# Rate of Decay of the Tail Dependence Coefficient and Modelling for Bivariate Financial Returns Distributions

Fung, Thomas
*Macquarie University, Department of Statistics*
*North Ryde 2109, Australia*
*E-mail: thomas.fung@mq.edu.au*

Seneta, Eugene
*University of Sydney, School of Mathematics and Statistics*
*Sydney 2006, Australia*
*E-mail: eseneta@maths.usyd.edu.au*

## Introduction

Our motivation arose out of a multivariate model for asset price moments over time, $t \geq 0$. Let $X_t^{(i)}$, $i = 1, \dots, n$, be the marginal log-price increment ("return") over a unit time period, say $X_t^{(i)} = \log S_t^{(i)} - \log S_{t-1}^{(i)}$. The underlying idea in its univariate form is the comparison of the Variance Gamma (VG) and $t$ models for returns, which both arise from subordinator models alternative to the classical geometric Brownian motion. A subordinator model specifies that the return, i.e. $X_t^{(i)}$, is driven by Brownian motion $B^{(i)}(t)$ evaluated at a random time-change, the so called "activity time" $T_t$, for $t \geq 0$. That is, for returns we assume

$$X_t^{(i)} = \mu_i + \theta_i(T_t - T_{t-1}) + \sigma_i(B^{(i)}(T_t) - B^{(i)}(T_{t-1})).$$

for $\mu_i$, $\theta_i \in \mathbb{R}$, $\sigma_i > 0$, where $\mu_i$ and $\sigma_i$ correspond to the drift and diffusion coefficients of Brownian motion respectively. $\{T_t\}$, which is shared between all the marginal return processes, is assumed to be an increasing process with $T_0 = 0$ independent of the Brownian motion, and having stationary independent increments. Here $\{T_t\}$ has a practical interpretation as the information flow or trading activity. As the more frenzied trading becomes, or the more information released to the market on a given day, the faster the "time" flows. The extra parameter $\theta_i$ over the geometric Brownian motion allows us to incorporate skewness into the model. The correlations between returns arise from the Brownian motions with positive definite covariance matrix $\Sigma$ over unit time and hence the multivariate return $\mathbf{X}_t = (X_t^{(1)}, \dots, X_t^{(n)})^T$, $n \geq 1$ can be represented as

$$(1) \quad \mathbf{X}_t | \tau_t \sim N_n(\boldsymbol{\mu} + \boldsymbol{\theta}\tau_t, \tau_t\Sigma)$$

where $\tau_t = T_t - T_{t-1} > 0$. Hence the correlation structure is controlled by $\tau_t$ and $\Sigma$. The study of the dependence structure for such a model, using copulas, of the (symmetric) $t$ model by Demarta and McNeil (2005), and of the (skew) VG model by Luciano and Schoutens (2006), motivates our work.

A particularly important way in which the effect of the dependence structure is manifested is whether the bivariate distribution modelling joint returns possesses asymptotic lower tail dependence (see Coles, Heffernan and Tawn (1999) for an introduction) defined in terms of a bivariate vector $\mathbf{X} = (X_1, X_2)^T$ as

$$(2) \quad \lambda_L = \lim_{u \to 0^+} \lambda_L(u), \text{ where } \lambda_L(u) = P(X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u)),$$

where $F_i^{-1}(\cdot)$, $i = 1, 2$, are the quantile functions of the marginals. This quantity would seem to provide insight on the tendency for the distribution to generate joint extreme events since it measures the strength of dependence (or association) in the lower tails of a bivariate distribution. Since joint negative extreme price movement emulates a market crash, it is a particularly relevant concept in modelling prices and returns, or credit risk. Demarta and McNeil (2005) provide a comprehensive review on the matter of tail dependence for a various (bivariate) symmetric distributions. For instance, the Gaussian distributions are asymptotic independent in the tails while the symmetric $t$ does possess tail dependence. Notice that the symmetric $t$ can be defined by using the normal variance mixing representation, i.e. (1) with $\boldsymbol{\theta} = 0$ with the mixing distribution being inverse gamma.

