GHICA - Risk Analysis with GH Distributions and Independent Components

Ying Chen
Wolfgang Härdle
Vladimir Spokoiny

Institut für Statistik and Ökonometrie
CASE - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin

Weierstraß Institut
für Angewandte Analysis und Stochastik

http://ise.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://www.wias-berlin.de

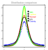
Measuring risk exposure

$$r(t) = b(t)^\top x(t) = b(t)^\top \Sigma_x^{1/2}(t)\varepsilon_x(t)$$

$r(t)$: portfolio returns
$b(t)$: trading strategie
$x_t \in \mathbb{R}^d$: individual returns with cov $\Sigma_x(t)$
$\varepsilon_x(t)$: stochastic term

$\mathrm{VaR}_{t,\mathrm{pr}} = -\mathrm{quantile}_{\mathrm{pr}}\{r(t)\}$
pr: $h = 1$-day or $h = 5$-day forecasted probability of $r(t)$.

Critical points: estimate $\Sigma_x(t)$
identify the distributional behavior of $\varepsilon_x(t)$



Popular risk management models

$$r(t) = b(t)^\top x(t) = b(t)^\top \Sigma_x^{1/2}(t)\varepsilon_x(t)$$

RiskMetrics

$\varepsilon_x(t) \sim \mathrm{N}(0, \mathrm{I}_d)$

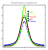
$\Sigma_x(t) = \varpi \Sigma_x(t-1) + (1-\varpi)x(t-1)x^\top(t-1)$

(**E**xponential **M**oving **A**verage)

t-deGARCH

$\varepsilon_x(t) \sim t(\mathrm{df})$

$\Sigma_x(t) = \varpi + \alpha_1 \Sigma_x(t-1) + \beta_1 x(t-1)x^\top(t-1)$ (**GARCH(1,1)**)



Limitations of the popular risk management models

- covariance estimation relies on a time-invariant form

$$
\Sigma_x(t) = \left\{ \sum_{m=0}^{\infty} \eta^m x(t-m-1) x^\top (t-m-1) \right\} / \left\{ \sum_{m=0}^{\infty} \eta^m \right\}
$$

$$
\eta \in [0, 1]
$$

$$
\begin{aligned}
\Sigma_x(t) &= \omega + \alpha x(t-1) x^\top (t-1) + \beta \Sigma_x(t-1) \\
&= \frac{\omega}{1-\beta} + \alpha \sum_{m=0}^{\infty} \beta^m x(t-m-1) x^\top (t-m-1)
\end{aligned}
$$

Limitations of the popular risk management models

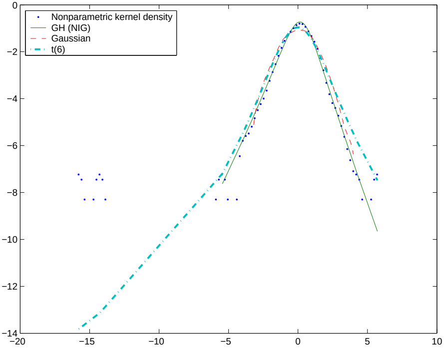
Example: Large loss in the US and European stock markets on **13 October 1989**.

time period	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$
1988/01/04-1989/10/13	8.63e-06 (6.36e-06)	0.07 (0.03)	0.87 (0.05)
1989/10/13-1991/08/07	6.54e-06 (2.95e-06)	0.17 (0.07)	0.61 (0.12)
1988/01/04-1991/08/07	1.61e-05 (6.93e-06)	0.12 (0.04)	0.83 (0.04)

Table 1: ML estimates of the GARCH(1,1) model on the base of the German stock Allianz. The standard deviation of the estimates are reported in parentheses.

- unrealistic distributional assumption

Example: Log-density of the DAX portfolio, $b(t) = \text{unit}(1/20)$.
Time interval: 1988/01/04 - 1996/12/30.
$\varepsilon_r(t) \sim \text{GH}(-0.5, 1.21, -0.21, 1.21, 0.24)$.
Data source: FEDC (http://sfb649.wiwi.hu-berlin.de)



Limitations of the popular risk management models

- numerical problems appear when applied to high-dimensional portfolios

  Example: Dynamic conditional correlation (DCC) model:

$$\Sigma_x(t) = D_x(t) R_x(t) D_x(t)^\top$$

$D_x(t)$: GARCH(1,1)
$R_x(t) = \tilde{R}_x(1-\theta_1-\theta_2)+\theta_1\{\varepsilon_x(t-1)\varepsilon_x^\top(t-1)\}+\theta_2 R_x(t-1)$
$\tilde{R}_x$: sample correlation
$\varepsilon_x \in \mathbb{R}^d$: standardized returns

GHICA

Generalized **H**yperbolic distribution + **I**ndependent **C**omponent **A**nalysis

$$\begin{aligned} r(t) &= b(t)^\top x(t) = b(t)^\top W^{-1} y(t) \\ &= b(t)^\top W^{-1} D_y^{1/2}(t) \varepsilon_y(t) \end{aligned}$$

$\varepsilon_{y_j}(t) \sim \mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu), \quad j = 1, \cdots, d$
W is a $d \times d$ nonsingular ICA matrix
$y(t) \in \mathbb{R}^d$ is (approximately) independent
$D_y(t) = \mathrm{diag}\big(\sigma^2_{y_1}(t), \cdots, \sigma^2_{y_d}(t)\big)$ is the covariance matrix of $y(t)$
$\sigma^2_{y_j}(t) = \left\{\sum_{m=0}^{\infty} \eta^m(\mathbf{t}) y^2(t-m-1)\right\} / \left\{\sum_{m=0}^{\infty} \eta^m(\mathbf{t})\right\}$

ICA example

$y_1(t)$: generalized hyperbolic variable GH$(1, 2, 0, 1, 0)$
$y_2(t)$: GH$(1, 1.7, 0, 0.5, 0)$
$y_3(t)$: GH$(1, 1.5, 0, 1, 0)$

$$A = W^{-1} = \begin{pmatrix} 1.31 & 0.14 & 0.18 \\ -0.42 & -1.26 & -1.25 \\ -0.03 & 0.41 & -0.49 \end{pmatrix} 10^{-2}$$

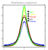
$$x(t) = A\, y(t)$$

Note: W is the estimated linear transformation matrix based on returns of three DAX components: ALLIANZ, BASF and BAYER from 1974/01/02 to 1996/12/30 (Data source: FEDC).

