \sum_{j=0}^{N} I_{t,j} \ln \widetilde{\pi}_{j,t|t-1} \tag{21}$$

and, since $\widetilde{\pi}_{j,t+1|t} = \widetilde{\pi}_{j,t}$, it may be computed from the recursions

$$\widetilde{\pi}_{j,t+1|t} = (1-\omega) I_{j,t} + \omega \widetilde{\pi}_{j,t|t-1}, \quad j = 1, ..., N, \quad t = 1, .., T, \tag{22}$$

with $\widetilde{\pi}_{j,1|0} = 1/N$, $j = 1, ..., N$. The binary likelihood, (9), is obtained when $N = 2$.

The filtered proportions of observations less than or equal to ξ_j, $j = 1, ..., N$ are

$$\widetilde{\tau}_{j,t+1|t} = \widetilde{\tau}_{j,t} = \Sigma_{i=1}^{j} \widetilde{\pi}_{i,t}, \qquad j = 1, ..., N, \quad t = 1, ..., T \tag{23}$$

Note that $\widetilde{\tau}_{N,t} = 1$ and that each $\widetilde{\tau}_{j,t+1|t}, j = 1, ..., N$, is also given by an EWMA of the form (22). Plots of the $\widetilde{\tau}'_{j,t}s$ or $\widetilde{\pi}'_{j,t}s$ may be useful in showing how the distribution changes over time. It might also be useful to estimate the probability of being below a certain threshold, such as VaR, at any point in time.

The above EWMAs may be replaced by other recursions. In particular, recursions appropriate for stationary processes may be defined as at the end of sub-section 2.3.

The EWMAs may be modified to allow for explanatory variables; see appendix B.

3.2 Time-varying quantiles

Time-varying quantiles may be extracted from the $\widetilde{\tau}'_{j,t}s$ as follows. Suppose an estimate of $\xi_t(\tau)$ is required, and that $\widetilde{\tau}_{k-1,t} \leq \tau \leq \widetilde{\tau}_{k,t}$ for some $k = 1,..,N$. Linear interpolation then yields the estimate

$$\widehat{\xi}_t(\tau) = \frac{\tau - \widetilde{\tau}_{k-1,t}}{\widetilde{\tau}_{k,t} - \widetilde{\tau}_{k-1,t}} \{\xi_k - \xi_{k-1}\} + \xi_{k-1}, \qquad t = 1, ..., T \tag{24}$$

where $\widetilde{\tau}_{0,t} = 0$ and $\widetilde{\tau}_{N,t} = 1$. The boundaries, ξ_0 and ξ_N, may need to be redefined and this requires some judgement. One possibility is to set $\xi_0 = y_{\min}$ and $\xi_N = y_{\max}$, but a more stable choice is the 1% and 99% quantiles.

By construction, $\widetilde{\tau}_{k-1,t} < \widetilde{\tau}_{k,t}$ so the time series of $\widehat{\xi}_t(\tau)'s$ cannot cross for different τ. Indeed they cannot even touch. Specifically, for finite T,

$$\widehat{\xi}_t(\tau_1) > \widehat{\xi}_t(\tau_2), \quad \tau_1 > \tau_2, \quad t = 1, ..., T$$

3.3 Different discount factors

With different discount factors, ω_j, $j = 1, ..., N$, the recursions are

$$\begin{aligned} a_{j,t|t-1} &= \omega_j a_{j,t-1}, \qquad j = 1, ..., N, \quad t = 1, ..., T & (25) \\ b_{j,t|t-1} &= \omega_j b_{j,t-1}, & (26) \end{aligned}$$

and

$$\begin{aligned} a_{j,t} &= a_{j,t|t-1} + I_{j,t}, \qquad j = 1, ..., N, \quad t = 1, ..., T & (27) \\ b_{j,t} &= b_{j,t|t-1} + 1, & (28) \end{aligned}$$

with the uniform prior, $a_{0j} = 1$ and $b_{0j} = N, j = 0, 1, ..., N$.

Solving the recursions shows that

$$\widehat{\pi}_{j,t|t} = \frac{a_{j,t}}{b_{j,t}} = \frac{\sum\limits_{i=0}^{t-1} \omega_j^i I_{j,t-i} + a_{j,0}\omega_j^t}{\sum\limits_{i=0}^{t-1} \omega_j^i + (a_{0,0} + a_{1,0} + ...a_{N,0})\omega_j^t}, \quad j = 0, 1, ..., N, \quad t = 1, 2, ... \tag{29}$$

and so $a_{j,t}$, $b_{j,t} > 0$, while $a_{j,t}/b_{j,t}$ is an EWMA as before. The log-likelihood function is as in (21).

When the $\omega' s$ are different, it is better to estimate the proportion in each category by

$$\widetilde{\pi}_{j,t|t} = \frac{\widehat{\pi}_{j,t|t}}{\Sigma_{i=0}^N \widehat{\pi}_{i,t|t}}, \qquad j = 1, ..., N, \quad t = 0, 1, ..., T \tag{30}$$

since $\Sigma_{i=0}^N \widehat{\pi}_{j,t|t}$ is not necessarily unity if the discount factors are different, and similarly for $\widetilde{\pi}_{j,t+1|t}$, which is no longer guaranteed to be the same as $\widetilde{\pi}_{j,t|t}$. We then define

$$\widetilde{\tau}_{j,t|t} = \Sigma_{i=0}^{j-1} \widetilde{\pi}_{i,t|t}, \qquad j = 1, ..., N, \quad t = 1, ..., T$$

and similarly for $\widetilde{\tau}_{j,t+1|t}$. The $\widetilde{\tau}'_{j,t}s$ lie in the range [0,1] by construction. Since $a_{j,t} > 0$, as is $b_{j,t}$, it follows that $\widetilde{\pi}_{j,t|t} > 0$, as is $\widetilde{\pi}_{j,t+1|t}$. Hence $\widetilde{\tau}_{j,t|t} > \widetilde{\tau}_{j-1,t|t}$ and, as in the constant ω case, the quantiles cannot cross.

When there many categories, allowing the discount factors to be different is only viable if some kind of restriction is placed on them. For example, we might assume that they fall on a quadratic, so $\omega(\tau) = a + b + c\tau^2$ and the likelihood function is maximized wrt a, b and c. The discounting is symmetric around the median if $\omega_i = \omega_{N-i+1}, i = 0, .., N/2 - 1$, and this constraint is easily imposed if required. (Symmetric discounting assumes that the categories are delineated symmetrically, so, for N even, the median is $\xi_{N/2}$ and $\xi_{N-i} = \xi_i$, $i = 0, .., N/2 - 1$.) There may sometimes be a case for setting $\omega = 1$ at the median. A likelihood ratio test of the null hypothesis that ω is constant is possible.