We first review some recent results on the tail dependence for various (bivariate) skew distributions obtained from (1) by Fung and Seneta (2010A, 2011A and 2011B). We then use the mixing representation to obtain maximum likelihood estimates and compare goodness of fit of the bivariate skew VG and $t$ as models, by using simulated and actual data, making convenient use of the statistical package `WinBUGS` as in Fung and Seneta (2010B). We thus establish there is little to choose between the distributions as models even through they contrast in their asymptotic lower tail dependence. We are led consequently to examine the rate of decay to zero of $\lambda_L(u)$ when the limit is zero, and present a new result on the rate of decay of the the tail dependence coefficient of the skew VG distribution.

## Skew models

We will now introduce the two skew models with focus on their mixing representations. These models are chosen not only because they are serious competitors in modelling in a financial context, but because they can also be expressed in similar mixing representations.

Luciano and Schoutens (2006) extended the multivariate symmetric VG model by using the normal mean-variance mixing representation. A r.v. $\mathbf{X}$ is said to have a multivariate skew VG model, denoted by $X \sim SVG_n(\boldsymbol{\mu}, \Sigma, \boldsymbol{\theta}, \nu)$, if

$$(3) \quad \mathbf{X}|Y \sim N_n(\boldsymbol{\mu} + \boldsymbol{\theta}Y, Y\Sigma),$$

where $Y \sim \Gamma(\frac{1}{\nu}, \frac{1}{\nu})$ with $\nu > 0$ i.e. has a Gamma distribution with density as

$$f(y) = \frac{1}{\nu^{\frac{1}{\nu}}\Gamma(\frac{1}{\nu})} y^{\frac{1}{\nu}-1} e^{-\frac{y}{\nu}}, \quad y > 0; \quad = 0, \quad \text{otherwise.}$$

$\boldsymbol{\mu} \in \mathbb{R}^n$ is the location vector, $\Sigma$ is an $n \times n$ positive definite symmetric scale matrix and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T \in \mathbb{R}^n$ controls the asymmetry of the model.

The multivariate skew $t$ model is introduced by Demarta and McNeil (2005). A random vector $\mathbf{X}$ is said to have a multivariate skew $t$ distribution if $\mathbf{X}$ is defined by a normal variance-mean mixture, parallel to those of skew VG, as

$$(4) \quad \mathbf{X}|Y \sim N_n(\boldsymbol{\mu} + \boldsymbol{\theta}Y^{-1}, Y^{-1}\Sigma),$$

where $Y \sim \Gamma(\frac{\eta}{2}, \frac{\eta}{2})$. We denote this skew $t$ model as $\mathbf{X} \sim St_n(\boldsymbol{\mu}, \Sigma, \boldsymbol{\theta}, \eta)$. $\eta > 0$, is the shape parameter and is usually termed as the degrees of freedom of the $t$ distribution.

## Tail dependence

Although both skew models are defined similarly by their mixing representation, they have vastly different asymptotic tail dependence. In this section, we review the corresponding results.

The discussion of the asymptotic lower tail dependence for the (bivariate) skew VG model is initiated by Luciano and Schoutens (2006); the analytical result is as follows and its proof can be found in Fung and Seneta (2011A).

**Theorem 1.** *If $\boldsymbol{X} = (X_1, X_2)^T$ has the bivariate skew VG distribution defined by (3) at $n = 2$ with $\Sigma$ is positive definite symmetric matrix, $\boldsymbol{\mu}, \boldsymbol{\theta} = (\theta_1, \theta_2)^T \in \mathbb{R}^2$ and $\nu > 0$, then there is no asymptotic lower tail dependence, i.e. $\lambda_L = \lim_{u \to 0^+} P(X_2 \leq F_2^{-1}(u)|X_1 \leq F_1^{-1}(u)) = 0$.*

Tail dependence for the classical skew $t$ distribution is very recently discussed in Banachewicz and van der Vaart (2008), in terms of the slightly more general skewed, grouped $t$ distribution. But it is sufficient for our modelling purpose to focus on its special case, the skew $t$ and the corresponding tail dependence results are summarised as the following theorem.