ICA example

The Mahalanobis transformation:

$$
\begin{aligned}
\widehat{\mathrm{cov}}_x^{-1/2} &= \begin{pmatrix} 0.91 & -0.09 & -0.12 \\ -0.09 & 1.03 & -0.41 \\ -0.12 & -0.41 & 1.04 \end{pmatrix} 10^2 \\
\neq \quad W &= \begin{pmatrix} 0.79 & 0.10 & 0.03 \\ -0.11 & -0.44 & 1.08 \\ -0.15 & -0.38 & -1.10 \end{pmatrix} 10^2
\end{aligned}
$$

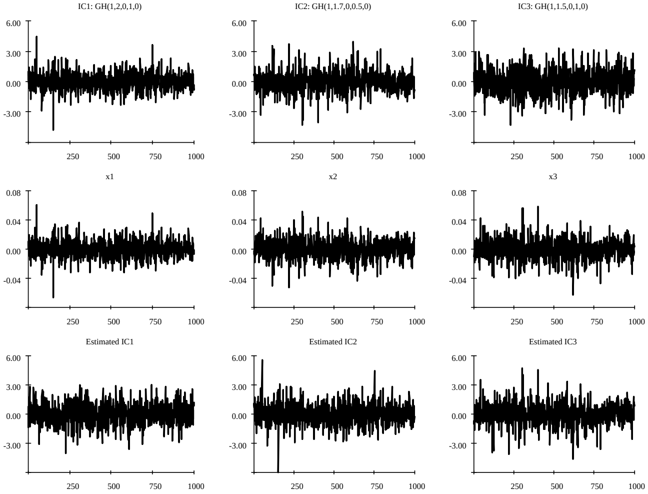


ICA example

Cross-cumulants:

Transformation	Mahalanobis	ICA
$E[y_1^2 y_3]$	0.04	-0.01
$E[y_2^2 y_3]$	0.14	0.00
$E[y_1^3 y_2]$	-0.17	0.00
$E[y_1 y_2^2 y_3]$	0.37	-0.03

Time plots of three ICs (top), $x(t)$ (middle) and $\hat{y}(t)$ (bottom).



GHICAsim.xpl

Procedure: GHICA

1. Implement ICA to get ICs.
2. Estimate variance of each IC by using the local exponential smoothing approach
3. Identify GH distributional parameters of the innovations of each IC
4. Estimate the density of portfolio returns using the FFT technique
5. Calculate risk measures

Outline

1. Motivation: ICA + GH = GHICA ✓
2. ICA: properties and estimation
3. Method: GH distribution, adaptive exponential smoothing and FFT
4. Simulation study
5. Empirical study
6. Conclusion

Definition

ICA model:

$$
\begin{aligned}
\begin{pmatrix} y_{1t} \\ \vdots \\ y_{dt} \end{pmatrix} &= \begin{pmatrix} w_{11} & \cdots & w_{1d} \\ \cdot & \cdots & \cdot \\ w_{d1} & \cdots & w_{dd} \end{pmatrix} \begin{pmatrix} x_{1t} \\ \vdots \\ x_{dt} \end{pmatrix} \\
y(t) &= Wx(t) = (w_1, \cdots, w_d)^\top x(t) \\
\text{equivalently } x(t) &= Ay(t)
\end{aligned}
$$

where $x(t)$ are d-dimensional observations, $y(t)$ are ICs and W the nonsingular linear transformation matrix: $W^{-1} = A$.

Properties of ICA

Scale identification: the scales of the ICs are not identifiable since both $y(t)$ and W are unknown:

$$x_{1t} = \sum_{j=1}^{d} a_{1j} y_{jt} = \sum_{j=1}^{d} \{\frac{1}{k_j} a_{1j}\}\{k_j y_{jt}\}$$

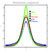
Hence: prewhiten $x(t)$ by the Mahalanobis transformation $\widehat{\text{cov}}(x)^{-1/2}$ and assume that each IC has unit variance: $\text{E}[y_j^2] = 1$.
From now on $x(t)$ is prewhitened!

Properties of ICA

Order identification: the order of the ICs is undetermined.

$$x(t) = Ay(t) = AP^{-1}Py(t)$$

where P is a permutation matrix and Py_t are the original ICs but in a different order.



Properties of ICA

ICs are necessarily non-Gaussian

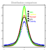
Consider two prewhitened Gaussian ICs y_1 and y_2 with pdf:

$$f(y_1, y_2) = |2\pi \mathrm{I}|^{-\frac{1}{2}} \exp(-\frac{y_1^2 + y_2^2}{2}) = \frac{1}{2\pi} \exp(-\frac{||y||^2}{2})$$

where $||y||$ is the norm of the vector $(y_1, y_2)^\top$.
The joint density of the observation x_1 and x_2 is given by:

$$f(x_1, x_2) = |2\pi \mathrm{I}|^{-\frac{1}{2}} \exp(-\frac{||Wx||^2}{2})|\det W| = \frac{1}{2\pi} \exp(-\frac{||x||^2}{2}).$$

Since A is an orthogonal matrix after prewhitening.



How to find ICs? - Minimize mutual information

$$
\begin{aligned}
I(W, y) &= \sum_{j=1}^{d} H(y_j) - H(y) \\
&= \sum_{j=1}^{d} H(y_j) - H(x) - \log |\det(W)| \\
\min \sum_{j=1}^{d} H(y_j) &\geq \sum_{j=1}^{d} \min H(y_j) \\
\hat{w}_j = \operatorname{argmin} H(y_j) &= \operatorname{argmax} J(w_j, y_j)
\end{aligned}
$$

where $H(\cdot)$ is the entropy and $J(\cdot)$ is the negentropy.

How to find ICs?

Jones and Sibson (1987): projection pursuit

- Cumulant based measure: e.g. skewness and excess kurtosis: sensitive to outliers.
- Negentropy: Gaussian variable has the maximal entropy given a fixed variance.

$$
\begin{aligned}
J(w, y) &= J(f_y) = H\{\mathrm{N}(0,1)\} - H(y) \\
\text{entropy: } H(y) &= H(f_y) = -\int f_y(u) \log f_y(u) du.
\end{aligned}
$$

Note that y is now a univariate and prewhitened variable. Negentropy requires the knowledge of f_y.

How to approximate univariate negentropy?

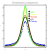
Given y univariate and prewhitened:

$$\text{argmax}\{J(f_y)\} = \text{argmin}\{H(f_y)\}.$$

Cover and Thomas (1991):
Fix sample expectations c_j with given functions $G_j(y)$

$$\text{E}[G_j(y)] = \int G_j(y) f(y) dy = c_j, \quad j = 1, \cdots, s.$$

Problem: $f(y)$ is not identifiable.



How to approximate univariate negentropy?

Given y univariate and prewhitened:

$$\operatorname{argmax}\{J(f_y)\} = \operatorname{argmin}\{H(f_y)\}.$$

Minimize the univariate entropy w.r.t. the density family:

$$f_0(y;a) = A\exp\{\sum_j a_j G_j(y)\} \quad (1)$$

How to approximate univariate negentropy?

Step 1: estimate pdf of $y(t)$ with the smallest entropy, i.e. search for non-Gaussian distributions:

$$\hat{f}(\cdot) = \text{argmax}_a[-H\{f_0(y;a)\}].$$

Include the following functions for standardization:

$$G_{s+1}(y) = y, c_{s+1} = 0 \quad G_{s+2}(y) = y^2, c_{s+2} = 1$$

make G_j an orthogonal system.

$$\hat{f}_y = \varphi(y)\{1 + \sum_{j=1}^{s} c_j G_j(y)\} \tag{2}$$

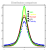
How to approximate univariate negentropy?