3.4 Comparison with time-varying quantiles

Changing quantiles may be estimated non-parametrically, as in Yu and Jones (1998), by minimizing a local check function to give an estimator, $\widehat{\xi}_t$, that satisfies

$$\sum_{j=-h}^{h} K(j/h) IQ(y_{t+j} - \widetilde{\xi}_t(\tau)) = 0 \tag{31}$$

where $K(.)$ is a weighting kernel, with $\sum_{j=-h}^{h} K(j/h) = 1$, h is a bandwidth and $IQ(y_{t+j} - \xi(\tau)) = \tau - I(y_{t+j} \leq \xi(\tau))$. If the same kernel and bandwidth are used for different quantiles, they cannot cross (though they may touch).

De Rossi and Harvey (2007) construct a similar estimator by assuming that $\xi(\tau)$ follows a random walk. They obtain an estimator that satisfies

$$\widetilde{\xi}_t(\tau) = \frac{1-\theta}{1+\theta} \sum_{j=-\infty}^{\infty} \theta^{|j|} [\widetilde{\xi}_{t+j}(\tau) + \kappa IQ(y_{t+j} - \widetilde{\xi}_{t+j}(\tau))] \tag{32}$$

where κ is a scaling constant and θ depends on the variance of the disturbance driving the random walk. If $\widetilde{\xi}_{t+j}$ in (32) is constant, it satisfies (31) with $K(j/h)$ replaced by $\theta^{|j|}$ so giving an (infinite) exponential decay. The time series model determines the shape of the kernel while the θ parameter plays a similar role to that of the bandwidth.

What is the relationship between the estimates computed from changing proportions as in sub-section 3.1 and those obtained directly from (31)? It follows from (20) and (23) that the filtered estimator of the proportion of observations less than or equal to a pre-defined value, ξ_j, is an EWMA, while, in the middle of a large sample, the associated smoothed estimator is

$$\widetilde{\tau}_{j,t|T} \simeq \frac{1-\omega}{1+\omega} \sum_j \omega^{|j|} I(y_t \leq \xi_j)$$

On the other hand, (31) implies that

$$\sum_{j=-h}^{h} K(j/h) I(y_t \leq \widetilde{\xi}_t(\tau)) = \tau.$$

Setting $K(j/h) = [(1-\omega)/(1+\omega)]\omega^j$ and letting h be large gives a similar structure to $\widetilde{\tau}_{j,t|T}$, with the only difference being that $\xi(\tau)$ is pre-set as ξ_j while τ is estimated. In terms of (32), $\omega = \theta$.

3.5 Example: Hong Kong stock returns

Figure 1 showed daily Hong Kong stock returns from 27/11/79 to 27/11/07. The time-varying quantiles estimated from filtered proportions are shown in figure 7 were computed with ω estimated as 0.991. The median lag in this case is 69, (for $\omega = .995$ it is 138) so the response to changes is relatively slow. Nevertheless the sharp rise in volatility after 20 Oct 1997 and, to a lesser extent, after October 19th, 1987, is quite clear. The estimates from smoothed proportions are shown in figure 8. Here the change in volatility shows up before the events of October 1997 and 1987 because of the two-sided weighting.

Figure 9 shows how volatility is captured by the interquartile range and the 90-10 quantile range. The ratio of the 90-10 range to the interquartile range, shown in figure 10 gives an indication of changing kurtosis.

The attack on Hong Kong dollar was a significant event in the so-called 'Asian crisis', the end of which is usually taken to be 31 August 1998; see Dungey et al (2005, p19). This date has also been added to the graphs, together with 15th July 2002, a date indicating a period when a number of markets where in turmoil and 25 July 2007, which roughly marks the beginning of the 'credit crisis' in the US. The 2002 date appears to be unimportant, but the filtered estimates show a sharp rise after the beginning of the credit crisis.

Both filtered and smoothed estimates show movements in the median and the stationarity tests of Busetti and Harvey (2007) are statistically significant. There are also noticeable asymmetries, particularly after the speculative attack on the Hong Kong dollar in 1997. This is particularly apparent in the sharp falls in the lower quantiles. For a symmetric distribution

$$\xi_t(\tau) + \xi_t(1-\tau) - 2\xi_t(0.5), \qquad \tau < 0.5$$

is zero for all $t = 1, .., T$. Hence a plot of $S_t(\tau)$, defined as the above contrast divided by the range as measured by $\xi_t(1-\tau) - \xi_t(\tau)$, shows how the asymmetry captured by the complementary quantiles, $\xi_t(\tau)$ and $\xi_t(1-\tau)$, changes over time. Figure 10 plots $S_t(0.25)$. The measure $S(0.25)$ was originally proposed by Bowley in 1920; the inter-quartile range scales the coefficient so that the maximum value is 1, representing extreme right (positive) skewness and the minimum value is -1, representing extreme left skewness ; see Kim and White (2004) for a recent discussion of measures of skewness and kurtosis. During the Asian crisis $S_t(0.25)$ is negative, while in the period immediately

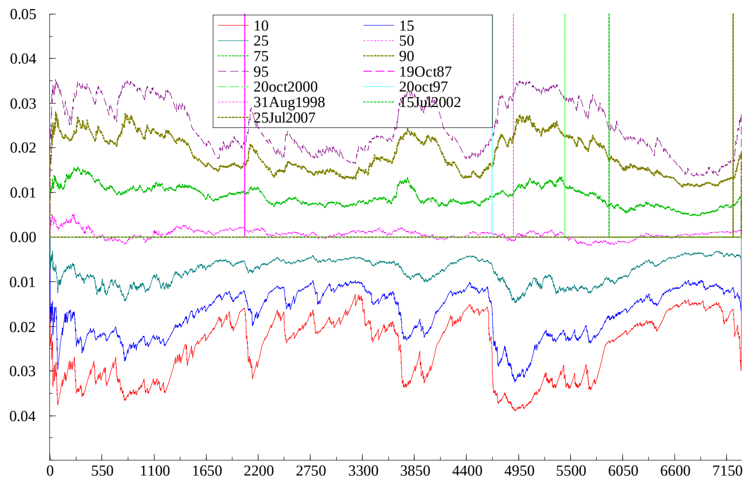

Figure 7: Time-varying quantiles for Hong Kong returns calculated from filtered proportions

before it is strongly positive. The skewness and kurtosis measures plotted in figure 10 are much more stable than corresponding measures constructed from weighted averages of third and fourth moments, as in Jondeau and Rockinger (2003, p1722).