**Theorem 2.** *The lower tail dependence coefficient for $\boldsymbol{X} = (X_1, X_2)^T$ defined in (4) at $n = 2$ with $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$, for $\boldsymbol{\mu} \in \mathbb{R}^2$ and $\eta > 0$, is given by*

1. *If $\theta_1 = \theta_2 = 0$ (i.e. the bivariate symmetric t), then $\lambda_L = 2P(t_{\eta+1} \leq -\sqrt{\frac{(\eta+1)(1-\rho)}{1+\rho}})$;*
2. *If $\theta_1$ and $\theta_2 > 0$, then $\lambda_L = 0$;*
3. *If $\theta_1$ and $\theta_2 < 0$, then $\lambda_L = 1$;*
4. *If $\theta_1 > 0$ and $\theta_2 < 0$, then $\lambda_L = 0$;*
5. *If $\theta_1 = 0$ and $\theta_2 > 0$, then $\lambda_L = 0$;*
6. *If $\theta_1 = 0$ and $\theta_2 < 0$, then $\lambda_L = \int_0^1 \left(1 - \Phi\left( (\frac{2^{\frac{\eta}{2}}\Gamma(\frac{\eta+1}{2})}{2\sqrt{\pi}})^{\frac{1}{\eta}} u^{\frac{1}{\eta}} \right)\right) du$,*

*where $\Phi(\cdot)$ is the standard normal distribution function, $t_{\eta+1}$ is the Student's $t$ distribution with $\eta + 1$ degrees of freedom, and $|\rho| = \left|\frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}\right| < 1$.*

It is noted in Banachewicz and van der Vaart (2008) that skewing the multivariate symmetric $t$ distribution using (4) leads to trivial values - 0 or 1 - of the asymptotic lower tail dependence is rather disappointing from a modelling perspective, since the symmetric $t$ has non-trivial values for asymptotic lower tail dependence which is Case (1) of Theorem 2.

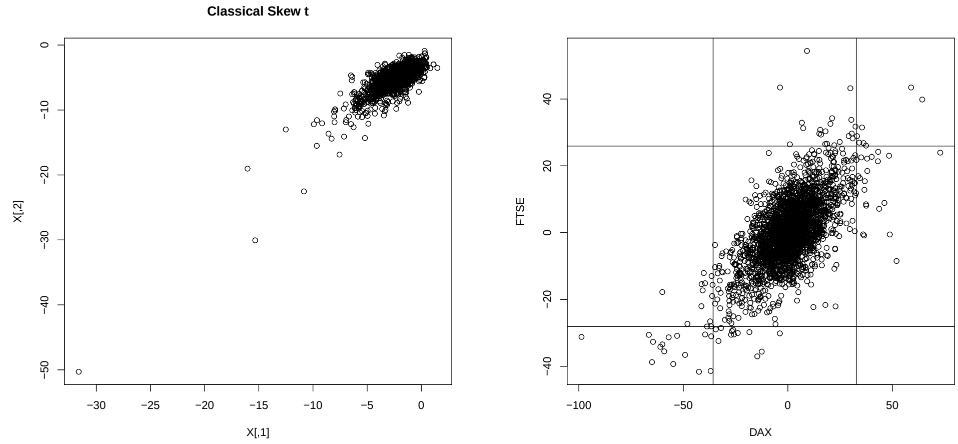
These contrasting situations raise the question to what extent the limit value describes tail dependence in practice and we try to explore this further by using simulated and real data as described in the next section.

## Modelling and Estimation

Since the two skew models have different asymptotic tail dependence, we are interested in how these models will perform against each other in practice. We use a program called `WinBUGS` to obtain maximum likelihood estimates for large samples. Its applications to univariate financial data have already been explored in Finlay and Seneta (2008). Here we apply it to some bivariate data sets.

