Multivariate heavy-tailed models for Value-at-Risk estimation

Carlo Marinelli* Stefano d'Addona† Svetlozar T. Rachev‡

15 May 2010

Abstract

For purposes of Value-at-Risk estimation, we consider three multivariate families of heavy-tailed distributions, which can be seen as multidimensional versions of Paretian stable and Student's t distributions allowing different marginals to have different tail thickness. After a discussion of relevant estimation and simulation issues, we conduct a backtesting study on a set of portfolios containing derivative instruments, using historical US stock price data.

1 Introduction

The purpose of this paper is to assess the performance of some classes of multivariate laws with heavy tails in the estimation of Value-at-Risk for nonlinear portfolios. The inadequacy of Gaussian laws, in one or several dimensions, to model the distribution of risk factors, especially in view of applications to risk modeling, is well-documented in the empirical literature (see e.g. [3, 6] and references therein). Here we concentrate on models for risk factors that are multivariate extensions of the classical α-stable and Student's t distributions. In particular, we consider multivariate laws whose marginals may have different indices of tail thickness, and/or whose structure allow for tail dependence (i.e., roughly speaking, extreme movements of several risk factors may happen together).

Let us briefly recall how VaR is usually estimated for nonlinear (i.e. containing derivative instruments) portfolios, and what kind of improvements have been proposed. In the simplest setting, one uses a linear approximations of losses with normally distributed risk factors: denoting by L the loss over a certain time period, one sets $L \approx \langle \Delta, X \rangle$, where $X \sim N(m, Q)$ is a d-dimensional vector of Gaussian risk factors, Δ is an element of $\mathbb{R}^d$, and $\langle \cdot, \cdot \rangle$ stands for the usual scalar product of two vectors. Then $\langle \Delta, X \rangle$ follows a one-dimensional Gaussian distribution with mean $\langle \Delta, m \rangle$ and variance $\langle Q\Delta, \Delta \rangle$, so

*Facoltà di Economia, Università di Bolzano, Piazza Università 1, I-39100 Bolzano, Italy

†Department of International Studies, University of Rome 3, Via G. Chiabrera, 199, I-00145 Rome; daddona@uniroma3.it, and Baruch College CUNY, One Bernard Baruch Way, 10010 New York.

‡Institut für Wirtschaftstheorie und Statistik Karlsruher Institut für Technologie Gebäude 20.12 (Kollegium am Schloß) D-76128 Karlsruhe, Germany

that (an approximation of) VaR can be obtained immediately. However, it is clear that such a scheme suffers of two major weaknesses: the linear approximation is inaccurate, as the payoff function of derivatives is usually highly non-linear, and the hypothesis that random factors are Gaussian is often inappropriate, as briefly mentioned above (the literature on this issue is very rich – see e.g. [4, 9, 8], to mention just a few classical references). Among the many improvements that have been suggested in literature, some focus on a better modeling of the nonlinear relation between L and X (e.g. by using quadratic approximations of the type $L \approx \langle \Delta, X \rangle + \langle \Gamma X, X \rangle$), but still assuming X Gaussian (see e.g. [6]), while others introduce alternative distributions of portfolio losses, often just in the univariate setting (see e.g. [15]). To the best of our knowledge, however, there are only a small number of studies devoted to models that take into account both non-linearities and non-normality in a multivariate setting: Duffie and Pan [7] and Glasserman et al. [10] adopt the quadratic approximation and non-Gaussian risk factors. In particular, risk factors include a jump component in the first work, and are modeled by multivariate t distributions (or a modification thereof) in the latter. However, both works are devoted to different issues (analytic approximations and efficient simulation techniques, respectively), therefore they do not address the statistical issues related to the implementation of their models, and do not measure their empirical performance on real data.

Our contributions are the following: we introduce a stable-like model for risk factors obtained by multivariate subordination of a Gaussian law on $\mathbb{R}^d$ (see §2), such that each marginal (i.e. each risk factor) can have a different index of tail thickness. We construct estimators for the parameters of this distribution and we study their asymptotic behavior. An analogous program is carried out for a multivariate t-like law (see §3). The statistical properties of an alternative multivariate t-like law, obtained by "warping" the marginals of a standard multivariate t, which was introduced in [10], are studied in §4. In §6 we provide an extensive back-testing study of the three parametric families of distributions using real data, on portfolios containing both standard and exotic options, relying both on full revaluation of the portfolio value and on its quadratic approximation.

2 Multivariate stable-like risk factors

2.1 Description of the model

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $X : \Omega \to \mathbb{R}^d$ be a random vector of risk factors such that

$$X = A^{1/2} G, \tag{1}$$

where $A = \mathrm{diag}(A_1, \dots, A_d)$ is a diagonal random matrix with independent entries,

$$A_i \sim S_{\alpha_i/2}\Big(\big(\cos\frac{\pi\alpha_i}{4}\big)^{2/\alpha_i}, 1, 0\Big) \qquad \forall i = 1, \dots, d, \tag{2}$$

and G is a $\mathbb{R}^d$-valued Gaussian random vector, independent of A, with mean zero and covariance matrix Q. In (2) we assume $\alpha_i \in]1, 2[$ for all $i = 1, \dots, d$.

Note that (1) and (2) imply that, for each $i = 1, \ldots, d$, the i-th marginal of X has distribution $S_{\alpha_i}(\sigma_i, 0, 0)$, where $\mathbb{E}[G_i^2] = 2\sigma_i^2$. In particular, risk factors are allowed to have different indices of tail thickness α_i, and they are dependent through the Gaussian component G. Other distributional properties of this class of random vectors are discussed in §2.4 below.

2.2 Estimation

Let $X(t)$, $t = 1, \ldots, n$ be independent samples from the distribution of X. For $p < \min_{1 \leq i \leq d}(\alpha_i)/2$, define the (improper) sample p-th moment as

$$M_p(n) = n^{-1} \sum_{t=1}^{n} X_i(t)^{\langle p \rangle} X_j(t)^{\langle p \rangle},$$

where $X^{\langle p \rangle} := |X|^p \operatorname{sgn}(X)$. Note that, by Cauchy-Schwarz' inequality, we have

$$\mathbb{E}|X_i X_j|^p \leq \big(\mathbb{E}|X_i|^{2p}\big)^{1/2} \big(\mathbb{E}|X_j|^{2p}\big)^{1/2} < \infty,$$

thus also, by Kolmogorov's strong law of large numbers,

$$\lim_{n \to \infty} M_p(n) = \mathbb{E}(X_i X_j)^{\langle p \rangle} \quad \text{a.s..}$$

Since the random matrix A and the random vector G are independent, one has

$$\mathbb{E}(X_i X_j)^{\langle p \rangle} = \Big(\mathbb{E}A_i^{p/2}\Big)\Big(\mathbb{E}A_j^{p/2}\Big)\mathbb{E}(G_i G_j)^{\langle p \rangle},$$

where (see e.g. [16, p. 18])

$$\mathbb{E}A_i^{p/2} = \frac{2^{p/2}\Gamma(1 - p/\alpha_i)}{p \int_0^\infty u^{-p/2-1} \sin^2 u \, du} \Big(1 + \tan^2 \frac{\alpha_i \pi}{4}\Big)^{\frac{p}{2\alpha_i}} \Big(\cos \frac{\pi \alpha_i}{4}\Big)^{\frac{p}{\alpha_i}} \cos \frac{p\pi}{4} =: C_{\alpha_i, p}.$$

The constant $C_{\alpha,p}$ can be computed explicitly, recalling that

$$\int_0^\infty u^{-p/2-1} \sin^2 u \, du = -2^{p/2-1} \cos \frac{\pi p}{4} \Gamma(-p/2).$$

Let us now define the function

$$\begin{aligned} f_p :]-1, 1[&\to \mathbb{R} \\ q &\mapsto \mathbb{E}[(Z_1 Z_2)^{\langle p \rangle}], \end{aligned}$$

where Z_1, Z_2 are jointly normal random variables with covariance matrix

$$\begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}.$$

For any given $p < \min_i(\alpha_i)/2$, matching the theoretical signed p-th moments of X_iX_j with their sample counterparts, we obtain the following estimator for the matrix Q:

$$\hat{Q}_{ij} = 2\sigma_i\sigma_j f_p^{\leftarrow}\left(\frac{n^{-1}\sum_{t=1}^n X_i(t)^{\langle p\rangle}X_j(t)^{\langle p\rangle}}{2^p\sigma_i^p\sigma_j^p C_{\alpha_i,p}C_{\alpha_j,p}}\right), \qquad i,\, j = 1,\dots,d,$$

where the superscript "$\leftarrow$" stands (here and henceforth) for the inverse function.

If $\{\sigma_i\}_i$ and $\{\alpha_i\}_i$ are not known a priori, but we rather have only consistent estimators $\{\hat{\sigma}_{in}\}_i$ and $\{\hat{\alpha}_{in}\}_i$, respectively, one can easily deduce (by several applications of the continuous mapping theorem), that the estimator of Q obtained replacing α_i with $\hat{\alpha}_{in}$ and σ_i with $\hat{\sigma}_{in}$ in the above expression is still consistent.

Remark 1. (i) As far as the estimation of the covariance matrix Q is concerned, the heavy tailed assumption does not imply any extra computation burden.

(ii) For our purposes, it is enough to choose $p = 1/2$, as we always assume $\alpha_i > 1$ for all i (as is well-known, this is equivalent to assuming that all returns have finite mean).