Step 2: approximate the negentropy:

$$H(y) \approx -\int \hat{f}_y(u) \log \hat{f}_y(u) du \approx H(y_{gauss}) - \frac{1}{2} \sum_{j=1}^{s} c_j^2 \quad (3)$$

$$J(y) = H(y_{gauss}) - H(y) \approx \frac{1}{2} \sum_{j=1}^{s} c_j^2 \quad (4)$$

Proof in Appendix.



How to approximate univariate negentropy?

Step 3: choose functions G_j:

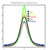
1. $\mathrm{E}[G_j(y)]$ should be easily computable and not sensitive to outliers
2. $G_j(y)$ should not grow faster than quadratically to ensure that $f_0(y)$ in (3) is integrable
3. $G_j(\cdot)$ should capture distributional features of $\log\{f_y(\cdot)\}$.

How to approximate univariate negentropy?

Two important features measure non-Gaussianity:

- Asymmetry - G_1 an odd function
- Tail behavior - G_2 an even function

$$
\begin{aligned}
J(y) &\approx \frac{1}{2}\sum_{j=1}^{s=2} c_j^2 \\
&\approx k_1\,\mathrm{E}\{G_1(y)\}^2 + k_2[\mathrm{E}\{G_2(y)\} - \mathrm{E}\{G_2(y_{gauss})\}]^2
\end{aligned}
$$



How to approximate univariate negentropy?

Example: Negentropy approximation

Approximation a: $k_1 = 36/(8\sqrt{3} - 9)$ and $k_2^a = 1/(2 - 6/\pi)$

$$\begin{aligned} J(y) &\approx k_1[\mathrm{E}\{y\exp(-y^2/2)\}]^2 + k_2^a[\mathrm{E}\{\exp(-y^2/2)\} - \sqrt{1/2}]^2 \\ G_1^a(y) &= y\exp(-y^2/2) \\ G_2^a(y) &= \exp(-y^2/2) \end{aligned}$$

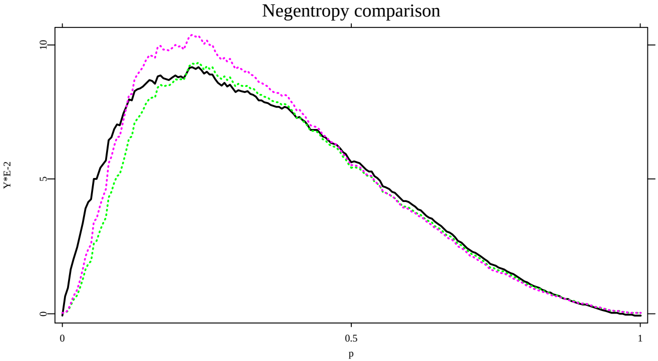
Approximation b: $k_1 = 36/(8\sqrt{3} - 9)$ and $k_2^b = 24/(16\sqrt{3} - 27/\pi)$

$$\begin{aligned} J(y) &\approx k_1[\mathrm{E}\{y\exp(-y^2/2)\}]^2 + k_2^b[\mathrm{E}\{|y|\} - \sqrt{2/\pi}]^2 \\ G_1^b(y) &= y\exp(-y^2/2) \\ G_2^b(y) &= |y| \end{aligned}$$

How to approximate univariate negentropy?

Comparison of the true negentropy (black) and its approximations (a: red, b: blue) of simulated Gaussian mixture variable: $pN(0,1) + (1-p)N(1,4)$ for $p \in [0,1]$.

GHICAnegentropyapp.xpl



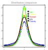
Negentropy approximations and FastICA

In the VaR context: tail behavior is more relevant than asymmetry. Therefore,

$$J(y) \approx C\{\mathsf{E}[G(y)] - \mathsf{E}[G\{\mathsf{N}(0,1)\}]\}^2.$$

$$\begin{aligned} G(y) &= \frac{1}{s}\log\cosh(sy), \quad 1 \le s \le 2 \\ g(y) \stackrel{\text{def}}{=} G'(y) &= \tanh(sy) \\ g'(y) &= s\{1 - \tanh^2(sy)\} \end{aligned}$$

very often, $s = 1$ is taken in this approximation.



FastICA

Objective function

$$\{\mathrm{E}\{G(WX)\} - \mathrm{E}[G\{\mathrm{N}(0,1)\}]\}\,\mathrm{E}\{Xg(WX)\} = 0 \tag{5}$$

A fast gradient method can be formulated under the constraint $W^\top W = \mathrm{I}_d$:

$$\mathrm{E}\{Xg(WX)\} + \chi W = 0 \tag{6}$$

The iteration of w_j with respect to y_j:

$$w_j^{(n+1)} = \mathrm{E}[Xg(w_j^{(n)}X) - \mathrm{E}\{g'(w_j^{(n)}X)\}w_j^{(n)}] \tag{7}$$

FastICA

Algorithm

1. Choose an initial vector w_j of unit norm, $W = (w_1, \cdots, w_d)^\top$.
2. Let $w_j^{(n)} = \mathrm{E}[g(w_j^{(n-1)}x)x] - \mathrm{E}[g'(w_j^{(n-1)}x)]w_j^{(n-1)}$. In practice, the sample mean is used to calculate $\mathrm{E}[\cdot]$.
3. Orthogonalization (decorrelated): $w_j^{(n)} = w_j^{(n)} - \sum_{k \neq j}(w_j^{(n)\top} w_k)w_k$.
4. Normalization: $w_j^{(n)} = w_j^{(n)}/||w_j^{(n)}||$.
5. If not converged, i.e. $||w_j^{(n)} - w_j^{(n-1)}|| \neq 0$, go back to 2.
6. Set $j = j + 1$. For $j \leq d$, go back to step 1.

Outline

1. Motivation: ICA + GH = GHICA ✓
2. ICA: properties and estimation ✓
3. Method: GH distribution, adaptive exponential smoothing and FFT
4. Simulation study
5. Empirical study
6. Conclusion

GH distribution

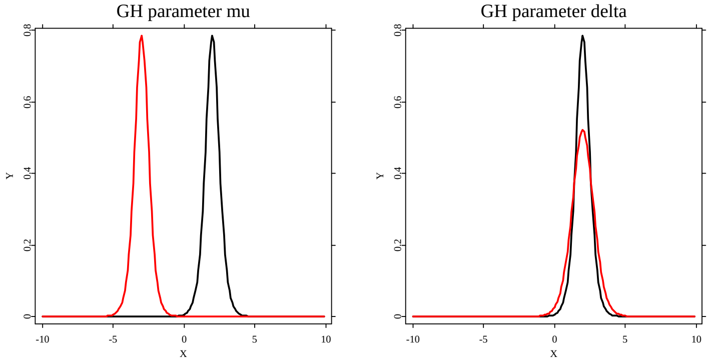
$X \sim GH$ with density:

$$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = \frac{(\iota/\delta)^{\lambda}}{\sqrt{2\pi} K_{\lambda}(\delta\iota)} \frac{K_{\lambda-1/2}\left\{\alpha\sqrt{\delta^2 + (x-\mu)^2}\right\}}{\left\{\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right\}^{1/2-\lambda}} \cdot \exp\{\beta(x-\mu)\}$$

Where $\iota^2 = \alpha^2 - \beta^2$, $K_{\lambda}(\cdot)$ is the modified Bessel function of the third kind with index λ: $K_{\lambda}(x) = \frac{1}{2}\int_0^{\infty} y^{\lambda-1} exp\{-\frac{x}{2}(y + y^{-1})\}\, dy$.