`WinBUGS` is a Windows programme for Bayesian estimation of parameters, where BUGS stands for **B**ayesian inference **U**sing **G**ibbs **S**ampling (See Lunn, Best and Spiegelhalter (2000)). The idea behinds `WinBUGS` is surprisingly simple. Observed iid (identical and independently distributed) data is from a specified distribution regarded as a conditional distribution in which the parameters are regarded as fixed values of random variables. The statistician specifies a prior distribution (usually the uniform or some other non-informative distribution) of the parameters, and uses the observed sample values $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N$ to modify the prior distribution to a posterior distribution of parameters. `WinBUGS` uses the Gibbs sampler to simulate realisations of a Markov chain whose states are parameters, and whose limiting distribution is the posterior distribution of parameters. Thus once a realisation of a chain has been allowed to settle down (the "burn-in" time), the Bayes estimators of parameters are obtained from averaging the Markov-dependent values of the Markov chain. For a large size N of the

***Figure 1: Scatter plots of the data sets.***




(a) Simulated Skew $t$ (b) DAX 30 and FTSE 100

iid sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, the Bayes estimators so obtained are *asymptotically equivalent to maximum likelihood estimators.* (For more details, see Gelman, Carlin, Stern and Rubin (2004).)

The advantage of `WinBUGS` is it allows us to use a mixing representation to specify the model. As a result, one can obtain MLE without specifying the likelihood function explicitly, as we know the mixing distribution, which makes the input code straightforward. For instance, we could simply use (4) to specify the skew $t$ model.

We now apply both the skew VG and the skew $t$ models to two bivariate data sets, one simulated and the other one real. For the simulated data set, we use `R` to generate a random samples of length $N = 2000$ for the skew $t$ model using (4) with

$$(5) \quad \boldsymbol{\mu} = (-1, -3)^T, \quad \boldsymbol{\theta} = (-1, -1.5)^T, \quad \Sigma = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}, \text{ and } \eta = 8,$$

This set of parameters are chosen so that the distribution does not have tails which are too heavy, is mildly positive correlated and most importantly both marginals are negatively skewed and thus guarantee that there will be complete asymptotic tail dependence as shown in Case (3) of Theorem 2. On the other hand, the real bivariate financial data set consists of the returns each multiplied by 1,000 of the daily closing prices of DAX 30 Composite index and the FTSE 100 Composite index between 1 January 1991 and 31 December 2001. We choose this particular data set because of the work of Fortin and Kuzmics (2002) in which the authors suggest this data set displays asymptotic tail dependence. Since the assets are listed in different stock exchanges, we define a pair of readings if they correspond to the closing prices on the same trading day. The data are collected from Yahoo Finance: Historical Prices. We have N = 2722 pairs of readings. Figure 1 shows the scatter plot of our data sets. The grid lines represent the 1% and 99% empirical quantiles for each of the marginal data sets.

In `WinBUGS`, we asked the program to initialise the parameters and generate 3 separate Markov chains for both skew VG and skew $t$ models. After 10,000 iterations of which the first 5,000 are treated as burn-in, we had 5,000 stable-state values from each chain, making 15,000 in all, from which to obtain the Bayes (equivalently maximum likelihood because of the large sample size) estimators.

Assessing the data fitting capability of the two models is our secondary aim, so for each data set both models were fitted and Deviance Information Criterion (DIC) is used to access and compare the fit. The computation of DIC is an option in `WinBUGS`. The model which returns the lowest DIC is said to provide the best fit for that data set. The results of the estimated parameters and their corresponding DICs are collected in Tables 1 and 2, where $\nu$ and $\eta$ are shape parameters for skew VG and the skew $t$ distributions respectively. The true parameters values for the simulated data are listed in (5).