(iii) Unfortunately we are not aware of an explicit expression for the function $q \mapsto f_p(q)$. However, it can be expressed as an integral with respect to a Gaussian measure in $\mathbb{R}^2$:

$$\begin{aligned} f_p(\rho) &= \frac{1}{2\pi\sqrt{\det Q}}\int_{\mathbb{R}^2}(x_1x_2)^{\langle p\rangle}e^{-\frac{1}{2}\langle Q^{-1}x,x\rangle}\,dx \\ &= \frac{1}{2\pi\sqrt{1-q^2}}\int_{\mathbb{R}^2}(x_1x_2)^{\langle p\rangle}e^{-\frac{1}{2(1-q^2)}(x_1^2-2qx_1x_2+x_2^2)}\,dx_1\,dx_2 \end{aligned} \tag{3}$$

which can be computed by numerical integration with essentially any accuracy. Figure 1 plots the function $f_{1/2}$ on the interval $[0, 1[$.

Let us consider a simplified case: $d = 2$, $\sigma_1 = \sigma_2 = 1/\sqrt{2}$, and α_1, α_2 given. The assumption $d = 2$ is harmless, as in any case the method works componentwise. The case of unknown α_i and σ_i can be dealt with replacing them with their corresponding consistent estimators, as discussed above.

Let us define

$$\hat{q}_n = f^{\leftarrow}\left(\frac{n^{-1}\sum_{t=1}^n X_1(t)^{\langle p\rangle}X_2(t)^{\langle p\rangle}}{C_{\alpha_1,p}C_{\alpha_2,p}}\right) \tag{4}$$

We first prove the following lemma:

Lemma 2. *The function* $f_p :]-1,1[\to \mathbb{R}$ *is bounded, continuously differentiable, concave increasing on* $]-1,0[$ *and convex increasing on* $]0,1[$.

Proof. Boundedness follows by concavity of the function $x \mapsto |x|^p$ for $p < 1$ and Jensen's inequality, that yield

$$|f_p(q)| = |\mathbb{E}Z_1^{\langle p\rangle}Z_2^{\langle p\rangle}| \le \mathbb{E}|Z_1Z_2|^p \le (\mathbb{E}|Z_1Z_2|)^p \le 1,$$

where the last inequality follows by Cauchy-Schwarz' inequality and $\mathbb{E}Z_1^2 = \mathbb{E}Z_2^2 = 1$. Continuous differentiability w.r.t. q is immediate by inspection of (3). Differentiating (3)

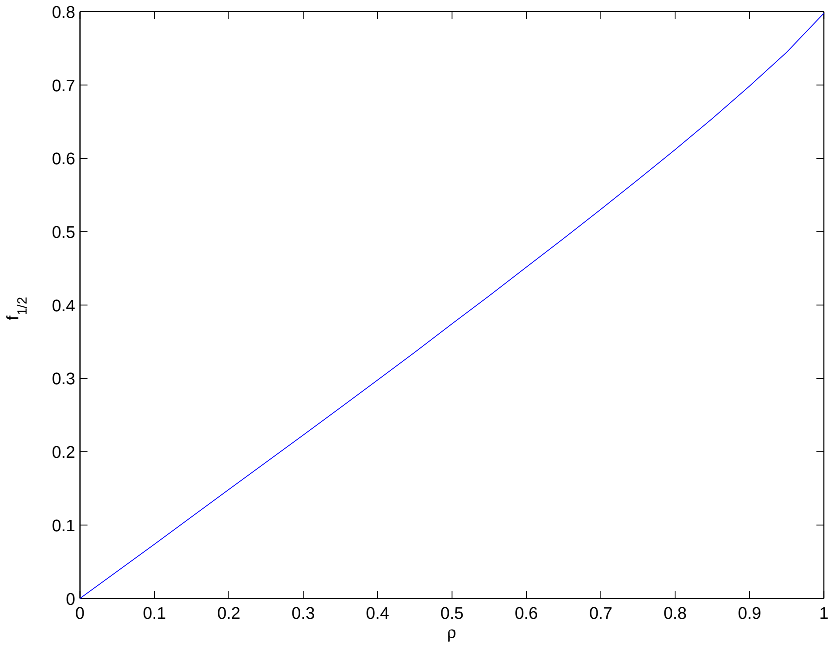


Figure 1: Plot of the function f_p, with $p = 1/2$.

w.r.t. q twice, one gets (after some cumbersome but elementary calculations) $f_p'(q) > 0$ for all $q \in]-1, 1[$, and $f_p''(q) < 0$ for $q < 0$, $f_p''(0) = 0$, $f_p''(q) > 0$ for $q > 0$. The lemma is thus proved. □

It is easy to prove that $\hat{q}_n$ is strongly consistent, i.e. that $\hat{q}_n \to q$ a.s. as $n \to \infty$. In fact, as above, since $p < (\min_i \alpha_i)/2$, by Kolmogorov's strong law of large numbers one has

$$f_p(\hat{q}_n) = \frac{n^{-1} \sum_{t=1}^{n} X_1(t)^{\langle p \rangle} X_2(t)^{\langle p \rangle}}{C_{\alpha_1,p} C_{\alpha_2,p}} \xrightarrow{n \to \infty} \mathbb{E}(Z_1 Z_2)^{\langle p \rangle} = f_p(q) \qquad \text{a.s.},$$

from which we can conclude thanks to the continuous mapping theorem and the continuity of $f^{\leftarrow}$.

We are now going to prove that the estimator (4) is asymptotically normal, under a more stringent assumption on the chosen value of p. Let us define the function $g_p : \mathbb{R}^2 \to \mathbb{R}$,

$$g_p : x = (x_1, x_2) \mapsto \frac{x_1^{\langle p \rangle} x_2^{\langle p \rangle}}{C_{\alpha_1,p} C_{\alpha_2,p}}.$$

It is clear that the estimator (4) can be defined as the solution of the equation

$$\mathbb{P}_n g_p := \frac{1}{n} \sum_{k=1}^{n} g_p(X(k)) = \mathbb{E}_q g_p(X) =: f_p(q), \tag{5}$$

where $\mathbb{P}_n$ stands for the (averaged) empirical measure of the sample $X(1), \ldots, X(n)$, i.e.

$$\mathbb{P}_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{X(k)}.$$

Proposition 3. *If $p < \min_i(\alpha_i)/4$, then $\hat{q}_n$ is asymptotically normal and satisfies*

$$\sqrt{n}(\hat{q}_n - q) \Rightarrow N\Big(0, f_p'(q)^{-2}\big(\mathbb{E}_q[g_p^2(X)] - f_p^2(q)\big)\Big). \tag{6}$$

Proof. We have proved in lemma 2 that $f_p(q) = \mathbb{E}_q g_p(X)$ is a bijection on the open set $]-1,1[$, it is continuously differentiable on its domain, and $f_p'(x) \neq 0$ for all $x \in]-1,1[$. Moreover, as it follows from (4) and (5), one can write

$$\sqrt{n}(\hat{q}_n - q) = \sqrt{n}\Big(f_p^{\leftarrow}(\mathbb{P}_n g_p) - f_p^{\leftarrow}(\mathbb{E}_q g_p(X))\Big). \tag{7}$$

We have, by the strong law of large numbers, $\mathbb{P}_n g_p \to \mathbb{E}_q g_p$ a.s. as $n \to \infty$. Recalling that by hypothesis $p < \min_i(\alpha_i)/4$, it follows that $\mathbb{E}_q g_p^2(X) < \infty$, hence, by the central limit theorem,

$$\sqrt{n}(\mathbb{P}_n g_p - \mathbb{E}_q g_p(X)) \Rightarrow N(0, \mathbb{E}_q g_p^2(X) - f_p^2(q)).$$

An application of the delta method, taking into account the inverse function theorem, now yields the result. □

A shortcoming of the asymptotic confidence interval implied by the above proposition is that the asymptotic variance depends on the parameter to be estimated itself. One can overcome this problem by a variance stabilizing transformation: let us define the function $\gamma :]-1,1[\to \mathbb{R}$,

$$\gamma_p(q) = \mathbb{E}_q g_p^2(X) - \big(\mathbb{E}_q g_p(X)\big)^2 = \frac{\mathbb{E}|X_1 X_2|^{2p}}{C_{\alpha_1,2p} C_{\alpha_2,2p}} - \frac{\big(\mathbb{E}X_1^{\langle p\rangle} X_2^{\langle p\rangle}\big)^2}{C_{\alpha_1,p}^2 C_{\alpha_2,p}^2}$$

and

$$\varphi_p(x) = \int_0^x \frac{f_p'(y)}{\gamma_p^{1/2}(y)}\, dy.$$

Then, again by the delta method, we obtain

$$\sqrt{n}(\varphi_p(\hat{q}_n) - \varphi_p(q)) \Rightarrow N\Big(0, \varphi_p'(q)^2 \frac{\gamma_p(q)}{f_p'(q)^2}\Big) = N(0,1),$$

and a corresponding asymptotic confidence interval for q as

$$q \in [\varphi_p^{\leftarrow}(\varphi_p(\hat{q}_n - z_\alpha/\sqrt{n}), \varphi_p^{\leftarrow}(\varphi_p(\hat{q}_n + z_\alpha/\sqrt{n})].$$

This asymptotic normality result for $\varphi_p(\hat{q}_n)$ would of course be better if we had an explicity expression for φ_p, which instead needs to be approximated numerically. However, since both f_p and γ_p are smooth functions (i.e. at least C^2), constructing a numerical approximation of φ_p is a rather simple task.

2.3 Simulation

In view of the results of the previous subsection, we assume that the covariance matrix Q is known, hence, with a slight but harmless abuse of notation, we shall write Q instead of $\hat{Q}$.

