Fat Tail Analysis and Package FatTailsR

patrice.kiener@inmodelia.com

9th R/Rmetrics Workshop 27June2015

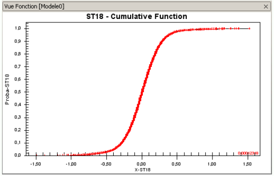
Villa Hatt – Zürich

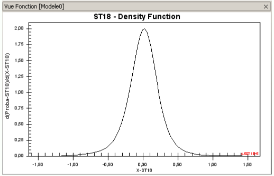
Outline

About InModelia

Since 2009: a small company located in Paris

- Consulting services in data analysis
  - Design of experiments
  - Multivariate nonlinear modeling
  - Neural networks
  - Times series
- Software development in R





June 2014:

- 8th R/Rmetrics workshop
- First version of R package **FatTailsR** introducing Kiener distributions with symmetric (3 parameters) and asymmetric (4 parameters) fat tails

June 2015:

- 9th R/Rmetrics workshop
- Second version of R package **FatTailsR** including 3 functions for parameter estimation
- R package **FatTailsRplot** including advanced plotting functions

Fat tails

It is now well established that financial markets exhibit **fat tails**. It occurs on all markets.

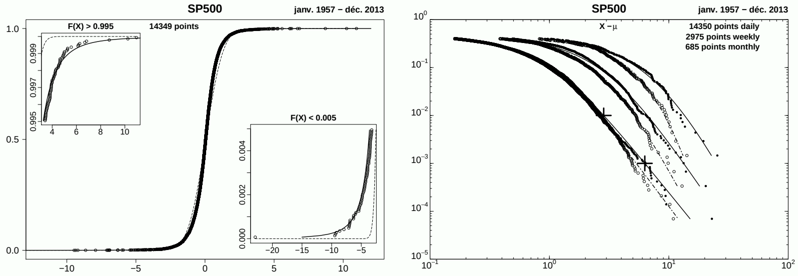


Figure : (a) SP500 daily log-returns (b) SP500 daily, weekly and monthly log-returns in log-log scale
black dots = negative returns, white circles = positive returns

$$\lim_{x \to -\infty} F(x) = |x|^{-\alpha} \qquad \lim_{x \to +\infty} 1 - F(x) = x^{-\omega} \qquad \alpha, \omega \approx 3.5$$

Symmetric distributions K1. Asymmetric distributions K2, K3, K4

Last year [1], we introduced a family of four probability distributions which deal with symmetric and asymmetric fat tails. These distributions were named Kiener distributions or, simply, distributions K1, K2, K3, K4. They combine two log-logistic quantile functions, one for each tail.

The parameters are:

- median μ, scale γ, left tail α, right tail ω, shape parameter $\kappa = \frac{2\alpha\omega}{\alpha+\omega}$ or harmonic mean $\frac{1}{\kappa} = \frac{1}{2}\left(\frac{1}{\alpha} + \frac{1}{\omega}\right)$
- distorsion δ and eccentricity ϵ defined by $\delta = \frac{\epsilon}{\kappa} = \frac{1}{2}\left(-\frac{1}{\alpha} + \frac{1}{\omega}\right)$ with bonds $-\frac{1}{\kappa} < \delta < \frac{1}{\kappa}$ and $-1 < \epsilon < 1$

Quantile functions of distributions K2(μ,γ,α,ω), K3(μ,γ,κ,δ), K4($\mu,\gamma,\kappa,\epsilon$), K1($\mu,\gamma,\kappa$)

$$\text{(K2)} \quad x(p) = \mu + \gamma\kappa\left[-\left(\frac{p}{1-p}\right)^{-\frac{1}{\alpha}} + \left(\frac{p}{1-p}\right)^{\frac{1}{\omega}}\right] = \mu + \gamma\kappa\left(-e^{-\frac{\text{logit}(p)}{\alpha}} + e^{\frac{\text{logit}(p)}{\omega}}\right)$$

$$\text{(K3)} \quad x(p) = \mu + \gamma\kappa\left[-\left(\frac{p}{1-p}\right)^{-\frac{1}{\kappa}} + \left(\frac{p}{1-p}\right)^{\frac{1}{\kappa}}\right]\left(\frac{p}{1-p}\right)^{\delta} = \mu + 2\gamma\kappa\sinh\left(\frac{\text{logit}(p)}{\kappa}\right)e^{\delta\,\text{logit}(p)}$$