3.6 Smoothing the distribution

The estimated distributions may have a somewhat uneven appearance. Kernel smoothing may be carried out on the time-varying histogram, smoothed or filtered, at any point in time. Alternatively the distribution function may be estimated; see Azzalani (1981).

Another possibility, provided the discount factors are the same, is to give each observation a weight of

$$w_{t,i} = \frac{\omega^{t-i} + \omega^{t}}{\sum\limits_{i=0}^{t-1} \omega^{i} + (N+1)\omega^{t}} \simeq (1-\omega)\omega^{t-i}, \qquad i = 1, ..., t. \tag{33}$$

Then, when the kernel is applied at time t, the kernel weight for the observation at time $t-i$ is multiplied by $(1-\omega)\omega^{t-i}$; see appendix C. Updating

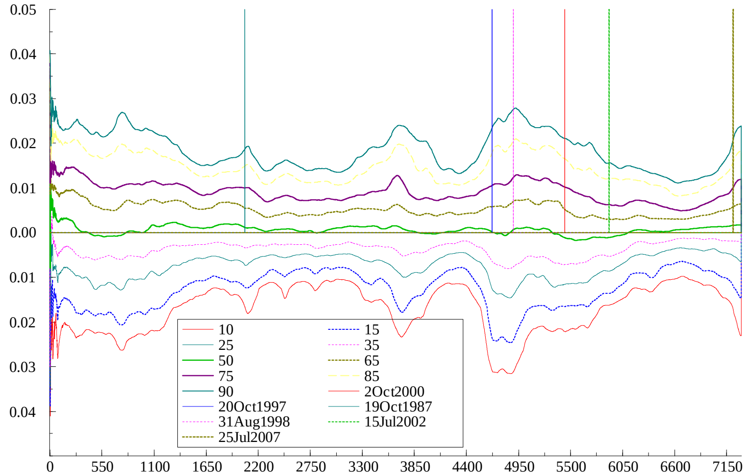

Figure 8: Time-varying quantiles for Hong Kong returns calculated from smoothed proportions

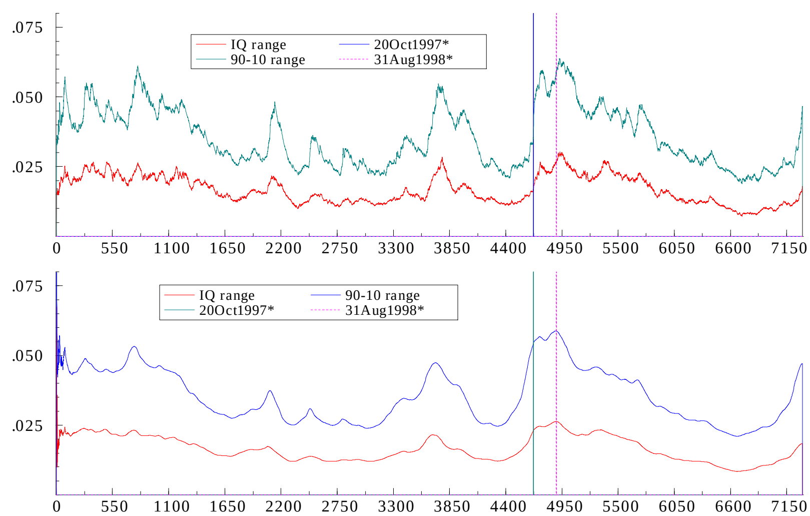

Figure 9: Interquartile range and interdecile (90-10) range from filtered (top panel) and smoothed quantiles for Hong Kong.

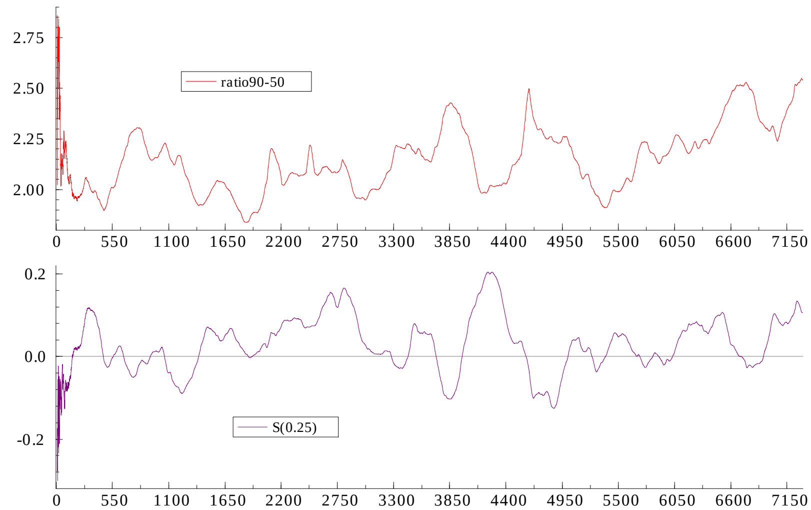

Figure 10: Ratio of interdecile range to interquartile range for Kong Hong and Bowley coefficient of asymmetry.

the weights is straightforward; multiplying the weights at time t by ω gives the weights at $t+1$. For smoothing (as opposed to filtering) in the middle of the sample, weight by

$$w_{t,i} = \frac{1-\omega}{1+\omega}\omega^{|t-i|}, \qquad i = 1, ..., T.$$

Figure 11 shows kernel smoothing of the observations for Hong Kong at the mid-point in the series (3652) weighted with $\omega = 0.99$. The kernel is Epanechnikov with the bandwidth determined by the rule of thumb in Silverman (1986, p 45-8), given as expression (39) in appendix C. Formula (12) suggests an effective sample size for the filtered observations of $T(\omega) = (2-\omega)/(1-\omega) \simeq 1/(1-\omega)$ when ω is close to one. For smoothed observations the suggestion is $T(\omega) = (1+\omega)/(1-\omega) \simeq 2/(1-\omega)$. Figure 12 shows the smoothed distribution at point 4300 where figure 10 indicates that there is strong positive skewness. Figure 13, at point 4700, shows the negative skewness induced by the large falls following the attack on the Hong Kong dollar.