***Table 1: Estimated parameters for the simulated bivariate skew $t$ data set with $N = 2000$.***

| Model | $(\mu_1, \mu_2)$ | $(\sigma_{11}, \sigma_{12}, \sigma_{22})$ | $(\theta_1, \theta_2)$ | $\nu$ or $\eta$ | DIC |
|---|---|---|---|---|---|
| Skew VG | (-1.27, -3.47) | (1.52,1.07,1.80) | (-1.10,-1.57) | 0.49 | 11796.40 |
| Skew $t$ | (-1.13, -3.20) | (1.02,0.68,1.13) | (-0.89,-1.31) | 7.73 | 12075.50 |

***Table 2: Estimated parameters for the $\boldsymbol{X}^T = (X_1, X_2) =$ (DAX 30, FTSE 100) data sets with $N = 2722$.***

| Model | $(\mu_1, \mu_2)$ | $(\sigma_{11}, \sigma_{12}, \sigma_{22})$ | $(\theta_1, \theta_2)$ | $\nu$ or $\eta$ | DIC |
|---|---|---|---|---|---|
| Skew VG | (1.424, 0.7247) | (160.0, 77.97,96.37) | (-0.9363,-0.3944) | 0.4594 | 39260.4 |
| Skew $t$ | (1.454,0.8074) | (104.7,50.81,63.58) | (-0.6247,-0.3078) | 5.636 | 39786.9 |

By comparing the results on Table 1 with the true values used to generate the data sets in (5), we conclude that the algorithms are functioning properly, since the estimated parameters for the skew $t$ model are close to the true values. The skew VG, which has no asymptotic tail dependence as shown in Theorem 1, provides the best fit in both simulated and real data in terms of DIC, even though the simulated data set is generated by a heavy-tailed distribution with complete asymptotic tail dependence. This sort of difference means tail dependence has to be interpreted as a property of the model rather than reflecting actual tail dependence structure in the data set. Thus one might infer that positivity of $\lambda_L$, as defined in (2), is too extreme a measure, as substantial lower tail dependence may be present even when the limit is zero. As a result, we are led to examine the rate of decay of the tail dependence coefficient for the Gaussian and skew VG distributions in the next section. These results represent an initial step of our investigation in explaining the discrepancy between zero asymptotic tail dependence coefficient and mass in the tail of a joint distribution.

**Rate of Decay**

For the bivariate normal, the asymptotic tail dependence coefficient (See Embrechts, McNeil and Strumann (2001)) is

$$\lambda_L = \lim_{u \to 0^+} \lambda_L(u) = 0 \text{ where } \lambda_L(u) = 2\Phi(\Phi^{-1}(u)\sqrt{\frac{1-\rho}{1+\rho}}),$$

where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are the distribution function and the quantile function of the standard normal distribution respectively. This indicates that there is no asymptotic tail dependence irrespective of the size of $\rho$. Fung and Seneta (2011B) showed that

$$\lambda_L(u) \sim u^{\frac{1-\rho}{1+\rho}} L(u), \text{ where } L(u) \sim 2\sqrt{\frac{1+\rho}{1-\rho}}(-4\pi \log u)^{\frac{-\rho}{1+\rho}} \text{ as } u \to 0^+.$$

We can see that $\lambda_L(u)$ is a regularly varying function with index of $\frac{1-\rho}{1+\rho}$ which tends to zero as $\rho \to 1$. In other words, as the linear correlation $\rho$ increases, we will see more and more weight at the tail of the joint distribution, so $\lambda_L(u)$ largely only reflects the size of the measure of linear correlation.

Finally, we present a bound on the rate of decay of the tail dependence coefficient for the bivariate skew VG distribution. From Theorem 1, we recall that the bivariate skew VG distribution has no asymptotic tail dependence irrespective of the value of its parameters. Under the condition