$$\text{(K4)} \quad x(p) = \mu + \gamma\kappa\left[-\left(\frac{p}{1-p}\right)^{-\frac{1}{\kappa}} + \left(\frac{p}{1-p}\right)^{\frac{1}{\kappa}}\right]\left(\frac{p}{1-p}\right)^{\frac{\epsilon}{\kappa}} = \mu + 2\gamma\kappa\sinh\left(\frac{\text{logit}(p)}{\kappa}\right)e^{\frac{\epsilon}{\kappa}\,\text{logit}(p)}$$

$$\text{(K1)} \quad x(p) = \mu + \gamma\kappa\left[-\left(\frac{p}{1-p}\right)^{-\frac{1}{\kappa}} + \left(\frac{p}{1-p}\right)^{\frac{1}{\kappa}}\right] = \mu + 2\gamma\kappa\sinh\left(\frac{\text{logit}(p)}{\kappa}\right) = \mu + \gamma\,\text{kogit}(p,\kappa)$$

[1] P. Kiener, Explicit models for bilateral fat-tailed distributions and applications in finance with the package FatTailsR, 8th R/Rmetrics Workshop, Paris, 2014. http://www.inmodelia.com/fattailsr-en.html

Distribution K1 and some properties

Probability and density functions of distribution K1(μ,γ,κ), $t = \text{asinh}\left(\frac{x-\mu}{2\gamma\kappa}\right)$

$$F(x) = \frac{1}{1+e^{-\kappa\,\text{asinh}\left(\frac{x-\mu}{2\gamma\kappa}\right)}} \qquad f(x) = \frac{1}{4\gamma\cosh(t)\,(1+\cosh(\kappa t))} \qquad f(0) = \frac{1}{8\gamma}$$

Logit-probability:

$$\text{logit}(F(x)) = \kappa\,\text{asinh}\left(\frac{x-\mu}{2\gamma\kappa}\right)$$

Density as a function of the probability:

$$f(x) = \frac{1}{2\gamma}\,\text{sech}\left(\frac{\text{logit}(p)}{\kappa}\right) p(1-p)$$

Conversion between parameters α, ω, κ, δ, ϵ

$$\frac{1}{\kappa} = \frac{1}{2}\left(\frac{1}{\alpha}+\frac{1}{\omega}\right) \quad \delta = \frac{\epsilon}{\kappa} = \frac{1}{2}\left(-\frac{1}{\alpha}+\frac{1}{\omega}\right) \quad \frac{1}{\alpha} = \frac{1}{\kappa}-\delta \quad \frac{1}{\omega} = \frac{1}{\kappa}+\delta \quad \epsilon = \frac{\alpha-\omega}{\alpha+\omega} \quad \alpha = \frac{\kappa}{1-\epsilon} \quad \omega = \frac{\kappa}{1+\epsilon}$$

Property 1: Pareto aymptotic functions

$$\lim_{x\to-\infty} F(x) = \left|\frac{x-\mu}{\gamma\kappa}\right|^{-\alpha} \qquad \lim_{x\to+\infty} 1-F(x) = \left|\frac{x-\mu}{\gamma\kappa}\right|^{-\omega}$$

Property 2: Karamata theorem about slowly α-varying and ω-varying functions

$$\lim_{x\to-\infty} \frac{x f(x)}{F(x)} = -\alpha \qquad \lim_{x\to+\infty} \frac{x f(x)}{1-F(x)} = \omega$$

Comparaison with other distributions

Distributions K1, K2, K3, K4 are similar, but more tractable and easier to use than Generalized Tukey lambda distribution and Tadikamalla and Johnson L_U distribution[2].

K2 and K4 distributions (Kiener, 2014):

$$x(p) = \mu + \gamma\kappa \left[-\left(\frac{p}{1-p}\right)^{-1/\alpha} + \left(\frac{p}{1-p}\right)^{1/\omega} \right] = \mu + 2\gamma\kappa \sinh\left(\frac{\text{logit}(p)}{\kappa}\right) e^{\frac{\epsilon}{\kappa}\,\text{logit}(p)}$$

Generalized Tukey lambda distribution (Ramberg et Schmeiser, 1972):

$$x(p) = \lambda_1 + \frac{1}{\lambda_2}\left[-(1-p)^{\lambda_4} + p^{\lambda_3} \right]$$

L_U distribution (Tadikamalla and Johnson, 1982):

$$x(p) = \xi + \lambda \sinh\left(-\frac{\gamma}{\delta} + \frac{1}{\delta}\,\text{logit}(p) \right)$$

[2] Non-central Student distribution and Cauchy distribution were already discussed last year