Random vectors from the distribution of X can be simulated by the following simple algorithm:

(i) generate d independent random variables $Z_i \sim N(0,1)$, $i = 1, \ldots, d$, and form the random vector $Z = (Z_1, \ldots, Z_d) \sim N(0, I)$, so that $Q^{1/2}Z \sim N(0, Q)$;

(ii) independently from Z, generate d independent random variables from the distribution of A_i, $i = 1, \ldots, d$, as defined in (2);

(iii) setting $A = \operatorname{diag}(A_1, \ldots, A_d)$, one has that $A^{1/2}Q^{1/2}Z$ is a sample from the d-dimensional law of X

Note that the only computational overhead with respect to the simulation of a Gaussian vector is the simulation of the stable subordinators, for which nonetheless efficient algorithms are available (see e.g. [16]).

2.4 Distributional properties

We shall now prove some distributional properties of the random vector of risk returns X. We begin showing that X can be regarded as a particular case of multivariate subordination, with a construction analogous to that of [1]. Let us recall that a subordinator in $\mathbb{R}^d$ is an $\mathbb{R}^d_+$-valued Lévy process with $T(0) = 0$ a.s. and $T_k(t)$ increasing a.s. for all $k = 1, \ldots, d$. Then one has, for $\gamma \in \mathbb{R}^d_+$,

$$\mathbb{E}e^{-\langle \gamma, T(t)\rangle} = \exp\left[t \int_{\mathbb{R}^d_+} (e^{i\langle \gamma, x\rangle} - 1)\nu(dx) + it\langle \gamma, c\rangle\right] =: e^{-t\psi(\gamma)}, \tag{8}$$

where $c \in \mathbb{R}^d_+$ and ν is a σ-finite measure on $\mathbb{R}^d$ such that $\operatorname{supp}\nu \subseteq \mathbb{R}^d_+ \setminus \{0\}$ and

$$\int_{\mathbb{R}^d_+} (1 \wedge |x|)\, \nu(dx) < \infty.$$

Given a process $Y(t)$, $t \geq 0$, with values in $\mathbb{R}^d$, we define the process subordinate to Y by T as

$$Y(T(t)) := (Y_1(T_1(t)), \ldots, Y_d(T_d(t))).$$

Proposition 4. *There exists a multivariate subordinator $T(t)$, $t \geq 0$, such that*

$$X \overset{d}{=} W(T(1)),$$

with W an $\mathbb{R}^d$-valued Wiener process with covariance matrix Q.

Proof. In fact, taking $T(t) = (A_1(t), \dots, A_d(t))$, with $A_i(t)$ independent $\alpha_i/2$-stable subordinator for all $i = 1, \dots, d$, by subordinating the Wiener process $W(t)$ we obtain

$$\begin{aligned}(W_1(A_1(t)), \dots, W_d(A_d(t))) &\stackrel{d}{=} (A_1(t)^{1/2} W_1(1), \dots, A_d(t)^{1/2} W_d(1)) \\ &\stackrel{d}{=} A^{1/2} G \stackrel{d}{=} X,\end{aligned}$$

choosing $t = 1$. □

As is well known, classical subordination (i.e. when subordinators are one dimensional increasing Lévy processes) is closely related to the class of type G laws. This remains true in the multivariate case, and the following proposition establishes a connection with a class of multivariate laws recently studied in [1].

Proposition 5. *The law of the random vector X is of type* mult$G$.

Proof. We have $X \stackrel{d}{=} A^{1/2} Q^{1/2} Z$, with $Z \sim N(0, I)$, hence also $X \stackrel{d}{=} S^{1/2} Z$, with $S = A^{1/2} Q^{1/2} (A^{1/2} Q^{1/2})^*$. The random matrix S is clearly symmetric positive definite, and its infinite divisibility follows immediately by

$$S = A^{1/2} Q A^{1/2} = n\Big(A^{1/2} \frac{Q}{n} A^{1/2}\Big),$$

therefore X is of class mult$G$ by definition. □

Since S is infinitely divisible as an element of the space of positive definite symmetric $d \times d$ matrices $L_1^+(\mathbb{R}^d)$, its law μ_S admits a Lévy measure m_S.

Proposition 6. *The law of the random vector X*

(i) *is infinitely divisible with characteristic triplet $[0, 0, m_X]$, where $m_X(dx) = u(x)\, dx$,*

$$u(x) = \int_{L_1^+(\mathbb{R}^d)} \frac{dN_{0,R}(x)}{dx}\, m_S(dR);$$

(ii) *is absolutely continuous with respect to Lebesgue measure with probability density function*

$$f_X(x) = \int_{L_1^+(\mathbb{R}^d)} \frac{dN_{0,R}(x)}{dx}\, \mu_S(dR);$$

(iii) *admits the characteristic function*

$$\mathbb{E} e^{i\langle \xi, X\rangle} = \mathbb{E} \exp\Big(-\frac{1}{2} \mathrm{Tr}(\xi\xi^* S)\Big) = \exp \int_{L_1^+(\mathbb{R}^d)} (e^{-\mathrm{Tr}(\xi\xi^* R)} - 1)\, \mu_S(dR),$$

where ξ is treated as a column vector.

Proof. Infinite divisibility of X follows immediately noting that for any $n \in \mathbb{N}$ we have

$$X \stackrel{d}{=} A^{1/2} G \stackrel{d}{=} \sum_{k=1}^{n} X_k \stackrel{d}{=} \sum_{k=1}^{n} A^{1/2} G_k,$$

with $\sum_{k=1}^{n} G_k \stackrel{d}{=} G$. The latter decomposition obviously always exists because Gaussian vectors are infinitely divisible. The characteristic triplet of X is of the type $[0, 0, m_X]$ because subGaussian random variables have no Gaussian component and mean zero (see e.g. [16]). Moreover, since we assumed that Q has full rank, and the laws of stable subordinators are absolutely continuous with respect to Lebesgue measure and their support coincides with $\mathbb{R}_+$, it follows that the Lévy measure m_S is supported on the space of strictly positive definite $d \times d$ matrices. Then the result follows by Proposition 3.1 in [1]. □

2.5 Extensions

Let us remark that the model (1) for the vector of risk factors can be extended to allow for asymmetries. In particular, setting

$$\tilde{X} = A^{1/2} G + B,$$

where B is a random vector with independent components B_i with law $S_{\alpha_i}(\sigma_{Bi}, \beta_{Bi}, 0)$, we have that the i-th marginal of the vector $\tilde{X}$ has distribution $S_{\alpha_i}(\tilde{\sigma}_i, \tilde{\beta}_i, 0)$, where

$$\tilde{\sigma}_i = (\sigma_i^{\alpha_i} + \sigma_{Bi}^{\alpha_i})^{1/\alpha_i}, \qquad \tilde{\beta}_i = \beta_{Bi} \frac{\sigma_{Bi}^{\alpha_i}}{\sigma_i^{\alpha_i} + \sigma_{Bi}^{\alpha_i}}.$$

One can then estimate the parameters $\tilde{\sigma}_i$ and $\tilde{\beta}_i$ fitting a general Paretian stable law to observed data, and obtain (in general in a non-unique way) values of σ_i, σ_{B_i}, and β_{B_i}. A common choice is $\beta_{B_i} = 1$, so that

$$\sigma_{B_i} = \tilde{\beta}_i^{1/\alpha_i} \tilde{\sigma}_i, \qquad \sigma_i = \big(1 - \tilde{\beta}_i\big)^{1/\alpha_i} \tilde{\sigma}_i.$$

Moreover, model (1) does not allow for tail dependence among different risk factors. As a remedy, one may use the series representation of stable subordinators (see e.g. [16]), setting

$$A_i = \sum_{k=0}^{\infty} \gamma_k^{2/\alpha_i}, \qquad i = 1, \ldots, n,$$

where $(\gamma_k)_{k \geq 0}$ is a (fixed) sequence of independent standard Gamma random variables. The analysis of this model, however, is considerably more involved, and we plan to elaborate on these issues in a future work.

3 Multivariate t-like risk factors

3.1 Description

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let us consider a d-dimensional random vector of risk factors X such that

$$X_k = \frac{G_k}{\sqrt{V_k/\nu_k}}, \qquad k = 1, \ldots, d, \tag{9}$$

where $G \sim N(0, Q)$ and $V_1, \ldots, V_d$ are independent one-dimensional χ^2-distributed random variables with parameters $\nu_1, \ldots, \nu_d$, respectively. We also assume that G and $(V_1, \ldots, V_d)$ are independent. Then, for each $k = 1, \ldots, d$, the k-th marginal of X is distributed according to a Student's t distribution with parameter ν_k, multiplied by $\sigma_k := (\mathbb{E}G_k^2)^{1/2}$. In particular, as in the case of the previous section, risk factors may have different indices of tail thickness (measured by ν_k), and their dependence comes from the Gaussian component G.