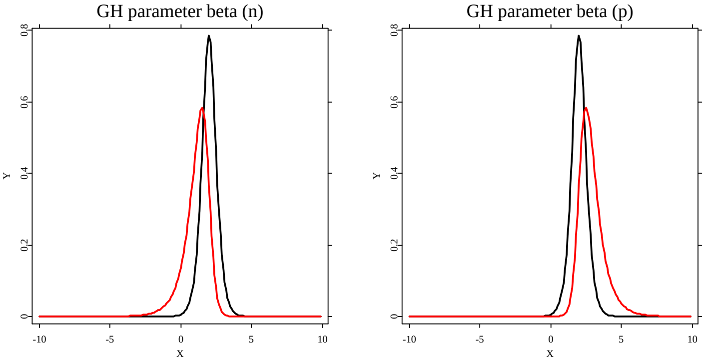
Parameters of GH distribution

Parameters μ *and* δ*: pdf of GH*$(-0.5, 3, 0, \mathbf{1}, \mathbf{2})$ *(black). On the left is the pdf of GH*$(-0.5, 3, 0, 1, \mathbf{-3})$ *and on the right is GH*$(-0.5, 3, 0, \mathbf{2}, 2)$.



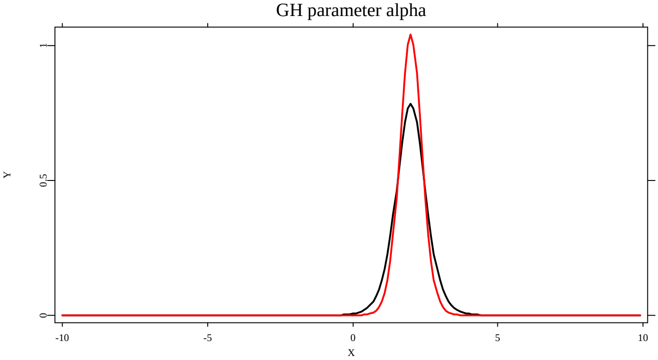
Parameters of GH distribution

Parameter β*: pdf of* $GH(-0.5, 3, \mathbf{0}, 1, 2)$ *(black). On the left is the pdf of* $GH(-0.5, 3, \mathbf{-2}, 1, 2)$ *and on the right is* $GH(-0.5, 3, \mathbf{2}, 1, 2)$*.*



Parameters of GH distribution

Parameter α: pdfs of $GH(-0.5, \mathbf{3}, 0, 1, 2)$ *(black) and* $GH(-0.5, \mathbf{6}, 0, 1, 2)$ *(red).*



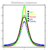
Subclass of GH distribution

The parameters $(\mu, \delta, \beta, \alpha)^\top$ *can be interpreted as trend, riskiness, asymmetry and the likeliness of extreme events.*

Normal-inverse Gaussian (NIG) distributions: $\lambda = -1/2$**,**

$$f_{NIG}(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta}{\pi} \frac{K_1\left\{\alpha\sqrt{\delta^2 + (x-\mu)^2}\right\}}{\sqrt{\delta^2 + (x-\mu)^2}} \exp\{\delta\iota + \beta(x-\mu)\},$$

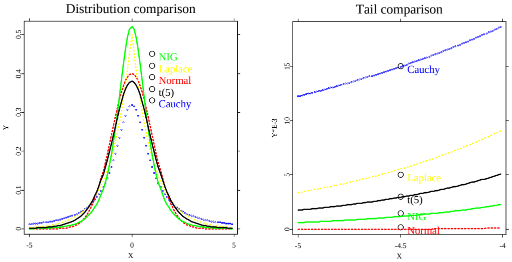
where x, $\mu \in \mathbb{R}$, $0 < \delta$ and $|\beta| \leq \alpha$.



Tail behavior of GH distribution

$$f_{GH}(x;\lambda,\alpha,\beta,\delta,\mu=0)\sim x^{\lambda-1}e^{(\mp\alpha+\beta)x}\text{ as }x\to\pm\infty,$$

Tail behaviors of five normalized distributions: NIG, standard normal, Laplace and Cauchy distributions.



Adaptive exponential smoothing

Chen and Spokoiny (2006)

$$y(t) = \sigma(t)\varepsilon(t)$$

$\varepsilon(t) \sim \text{NIG}$
$\hat{\sigma}(t)$: the "best" local estimate from $\left\{\tilde{\sigma}^{(k)}(t)\right\}$ for $k = 1, \cdots, K$

$$\tilde{\sigma}^{(k)}(t) = \left[\{\sum_{m=0}^{M_k} \eta_k^m y^2(t-m-1)\} / \{\sum_{m=0}^{M_k} \eta_k^m\} \right]^{1/2}$$

$$\text{s.t. } \eta_k^{M_k+1} \leq c \to 0$$

Adaptive exponential smoothing

$\varepsilon(t) \sim$ NIG: quasi ML estimation
Power transformation with $0 \leq p < 0.5$ guarantees $\mathrm{E}[\exp\{\rho\varepsilon^2(t)\}]$ exists:

$$
\begin{aligned}
y_p(t) &= \operatorname{sign}\{y(t)\}|y(t)|^p \\
\theta(t) &= \operatorname{var}\{y_p(t)|\mathcal{F}_{t-1}\} = \mathrm{E}\{|y(t)|^{2p}|\mathcal{F}_{t-1}\} \\
&= \sigma^{2p}(t)\,\mathrm{E}\,|\varepsilon(t)|^{2p} = \sigma^{2p}(t)C_p \\
\tilde{\theta}^{(k)}(t) &= \{\sum_{m=0}^{M_k} \eta_k^m |y(t-m-1)|^{2p}\}/\{\sum_{m=0}^{M_k} \eta_k^m\}
\end{aligned}
$$

Adaptive exponential smoothing

Localization:

- decreasing variation: $\frac{N_{k+1}}{N_k} \approx \frac{1-\eta_k}{1-\eta_{k+1}} = a > 1$
  where $N_k = \sum_{m=0}^{M_k} \eta_k^m$
- the first local estimate ($k = 1$) is automatically accepted as $\hat{\theta}^{(k)}(t)$. The consequent local estimate would be accepted if the fitted Gaussian log-likelihood ratio L is bounded by the critical value $\mathfrak{z}_k$:

$$L\left(\eta_k, \tilde{\theta}^{(k)}(t), \hat{\theta}^{(k-1)}(t)\right) = L\big(\eta_k, \tilde{\theta}^{(k)}(t)\big) - L\big(\eta_k, \hat{\theta}^{(k-1)}(t)\big)$$