Figures related to symmetric distribution K1

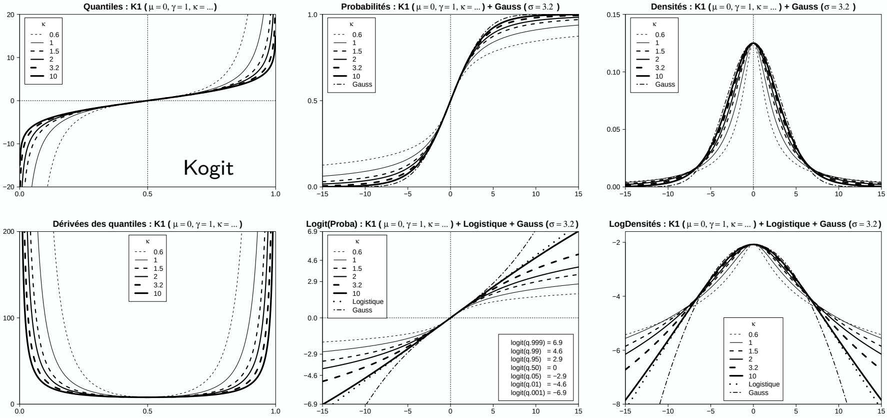


Figure : Distribution K1 : (a) Quantiles - (b) Cumulative functions (c) Densities (d) Quantile derivatives - (e) Logit of the cumulative functions - (f) Logdensities

Figures related to asymmetric distributions K2, K3, K4

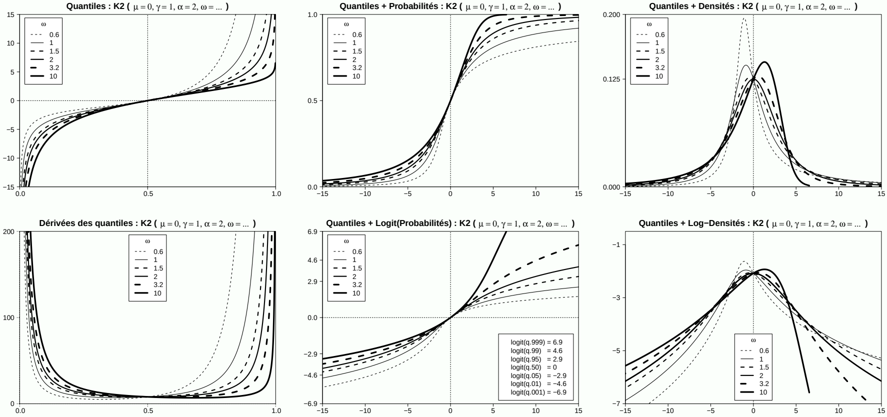


Figure : Distribution K2 : (a) Quantiles - (b) Cumulative functions (c) Densities (d) Quantile derivatives - (e) Logit of the cumulative functions - (f) Logdensities

Limit points

When $\alpha = \infty$ or $\omega = \infty$, the quantile function becomes a single log-logistic quantile function with left or right limit points

Property 4: Left limit point ($\alpha = \infty$), Right limit point ($\omega = \infty$)

$$\alpha = \infty,\ \kappa = 2\omega,\ \delta = \frac{1}{2\omega},\ \epsilon = 1 \quad \Rightarrow \quad x(p) = \mu - 2\gamma\omega + 2\gamma\omega\left(\frac{p}{1-p}\right)^{-1/\omega} \xrightarrow[p \to 0]{} \mu - 2\gamma\omega$$

$$\omega = \infty,\ \kappa = 2\alpha,\ \delta = -\frac{1}{2\alpha},\ \epsilon = -1 \quad \Rightarrow \quad x(p) = \mu + 2\gamma\alpha - 2\gamma\alpha\left(\frac{p}{1-p}\right)^{-1/\alpha} \xrightarrow[p \to 1]{} \mu + 2\gamma\alpha$$

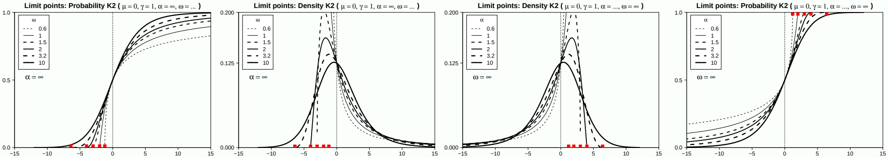


Figure : Cumulative functions K2: (a) $\alpha = \infty$ (b) $\omega = \infty$ Density functions K2: (c) $\alpha = \infty$ (d) $\omega = \infty$

Raw and central moments — K2, K1

Mean exists if $\min(\alpha,\omega) > 1$. Variance exists if $\min(\alpha,\omega) > 2$.
Skewness exists if $\min(\alpha,\omega) > 3$. Kurtosis exists if $\min(\alpha,\omega) > 4$.