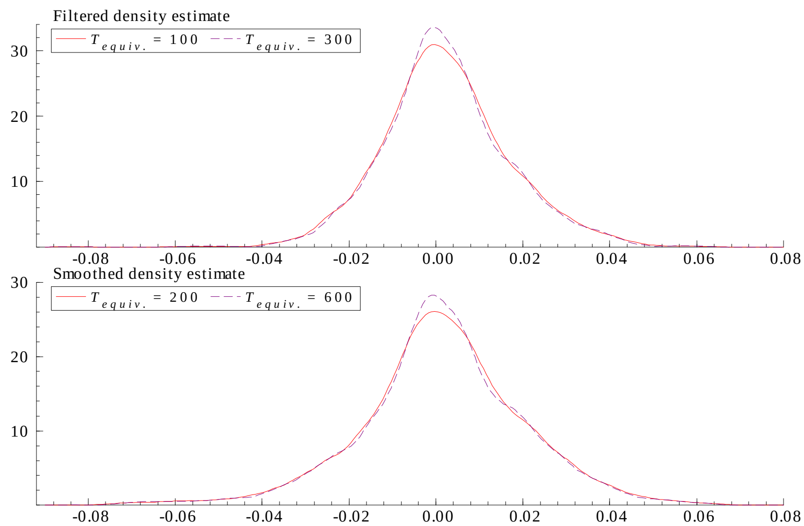

Figure 11: Kernel smoothing of weighted observations at mid-point (3652) of Hong Kong time series using Silverman's rule of thumb for the bandwidth in an Epanechnikov kernel. Upper panel uses filtered observations.

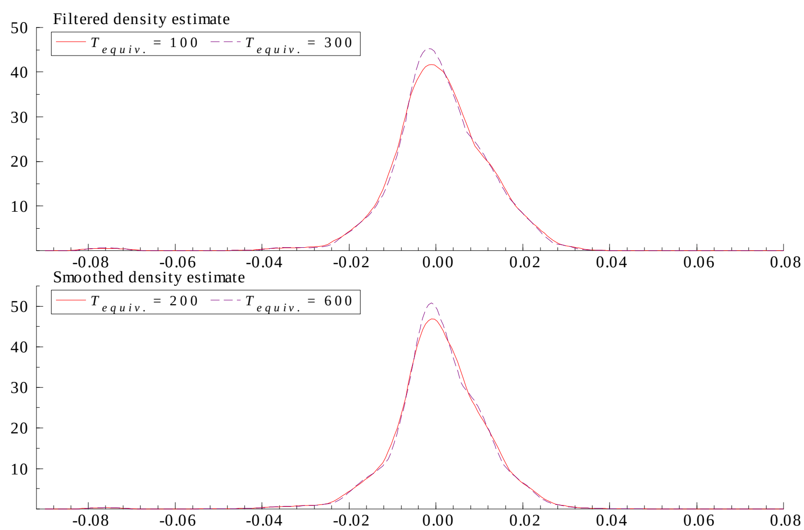

Figure 12: Kernel smoothing of weighted observations at observation 4300 of Hong Kong time series using Silverman's rule of thumb for the bandwidth in an Epanechnikov kernel. Upper panel uses filtered observations.

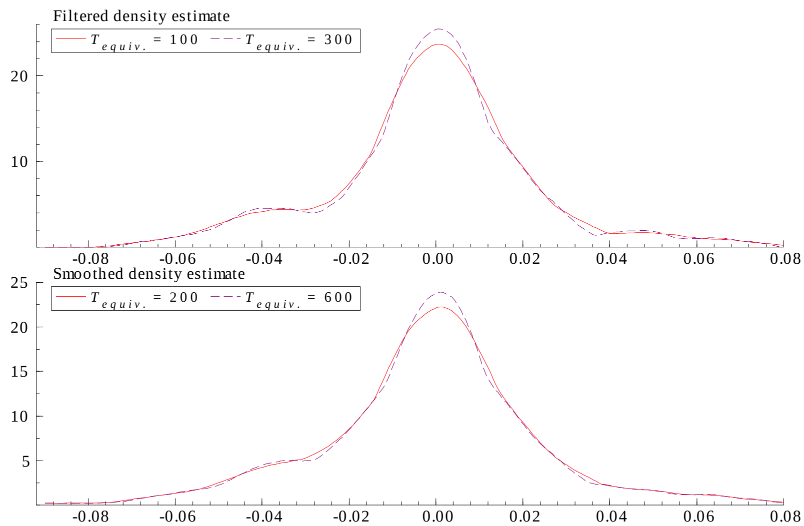

Figure 13: Kernel smoothing of weighted observations at mid-point (4700) of Hong Kong time series using Silverman's rule of thumb for the bandwidth in an Epanechnikov kernel. Upper panel uses filtered observations.

4 Estimation of copulas with time-varying marginal distributions

A time-varying copula with time-varying marginals may be estimated[6] as in section 2 by defining the indicator variables, (15), in terms of the quantiles, estimated as described in section 3. Smoothed estimates of the copula are computed with smoothed quantiles and filtered estimates with filtered quantiles.

Busetti and Harvey (2008) propose testing for the constancy of a copula by means of stationarity tests applied to indicators constructed from the two series. Such indicators typically take a value of one if both observations are below pre-assigned quantiles and zero otherwise. The tests have power against breaks as well as slowly evolving changes. A rejection of the null hypothesis of constancy leads one to consider tracking the copula over time.

4.1 Multivariate estimation of the copula

The copula can be estimated by using the same method adopted for univariate series. The domain of the copula is broken down into a grid and indicators are defined for each square. A single discount parameter can be estimated from a multinomial distribution and when the copula is constructed from the frequency copula the level curves[7] cannot cross.

Define τ_{1j} and τ_{2k} so as to partition the unit interval on the u_1 and u_2 axes into n sub-intervals, $0 = \tau_{10} < \tau_{11} < \tau_{12} < ... < \tau_{1,n} = 1$ and similarly for τ_{2k}. To simplify matters it will be assumed that the sub-intervals are equal. The estimates of the copula for the whole grid are then denoted by $\widehat{C}_t(\tau_{1j}, \tau_{2k})$, $j, k = 1, ..., n$, but with $\widehat{C}_t(\tau_{1,n}, \tau_{2,n}) = 1$ by definition. The corresponding estimates of the copula frequency on the squares defined by $\tau_{1j}, \tau_{1,j-1}$ and $\tau_{2k}, \tau_{2,k-1}$ are given by

$$\widehat{c}_t(\tau_{1j}, \tau_{2k}) = \widehat{C}_t(\tau_{1j}, \tau_{2k}) - \widehat{C}_t(\tau_{1,j-1}, \tau_{2k}) - \widehat{C}_t(\tau_{1j}, \tau_{2,k-1}) + \widehat{C}_t(\tau_{1,j-1}, \tau_{2,k-1}),$$

for $j, k = 1, ..., n$.

The copula frequency may be estimated directly by defining n^2 indicators

$$I_t(\tau_{1j}, \tau_{2k}) = I(y_{1t} > \xi_1(\tau_{1,j-1})).I(y_{1t} \leq \xi_1(\tau_{1j})).I(y_{2t} > \xi_2(\tau_{2,k-1})).I(y_{2t} \leq \xi_2(\tau_{2,k}))$$

[6] A common discount factor could be estimated by adding the two likelihood functions.