Algorithm

1. Initialization: $\hat{\theta}^{(1)}(t) = \tilde{\theta}^{(1)}(t)$.
2. Loop: for $k \geq 2$
   $\hat{\theta}^{(k)}(t) = \tilde{\theta}^{(k)}(t)$, if $L\left(\eta_k, \tilde{\theta}^{(k)}(t), \hat{\theta}^{(k-1)}(t)\right) \leq \mathfrak{z}_k$
   $\hat{\theta}^{(k)}(t) = \hat{\theta}^{(s)}(t) = \tilde{\theta}^{(k-1)}(t)$ for $k \leq s \leq K$, otherwise
3. Final estimate: if $k = K$, $\hat{\theta}(t) = \hat{\theta}^{(K)}(t)$.
4. Save the selected local parameter $\hat{\eta}(t)$. Since C_p is only a constant, the volatility estimate is:
   $\hat{\sigma}^{(k)}(t) = \left[\{\sum_{m=0}^{\hat{M}_k} \hat{\eta}^m(t) y^2(t-m-1)\} / \{\sum_{m=0}^{\hat{M}_k} \hat{\eta}^m(t)\}\right]^{1/2}$

Parameter choice

- Initial values: $\eta_1 = 0.60$, $c = 0.01$, $a = 1.25$ and $p = 0.25$
- Critical values: Monte Carlo simulation.
  - apply the general critical values under the normal distributional assumption since the transformed variable is close to Gaussian
  - estimate $\hat{C}_p$ based on the estimates $\hat{\theta}(t)$ such that $\mathrm{var}\{\hat{\varepsilon}(t)\} = \mathrm{var}\left[y(t)\{\hat{C}_p/\hat{\theta}(t)\}^{\frac{1}{2p}}\right] = 1.$
  - estimate the NIG distributional parameters of $\hat{\varepsilon}(t) = y(t)/\hat{\sigma}(t)$ where $\hat{\sigma}(t) = \{\hat{\theta}(t)/\hat{C}_p\}^{\frac{1}{2p}}$
  - calculate the critical values based on the identified NIG variables.

Characteristic function of portfolio returns

The characteristic function of the NIG variable is:

$$\varphi_y(z) = \exp\left[\mathbf{i}z\mu + \delta\{\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+\mathbf{i}z)^2}\}\right]$$

The scaling transformation of NIG r.v. $y' = cy$:

$$f_{\mathrm{NIG}}(y';\alpha',\beta',\delta',\mu') = f_{\mathrm{NIG}}(cy;\alpha/|c|,\beta/c,|c|\delta,c\mu)$$

Given $r(t) = b(t)^\top W^{-1} D_y(t)^{1/2} \varepsilon_y(t) = a(t)\varepsilon_y(t)$,
$a_j(t)\varepsilon_j(t) \sim \mathrm{NIG}(\breve{\alpha}_j, \breve{\beta}_j, \breve{\delta}_j, \breve{\mu}_j)$ with $j = 1, \cdots, d$:

$$\mathrm{NIG}(\breve{\alpha}_j, \breve{\beta}_j, \breve{\delta}_j, \breve{\mu}_j) = \mathrm{NIG}(\alpha_j/|a_j(t)|, \beta_j/a_j(t), |a_j(t)|\delta_j, a_j(t)\mu_j)$$

Density estimation by using FFT

The characteristic function of the portfolio return at time t is:

$$\varphi_r(z) = \prod_{j=1}^{d} \varphi_{\zeta_j}(z) = \exp\left(\mathbf{i}z \sum_{j=1}^{d} \breve{\mu}_j\right)$$
$$\cdot \exp\left[\sum_{j=1}^{d} \breve{\delta}_j \{\sqrt{\breve{\alpha}_j^2 - \breve{\beta}_j^2} - \sqrt{\breve{\alpha}_j^2 - (\breve{\beta}_j + \mathbf{i}z)^2}\}\right]$$

The density function is approximated by using the FFT:

$$f(r) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-\mathbf{i}tr)\psi(z)dt \approx \frac{1}{2\pi}\int_{-s}^{s} \exp(-\mathbf{i}tr)\psi(z)dt$$

Procedure: GHICA

1. Implement ICA to get ICs.
2. Estimate variance of each IC by using the local exponential smoothing approach
3. Identify GH distributional parameters of the innovations of each IC
4. Estimate the density of portfolio returns using the FFT technique
5. Calculate risk measures

Simulation study on covariance estimation

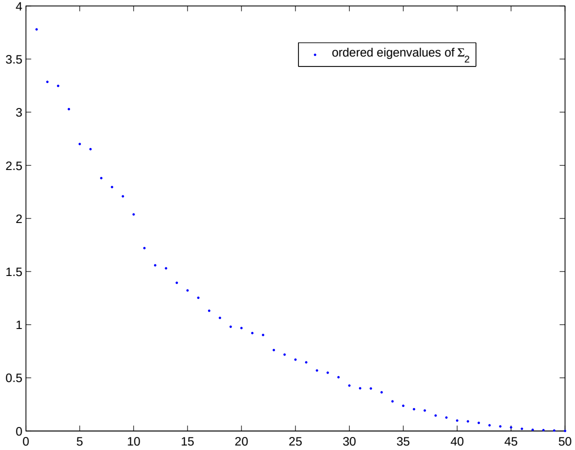
Goal GHICA versus DCC:

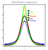
$$
\begin{aligned}
\Sigma_x(t) &= W^{-1} D_y(t) W^{-1\top} \\
\Sigma_x(t) &= D_x(t) R_x(t) D_x(t)^\top
\end{aligned}
$$

Design

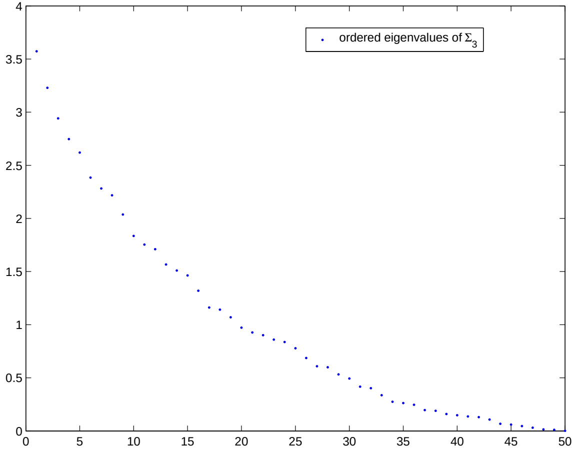
- $d = 50$ centered and symmetric $\mathrm{NIG}(\alpha_j, 0, \delta_j, 0)$ where $\alpha_j \sim U[1,2]$ and $\alpha_j = \delta_j$ to guarantee standardization
- sample size $T = 1900$, $N = 100$ simulations
- covariance shifts $\Sigma_1 = I_d$, Σ_2 and Σ_3 are self-correlated

Ordered eigenvalues of the generated covariance Σ_2.