Distribution K2$(\mu,\gamma,\alpha,\omega)$

$$m_1 = E(X) = \mu + \gamma\kappa\left(-\beta\left(1-\frac{1}{\alpha},1+\frac{1}{\alpha}\right)+\beta\left(1-\frac{1}{\omega},1+\frac{1}{\omega}\right)\right) = \mu+\gamma\kappa\nu$$

$$\mu_2 = E\left((X-E(X))^2\right) = \gamma^2\kappa^2\left[\beta\left(1-\frac{2}{\alpha},1+\frac{2}{\alpha}\right)+\beta\left(1-\frac{2}{\omega},1+\frac{2}{\omega}\right)-2\beta\left(1+\frac{1}{\alpha}-\frac{1}{\omega},1-\frac{1}{\alpha}+\frac{1}{\omega}\right)\right]$$
$$+\ \gamma^2\kappa^2\left[-\beta^2\left(1-\frac{1}{\alpha},1+\frac{1}{\alpha}\right)-\beta^2\left(1-\frac{1}{\omega},1+\frac{1}{\omega}\right)+2\beta\left(1-\frac{1}{\alpha},1+\frac{1}{\alpha}\right)\beta\left(1-\frac{1}{\omega},1+\frac{1}{\omega}\right)\right]$$

$$\beta_1 = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{\sum_{i=0}^{3}\sum_{j=0}^{i}\binom{3}{i}\binom{i}{j}(-\nu)^{3-i}(-1)^j\beta\left(1+\frac{j}{\alpha}-\frac{i-j}{\omega},1-\frac{j}{\alpha}+\frac{i-j}{\omega}\right)}{\left[\sum_{i=0}^{2}\sum_{j=0}^{i}\binom{2}{i}\binom{i}{j}(-\nu)^{2-i}(-1)^j\beta\left(1+\frac{j}{\alpha}-\frac{i-j}{\omega},1-\frac{j}{\alpha}+\frac{i-j}{\omega}\right)\right]^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{(\mu_2)^{2}} = \frac{\sum_{i=0}^{4}\sum_{j=0}^{i}\binom{4}{i}\binom{i}{j}(-\nu)^{4-i}(-1)^j\beta\left(1+\frac{j}{\alpha}-\frac{i-j}{\omega},1-\frac{j}{\alpha}+\frac{i-j}{\omega}\right)}{\left[\sum_{i=0}^{2}\sum_{j=0}^{i}\binom{2}{i}\binom{i}{j}(-\nu)^{2-i}(-1)^j\beta\left(1+\frac{j}{\alpha}-\frac{i-j}{\omega},1-\frac{j}{\alpha}+\frac{i-j}{\omega}\right)\right]^{2}}$$

Distribution K1(μ,γ,κ)

$$m_1 = \mu \qquad \mu_2 = 2\gamma^2\kappa^2\left(\beta\left(1-\frac{2}{\kappa},1+\frac{2}{\kappa}\right)-1\right) \qquad \beta_1 = 0 \qquad \beta_2 = \frac{\beta\left(1-\frac{4}{\kappa},1+\frac{4}{\kappa}\right)-4\beta\left(1-\frac{2}{\kappa},1+\frac{2}{\kappa}\right)+3}{2\left(\beta\left(1-\frac{2}{\kappa},1+\frac{2}{\kappa}\right)-1\right)^2}$$

Pearson chart (β_1^2, β_2)

We plot the skewness β_1 and kurtosis β_2 of distributions K1 and K2, K3, K4 in chart (β_1^2, β_2).
K1 is on the vertical β_2 axis. K2, K3 and K4 are located under the log-logistic curve.
Some remarkable points at the limits:

- $(\alpha = \infty, \omega = \infty) \Rightarrow (\kappa = \infty, \epsilon = 0) \Rightarrow (\beta_1 = 0, \beta_2 = 4.2)$
- $(\alpha = 4.1, \omega = 4.1) \Rightarrow (\kappa = 4.1, \epsilon = 0) \Rightarrow (\beta_1 = 0, \beta_2 = 64.8)$
- $(\alpha = \infty, \omega = 4.1) \Rightarrow (\kappa = 8.2, \epsilon = 1) \Rightarrow (\beta_1 = 3.97, \beta_1^2 = 15.8, \beta_2 = 340)$

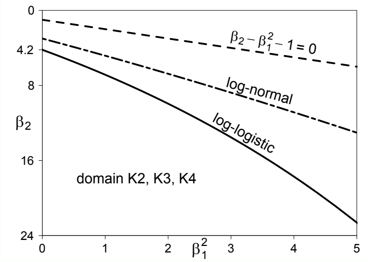


Figure : Skewness β_1 and kurtosis β_2 of distributions K1 and K2, K3, K4

The domain is identical to the domain associated to Tadikamalla and Johnson L_U distribution.