3.2 Estimation

Assuming for the time being ν_k, $k = 1, \ldots, d$, to be known, let us estimate the covariance matrix Q by the method of moments. We shall assume from now on that $\nu_k > 2$ for all k, which implies in particular that $\mathbb{E}X_k^2 < \infty$ for all k. One has

$$\begin{aligned}\mathbb{E}X_h X_k &= \sqrt{\nu_h \nu_k}\, \mathbb{E}G_h G_k \, \mathbb{E}V_h^{-1/2} \, \mathbb{E}V_k^{-1/2} \\ &= Q_{hk} \sqrt{\nu_h \nu_k}\, \mathbb{E}V_h^{-1/2} \, \mathbb{E}V_k^{-1/2}\end{aligned}$$

for all $h \neq k$, and

$$\mathbb{E}X_k^2 = Q_{kk}\, \nu_k \, \mathbb{E}V_k^{-1} = \sigma_k^2 \, \nu_k \, \mathbb{E}V_k^{-1}.$$

Denoting, for simplicity, a random variable with $\chi^2(\nu)$ distribution by V, the density of V is given by

$$f_\nu(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-\frac{x}{2}},$$

so that

$$\begin{aligned}\mathbb{E}V^{-1/2} &= \int_0^\infty x^{-1/2} f_\nu(x)\, dx = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty x^{\nu/2-3/2} e^{-\frac{x}{2}}\, dx \\ &= \frac{\Gamma(\nu/2 - 1/2)}{\sqrt{2}\,\Gamma(\nu/2)}\end{aligned}$$

and, similarly,

$$\mathbb{E}V^{-1} = \int_0^\infty x^{\nu/2-2} e^{-\frac{x}{2}}\, dx = \frac{1}{2} \frac{\Gamma(\nu/2 - 1)}{\Gamma(\nu/2)} = \frac{1}{\nu - 2}.$$

Here we have used the definition of Gamma function,

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt, \qquad z > 0,$$

and its "factorial" property $\Gamma(z+1) = z\Gamma(z)$. The above calculations yield

$$Q_{hk} = \frac{2}{\sqrt{\nu_h \nu_k}} \frac{\Gamma(\nu_h/2)}{\Gamma(\nu_h/2-1)} \frac{\Gamma(\nu_k/2)}{\Gamma(\nu_k/2-1)} \mathbb{E} X_h X_k, \qquad h \neq k,$$

and

$$Q_{kk} = \sigma_k^2 = \frac{\nu_k - 2}{\nu_k} \mathbb{E} X_k^2.$$

We have thus obtained the following moment estimator for Q:

$$\hat{Q}_{hk} = \frac{2}{\sqrt{\nu_h \nu_k}} \frac{\Gamma(\nu_h/2)}{\Gamma(\nu_h/2-1)} \frac{\Gamma(\nu_k/2)}{\Gamma(\nu_k/2-1)} \frac{1}{n} \sum_{t=1}^{n} X_h(t) X_k(t), \qquad h \neq k,$$

and

$$\hat{Q}_{kk} = \hat{\sigma}_k^2 = \frac{\nu_k - 2}{\nu_k} \frac{1}{n} \sum_{t=1}^{n} X_k(t)^2.$$

Note that, for each k, ν_k can be estimate by one-dimensional maximum likelihood on the k-th marginal, thus obtaining a family of consistent estimators $\hat{\nu}_k$, $k = 1, \ldots, d$. Therefore, the corresponding estimator of Q obtained by substituting in the previous expressions each ν_k with $\hat{\nu}_k$, for each k, is still consistent.

We can now prove that $\hat{Q}_{hk}$ is asymptotically normal. For compactness of notation, we shall set

$$C_\nu := \frac{\sqrt{2}}{\sqrt{\nu}} \frac{\Gamma(\nu/2)}{\Gamma(\nu/2-1)},$$

and we shall consider only the case $h \neq k$. The asymptotic normality of the estimators $\hat{\sigma}_k$ can be established analogously (see also §3.4).

Proposition 7. *Let* $d = 2$*,*

$$Q = \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix},$$

and

$$\hat{q}_n := C_{\nu_1} C_{\nu_2} \frac{1}{n} \sum_{t=1}^{n} X_1(t) X_2(t).$$

Then one has

$$\sqrt{n}\big(\hat{q}_n - q\big) \Rightarrow N\big(0, v_{\nu_1,\nu_2}(q)\big),$$

where

$$v_{\nu_1,\nu_2}(q) = \frac{\nu_1 C_{\nu_1}^2}{\nu_1 - 2} \frac{\nu_2 C_{\nu_2}^2}{\nu_2 - 2} (2q^2 + 1) - q^2$$

Proof. We have $\operatorname{Var} q_n = \mathbb{E} q_n^2 - q^2$ and

$$\begin{aligned} \mathbb{E} \hat{q}_n^2 &= C_{\nu_1}^2 C_{\nu_2}^2 \, \mathbb{E} X_1^2 X_2^2 = \nu_1 \nu_2 C_{\nu_1}^2 C_{\nu_2}^2 \, \mathbb{E} G_1^2 G_2^2 \, \mathbb{E} V_1^{-1} \, \mathbb{E} V_2^{-1} \\ &= \frac{\nu_1 C_{\nu_1}^2}{\nu_1 - 2} \frac{\nu_2 C_{\nu_2}^2}{\nu_2 - 2} \, \mathbb{E} G_1^2 G_2^2, \end{aligned}$$

where we have used the identity $\mathbb{E}V^{-1} = (\nu-2)^{-1}$. To compute $\mathbb{E}G_1^2G_2^2$, let us write

$$\begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \stackrel{d}{=} \begin{bmatrix} 1 & 0 \\ q & \sqrt{1-q^2} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix},$$

with $(Z_1, Z_2) \sim N(0, I)$. This yields, recalling that the fourth moment of a standard Gaussian measure is equal to 3,

$$\mathbb{E}G_1^2G_2^2 = q^2\mathbb{E}Z_1^4 + 1 - q^2 = 2q^2 + 1, \tag{10}$$

thus also

$$\operatorname{Var} \hat{q}_n = \frac{\nu_1 C_{\nu_1}^2}{\nu_1 - 2} \frac{\nu_2 C_{\nu_2}^2}{\nu_2 - 2}(2q^2+1) - q^2,$$

whence the result follows by the central limit theorem. □

Remark 8. One could derive from this asymptotic normality result an asymptotic confidence interval using a variance stabilizing transformation, as shown in the previous section.

3.3 Simulation

Generating random vectors from the distribution of a multivariate t-like distribution is a straightforward modification of the procedure outlined in §2.3 above.

3.4 Extensions

Since marginals of the random vector X follow a univariate t distribution, they are symmetric. In order to allow for asymmetric marginals, one may posit $X = (X_1, \ldots, X_d)$,

$$X_k := \tilde{X}_k - \eta_k := \frac{G_k + m_k}{\sqrt{V_k/\nu_k}} - \eta_k, \qquad k = 1, \ldots, d,$$

where $G \sim N(0, Q)$, and $m = (m_1, \ldots, m_d)$, $\eta = (\eta_1, \ldots, \eta_d) \in \mathbb{R}^d$. Then for the k-th marginal one has that $X_k + \eta_k$ follows a noncentral t-distribution. The reason for subtracting the vector η from $\tilde{X}$ is that $\mathbb{E}\tilde{X} \neq 0$, unless $m = 0$, and it is common to assume that risk factors have mean zero. Unfortunately the density of the noncentral t law is expressed in terms of a definite integral depending on parameters (see e.g. [17]), hence maximum likelihood estimation on the marginals becomes numerically quite involved. On the other hand, assuming $\nu_k > 4$ for all k, one can use the method of moments to construct estimators for $\nu = (\nu_1, \ldots, \nu_d)$, m, η and Q. In fact, considering k fixed and equal to 1 for the sake of simplicity, the constraint $EX_1 = 0$ translates into the relation

$$\eta_1 = m_1\sqrt{\nu_1}\,\mathbb{E}V_1^{-1/2} = m_1\sqrt{2\nu_1}\,\frac{\Gamma((\nu_1+3)/2)}{\Gamma(\nu_1/2)}.$$

Since we need to estimate four parameters, we need other three equations. These can be obtained by matching the second, third, and fourth sample moments to the corresponding theoretical moments, which are known in closed form (see e.g. [11]).

We should also observe that in general it is not necessary to match moments of integer order to obtain consistent and asymptotically normal estimators. One may also use fractional moments, as it has been done in the previous section, thus relaxing the assumptions on the parameters ν_k. For instance, let X be as in (9), $d = 2$, $Q = \left[\begin{smallmatrix} 1 & q \\ q & 1 \end{smallmatrix}\right]$, and consider the problem of estimating q. Setting $g_p(x_1, x_2) = x_1^{\langle p \rangle} x_2^{\langle p \rangle}$, we can write

$$\mathbb{E}g_p(X) = (\nu_1 \nu_2)^{p/2} \mathbb{E}V_1^{-p/2} \mathbb{E}V_2^{-p/2} \mathbb{E}(G_1 G_2)^{\langle p \rangle}.$$

Note that $\mathbb{E}(G_1 G_2)^{\langle p \rangle} = f_p(q)$, where f_p is the function introduced and studied in §2.2, and, in analogy to a calculation already encountered in this section,

$$\mathbb{E}V_k^{-p/2} = \frac{\Gamma(\nu_k/2 - p/2)}{2^{p/2}\Gamma(\nu_k/2)}, \qquad k = 1, 2. \tag{11}$$

This relation can be used as a basis for a moment estimator, as in §2.2. Choosing p small enough, one does not need to assume $\nu_k > 2$.