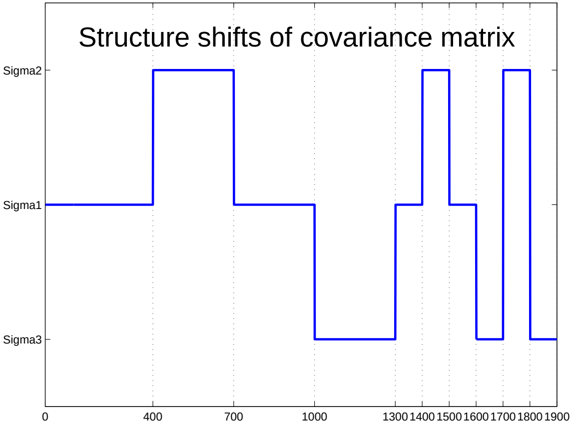


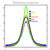


Ordered eigenvalues of the generated covariance Σ_3.

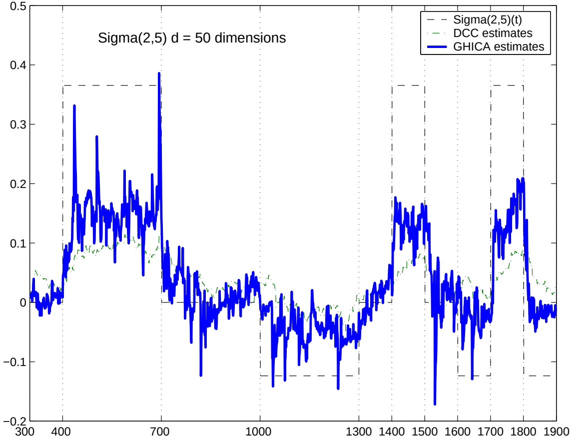


Structure shifts of the generated covariance through time.

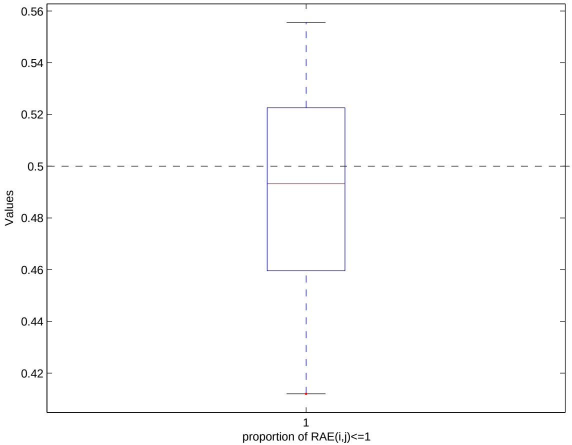




Realized estimates of $\Sigma(2,5)$ based on the GHICA and DCC methods.



Boxplot of the proportion $\frac{\sum_i \sum_j \mathbf{1}(\text{RAE}(i,j) \leq 1)}{d \times d}$ for $i, j = 1, \cdots, d$ over 100 simulations, where $\text{RAE}(i,j) = \frac{\sum_{t=301}^{T} |\hat{\Sigma}_{(i,j)}^{\text{GHICA}}(t) - \Sigma_{(i,j)}(t)|}{\sum_{t=301}^{T} |\hat{\Sigma}_{(i,j)}^{\text{DCC}}(t) - \Sigma_{(i,j)}(t)|}$



Risk measures and requirements

- Regulatory: to ensure the adequacy of capital and restrict the happening of large losses of financial institutions.

  $\mathrm{VaR}_{t,\mathrm{pr}} = -\mathrm{quantile}_{\mathrm{pr}}\{r(t)\}$,
  where pr is the $h = 1$-day forecasted probability of the portfolio returns

  $\text{Risk charge}_t = \max\left(M_f \frac{1}{60}\sum_{i=1}^{60} \mathrm{VaR}_{t-i,1\%}, \mathrm{VaR}_{t,1\%}\right)$,
  where M_f relies on the number of exceptions $(-r(t) > \mathrm{VaR}_{t,\mathrm{pr}})$ over last 250 days and identifies according to the "traffic light" rule.

No. exceptions	Increase of M_f	Zone
0 bis 4	0.00	**green**
5	0.40	**yellow**
6	0.50	**yellow**
7	0.65	**yellow**
8	0.75	**yellow**
9	0.85	**yellow**
More than 9	1.00	**red**

Table 2: Traffic light as a factor of the exceeding amount, cited from Frank, Härdle and Hafner (2004).

Risk measures and requirements

- Minimum requirement of regulatory:
  $\hat{\mathrm{pr}} \leq \frac{4}{250}$ (green zone)
  small amount of risk charge:
  Risk charge (RC) $=$ mean $(\mathrm{VaR}_{t,\mathrm{pr}})$

Risk measures and requirements

- Investors: suffer loss (at least the amount of the expected shortfall) once bankruptcy happens

  Expected shortfall (ES) measures the expected size of loss:
  $\mathrm{ES} = \mathsf{E}\{-r(t) | -r(t) > \mathsf{VaR}_{t,\mathrm{pr}}\}$
  ES as small as possible

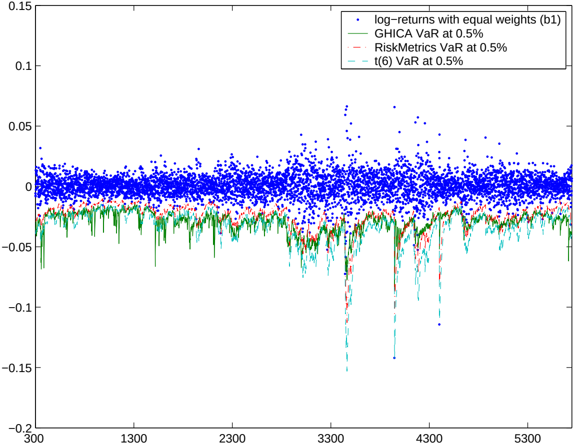
- Internal supervisory: exactly measure the market risk exposures

  $\hat{\mathrm{pr}} = \frac{\text{No. exceptions}}{\text{No. total observations}}$
  $\hat{\mathrm{pr}}$ close to pr

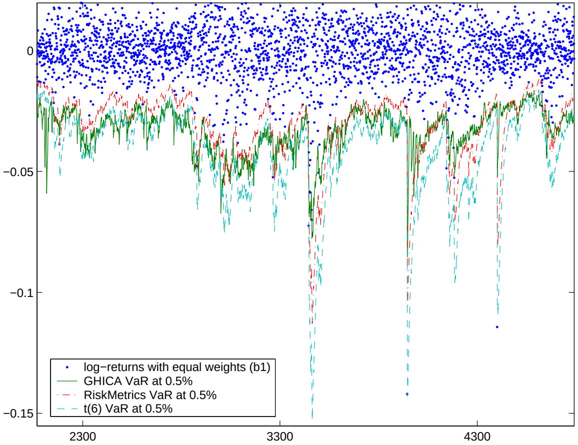
DAX portfolio

- Data: 20 DAX stocks 1974/01/02 - 1996/12/30 (5748 observations). All are heavy-tailed distributed (kurtosis>3). The smallest correlation coefficient is 0.3654
- Static trading strategies: $b(t) = b^{(1)} = (1/d, \cdots, 1/d)^\top$ and $b(t) = b^{(2)} \sim U[0,1]$
- Goal: GHICA versus RiskMetrics and ES $t(6)$ (exponential smoothing with $t(6)$ distributional assumption)

One day log-returns of the DAX portfolio with the static trading strategy $b(t) = b^{(1)}$. The VaRs are from 1975/03/17 to 1996/12/30 at $\mathrm{pr} = 0.5\%$ w.r.t. three methods.