Experimental and theoretical variance

Large datasets are required to reach the theoretical variance or the standard deviation K1

$$\sigma = \sqrt{\frac{1}{N}\sum_i^N (x(p_i)-\mu)^2} \quad \text{(coloured lines) vs (black line)} \quad \sigma = \gamma\kappa\sqrt{2\beta\left(1-\frac{2}{\kappa}, 1+\frac{2}{\kappa}\right)-2}$$

$$x(p_i) = \mu + 2\gamma\kappa \sinh\left(\frac{1}{\kappa}\,\mathrm{logit}\left(\frac{i}{N+1}\right)\right) \quad \text{with } \mu=0,\ \gamma=1,\ N=(20, 50, 100, 250, 1000, 10.000)$$

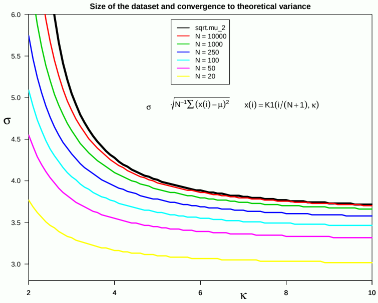


Figure : Theoretical variance and experimental variance related to the size of the dataset

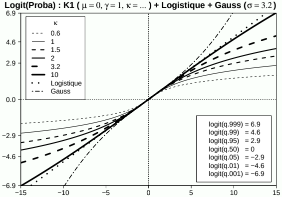
Explanation: There is a lot of information carried by the tails and not seen in small datasets!
In most cases, standard deviation should not be used with fat tails.

Parameter estimation

- Mathematics in **FatTailsR** package
- Plot and figures in **FatTailsRplot** package

2 algorithms and 3 functions are proposed in FatTailsR to estimate the parameters

1. **Nonlinear regression**
   - Functions **regkienerLX** (→*list*), **fitkienerLX** (→*data.frame*)
   - Work with K1, K2, K3, K4
   - Use the whole dataset of size N
   - Perform a nonlinear regression from $\text{logit}(F(x))$ to x with $F(x_i) = p_i = \frac{i}{N+1}$
   - Nonlinear least squares
   - Levenberg-Marquardt algorithm (package *minpack.lm*)
   - Examples: slide 16, videos of slide 17



2. **Quantile estimation**
   - Function **estimkienerX** (→*data.frame*)
   - Works with K2, K3, K4
   - Requires 5 quantiles only!
   - Very fast!
   - Examples: slides 19–20

Fat tail characterization

Pair (κ, δ) describes the tails. Let us decompose the quantile function $x(p)$ as:

- the median μ and scale γ $\Longrightarrow$ linear part
- the tail as a function of κ and δ $\Longrightarrow$ nonlinear part, asymmetric curvature

K3, C3 — Corrective tail function $C(p, \kappa, \delta)$, $g = \mathrm{logit}(p)$

$$C(p,\kappa,\delta) = \frac{\kappa}{g}\sinh\left(\frac{g}{\kappa}\right)\exp(g\,\delta) \quad \Longleftrightarrow \quad x(p) = \mu + 2\,\mathrm{logit}(p)\,\gamma\; C(p,\kappa,\delta)$$

$\Uparrow$ $\Uparrow$

$C(p, \kappa, \delta)$ acts as a multiplier of the logistic asymptote and incorporates the imbalance between negative and positive tails. For simplicity, let us consider:

- $c01 = C(0.01, \kappa, \delta)$ $\Longrightarrow$ Risk over long periods
- $c05 = C(0.05, \kappa, \delta)$ $\Longrightarrow$ Risk during trading