4 Warped multivariate t risk factors

4.1 Description

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let G and V be a d-dimensional random vector with law $N(0, Q)$ and an independent one-dimensional random variable with χ^2 distribution with ν_0 degrees of freedom, respectively. We shall call the law of the random vector $X' = G/\sqrt{V/\nu_0}$ a multivariate t distribution (with parameters ν_0 and Q). There are other possible multivariate generalizations of Student's t distribution (see e.g. [13]), but we shall concentrate exclusively on this definition. In contrast to the models considered in the previous two sections, the marginals of X have the same tail thickness (as measured by ν_0), but one expects nontrivial tail dependence between any two marginals. In order to allow for different tail behavior along different coordinates, one may set $X = f(X')$, where f is a deterministic nonlinear injective function, for instance defined through the marginals of X' and a copula function. The typical situation (see e.g. [10]) is to set

$$X_k = \delta_k F_{\nu_k}^{\leftarrow}(F_{\nu_0}(X'_k)), \qquad k = 1, \dots, d,$$

where $\delta_k > 0$ for all k, F_ν denotes the distribution function of a one dimensional t law with ν degrees of freedom, and $Q_{kk} = 1$ for all k. It is clear that the k-th marginal is t distributed with ν_k degrees of freedom, thus overcoming the problem of having all marginals with the same tail thickness.

4.2 Estimation

From the statistical point of view (hence also from an implementation perspective) the model suffers from a major drawback. In fact, once one has estimated the parameters ν_k and δ_k, $k = 1, \dots, d$ (for instance by maximum likelihood), in order to estimate Q we would need to know the "true" value of ν_0. In other words, the model is *not* identifiable given observations of X, as it requires to arbitrarily choose the value of a parameter a priori.

We shall assume from now on that ν_0 is fixed. We shall also treat ν_k and δ_k as known, although the latter assumption comes at no loss of generality, since the two parameters can indeed be estimated. Let us show how one can estimate the covariance matrix Q. As we have already done several times, we are going to consider only the case $d = 2$. Since

$$F_{\nu_0}^{\leftarrow}\big(F_{\nu_k}(X_k/\delta_k)\big) \stackrel{d}{=} \frac{G_k}{\sqrt{V/\nu_k}}, \qquad k = 1, 2, \tag{12}$$

we have, recalling that $\mathbb{E}V^{-1} = (\nu_0 - 2)^{-1}$,

$$q := \mathbb{E}G_1G_2 = \frac{\nu_0 - 2}{\nu_0}\mathbb{E}\big[F_{\nu_0}^{\leftarrow}\big(F_{\nu_1}(X_1/\delta_1)\big)F_{\nu_0}^{\leftarrow}\big(F_{\nu_2}(X_2/\delta_2)\big)\big].$$

It is then natural to define the estimator

$$\hat{q}_n := \frac{\nu_0 - 2}{\nu_0}\frac{1}{n}\sum_{t=1}^{n} F_{\nu_0}^{\leftarrow}\big(F_{\nu_1}(X_1(t)/\delta_1)\big)F_{\nu_0}^{\leftarrow}\big(F_{\nu_2}(X_2(t)/\delta_2)\big),$$

which is easily seen to be consistent. Since the function $(x, \nu) \mapsto F_\nu(x)$ is continuous as a map from $\mathbb{R}^2$ to $\mathbb{R}$, one still has $\hat{q}_n \to q$ in probability as $n \to \infty$ if we replace ν_1, ν_2, δ_1 and δ_2 with corresponding consistent estimators in the previous display.

Asymptotic normality of $\hat{q}_n$ follows easily by the central limit theorem. In particular, by (12) one easily obtains

$$\mathbb{E}\hat{q}_n^2 = (\nu_0 - 2)^2\mathbb{E}V^{-2}\,\mathbb{E}(G_1G_2)^2.$$

Recalling (11), the factorial property of the Gamma function, and taking (10) into account, we get

$$\operatorname{Var}\hat{q}_n = \frac{\nu_0 - 2}{\nu_0 - 4}(2q^2 + 1) - q^2.$$

We have thus proved that, if $\nu_0 > 4$, one has

$$\sqrt{n}(\hat{q}_n - q) \Rightarrow N\Big(0, \frac{\nu_0 - 2}{\nu_0 - 4}(2q^2 + 1) - q^2\Big).$$

In analogy to §2.2, one can use fractional moments of order p, with p sufficiently small, to relax the restriction on ν_0 and still obtain asymptotic normality of $\hat{q}_n$. Similarly, by a variance-stabilizing transformation, one can easily obtain an asymptotic confidence interval independent of the "true" value of q.

Let us discuss more closely the issue of non-identifiability of the model. In particular, in view of the application to the parametric estimation of VaR (see next section), it is natural to ask how sensitive the estimates of Q and of the quantiles can be with respect to the chosen value of ν_0. Even though it seems unlikely to be able to give a general answer in analytic terms, it is not difficult to obtain qualitative informations through a numerical study. In particular, let X be a sample from the warped multivariate t distribution in $\mathbb{R}^2$ with parameters ν_0, ν_1, ν_2 and Q. We keep ν_1, ν_2 and Q fixed, and let ν_0 vary in the simulation as well as in the estimation step. We also compute the α-quantiles (with $\alpha \in \{0.95, 0.99\}$) for the random variable $Y := w_1X_1 + w_2X_2$, with $w_1 + w_2 = 1$, again for varying values of ν_0 in the simulation and the estimation step. The results are plotted in figure 2 and figure 3, respectively.

One can see that the estimates of Q and of the quantiles of Y are very sensitive to the choice of a test ν_0 (by this we mean the value of ν_0 chosen for estimation purposes), unless both the true and the test value of ν_0 are sufficiently large. Moreover, considering a fixed test value of ν_0, the estimates of the quantiles of Y are essentially insensitive to the true value ν_0. One may thus conclude that

(i) the estimates of the covariance matrix of a warped multivariate t law are not reliable;

(ii) from the perspective of estimating quantiles, since they are essentially insensitive to the true value of ν_0, one could as well take $\nu_0 = \infty$, thus reducing to the case of X' being Gaussian.

4.3 Simulation

Random samples from the distribution of X can be generated again by a rather straightforward modification of the procedure outlined in §2.3. In fact, the distribution function of the univariate t distribution, as well as its inverse, are implemented in several software packages (such as Octave), even though they do not admit a closed-form representation.[1]

5 Estimation of Value-at-Risk by simulation

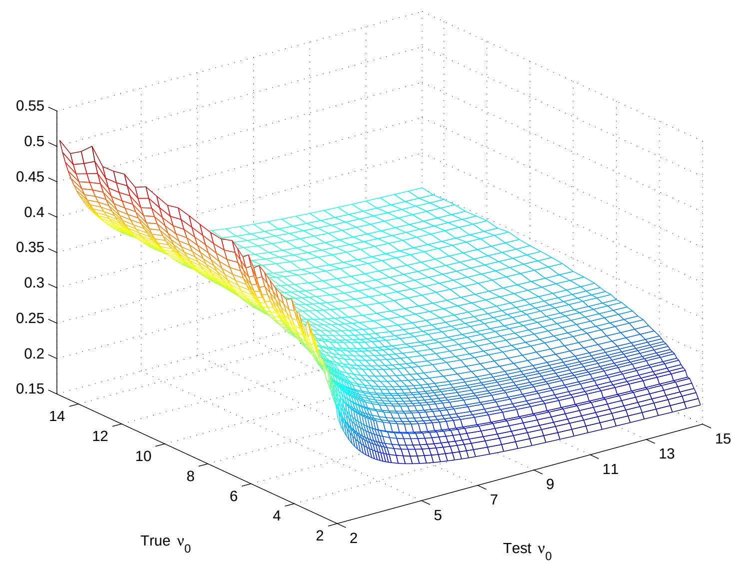
We shall denote by L the loss of a portfolio depending on the vector of risk factors X. Recall that the Value-at-Risk (VaR) of a portfolio at confidence level β (usually $\beta = 0.95$ or $\beta = 0.99$) is simply the β quantile of the distribution of portfolio losses. Since it is in general very difficult, if not impossible, to obtain analytically tractable expressions for the distribution function of the random variable L (even if the density function, or the characteristic function, of the vector X is known in closed form), one usually estimates quantiles of L by generating random samples from its distribution and computing the corresponding empirical quantiles. We shall exclusively deal with the so-called parametric (estimated) VaR, in the sense that we fit to observed data the parameters of a given family of distributions for the vector X of random factors, and we generate random samples from the law of X. In order to obtain a sample from the law of L we should know the functional relation between L and X. For a linear portfolio (roughly, a portfolio without derivative instruments), one simply has $L = \langle w, X \rangle$, where $w \in \mathbb{R}^d$. In the more interesting case of a portfolio containing derivatives, one has

[1] It might be better to say that they do, but in terms of hypergeometric functions.

Figure 2: Correlation estimates

This figure plots the correlation estimates from the warped multivariate t distribution. The parameters ν_1, ν_2 and Q are kept fixed to the values 4, 6, and 0.3, respectively, while we let ν_0 vary through the set $\{2.1, 15\}$ both in the simulation and in the estimation step. Panel A shows the whole simulation grid, while panel B shows the subset $\nu_0 \in \{2.1, 7\}$. Both panels are obtained with simulated samples of $100,000$ replications.

