B. INFERENCE IN COPULA MODELS

B.I. Empirical copulas

Analogue of empirical cdf

B.II. Smoothed copulas

B.II.1. Nonparametric estimation of densities

B.II.2. Nonparametric estimation of copulas

Use of nonparametric kernel methods to smooth empirical copulas

$\Rightarrow$ simple graphical device

B.III. Parametric estimation

Maximum likelihood methods
and semiparametric methods

B.IV. Conclusions

B.I. Empirical copulas

Data :

$$\{Y_t = (Y_{1t},\ldots,Y_{nt})', t = 1,\ldots,T\}$$

= i.i.d. observations
(observed returns or losses)

Distributions :

$f(y), F(y)$ = joint pdf and cdf of Y_t

$f_j(y_j), F_j(y_j)$ = pdf and cdf of margins

Empirical cdfs:

$$\hat{F}(y_i) = \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{n} I[Y_{jt} \leq y_{ij}], i = 1,\ldots,d$$

$$\hat{F}_j(y_{ij}) = \frac{1}{T}\sum_{t=1}^{T} I[Y_{jt} \leq y_{ij}], i = 1,\ldots,d$$

To build empirical copulas, we use ranks instead of the original observations

Let us build the grid :

$$\left\{\left(\frac{j_1}{T},\dots,\frac{j_n}{T}\right); j_i = 0,\dots,T; i = 1,\dots,n\right\}$$

The empirical copula is given by

$$\hat{C}\left(\frac{j_1}{T},\dots,\frac{j_n}{T}\right) = \frac{1}{T}\sum_{t=1}^{T}\prod_{i=1}^{n} I\left[R_{it} \le j_i\right]$$

where R_{it} corresponds to the rank of Y_{it}.

⇒ We obtain step functions which are not differentiable.

B.II. Smoothed copulas

B.II.1. Nonparametric estimation of densities

The moments of a random variable are a summary of its distributional behavior.

A full information is provided by its distribution.

The *cumulative distribution function* for a single asset or loss i corresponds to

$$F_i(y_i) = P(Y_{it} \le y_i),$$

while for two assets i and j, we have

$$F_{ij}(y_i, y_j) = P(Y_{it} \le y_i, Y_{jt} \le y_j).$$

A cdf may be expressed as an expectation:

$$F_i(y_i) = \int_{-\infty}^{y_i} f_i(u)du = \int_{-\infty}^{+\infty} 1_{u \le y_i} f_i(u)du$$
$$= E\left[1_{Y_{it} \le y_i}\right]$$

where $1_{Y_{it} \le y_i}$ = indicator function of the set $\{Y_{it} : Y_{it} \le y_i\}$

$$\begin{aligned} 1_{Y_{it} \le y_i} &= 1 \quad \text{if} \quad Y_{it} \le y_i, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

In order to estimate expectations, we need to replace E by an empirical average:

$$\hat{F}_i(y_i) = \frac{1}{T}\sum_{t=1}^{T} 1_{Y_{it} \le y_i} \,,$$

$$\hat{F}_{ij}(y_i, y_j) = \frac{1}{T}\sum_{t=1}^{T} 1_{Y_{it} \le y_i, Y_{jt} \le y_j} \,,$$

⇒ We obtain step functions which are *not* differentiable.

⇒ We cannot build empirical counterparts of densities

$$f_i(\zeta_i) = \frac{dF_i(y_i)}{dy_i}\Big|_{y_i=\zeta_i}$$

In order to do so we have to rely on a kernel estimation of univariate densities.

<u>Idea behind</u>:

We start from the histogram,

$$\hat{f}_i(\zeta_i) = \frac{1}{T}\sum_{t=1}^{T} 1_{y_{i,t}=\zeta_i}$$

and replace bars by smooth bumps

$$\hat{f}_i(\zeta_i) = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{y_{i,t}-\zeta_i}{h}\right)$$

The bump K is called a *Kernel*.

It should integrate to one and be symmetric.

Example:

Gaussian Kernel = Gaussian density

$$K\left(\frac{y_{i,t}-\zeta_i}{h}\right)=\frac{1}{\sqrt{2\pi}}\exp-\frac{1}{2}\left(\frac{y_{i,t}-\zeta_i}{h}\right)^2$$

The smoothing parameter h is called the *bandwidth*.

The bandwidth h plays the same role as the class length for histograms.