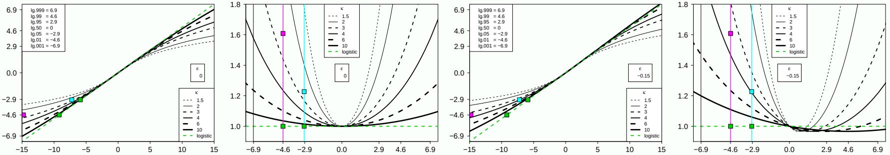


Figure : (a) Logit-Proba K4$(0, 1, \kappa, \epsilon = 0)$ (b) C4$(p, \kappa, \epsilon = 0)$ in logit scale
(c) Logit-Proba K4$(0, 1, \kappa, \epsilon = -0.15)$ (d) C4$(p, \kappa, \epsilon = -0.15)$ in logit scale

Example 1: SP500

SP500 from 1 January 1957 to 31 December 2013:

Days: 14349 points, Weeks: 2975 points, Months: 685 points

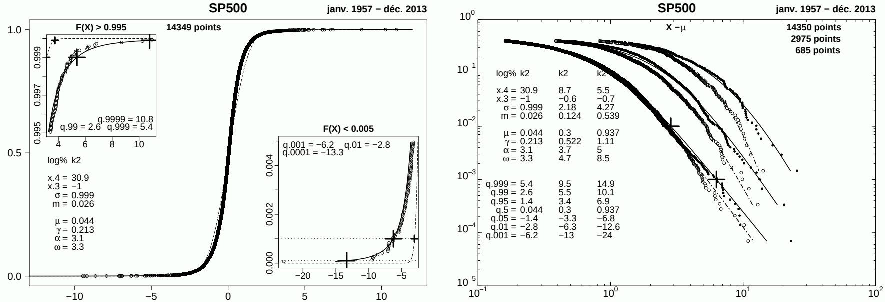


Figure : (a) SP500 daily log-returns (b) SP500 daily, weekly and monthly log-returns in log-log scale
black dots = negative returns, white circles = positive returns

Example 2: Videos SP500 — Rolling windows and Garch

SP500 from January 2005 to December 2013, rolling windows 252 points rebalanced every month

- **(left)** without Garch
- **(right)** K2 applied on Garch(1,1) residuals

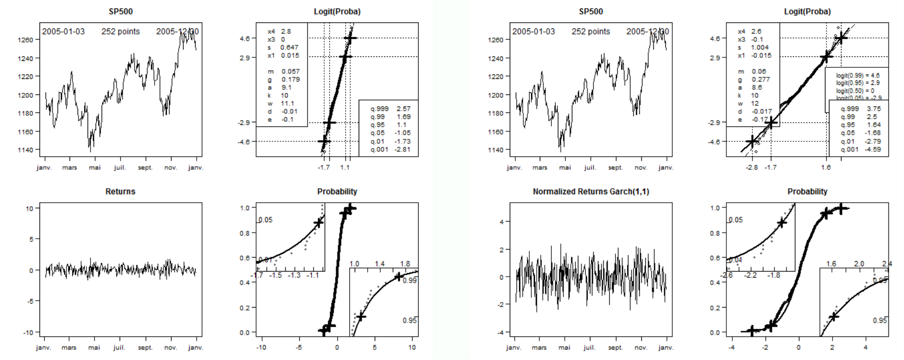


Figure : (a) SP500 log-returns + K2 (b) SP500 log-returns + Garch(1,1) + K2

Click on the links to watch the videos online: url: SP500 url: SP500-Garch url: VIX url: VIX-Garch

SP500 in 2011 and 2012 — Legend of next figures

Rolling windows 21, 41, 101 days rebalanced every day on **SP500** index

Plots:

- Left: Year 2011
- Right: Year 2012

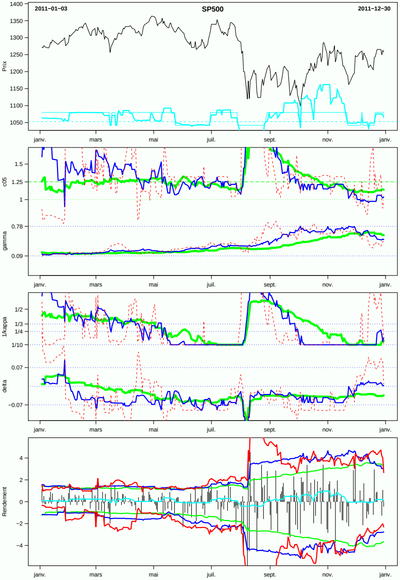
Plots:

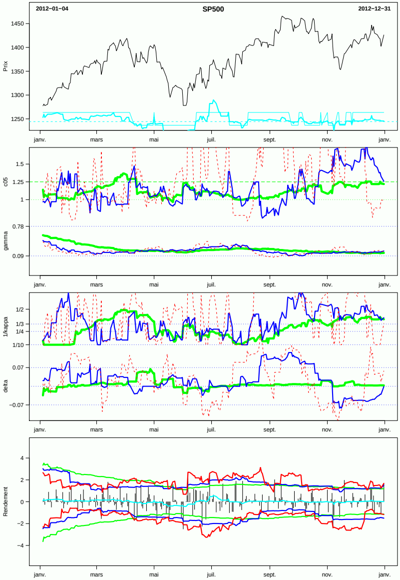
1. Top: Prices + rolling median (dots = 0)
2. Middle top: Rolling parameters c05 and γ
3. Middle bottom: Rolling parameters $1/\kappa$ and δ
4. Bottom: Returns + rolling quantiles at 5% and 95 % $\Rightarrow$ justifies c05 rather than c01

Colours:

- Black: Prices and returns on a daily basis
- Green: Rolling windows 101 days rebalanced every day
- Blue: Rolling windows 41 days rebalanced every day
- Red: Rolling windows 21 days rebalanced every day
- Cyan: Moving median 21 days rebalanced every day

Example 3: SP500 in 2011 and 2012 — Rolling 21, 41, 101 days





SP500 — Evolving parameters c05, q05. Rolling 41 days

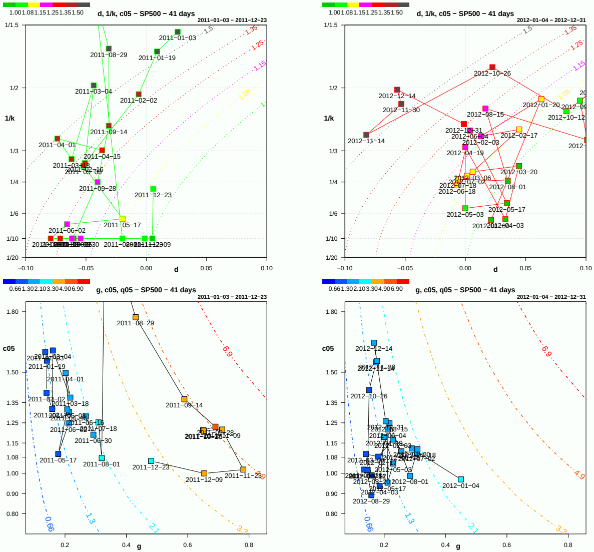


Figure : Parameters c05 (top) and q05 (bottom) from January 2011 (left) to December 2012 (right). SP500: Rolling periods of 41 days rebalanced every 10 days

Evolving parameter c01 — Rolling 88 days

$c01(\kappa,\delta) = \frac{\kappa}{4.6}\sinh\left(\frac{4.6}{\kappa}\right)\exp(-4.6\,\delta)$ changes over time. We can track it using plot $(\delta, 1/\kappa)$.

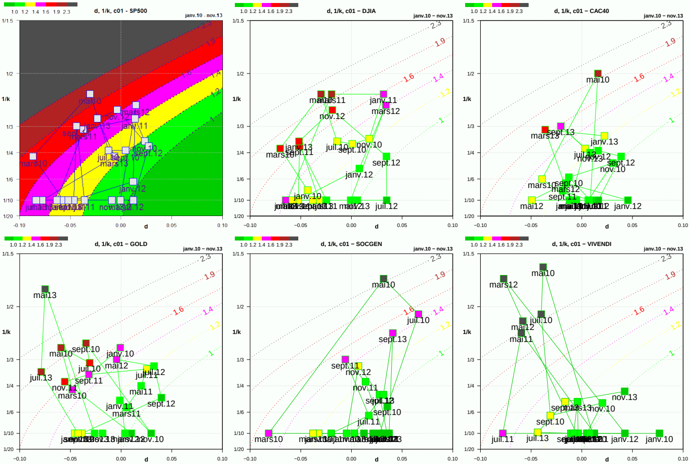


Figure : Parameter c01 from October 2009 (January 2010) to November 2013.
Rolling periods of 4 months (≈ 88 days) rebalanced every 2 months: SP500, DJIA, CAC40, Gold, Société Générale, Vivendi.