Enlarged part



Risk analysis of the DAX portfolios with two static trading strategies. The concerned forecasting interval is $h = 1$ or $h = 5$ days. The best results to fulfill the regulatory requirement are marked by ^{r}. The method preferred by investor is marked by ^{i}. For the internal supervisory, the method marked by ^{s} is recommended.

			GHICA			RiskMetrics N(μ, σ^2)			Exponential smoothing $t(6)$		
h	$b(t)$	pr	$\hat{\text{pr}}$	RC	ES	$\hat{\text{pr}}$	RC	ES	$\hat{\text{pr}}$	RC	ES
1	$b^{(1)}$	1%	0.55%	0.0264	0.0456	1.18%^{s}	**0.0229**^{r}	0.0279	0.40%	0.0292	**0.0269**^{i}
	$b^{(1)}$	0.5%	0.44%^{s}	0.0297	**0.0472**^{i}	0.75%	0.0254	0.0317	0.23%	0.0345	0.0506
	$b^{(2)}$	1%	0.59%	0.0265	0.0448	1.03%^{s}	**0.0231**^{r}	0.0288	0.38%	0.0294	**0.0406**^{i}
	$b^{(2)}$	0.5%	0.42%^{s}	0.0298	**0.0476**^{i}	0.71%	0.0256	0.0315	0.21%	0.0347	0.0514
5	$b^{(1)}$	1%	0.83%	0.0550	0.0841	1.15%^{s}	**0.0481**^{r}	0.0602	0.19%	0.0665	**0.0833**^{i}
	$b^{(1)}$	0.5%	0.51%^{s}	0.0612	**0.0939**^{i}	0.64%	0.0536	0.0683	0.09%	0.0784	0.1067
	$b^{(2)}$	1%	0.83%^{s}	0.0554	**0.0828**^{i}	1.18%	**0.0488**^{r}	0.0613	0.16%	0.0673	0.0852
	$b^{(2)}$	0.5%	0.50%^{s}	0.0617	**0.0943**^{i}	0.63%	0.0543	0.0676	0.07%	0.0794	0.1218

Foreign exchange rate portfolio

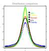
- Data: 7 FX rate 1997/01/02 to 2006/01/05 (2332 observations).
- Dynamic trading strategies: $b^{(3)}(t) = \frac{x(t-1)}{\sum_{j=1}^{d} x_j(t-1)}$, where $x(t) = \{x_1(t), \cdots, x_d(t)\}^\top$. EUR/USD and EUR/SGD rates are most correlated with the coefficient 0.6745
- Goal: GHICA versus DCCN (DCC with the Gaussian distributional assumption)

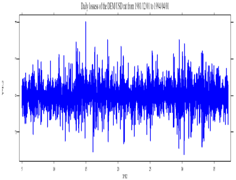
Risk analysis of the dynamic exchange rate portfolio. The best results to fulfill the regulatory requirement are marked by ^{r}. The recommended method to the investor is marked by ^{i}. For the internal supervisory, we recommend the method marked by ^{s}.

			GHICA			DCCN		
h	$b(t)$	pr	$\hat{\text{pr}}$	RC	ES	$\hat{\text{pr}}$	RC	ES
1	$b^{(3)}(t)$	1%	$1.28\%^s$	$\mathbf{0.0453}^r$	0.0778	1.59%	0.0494	$\mathbf{0.0254}^i$
	$b^{(3)}(t)$	0.5%	$0.59\%^s$	0.0493	$\mathbf{0.1944}^i$	0.94%	0.0547	0.0289
5	$b^{(3)}(t)$	1%	$1.53\%^s$	$\mathbf{0.0806}^r$	$\mathbf{0.2630}^i$	4.17%	0.0993	0.1735
	$b^{(3)}(t)$	0.5%	$0.79\%^s$	0.1092	$\mathbf{0.2801}^i$	3.44%	0.1100	0.1389

Conclusion and Outlook

- GHICA ✓
- Advanced ICA 1:
  Gaussian ICs ($\in \mathbb{R}^G$) + non-Gaussian ICs ($\in \mathbb{R}^{NG}$) with $G >> NG$
- Advanced ICA 2:
  Localization of ICA: $y(t) = W(\mathbf{t})x(t)$





Econometrics

Information Science

GHICA

Finance

Mathematics

Derivation in Negentropy Approximation

$$
\begin{aligned}
& \max\{-H(f_y)\} && \leftarrow \text{theory} \\
\text{s.t.} \quad & \int G_j(y) f_y \, dy = c_j && \leftarrow \text{data} \\
& \int \varphi(y) G_i(y) G_j(y) \, dy = 1 \text{ if } i = j && \leftarrow \text{orthogonality} \\
& = 0 \text{ otherwise} && \\
& \int \varphi(y) G_j(y) y^k \, dy = 0, \quad k = 0, 1, 2 &&
\end{aligned}
$$

Equation (4): $\hat{f}_y = \varphi(y)\{1 + \sum_{j=1}^{s} c_j G_j(y)\}$

Equation (5): $H(y) \approx H(y_{gauss}) - \frac{1}{2} \sum_{j=1}^{s} c_j^2$

$$
\begin{aligned}
f_0(y;a) &= A\exp\{\sum_{j=1}^{s+2} a_j G_j(y)\} \\
&= A\exp\{-\frac{y^2}{2} + a_{s+1}y + (a_{s+2} + \frac{1}{2})y^2 + \sum_{j=1}^{s} a_j G_j(y)\} \\
&= A\exp(-\frac{y^2}{2})\exp\{a_{s+1}y + (a_{s+2} + \frac{1}{2})y^2 + \sum_{j=1}^{s} a_j G_j(y)\} \\
&= \tilde{A}\varphi(y)\{1 + a_{s+1}y + (a_{s+2} + \frac{1}{2})y^2 + \sum_{j=1}^{s} a_j G_j(y)\}
\end{aligned}
$$

with $\tilde{A} = \sqrt{2\pi}A$ and $\varphi(y) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2})$.

Functions G_j are orthogonal:

$$
\begin{aligned}
\int f_0(y;a)dy &= \int \tilde{A}\varphi(y)\{1 + a_{s+1}y + (a_{s+2} + \frac{1}{2})y^2 + \sum_{j=1}^{s} a_j G_j(y)\}dy \\
&= \tilde{A}\{1 + (a_{s+2} + \frac{1}{2})\} = 1
\end{aligned}
$$

$$
\begin{aligned}
\int y f_0(y;a)dy &= \tilde{A}a_{s+1} = 0 \\
\int y^2 f_0(y;a)dy &= \tilde{A}\{1 + 3(a_{s+2} + \frac{1}{2})\} = 1 \\
\int G_j(y) f_0(y;a)dy &= \tilde{A}a_j = c_j, \quad j = 1, \cdots, s.
\end{aligned}
$$

Solution: $\tilde{A} = 1$, $a_{s+1} = 0$, $a_{s+2} = -\frac{1}{2}$ and $a_j = c_j$, $\Rightarrow$ (4).