If h is too large (large class), we get *oversmoothing*.

If h is too small (small class), we get *undersmoothing*.

Rule of thumb to select the bandwidth:

$$h = \hat{\sigma} T^{-1/5}$$

where $\hat{\sigma}$ is empirical standard deviation of the data.

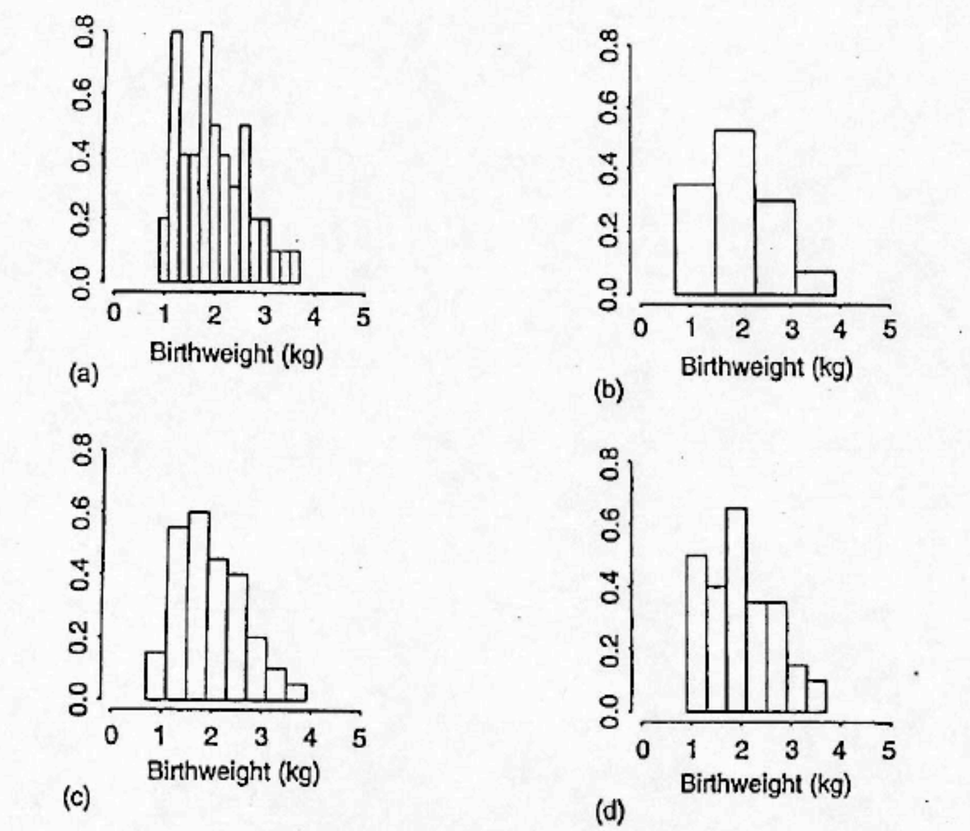


Figure 1.3. *Histograms of birthweight data. Figures (a) and (b) are based on binwidths of 0.2 and 0.8 respectively. Figures (c) and (d) are each based on a binwidth of 0.4 but with left bin edge at 0.7 and 0.9 respectively.*

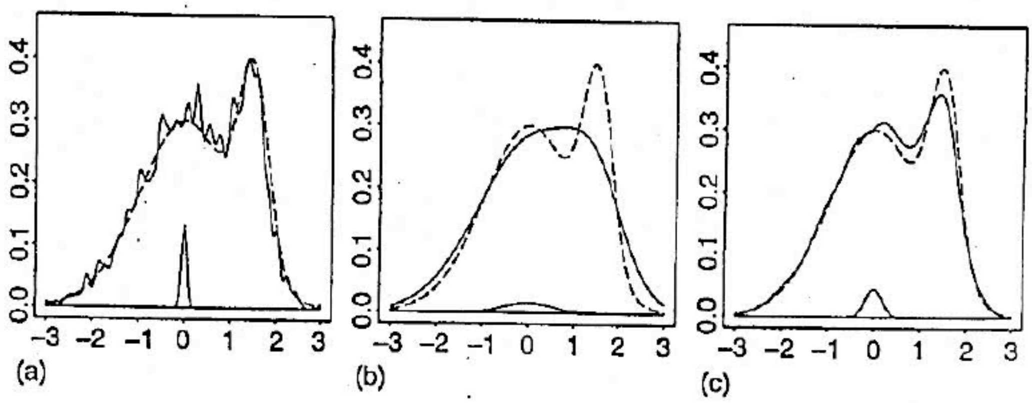


Figure 2.3. *Kernel density estimates based on a sample of* $n = 1000$ *observations from the normal mixture distribution* f_1 *described in the text. The solid line is the estimate, the broken line is the true density. The bandwidths are (a)* $h = 0.06$, *(b)* $h = 0.54$ *and (c)* $h = 0.18$. *The kernel weight for each estimate is illustrated by small kernels at the base of each figure.*