Panel A



Panel B

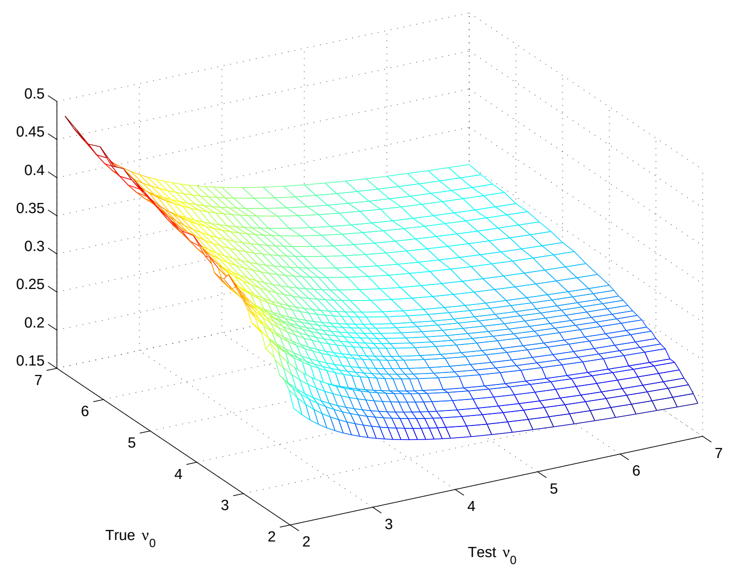
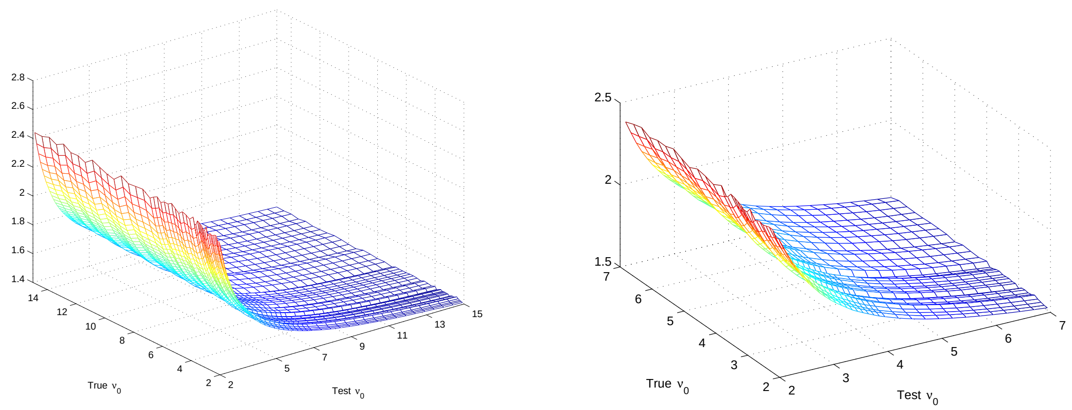


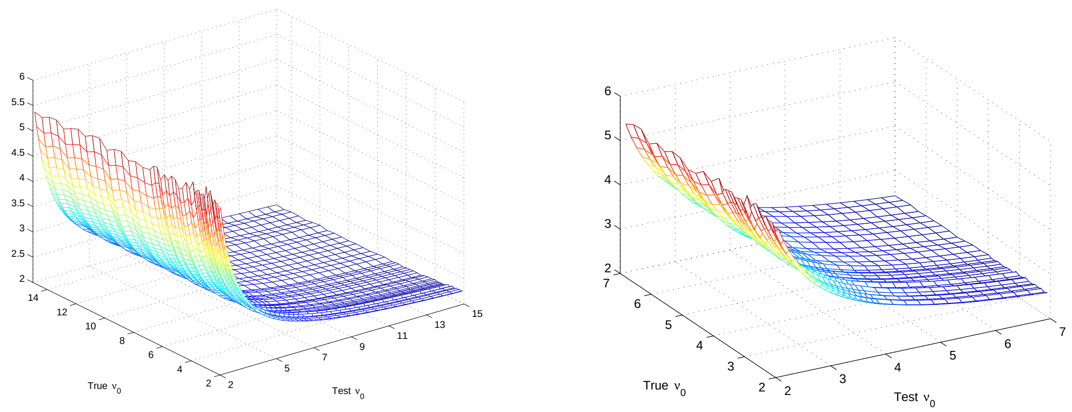
Figure 3: Quantiles estimates

This figure plots the quantiles estimates from the warped multivariate t distribution. The parameters ν_1, ν_2 and Q, as well as the values of ν_0, are exactly as in figure 2. Panel A plots the 95% quantiles of Y, while panel B plots the 99% quantiles. Both panels are obtained with simulated samples of $100,000$ replications.

Panel A: 95% quantiles



Panel B: 99% quantiles



$L = f(X)$, where $f : \mathbb{R}^d \to \mathbb{R}$ is a nonlinear function. Unless the derivatives in the portfolio are very simple, the function f may not admit a closed-form representation, or could just be obtained by nontrivial numerical procedures, that would have to be carried out for each random sample of X. For this reason one usually relies on approximations of the function f of the form

$$L \approx f(0) + \langle f'(0), X \rangle + \frac{1}{2} \langle f''(0)X, X \rangle,$$

which is obviously motivated by the second-order Taylor expansion of the function $\mathbb{R}^d \ni x \mapsto f(x)$ around zero. The values of $f'(0)$ and $f''(0)$ are in general determined by the so-called greeks (in this case, Delta, Gamma and Theta) of the derivatives in the portfolio. Note that in the above approximation the possible dependence of f on time can be taken into account by including time in the set of risk factors.

The analytic computation, or just approximation, of quantiles of quadratic forms in random vectors (other than Gaussian) is in general a very difficult task. Simulation is hence a viable alternative, as long as one can generate samples from the distribution of X.

We are going to perform a backtesting study on the three classes of parametric models for the distribution of risk factors introduced in Sections 2-4, to which we refer for the corresponding estimation and simulation procedures. Value-at-Risk is just estimated by empirical quantiles of random samples of L, either obtained by full revaluation, or by the above quadratic approximation. In particular, we do not focus on efficient simulation methods for quantile estimation, but we are rather interested on the relative performance of different distributional hypotheses for risk factors, when tested on real data.

6 Empirical tests

6.1 The data set

We tried to mimick a (US) domestic investor with a potentially highly correlated stock exposure. To do so we chose two US stocks from each of four different industries[2]. The raw price series are freely available on the web, and the returns are calculated as daily log-differences[3]. The data set covers the time period from 2-Jan-1991 until 31-Dec-2008.

Let us provide a few descrptive statistics of the data set. Table 1 collects mean, standard deviation, skewness and kurtosis for each stock return. Note that the sample kurtosis is larger than 3 for most return series, which could be interpreted as evidence of tail-thickness of the underlying distribution. Table 2 reports the (whole sample) correlation matrix for the eight return series. The correlation coefficients are all positive

[2]The selected stocks are Apple, Bank of America, Chevron, Citigroup Conoco, Microsoft Johnson and Johnson, and Pfitzer.

[3]We restrict ourselves to consider daily data for two reasons: the first and most important is that the industry and regulatory standard is to compute VaR and related risk measures on a daily basis. On the other hand, studying lower frequencies (such as weekly or monthly) would considerably decrease the size of our samples, possibly invalidating the asymptotic properties of the proposed estimators.

Table 1: Descriptive statistics of financial series

This table reports some descriptive statistics for the log-returns of the analyzed time series.

	Descriptive statistics			
	Mean	Standard deviation	Skewness	Kurtosis
Microsoft	0.000	0.022	0.027	5.184
Chevron	0.000	0.016	0.153	10.488
Apple	0.000	0.033	-2.196	54.773
Conoco	0.000	0.018	-0.253	5.372
Bank of America	0.000	0.023	-0.459	22.909
Citigroup	0.000	0.026	0.412	35.400
Johnson & Johnson	0.000	0.015	-0.153	6.733
Pfitzer	0.000	0.019	-0.161	2.912

Table 2: Correlation matrix of financial series

This table reports the sample correlation matrix for the log-returns of the analyzed time series.

	Correlation matrix							
	Msft	Chevron	Apple	Conoco	BoA	Citigroup	J&J	Pfitzer
Msft	1.00	0.26	0.35	0.22	0.33	0.36	0.27	0.30
Chevron	0.26	1.00	0.14	0.68	0.30	0.31	0.27	0.29
Apple	0.35	0.14	1.00	0.16	0.24	0.24	0.12	0.14
Conoco	0.22	0.68	0.16	1.00	0.28	0.27	0.23	0.25
BoA	0.33	0.30	0.24	0.28	1.00	0.70	0.27	0.32
Citigroup	0.36	0.31	0.24	0.27	0.70	1.00	0.30	0.35
J&J	0.27	0.27	0.12	0.23	0.27	0.30	1.00	0.52
Pfitzer	0.30	0.29	0.14	0.25	0.32	0.35	0.52	1.00

in a range from 0.12 to 0.70. Let us assess the significance of the correlation amongst the selected return series using Bartlett's test (see [2]). Denoting the sample size by n, under the null hypothesis that the correlation matrix $C \in \mathbb{R}^{d\times d}$ is equal to the identity, the statistics

$$-\big((n-1)-(2d+5)/6\big)\log(\det C) \tag{13}$$

is distributed as a χ^2 random variable with $d(d-1)/2$ degrees of freedom. Calculated on our sample of eight stocks, the value of the statistic is 9958, that is well above the value of $\chi^2_{99\%}(28) = 48.278$. So we can safely reject the hypothesis that the returns of the selected stocks are uncorrelated.

6.2 Test portfolios

We construct three test portfolios adding to a basic linear portfolio containing only the eight "underlyings", in equally value-weighted proportions, the following positions in options:

NLL long 10 calls and 5 puts on each asset ("NonLinear Long");

NLS short 5 calls and 10 puts on each asset ("NonLinear Short");

NLDC short 10 down-and-out calls with barrier equal to 95% of the asset price, and short 5 cash-or-nothing put with cash payoff equal to the strike price ("NonLinear Down and Cash").

All options are European, at-the-money, and with time to expiration equal to 6 months.