[7] $C(\tau_1, \tau_2)$ is constant as τ_1 and τ_2 change.

for $j, k = 1, ..., n$. The discount factor may be estimated as in section 3. At any given point in time, $\widehat{c}_t(\tau_{1j}, \tau_{2k})$ is an EWMA of past indictors and $\sum_{j,k} \widehat{c}_t(\tau_{1j}, \tau_{2k}) = 1$. The recursions are as in (22) with $\widehat{c}_{1|0}(\tau_{1j}, \tau_{2k}) = 1/n^2$, for $j, k = 1, ..., n$, corresponding to both independence and a non-informative prior. When there is no discounting $\sum_j \widehat{c}_T(\tau_{1j}, \tau_{2k}) = \sum_k \widehat{c}_T(\tau_{1j}, \tau_{2k}) = 1/n$. With discounting the expected values of these summations are $1/n$ for all $t = 1, ..., T$.

The copula may be estimated (recursively if desired) from the estimates of the copula density and modified estimates, of the form (16), may be constructed.

Explanatory variables can be included as descibed appendix B.

4.2 Measures of association

Estimators of Kendall's Tau and Spearman's ρ_S may be obtained from the time-varying estimates of the copula and its density, $\widehat{c}_t(\tau_{1j}, \tau_{2k})$. Thus for Spearman's ρ_S

$$\widetilde{\rho}_S = \frac{\sum_{j,k} \widehat{c}_t(\tau_{1j}, \tau_{2k})\tau_{1j}\tau_{2k} - n^2(\sum_j \widehat{c}_j \tau_{1j}/n)(\sum_k \widehat{c}_k \tau_{2k}/n)}{\sqrt{\sum_j \widehat{c}_j \tau_{1j}^2 - n^2(\sum_j \widehat{c}_j \tau_{1j}/n)^2}.\sqrt{\sum_k \widehat{c}_k \tau_{1k}^2 - n^2(\sum_k \widehat{c}_k \tau_{1k}/n)^2}}$$

where $\widehat{c}_j = \sum_k \widehat{c}_t(\tau_{1j}, \tau_{2k})$ and $\widehat{c}_k = \sum_j \widehat{c}_t(\tau_{1j}, \tau_{2k})$.

4.3 Estimation of parametric copulas

In canonical maximum likelihood (CML), the copula parameter or parameters, $\boldsymbol{\psi}$, are estimated without specifying the marginals by maximizing

$$\ln L(\boldsymbol{\psi}) = \sum_{t=1}^{T} \ln c(r_{1,t}/T, r_{2,t}/T; \boldsymbol{\psi})$$

where $r_{1,t}$ and $r_{2,t}$ are the ranked observations. With grouped data, the copula parameters at time t, denoted $\boldsymbol{\psi}_t$, may be estimated by maximizing a function, analogous to $\ln L(\boldsymbol{\psi})$, in which the value of the copula at the mid-point of a square is weighted by the estimated proportion of observations in that square. Thus the function to be maximized is

$$\ln L_t(\boldsymbol{\psi}_t) = \sum_{k=1}^{n} \sum_{j=1}^{n} \widehat{c}_t(\tau_{1j}, \tau_{2k}) \ln c_t(\tau_{1j}^*, \tau_{2k}^*; \boldsymbol{\psi}_t)$$

where $\tau_{1j}^* = (\tau_{1j} - \tau_{1,j-1})/2$, and similarly for τ_{2k}^*.

5 Applications

One aspect of contagion is dependence as measured by the copula. Such changes are to be contrasted with changes in the joint distribution where changes in volatility play an important role. It is conceivable that contagion could affect volatility without changing the strength or pattern of dependence in the copula.

Rodriguez (2007), in his study of Asian and Latin American stock indices, finds evidence of changing tail dependence during periods of turmoil and concludes as follows. 'Changes in tail dependence should be taken into account in the design of any sound asset allocation strategy. Failing to do so can be expensive, as recent theoretical literature has demonstrated. Moreover, it is important to note that these changes are not necessarily captured by correlation shifts.' Das and Upal (2004) highlight the costs of ignoring regime shifts for asset allocation.

The conclusions reached by Rodriguez (2007) are based on fitting parametric copulas in different time periods, the dates of which are determined by a switching model. Here the emphasis is on tracking the copula and then relating any movements to known events.

5.1 Hong Kong and Korea

Here the relationship between the Hong Kong (Hang Seng) and Korean (SET) stock markets is examined for the daily data, from 27/11/79 to 27/11/07, plotted in figures 1 and 2.

The analysis of Hong Kong returns in section 3 indicated that the median changes over time and there are noticeable asymmetries. The same is true for Korea. Hence correcting for changing volatility is not sufficient to render the marginal distributions constant over time. Pre-filtering is therefore carried out using time-varying quantiles. The estimated discount factor for Korea is 0.989, very similar to the one for Hong Kong. For the copula the estimate of omega is 0.993.

Figure 14 shows smoothed tail dependence (TD) for $\tau = 0.10, 0.25, 0.5,$ 0.75 and 0.90. TD is defined as upper tail dependence for $\tau > 0.5$ and is the same as quadrant association (QA) for $\tau = 0.5$. Filtered estimates for TD are displayed in figure 17. The filtered estimates are based on filtered estimates of the quantiles, while the smoothed estimates use smoothed estimates of the quantiles.

Figure 15 shows the average tail dependence, that is the average of the lower TD at .25 and the upper TD at .75, and similarly for .10 and .90. Figure 16 shows the measure of asymmetry calculated as the difference between lower TD and the complementary upper TD, for example .25 minus .75.

If series are independent, TD for .25 and .75 is .25. In other words, if an observation from one series is in the lower quartile, there is a 0.5 chance that the corresponding observation from the other series is in the lower quartile. The graphs show TD for both .25 and .75 moving from near independence before October 1997 to around .5 afterwards. TD for .10 and .90 is more variable, as might be expected, but shows an even bigger jump from around 0.10 to around 0.4 for .90 and around 0.5 for 0.10. However, these changes do not take place immediately. This may be due to the fact that the authorities in Hong Kong and Korea took different measures to combat the crisis.

It is interesting that the higher dependence remains after the end of the Asian crisis, whereas volatility returns to its pre-crisis level. Volatility peaks around 31st August, 1998, the end of the crisis, whereas TD, and of course QA and $C(\tau, \tau)$, is still increasing. By 2005 volatility is, if anything, below its immediate pre-crisis level. TD increases until the end of 2002 and, although it falls somewhat in 2005, mid-2006 shows a returns to the 2002 level.