Table 4.5.2 *Some kernels and their efficiencies*

Kernel	$K(u)$	$D(K_{\text{opt}}, K)$
Epanechnikov	$(3/4)(-u^2+1)\ I(\|u\| \leq 1)$	1
Quartic	$(15/16)(1-u^2)^2\ I(\|u\| \leq 1)$	1.005
Triangular	$(1-\|u\|)\ I(\|u\| \leq 1)$	1.011
Gauss	$(2\pi)^{-1/2}\exp(-u^2/2)$	1.041
Uniform	$(1/2)\ I(\|u\| \leq 1)$	1.060

Note: The efficiency is computed as $\{V(K_{\text{opt}})B(K_{\text{opt}})/[V(K)B(K)]\}^{-1/2}$ for $k = 0$, $p = 2$.

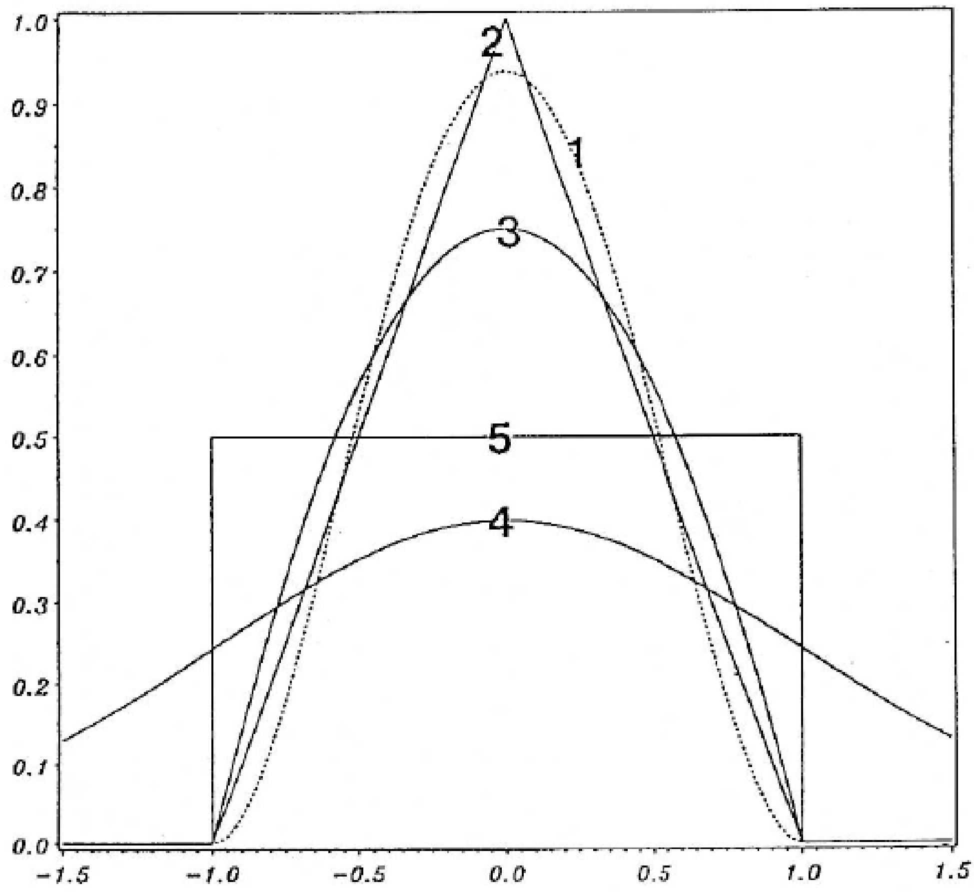


Figure 4.16. Positive kernels for estimating m (from Table 4.5.2). Label 1: quartic; label 2: triangular; label 3: Epanechnikov; label 4: Gauss; label 5: uniform.