The nonlinear part of the three test portfolios is "synthetic", in the sense that the option prices, unlike the stock prices, are computed on the basis of the information available on the corresponding underlying and the time series of (a proxy for) the risk-free rate, using Black-Scholes formula for the standard calls and puts, and its variants for the barrier and binary options[4]. Even though this procedure is incompatible with the non-Gaussian distributional assumptions we are going to test, this is nonetheless common practice (see e.g. [10] for a more thorough discussion of this issue).

6.3 Backtesting

Let us now turn to the analysis of the performance of the three parametric distributions for risk factors introduced above, when applied to the (predictive, i.e. out-of-sample) estimation of Value-at-Risk. More precisely, we fit each of the three multivariate distributions to a subset of the time series of stock returns (using a rolling window consisting of 250 observations), and we estimate the 0.95 and 0.99 quantiles of the distribution of losses by simulation, i.e. selecting the corresponding empirical quantiles from a simulated sample. In particular, once a random sample from the distribution of X is obtained, we translate it into a random sample from the distribution of portfolio losses either by a full revaluation of the portfolio value for each sample, or by the usual delta-gamma quadratic approximation (see §5). Let $[t-\ell, t]$ denote the time period over which the parametric families of distributions are estimated, where ℓ stands for the (fixed) length of the rolling window. Denoting by VaR_t the empirical quantiles of the simulated distribution of losses (with risk factors fitted over $[t-\ell, t]$), we form the statistic

$$\xi_{t+1} = \mathrm{sgn}^+(L_{t+1} - \mathrm{VaR}_t),$$

for all $t \in [\ell, T]$, where T denotes the length of the time series, L_t stands for the observed loss of portfolio value over the period $[t-1, t]$, and where $\mathrm{sgn}^+ x = 1$ if $x > 0$, and equals zero otherwise. This procedure produces a different set of $(\xi_t)_{\ell \le t \le T}$, for each combination of test portfolio, model for risk factors, quantile level (95% and 99%), and portfolio revaluation method (full vs. quadratic approximation).

To assess the accuracy of the VaR estimates, we perform a simple Proportion of Failure (PoF) test (cf. [14]), which mimicks a classical likelihood-ratio test. In particular, setting

$$\zeta = -2 \log \left(\frac{(1-\beta)^x \beta^{(T-\ell-x)}}{p^x (1-p)^{(T-\ell-x)}} \right), \tag{14}$$

[4]We provide formulas for prices and sensitivities of these exotic options in Appendix A.

where $\beta \in \{0.95,\, 0.99\}$,

$$x := \sum_{t=\ell+1}^{T} \xi_t, \qquad p := \frac{x}{T-\ell},$$

one expects ζ to be asymptotically χ^2 distributed with one degree of freedom. Therefore, the corresponding VaR model can be considered reliable with a 95% confidence level if $\zeta < \zeta_0 \approx 3.84$.

The results of the backtesting procedure with full revaluation are collected in table 3, where values of x are in the first column, p in the second column, and ζ in the third column. Note that we included, for comparison, VaR estimates obtained under the assumptions that risk factors are jointly Gaussian. Moreover, due to the identifiability problem of warped t distributions discussed in §4, we report the backtesting results for three choices of ν_0.

As one may expect, the "benchmark" Gaussian approach fails at 99% confidence level for all three test portfolios. On the other hand, as far as VaR estimates at 95% confidence level are concerned, the Gaussian approach is still satisfactory. The same observations apply to the multivariate t-like approach. The stable-like approach instead is rejected by the PoF test only once. We may therefore say that, between the two models constructed by multiplying the marginals of a Gaussian vector by a set of independent random variables (with suitable distribution), the stable-like approach is preferable. The clear winner, however, is the family of warped t laws, whose VaR estimates cannot be rejected for any one of the test portfolios. Surprisingly enough, the results essentially do not depend on the choice of the "test" value for the non-identifiable parameter ν_0. For any given test portfolio, quantile estimates are not sensitive to the choice of $\nu_0 \in \{3, 5, 7\}$ (although this cannot be seen directly from the tables, their relative difference is not above 1%). While this suggests that the distribution of risk factors cannot be a warped multivariate t law (cf. the discussion in §4), it seems that this class of models can be surprisingly robust for purposes of risk management. The main message, however, could be that models allowing for tail dependence may have an advantage with respect to models lacking this feature.

Completely analogous observations could be made for the estimates of VaR obtained by the delta-gamma quadratic approximation of portfolio losses, for which we refer to table 4. As in the case of full revaluation, the Gaussian approach performs remarkably well at the 95% confidence level. In this respect, it is probably worth recalling that obtaining the quantiles of a quadratic form in Gaussian vectors is particularly simple and can be done with very little computational effort. In this sense, the classical quadratic approximation with Gaussian risk factors could still be regarded as a useful tool.

A Prices and sensitivities of some exotic options

Throughout this appendix we place ourselves in a standard Black-Scholes model with one "underlying" stock, whose price process is denoted by S_t, $0 \leq t \leq T$, and whose (constant) volatility is denoted by σ. The risk-free rate will be denoted by r. We shall consider options written on the stock, denoting the exercise time by T, the strike price by K, and the barrier by H.

In the following table we collect the definitions, in terms of their payoff, of some barrier and binary options.

NAME	PAYOFF	
Down-and-In call	$\max(S_T - K, 0)$	if $\min_{0\le t\le T} S_t \le H$
Down-and-In put	$\max(K - S_T, 0)$	if $\min_{0\le t\le T} S_t \le H$
Down-and-Out call	$\max(S_T - K, 0)$	if $\min_{0\le t\le T} S_t \ge H$
Down-and-Out put	$\max(K - S_T, 0)$	if $\min_{0\le t\le T} S_t \ge H$
Up-and-In call	$\max(S_T - K, 0)$	if $\max_{0\le t\le T} S_t \ge H$
Up-and-In put	$\max(K - S_T, 0)$	if $\max_{0\le t\le T} S_t \ge H$
Up-and-Out call	$\max(S_T - K, 0)$	if $\max_{0\le t\le T} S_t \le H$
Up-and-Out put	$\max(K - S_T, 0)$	if $\max_{0\le t\le T} S_t \le H$
Cash-or-Nothing call	1	if $S_T \ge K$
Cash-or-Nothing put	1	if $S_T \le K$

We shall use C_{di} and P_{di} to denote the price (at time zero) of a down-and-in call and a down-and-in put, respectively. Completely analogous notation will be used for the remaining options, replacing the subscripts accordingly. The price of plain European call and put options will be denoted by C_{BS} and P_{BS}, respectively. The price at time zero of a European call option with strike K and exercise time T, written on an underlying whose price at time zero is S_0, will be denoted by $C_{BS}(S_0, K, T)$. The corresponding notation will be also used for European put options.

Setting

$$\lambda = \frac{2r}{\sigma^2} - 1, \qquad m = \frac{r}{\sigma^2} + \frac{1}{2}$$

and assuming $H < K$, one has (see e.g. [5]),

$$\begin{aligned} C_{di} &= H^{\lambda} S_0^{-\lambda} C_{BS}(H^2 S_0^{-1}, K, T), \\ P_{di} &= C_{di} + K H^{-1} P_{BS}(S_0, H, T) - (H S_0^{-1})^{2m-2} H K^{-1} C_{BS}(K H S_0^{-1}, K^2 H^{-1}, T), \\ C_{ui} &= C_{BS} \\ P_{ui} &= H^{\lambda} S_0^{-\lambda} P_{BS}(H^2 S_0^{-1}, K, T). \end{aligned}$$

By the obvious identities

$$C_{di} + C_{do} = C_{BS}, \qquad C_{ui} + C_{uo} = C_{BS},$$

and the corresponding ones for put options (i.e. those obtained replacing C with P), we obtain pricing formulas for all barrier options listed in the above table. By the well-known formulas for sensitivities of European call and put options, elementary calculus yields

$$\begin{aligned}\frac{\partial C_{di}}{\partial S_0} &= -\lambda H^{\lambda} S_0^{-\lambda-1} C_{BS}(H^2 S_0^{-1}, K, T) \\ &\quad - H^{\lambda+2} S_0^{-\lambda-2} \Delta_{BS}(H^2 S_0^{-1}, K, T), \\ \frac{\partial^2 C_{di}}{\partial S_0^2} &= \lambda(\lambda+1) H^{\lambda} S_0^{-\lambda-2} C_{BS}(H^2 S_0^{-1}, K, T) \\ &\quad + 2(\lambda+1) H^{\lambda+2} S_0^{-\lambda-3} \Delta_{BS}(H^2 S_0^{-1}, K, T) \\ &\quad + H^{\lambda+4} S_0^{-\lambda-4} \Gamma_{BS}(H^2 S_0^{-1}, K, T), \\ \frac{\partial C_{di}}{\partial T} &= H^{\lambda} S_0^{-\lambda} \Theta_{BS}(H^2 S_0^{-1}, K, T).\end{aligned}$$

Similar expressions can be derived for the sensitivities of the other binary options.