In the period before October 1987, the measures of tail dependence calculated with no discounting were .17, .29, .53, .27 and .15 for 0.1, 0.25, 0.5, 0.75 and 0.9 respectively. In the period between July 2002 and July 2007, the corresponding figures were .54, .53, .66, .52 and .47. Smoothed estimates of the quantiles, computed as before, were used to determine the indicators.

5.2 FTSE and Dow-Jones

Tail dependence for the FTSE and Dow-Jones indices from 2/1/84 to 27/11/07 is shown in figure 18. There are 6235 observations, starting from 3/1/84. The date on which the UK left the ERM, that is ‘Black Wednesday’, 16 Sept 1992, has been added. The 15th July 2002 was chosen because that date showed the biggest drop in the FTSE in the summer of 2002. There is a very short-term increase in volatility after the 87 crash. Dependence increases but the subsequent fall back to earlier levels takes place more slowly. Dependence levels after 1987 are at a similar level to those seen in Hong Kong and Korea

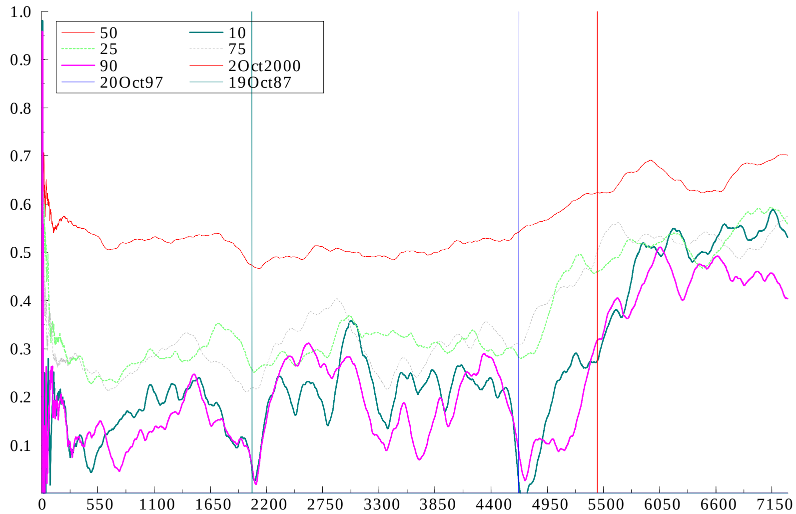

Figure 14: Smoothed tail dependence for $\tau = 0.10, 0.25, 0.5$, 0.75 and 0.90.

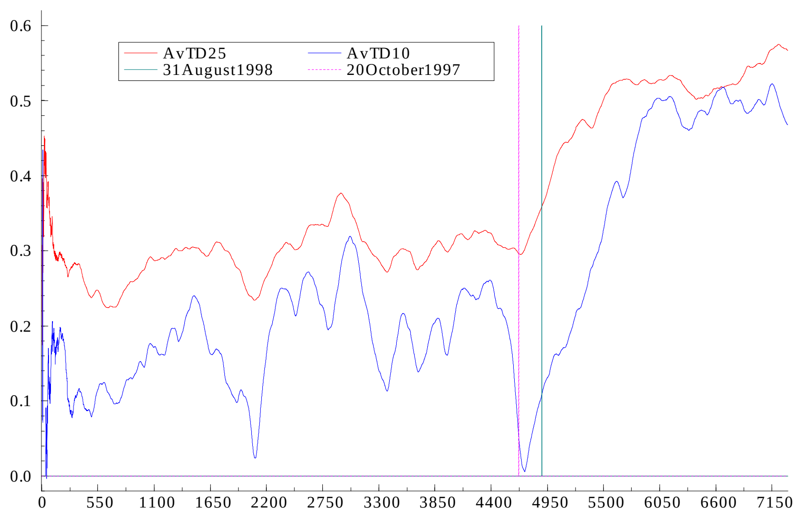

Figure 15: Average tail dependence for $\tau = .25$ and .10.

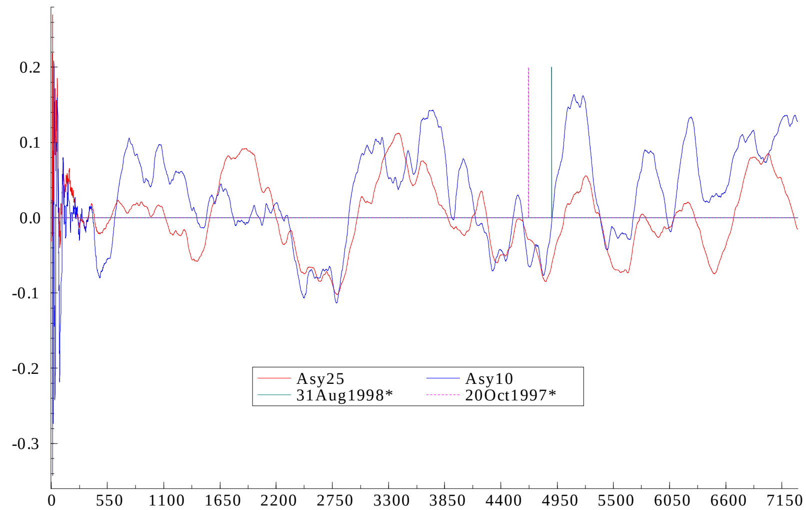

Figure 16: Asymmetry for $\tau = .25$ and $.10$

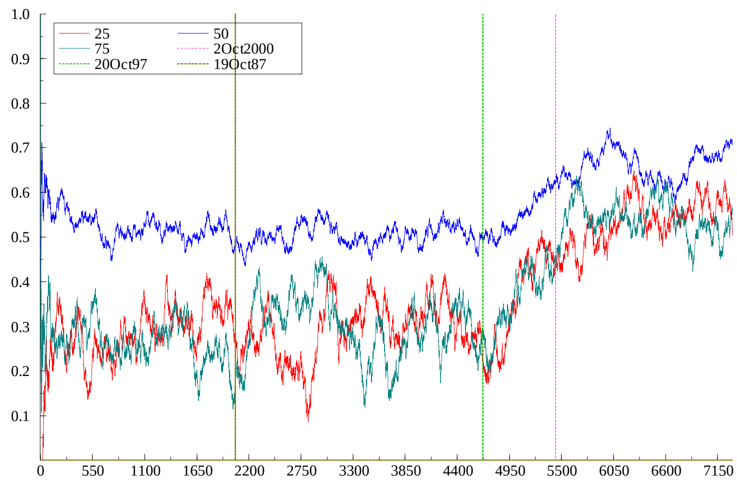

Figure 17: Filtered tail dependence for $\tau = 0.25, 0.5$ and 0.75.