It is possible to extend to higher dimensions and to the conditional case.

Kernel estimation of a bivariate density:

$$\hat{f}_{ij}(\zeta_i,\zeta_j) = \frac{1}{Th^2}\sum_{t=1}^{T} K\left(\frac{y_{i,t}-\zeta_i}{h}\right)K\left(\frac{y_{j,t}-\zeta_j}{h}\right)$$

Note that the *curse of dimensionality* appears when we are above five dimensions.

We need a lot of information (data) to get an accurate estimation of the high dimensional object to be estimated.

Kernel estimation of a conditional density:

Recall the definition (Bayes Theorem)

$$f(\zeta_i | y_{j,t} = \zeta_j) = \frac{f_{ij}(\zeta_i,\zeta_j)}{f_j(\zeta_j)}$$

⇒ we only need to replace the unknown quantities by their estimates

$$\hat{f}(\zeta_i \mid y_{j,t} = \zeta_j) = \frac{\hat{f}_{ij}(\zeta_i, \zeta_j)}{\hat{f}_j(\zeta_j)}$$

Extensions:

1) Zero boundary

Previous estimators have good properties when the data take values in $\Re$.

When data are bounded from below at zero (losses with a positive sign), they exhibit bias at the boundary (edge effect).

This *boundary bias* is due to weight allocation by the fixed symmetric kernel outside the density support when smoothing is carried out near the boundary.

One of the remedy consists in replacing symmetric kernels by asymmetric kernels, which never assigns weight outside the support.

The form of the estimators is the same

$$\hat{f}_i(\zeta_i) = \frac{1}{Th} \sum_{t=1}^{T} K\left(y_{i,t};\zeta_i,h\right)$$

but the symmetric kernel is replaced by an asymmetric kernel.

Examples:

Gamma Kernel:

$$K(y;\zeta,h) = \frac{y^{\varsigma/h} e^{-y/h}}{h^{\varsigma/h+1}\Gamma(\varsigma/h+1)}$$

where $\Gamma(x) = \int_0^{\infty} e^{-u} u^{x-1} du$

Reciprocal Inverse Gaussian Kernel:

$$K(y;\zeta,h) = \frac{1}{\sqrt{2\pi h y}} \exp\left(-\frac{\varsigma - h}{2h}\left(\frac{y}{\varsigma - h} - 2 + \frac{\varsigma - h}{y}\right)\right)$$

2) Compact support

When the data are defined on $\lfloor 0,1 \rfloor$, we face two boundaries.

It is then useful to use a kernel whose support is also $\lfloor 0,1 \rfloor$, for example the Beta kernel:

$$K(y;\zeta,h) = \frac{1}{B(\varsigma/h+1,(1-\varsigma)/h+1)} y^{\varsigma/h}(1-y)^{(1-\varsigma)/h}$$

where $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

Example:

This estimator is useful to analyze the distribution of recovery rates at default.

There is a renewed interest in LGD (loss given default), which is mainly prompted by Basle II and the explosion of the credit derivatives market.

Data are scarce, in particular outside the US.

The market standard to model LGD is a parametric assumption of beta distributed recoveries.

There are several measures of LGD

→ ultimate recoveries
→ trading price recoveries

These measures often give very different results.

Which one should be used depends who you are and what you do with your defaulted positions.