Setting

$$d_1 := \frac{\log(S_0/K) + (r+\sigma^2)T}{\sigma\sqrt{T}}, \qquad d_2 := d_1 - \sigma\sqrt{T},$$

we have (see e.g. [12])

$$C_{cn} = e^{-rT}\Phi(d_2), \qquad P_{cn} = e^{-rT}\Phi(-d_2),$$

where $\Phi(\cdot)$ stands for the distribution function of the Gaussian law on $\mathbb{R}$ with mean zero and unit variance. The sensitivities of binary options are just an exercise in elementary calculus. Let us include, for the sake of completeness, the sensitivities of the cash-or-nothing put, which is used in our portfolios:

$$\begin{aligned}\frac{\partial P_{cn}}{\partial S_0} &= \frac{-e^{-rT}\Phi'(-d_2)}{\sigma S_0 \sqrt{T}}, \\ \frac{\partial^2 P_{cn}}{\partial S_0^2} &= \frac{e^{-rT}\Phi'(-d_2)}{\sigma S_0^2 \sqrt{T}} + \frac{-d_2 e^{-rT-d_2^2/2}}{\sigma^2 T S_0^2 \sqrt{2\pi}}, \\ \frac{\partial P_{cn}}{\partial T} &= -re^{-rT}\Phi(-d_2) + \frac{re^{-rT}\Phi'(-d_2)\log(S_0/K)}{2\sigma T^{3/2}} - \frac{r-\sigma^2/2}{\sigma\sqrt{T}}.\end{aligned}$$

References

[1] O. E. Barndorff-Nielsen and V. Pérez-Abreu. Extensions of type G and marginal infinite divisibility. *Teor. Veroyatnost. i Primenen.*, 47(2):301–319, 2002.

[2] M. S. Bartlett. Test of significance in factor analysis. *The British J. of Psychology*, 3:77–85, 1950.

[3] J. Berkowitz and J. O'Brien. How accurate are Value-at-Risk models at commercial banks? *J. of Finance*, 57(3):1093–1111, 2002.

[4] R. Blattberg and N. Gonedes. A comparison of the stable and student distributions as statistical models of stock prices. *J. of Business*, 47:244–280, 1974.

[5] P. Carr, K. Ellis, and V. Gupta. Static hedging of exotic options. *J. of Finance*, 53(3):1165–1190, 06 1998.

[6] D. Duffie and J. Pan. An overview of value at risk. *J. of Deriv.*, 4:7–72, 1997.

[7] D. Duffie and J. Pan. Analytical value-at-risk with jumps and credit risk. *Finance Stoch.*, 5(2):155–180, 2001.

[8] E. Fama. The behaviour of stock market prices. *J. of Business*, 38, 1965.

[9] E. Fama. Portfolio analysis in a stable paretian market. *Management Sci.*, 3:404–419, 1965.

[10] P. Glasserman, Ph. Heidelberger, and P. Shahabuddin. Portfolio value-at-risk with heavy-tailed risk factors. *Math. Finance*, 12(3):239–269, 2002.

[11] D. Hogben, R. S. Pinkham, and M. B. Wilk. The moments of the non-central t-distribution. *Biometrika*, 48:465–468, 1961.

[12] J.C. Hull. *Option, futures, and other derivatives.* Pearson Prentice Hall, Upper Saddle River, New Jersey, seventh edition, 2009.

[13] S. Kotz and S. Nadarajah. *Multivariate t distributions and their applications.* Cambridge University Press, Cambridge, 2004.

[14] P. H. Kupiec. Techniques for verifying the accuracy of risk management models. *J. Deriv.*, 3:73–84, 1995.

[15] S. T. Rachev and S. Mittnik. *Stable Paretian Models in Finance.* John Wiley and Sons, NY, 2000.

[16] G. Samorodnitsky and M. S. Taqqu. *Stable non-Gaussian random processes.* Chapman & Hall, New York, 1994.

[17] A. van Aubel and W. Gawronski. Analytic properties of noncentral distributions. *Appl. Math. Comput.*, 141(1):3–12, 2003.

Table 3: Value-at-Risk backtesting

This table reports the results of a Value-at-Risk backtesting. Panels A and B report the results for the short and long portfolios, respectively, while Panel C reports the results for the down-and-out and cash-or-nothing portfolio.

Panel A: NLS			
	Violations	Percentage	LR
Warped $t_{95\%}(\nu_0 = 3)$	235	5.48%	2.02
Warped $t_{99\%}(\nu_0 = 3)$	35	0.82%	1.56
Warped $t_{95\%}(\nu_0 = 5)$	226	5.27%	0.65
Warped $t_{99\%}(\nu_0 = 5)$	39	0.91%	0.36
Warped $t_{95\%}(\nu_0 = 7)$	224	5.22%	0.44
Warped $t_{99\%}(\nu_0 = 7)$	42	0.98%	0.02
Multi t-like$_{95\%}$	225	5.25%	0.54
Multi t-like$_{99\%}$	64	1.49%	9.13^*
Stable-like$_{95\%}$	223	5.20%	0.36
Stable-like$_{99\%}$	49	1.14%	0.84
Gaussian$_{95\%}$	207	4.83%	0.27
Gaussian$_{99\%}$	63	1.47%	8.33^*
Panel B: NLC			
	Violations	Percentage	LR
Warped $t_{95\%}(\nu_0 = 3)$	220	5.13%	0.15
Warped $t_{99\%}(\nu_0 = 3)$	43	1.00%	0.00
Warped $t_{95\%}(\nu_0 = 5)$	213	4.97%	0.01
Warped $t_{99\%}(\nu_0 = 5)$	46	1.07%	0.22
Warped $t_{95\%}(\nu_0 = 7)$	208	4.85%	0.20
Warped $t_{99\%}(\nu_0 = 7)$	44	1.03%	0.03
Multi t-like$_{95\%}$	204	4.76%	0.54
Multi t-like$_{99\%}$	64	1.49%	9.13^*
Stable-like$_{95\%}$	205	4.78%	0.44
Stable-like$_{99\%}$	66	1.54%	10.81^*
Gaussian$_{95\%}$	191	4.45%	2.79
Gaussian$_{99\%}$	61	1.42%	6.84^*
Panel C: NLDC			
	Violations	Percentage	LR
Warped $t_{95\%}(\nu_0 = 3)$	235	5.48%	2.02
Warped $t_{99\%}(\nu_0 = 3)$	42	0.98%	0.02
Warped $t_{95\%}(\nu_0 = 5)$	231	5.39%	1.32
Warped $t_{99\%}(\nu_0 = 5)$	44	1.03%	1.03
Warped $t_{95\%}(\nu_0 = 7)$	224	5.22%	0.45
Warped $t_{99\%}(\nu_0 = 7)$	44	1.03%	0.03
Multi t-like$_{95\%}$	201	4.68%	0.90
Multi t-like$_{99\%}$	59	1.35%	5.48^*
Stable-like$_{95\%}$	207	4.83%	0.27
Stable-like$_{99\%}$	47	1.10%	0.39
Gaussian$_{95\%}$	212	4.94%	0.28
Gaussian$_{99\%}$	65	1.52%	9.95^*

Table 4: Value-at-Risk backtesting: quadratic approximation

This table reports the results of a Value-at-Risk backtesting using the quadratic approximation. Panels A and B report the results for the short and long portfolios, respectively, while Panel C reports the results for the down-and-out and cash-or-nothing portfolio.

	Violations	Percentage	LR
	Panel A: NLS		
	Violations	Percentage	LR
Warped $t_{95\%}(\nu_0 = 3)$	239	5.57%	2.87
Warped $t_{99\%}(\nu_0 = 3)$	39	0.91%	0.37
Warped $t_{95\%}(\nu_0 = 5)$	232	5.41%	1.48
Warped $t_{99\%}(\nu_0 = 5)$	42	0.98%	0.02
Warped $t_{95\%}(\nu_0 = 7)$	227	5.30%	0.77
Warped $t_{99\%}(\nu_0 = 7)$	47	1.10%	0.39
Multi t-like$_{95\%}$	228	5.32%	0.89
Multi t-like$_{99\%}$	66	1.54%	10.81*
Stable-like$_{95\%}$	232	5.41%	1.48
Stable-like$_{99\%}$	49	1.14%	0.84
Gaussian$_{95\%}$	215	5.01%	0.00
Gaussian$_{99\%}$	63	1.47%	8.33*
	Panel B: NLC		
	Violations	Percentage	LR
Warped $t_{95\%}(\nu_0 = 3)$	219	5.11%	0.10
Warped $t_{99\%}(\nu_0 = 3)$	42	0.98%	0.02
Warped $t_{95\%}(\nu_0 = 5)$	210	4.90%	0.10
Warped $t_{99\%}(\nu_0 = 5)$	42	0.98%	0.02
Warped $t_{95\%}(\nu_0 = 7)$	202	4.71%	0.77
Warped $t_{99\%}(\nu_0 = 7)$	42	0.98%	0.02
Multi t-like$_{95\%}$	203	4.73%	0.65
Multi t-like$_{99\%}$	63	1.47%	8.33*
Stable-like$_{95\%}$	204	4.75%	0.54
Stable-like$_{99\%}$	65	1.52%	9.95*
Gaussian$_{95\%}$	188	4.38%	3.56
Gaussian$_{99\%}$	61	1.42%	6.84*
	Panel C: NLDC		
	Violations	Percentage	LR
Warped $t_{95\%}(\nu_0 = 3)$	231	5.39%	1.32
Warped $t_{99\%}(\nu_0 = 3)$	41	0.96%	0.08
Warped $t_{95\%}(\nu_0 = 5)$	223	5.20%	0.36
Warped $t_{99\%}(\nu_0 = 5)$	43	1.00%	0.00
Warped $t_{95\%}(\nu_0 = 7)$	215	5.01%	0.00
Warped $t_{99\%}(\nu_0 = 7)$	44	1.03%	0.03
Multi t-like$_{95\%}$	185	4.31%	4.44*
Multi t-like$_{99\%}$	58	1.35%	4.85*
Stable-like$_{95\%}$	204	4.76%	0.54
Stable-like$_{99\%}$	45	1.05%	0.10
Gaussian$_{95\%}$	208	4.85%	0.20
Gaussian$_{99\%}$	64	1.49%	9.13*