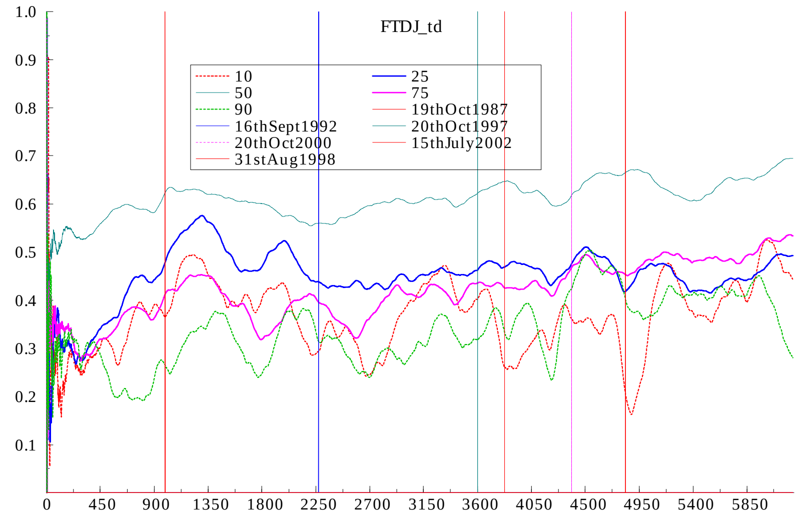

Figure 18: Tail dependence (smoothed) for FTSE and Dow-Jones

after 1998.

6 Conclusion

A time-varying histogram is estimated for a series by using simple EWMAs to estimate the probabilities of observations being in different categories. Estimates of time-varying quantiles are then obtained by linear interpolation. The discount parameter in the EWMA is estimated by maximum likelihood. but with a criterion function for the parameters that is based on ML. The method is therefore simply to apply and should appeal to practitioners. There is evidence of time variation in medians, asymmetry and the tails of distributions and so pre-filtering by fitting a GARCH or SV model may not be sufficient to make a distributions constant over time.

Indicator variables for two series are defined with respect to filtered or smoothed estimates of time-varying quantiles. A time-varying copula frequency is then estimated by EWMAs applied to the indicator variables. The same technique could be used to track the joint distribution (ie without pre-filtering). This may be useful but it does not separate volatility and dependence. Filters other than EWMAs may also be entertained; in particular we might adopt a filter appropriate to movements around a constant level.

The estimated copula probabilities are relatively robust. For example if our interest is in lower tail dependence, the estimates are not adversely affected by movements in the upper tail as they might be in a misspecified model for the copula. On the other hand, estimating small probabilities is difficult without a parametric model.

Tracking the marginal distributions and copula has the advantage that no prior decisions are made on the dating of regimes during which parameters are assumed to be constant. Thus for Hong Kong and Korea, higher dependence is observed but only some time after the start of the Asian crisis. The same level of dependence then continues to the present. This behaviour contrast with that of volatility, which immediately increases after the start of the crisis, but then returns to its pre-crisis level.

APPENDICES

A Smoothing in the local level model

The Gaussian random walk plus noise (local level) model is

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

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

The disturbances ε_t and η_t are mutually independent and the notation $NID\left(0, \sigma^2\right)$ denotes normally and independently distributed with mean zero and variance σ^2. The signal-noise ratio is $q = \sigma_\eta^2/\sigma_\varepsilon^2$.

Smoothing

The smoothed estimates can be computed from Kalman filter and smoother for the Gaussian local level model. The filter is

$$m_{t+1|t} = \left(1 - k_t\right) m_{t|t-1} + k_t y_t$$

where $k_t = p_{t|t-1}/\left(p_{t|t-1} + 1\right)$ is the gain, and

$$p_{t+1|t} = p_{t|t-1} - \left[p_{t|t-1}^2/\left(1 + p_{t|t-1}\right)\right] + q, \quad t = 1, ..., T \tag{34}$$

With a diffuse prior, $m_{2|1} = y_1$ and $p_{2|1} = 1 + q$.

The innovations and Kalman gains are saved and used in the backward smoothing recursion

$$r_{t-1} = (1 - k_t)r_t + (1 - k_t)\nu_t, \qquad t = T, ..., 2,$$

where $\nu_t = y_t - m_{t|t-1}$ and $r_T = 0$, followed by

$$\begin{aligned} m_{t|T} &= m_{t|t-1} + p_{t|t-1}r_{t-1}, \qquad t = 1, ...T, \\ &= m_{t|t-1} + k_t(r_t + \nu_t) \end{aligned}$$

Since $r_0 = (1 - k_1)r_1 + (1 - k_1)\nu_1$, initializing with a diffuse prior will give $m_{1|T} = (p_{1|0}/(p_{1|0} + 1))(r_1 + y_1)$ which goes to $r_1 + y_1$ as $p_{1|0}$ goes to infinity. The following forward recursion can also be used

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

with $m_{1|T} = r_1 + y_1$; see Koopman (1993).

The algorithm in the text assumes a steady state for $p_{t|t-1}$, and sets $1 - k_t = \omega$ in view of the relationship between the steady state value of $p_{t|t-1}$ and the smoothing constant ($\lambda = 1 - \omega$) in the EWMA; see Harvey(1989, p 175).

B Explanatory variables

Let $\pi_t^\dagger = \pi_t/(1+\exp(-\mathbf{x}_t'\boldsymbol{\beta}_j))$, where $(1+\exp(-\mathbf{x}_t'\boldsymbol{\beta}_j))^{-1}$ is the logit function. Let $a_{j,t}^\dagger = a_{j,t}/(1 + \exp(-\mathbf{x}_t'\boldsymbol{\beta}_j))$. Then, given $a_{j,t-1}$ and $b_{j,t-1}$,

$$\begin{aligned} a_{j,t|t-1}^\dagger &= \omega_j a_{j,t-1}/(1 + \exp(-\mathbf{x}_t'\boldsymbol{\beta}_j)), \qquad j = 1, ..., N, \quad t = 1, ..., T \\ b_{j,t|t-1} &= \omega_j b_{j,t-1}, \end{aligned} \tag{35}$$

and $\widetilde{\pi}_{t|t-1}^\dagger = a_{j,t|t-1}^\dagger/b_{j,t|t-1}$ is used in the likelihood function. The updating equations are

$$\begin{aligned} a_{j,t}^\dagger &= a_{j,t|t-1}^\dagger + I_{j,t}, \qquad j = 1, ..., N, \quad t = 1, ..., T \\ b_{j,t} &= b_{j,t|t-1} + 1, \end{aligned} \tag{36}$$

but since $a_{j,t} = a_{j,t}^\dagger(1 + \exp(-\mathbf{x}_t'\boldsymbol{\beta}_j))$, the first of these can be amended to

$$a_{j,t} = a_{j,t|t-1} + I_{j,t}(1 + \exp(-\mathbf{x}_t'\boldsymbol{\beta}_j)), \qquad j = 1, ..., N, \quad t = 1, ..., T$$

In other words the observation is divided by the logit. The recursions can be implemented as an EWMA

$$\widetilde{\pi}_{j,t+1|t} = (1-\omega)I_{j,t}(1+\exp(-\mathbf{x}_t'\boldsymbol{\beta}_j)) + \omega\widetilde{\pi}_{j,t|t-1}, \quad j=1,...,N, \quad t=1,..,T,$$

with $\widetilde{\pi}^{\dagger}_{j,t+1|t} = \widetilde{\pi}_{j,t+1|t}/(1+\exp(-\mathbf{x}_t'\boldsymbol{\beta}_j))$. ML estimation requires numerical optimization wrt the $\boldsymbol{\beta}_j' s$ are well as the $\omega_j' s$.