$\frac{\theta_1}{\sqrt{\Sigma_{11}}} = \frac{\theta_2}{\sqrt{\Sigma_{22}}} = \theta$, where $\theta_1$ and $\theta_2$ are the $\boldsymbol{\theta}$ component parameters before standardisation, we have however found that

$$\begin{aligned}
\lambda_L(u) &\geq u^\tau L(u), \text{ as } u \to 0^+ \\
\text{where } \tau &= \frac{\sqrt{\left(\theta^2 + \frac{2}{\nu} + \frac{\theta^2(1-\rho)}{1+\rho}\right)\left(\frac{2}{1+\rho}\right)} + \theta\left(\frac{1-\rho}{1+\rho}\right) - \sqrt{\theta^2 + \frac{2}{\nu}}}{\sqrt{\theta^2 + \frac{2}{\nu}} + \theta}, \\
L(u) &\sim \frac{2\sqrt{1-\rho^2}\left(\frac{2}{1+\rho}\right)^{\frac{1}{2}\left(\frac{1}{\nu}-\frac{1}{2}\right)}\left(\theta^2 + \frac{2}{\nu}\right)^{\frac{1}{2\nu}}}{\sqrt{2\pi}(1-\rho)\left(\theta^2 + \frac{2}{\nu} + \frac{\theta^2(1-\rho)}{1+\rho}\right)^{\frac{1}{2\nu}+\frac{1}{4}}} \left| \frac{\log\left(c\,u\left(\frac{|\log u|}{\sqrt{\theta^2+\frac{2}{\nu}}+\theta}\right)^{1-\frac{1}{\nu}}\right)}{\sqrt{\theta^2 + \frac{2}{\nu}} + \theta} \right|^{-\frac{1}{2}} \left(c\left(\frac{|\log u|}{\sqrt{\theta^2 + \frac{2}{\nu}} + \theta}\right)^{1-\frac{1}{\nu}}\right)^{\tau}, \\
\text{with } c &= \nu^{\frac{1}{\nu}}\Gamma\left(\frac{1}{\nu}\right)\left(\theta^2 + \frac{2}{\nu}\right)^{\frac{1}{2\nu}}\left(\sqrt{\theta^2 + \frac{2}{\nu}} + \theta\right), \text{ and } \rho = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}.
\end{aligned}$$

Therefore, the tail dependence coefficient of the skew VG distribution decays at least as slowly as a regularly varying function with index $\tau$. This rather slow rate of decay explains the excellent performance of the VG model in financial contexts such as reported in Luciano and Schoutens (2006) even though it lacks asymptotic tail dependence.

## REFERENCES


Banachewicz, K, and van der Vaart, A. (2008). Tail dependence of skewed grouped $t$-distributions. *Statistics and Probability Letters* **78**, 2388-2399.

Coles, S., Heffernan, J. and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes* **2**, 339-365.

Demarta, S. and McNeil, A.J. (2005). The $t$ copula and related copulas. *International Statistical Review* **73**, 111-129.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In, Dempster, M. , ed. *Risk Management: Value at Risk and Beyond.* Cambridge University Press, Cambridge, pp. 176-223.

Fortin, I. and Kuzmics, C. (2002). Tail-dependence in stock-returns pairs. *International Journal of Intelligent Systems in Accounting. Finance & Management* **11**, 89-107.

Finlay, R. and Seneta, E. (2008). Stationary-increment Variance-Gamma and $t$ models: Simulation and parameter estimation. *International Statistical Review* **76**, 167-186.

Fung, T. and Seneta, E. (2010A). Tail dependence for two skew $t$ distributions. *Statistics and Probability Letters* **80**, 784-791.

Fung, T. and Seneta, E. (2010B). Modelling and Estimation for Bivariate Financial Returns. *International Statistical Review* **78**, 117-133.

Fung, T. and Seneta, E. (2011A). Tail dependence and skew distributions. *Quantitative Finance* **11**, 327 - 333.

Fung, T. and Seneta, E. (2011B). The bivariate normal copula function is regularly varying. Under revision for *Statistics and Probability Letters.*

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis, 2nd Ed.* Chapman & Hall, New York.

Luciano, E. and Schoutens, W. (2006). A multivariate jump-driven financial asset model. *Quantitative Finance* **6**, 385-402.

Lunn, D.J., Best N. and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**, 325-337.