Evolving parameter c01 — Table

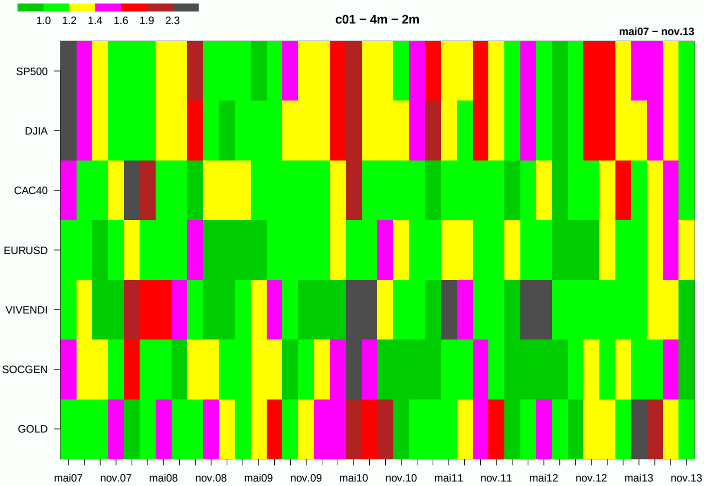


Figure : Parameter c01 from February 2007 (May 2007) to November 2013.
Rolling periods of 4 months (≈ 88 days) rebalanced every 2 months: SP500, DJIA, CAC40, Gold, Société Générale, Vivendi.

Conclusion (1)

A lot of knowledge on distributions with fat tails has been gained during the last 12 months

- K1, K2, K3, K4 distributions are good candidates for describing distributions with fat tails
  - They compare favourably to Tukey Generalized Lambda distribution, Tadikamalla and Johnson distribution, Student distribution, Cauchy distribution (see presentation last year for these last two).
  - They cover a vast domain under the log-logistic curve in Pearson $\left(\beta_1^2, \beta_2\right)$ plot.
  - Their moments fall exactly at the expected values of the tail parameters α, ω, κ.
  - Their parameters are well separated: median μ, scale γ, shape κ, distorsion δ or eccentricity ϵ.
  - Their parameters can be estimated by 2 different methods: nonlinear regression, quantile estimation.

- Effective tools to measure risk
  - Risk is split in scale risk (γ amplification) and asymmetric tail risk (c01, c05).
  - Approximation is good for datasets of various sizes: from 21 points to ten of thousands points.
  - SP500 has skewed and fat tails!

- Potential developments
  - Market risk (Bâle III, bcbs240, Jan-Fev.2013) $\Longrightarrow$ Include K3 distribution in the regulation?
  - Portfolio management $\Longrightarrow$ Extend the results to multidimensional distributions.
  - Derivatives $\Longrightarrow$ New process $\mu\gamma\kappa\delta$-Garck (Garck to differentiate from Gaussian Garch)?
  - Bayesian change points $\Longrightarrow$ Combine K1..K4 distributions with BCP?
  - Stability index $\Longrightarrow$ Combine K1..K4 distributions (κ, $c01$, $c05$) with SI?

- Two R packages: **FatTailsR** (license GPL-2, on CRAN) and **FatTailsRplot** (contact me)

Conclusion (2)

A remark rather than a conclusion:

During this presentation, I have talked a lot about risks but I never used the words « volatility » and « standard deviation » in the restrictive Gaussian meaning.

Distributions K1, K2, K3 and K4 introduce a new way of thinking of risk.
They deal with a 4 dimension space which is much larger than
the 2 dimension space of the Gaussian distribution.

With these new distributions, standard deviation is <u>not</u> the appropriate measure of risk.

Please, think about it.

References

Literature

1. N. L. Johnson, S. Kotz, N. Balakrishnan, Continuous univariate distributions, 2nd ed., Wiley, vol. 1, 1994 and vol. 2, 1995.
2. P. Kiener, Explicit models for bilateral fat-tailed distributions and applications in finance with the package FatTailsR, 8th R/Rmetrics workshop and summer school, Paris, 27 June 2014, http://www.inmodelia.com/fattailsr-en.html
3. C. Ley, Flexible modelling in statistics: past, present and future, J. Soc. Fran. Stat., Vol. 156, No. 1, 76-96, 2015. http://journal-sfds.fr/index.php/J-SFdS/article/download/415/394
4. The R project for statistical computing, http://www.r-project.org

Software

- Package **FatTailsR**, version 1.2-0, 18 June 2015, available on CRAN, license GPL-2, http://cran.r-project.org/web/packages/FatTailsR/index.html
- Package **FatTailsRplot**, version 1.2-0, 18 June 2015, http://www.inmodelia.com/fattailsr-en.html

To cite and download this document

- P. Kiener, Fat tail analysis and package FatTailsR, 9th R/Rmetrics workshop and summer school, Zurich, 27 June 2015, http://www.inmodelia.com/fattailsr-en.html

Thank you for your attention !

patrice.kiener@inmodelia.com

Tél. : +33.9.53.45.07.38