C Kernel density estimation

At each point y, the kernel estimator is given by

$$\hat{f}_t(y) = \frac{1}{h}\sum_{i=1}^{t} K\left(\frac{y-y_i}{h}\right) w_i, \tag{37}$$

where $K(.)$ is the Epanechnikov kernel

$$K(z) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1-\frac{1}{5}z^2\right) & \text{for } -\sqrt{5} \leq z \leq \sqrt{5} \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

and h is the bandwidth. The weight, w_i, is as defined in (33).

The rule-of-thumb bandwidth suggested in Silverman (1986) is

$$h_{opt} = 1.06 \times \min\{\hat{\sigma}, \frac{\widehat{IQR}}{1.34}\} \times T^{-1/5}, \tag{39}$$

where $\hat{\sigma}$ is the sample standard deviation, and $\widehat{IQR}$ is the sample interquartile range.

Acknowledgements

I am grateful Tirthankar Chakravarty and Vitaliy Oryshchenko for research assistance. I would like to thank Jesus Gonzalo, Loran Chollette, Alessio Sancetta, Neil Shephard and others for helpful comments. Any errors are my responsibility. I would also like to thank Mardi Dungey for suggesting the application and supplying the data. Earlier versions of this paper were presented at Carlos III in Madrid, Liverpool and at the Financial Econometrics Conference at Imperial College and at the London-Oxbridge Time Series

Workshop, both in May 2008. Financial support from the ESRC under the grant Time-Varying Quantiles, RES-062-23-0129, is gratefully acknowledged.

REFERENCES

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X. (2006). Volatility and correlation forecasting. In: Elliot, G., Granger, C., Timmerman, A. (Eds.), Handbook of Economic Forecasting, 777-878. Amsterdam: North Holland

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika 68, 326–328.

Bae, K-H., Karolyi, G.A. and R.M.Stulz (2003). A new approach to measuring financial contagion. The Review of Financial Studies, 16, 717-63.

Busetti, F. and A. C. Harvey (2007) Tests of time-invariance based on quantile indicators. Cambridge, Faculty of Economics, CWPE 0657

Busetti, F. and A. C. Harvey (2008) When is a copula constant? A test for changing relationships. Mimeo.

Cherubini, U., Luciano, E. and W. Vecchiato (2004). Copula methods in Finance. John Wiley and Sons: Chichester.

Das, S. and R. Upal. (2004) Systematic risk and international portfolio choice. Journal of Finance, 59, 2809-34.

De Rossi, G. and Harvey, A C (2006). Time-varying quantiles. Cambridge, Faculty of Economics, CWPE 0649.

De Rossi, G. and Harvey, A C (2008). Quantiles, expectiles and splines. Journal of Econometrics (to appear).

Doornik, J. A. (1999). *Ox: An Object-Oriented Matrix Language.* 3rd edition. Timberlake Consultants Press: London.

Dungey, M., Fry, R., Gonzalez-Hermosillo, B. and V.L.Martin (2005). Empirical modelling of contagion: a review of methodologies. Quantitative Finance, 5, 9-24.

Embrechts P, Lindskog F, and A.J. McNeil (2003) Modelling dependence with copulas and applications to risk management. In: Handbook of Heavy Tailed Distributions in Finance, ed. S. Rachev, Elsevier, Chapter 8, pp. 329-384

Engle, R.F., and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles, *Journal of Business and Economic Statistics*, 22, 367–381.

Forbes, K. and R. Rigobon, (2002). No contagion, only interdependence: measuring stock market movements. Journal of Finance, 57, 2223-61.

Fermanian, J-D. and O. Scaillet (2003). Nonparametric estimation of copulas for time series), Journal of Risk 5, No 4, 25-54.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and Kalman Filter.* Cambridge: Cambridge University Press.

Harvey A.C. and C. Fernandes (1989). Time series models for count data or qualitative observations. *Journal of Business and Economic Statistics,* 1989, 7, 409-422.

Harvey A.C., E. Ruiz and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies* 61: 247-64.

Harvey, C., and A. Siddique (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis,* 34, 465-87.

Jondeau, E. and M. Rockinger (2003). Conditional volatility, skewness and kurtosis: existence, persistence and comovements. *Journal of Economic Dynamics and Control,* 27, 1699-1737.

Kim, Tae-Hwan and Halbert White (2004). On more robust estimation of skewness and kurtosis. Finance Research Letters, 1, 56-73

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* 80: 117-26.

Koopman, S.J., N. Shephard and J. Doornik (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal,* 2, 113-66.

Kruskal, W.H. (1958). Ordinal measures of association. *Journal of the American Statistical Association,* 53, 814-61.

Laurent, S. (2007). GARCH5. Timberlake Consultants Ltd., London.

Linton, O. and Y.-J.Whang (2007). The quantilogram: with an application to evaluating directional predictability. *Journal of Econometrics,* 141, 250-82.

McNeil, A.J., Frey, R. and P.Embrechts (2005). Quantitative Risk Management. Princeton Series in Finance.

Nelsen (1999). An introduction to copulas. Springer-Verlag. NewYork.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47, 527-56.

Rodriguez, J.C. (2007). Measuring financial contagion: a copula approach. *Journal of Empirical Finance*, 14, 401-23.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.

Smith, J.Q. (1981). The multiparameter steady model. *Journal of the Royal Statistical Society, Series B* 43: 256-60.

Van den Goorbergha, R.W.J., Genest, C. and B.J.M. Werkerc (2005). Bivariate option pricing using dynamic copula models, *Insurance: Mathematics and Economics* 37, 101–114.

Yu, K. and M.C. Jones (1998). Local linear quantile regression, *Journal of the American Statistical Association,* 93, 228-37.