Set $B = \sum_{j=1}^{s} c_j G_j(y)$, then $\hat{f}_y = \varphi(y)(1+B)$

$$
\begin{aligned}
H(y) &\approx -\int \hat{f}_y \log \hat{f}_y dy \\
&\approx -\int \varphi(y)(1+B)[\log\{\varphi(y)\} + \log(1+B)]dy \\
&= -\int \varphi(y)(1+B)\log\{\varphi(y)\}dy \\
&\quad -\int \varphi(y)(1+B)\log(1+B)dy \\
&\approx -\int \varphi(y)\log\{\varphi(y)\}dy - \int B\varphi(y)\log\{\varphi(y)\}dy \\
&\quad -\int \varphi(y)[B + \frac{1}{2}B^2 + \mathcal{O}(B^2)] \quad \text{(Taylor expansion)} \\
&= H(y_{gauss}) - \frac{1}{2}\sum c_j^2 + \mathcal{O}(\sum c_j^2) \;\Rightarrow (5)
\end{aligned}
$$

Properties of FastICA

Consistency: Assume that the data follows the ICA model and G is a sufficiently smooth even function. Then the set of local maxima of $J(y)$ of corresponding IC y_j fulfills:

$$\mathrm{E}\{y_j g(y_j) - g'(y_j)\}[\mathrm{E}\{G(y_j)\} - \mathrm{E}\{G(\mathrm{N}(0,1))\}] > 0.$$

Asymptotic variance: The trace of the asymptotic (co)variance of $\hat{W}$ is minimized when G is of the form:

$$G_{\mathrm{opt}}(u) = c_1 \log f_y(u) + c_2 u^2 + c_3.$$

Modified Bessel functions

- Modified Bessel functions of the first kind:

$$K_\lambda^{(1)}(x) = \frac{1}{2\pi i} \int \exp\{(x/2)(t + 1/t)\} t^{-\lambda-1} dt$$

- Modified Bessel functions of the second kind:

$$K_\lambda^{(2)}(x) = \frac{\Gamma(\lambda + 0.5)(2x)^\lambda}{\sqrt{\pi}} \int_0^\infty \frac{\cos t}{(t^2 + x^2)^{\lambda+0.5}} dt$$

Backtesting

- Risk level test: $H_0 : \mathrm{E}[N] = Ta$

  $$\mathrm{LR1} = -2\log\left\{(1-a)^{T-N}a^N\right\} + 2\log\left\{(1-N/T)^{T-N}(N/T)^N\right\}$$

  is asymptotically $\chi^2(1)$ distributed, where N the sum of exceedances happend in the interval $[1, T]$. a is the expected risk level.

- Clustering test: $H_0 : \pi_{00} = \pi_{10} = \pi, \pi_{01} = \pi_{11} = 1 - \pi$

  $$\mathrm{LR2} = -2\log\left\{\hat{\pi}^{n_0}(1-\hat{\pi})^{n_1}\right\} + 2\log\left\{\hat{\pi}_{00}^{n_{00}}\hat{\pi}_{01}^{n_{01}}\hat{\pi}_{10}^{n_{10}}\hat{\pi}_{11}^{n_{11}}\right\}$$

  is asymptotically $\chi^2(1)$ distributed, where
  $\pi_{ij} = P(I_t = j | I_{t-1} = i), i, j = 0, 1$ is the transition probability,
  and $n_{ij} = \sum_{t=1}^{T} I(I_t = j | I_{t-1} = i), i, j = 0, 1$.
  $\hat{\pi}_{ij} = n_{ij}/(n_{ij} + n_{i,1-j})$, $n_j = n_{0j} + n_{1j}$, and $\hat{\pi} = n_0/(n_0 + n_1)$.

Reference

Back, A. and Weigend, A.
A first application of independent component analysis to extracting structure from stock returns.
International Journal of Neural Systems, 8: 473-484, 1998

Barndorff-Nielsen, O.
Exponentially decreasing distributions for the logarithm of particle size.
Proceedings of the Royal Society of London, A 353: 401-419, 1977.

Reference

- Barndorff-Nielsen, O. and Blæsild, P.
  Hyperbolic distribution and ramifications: Contributions to theory and applications.
  Statistical Distributions in Scientific Work, C. Taillie, P. Patil and A. Baldessari (eds).
  Vol. 4, D. Reidel, pp. 19-44, 1981.

- Barndorff-Nielsen, O. and and Shephard, N.
  Modelling by Lévy processes for financial econometrics.
  Lévy Processes : Theory and Applications, Barndorff-Nielsen, T. Mikosch and S. Resnik (eds).
  Birkhauser Boston, 2001.

Reference

Chen, Y., Härdle, W. and Jeong, S.
Nonparametric risk management with generalized hyperbolic distributions.
SFB 649, discussion paper 2005-001, 2005, submitted.

Chen, Y., Härdle, W. and Spokoiny, V.
Portfolio Value at Risk based on independent components analysis.
accepted by the Journal of Computational and Applied Mathematics, 2005.

Chen, Y. and Spokoiny, V.
Local exponential smoothing with applications to volatility estimation and risk management.
working paper, 2006.

Reference

Cover, T. and Thomas, J.
Elements of Information Theory
John Wiley & Sons, 1991.
Single-trial variability in event-related bold signals.
NeuroImage, 15: 823-835, 2002.

Eberlein, E. and Keller, U.
Hyperbolic distributions in finance.
Bernoulli, 1: 281-299, 1995.

Franke, J., Härdle, W. and Hafner, C.
Introduction to Statistics of Financial Markets
Springer- Verlag Berlin Heidelberg New York, 2004.

Reference

Gallant, A.R. and Tauchen, G.
Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications.
Econometrica, 57, 1091-1120, 1989

Härdle, W., Kleinow, T. and Stahl, G.
Applied Quantitative Finance
Springer- Verlag Berlin Heidelberg New York, 2002.

Hyvärinen, A.
New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit
MIT Press, pp. 273-279, 1998.

Reference

Hyvärinen, A. and Oja, E.
Independent component analysis: Algorithms and applications
Neural Networks, 13: 411-430, 1999.

Hyvärinen, A., Karhunen, J. and Oja, E.
Independent Component Analysis
John Wiley & Sons, Inc. 2001.

Jaschke, S. and Jiang, Y.
Approximating value at risk in conditional gaussian models
Applied Quantitative Finance, W. Härdle, T. Kleinow and G. Stahl (eds).
Springer Verlag, 2002.

Reference

Jones, M. C. and Sibson, R.
What is projection pursuit? (with discussion)
Journal of the Royal Statistical Society, A 150(1): 1-36, 1987.

Jorion, P.
Value at Risk
McGraw-Hill, 2001.

Mercurio, D. and Spokoiny, V.
Statistical inference for time inhomogeneous volatility models.
The Annals of Statistics, 32: 577-602, 2004.