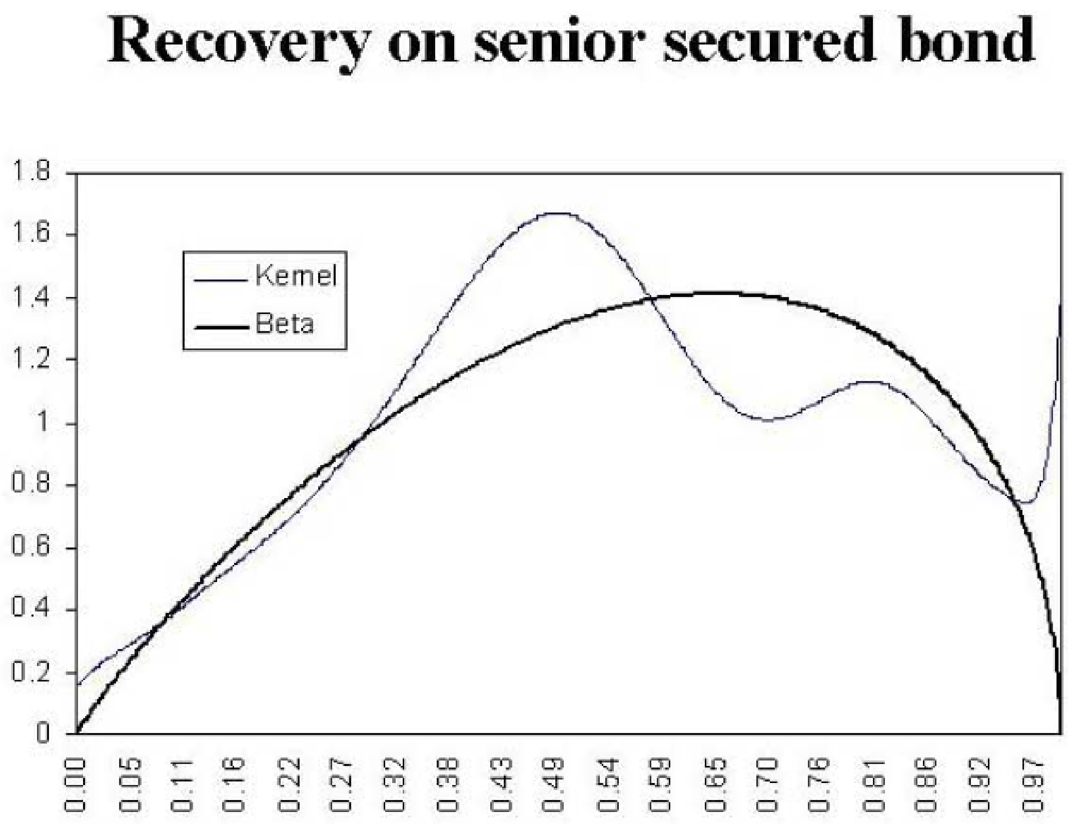
The data concern 623 US defaulted bond issues spanning from 1981 to end 1999.

These are trading price recoveries which are classified by industry and seniority.

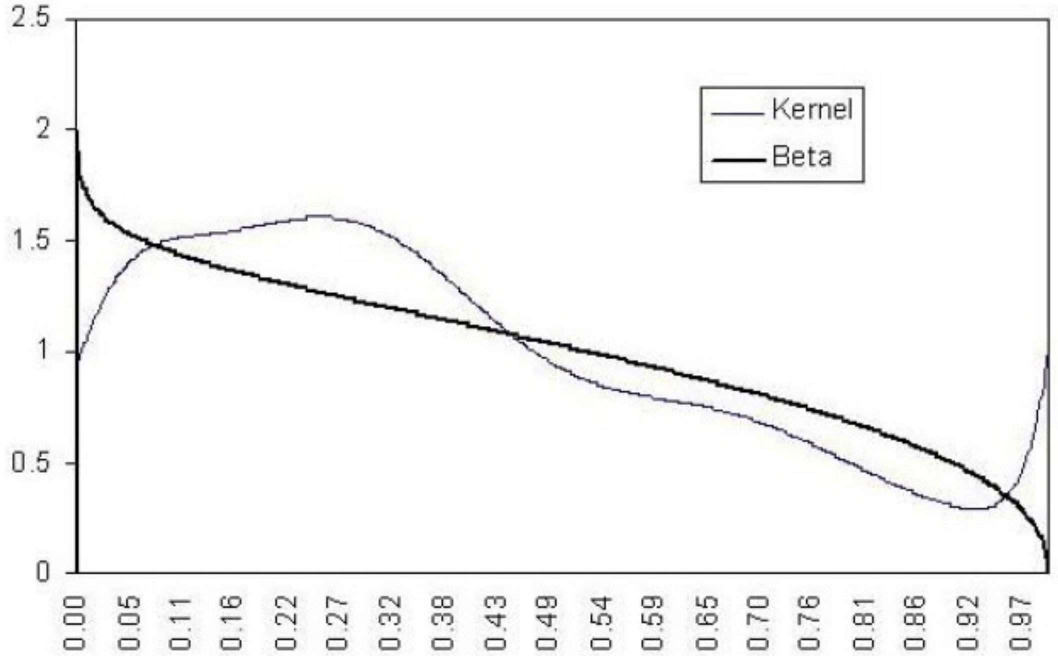
The data comes from the S&P/PMD database.

The market assumption of a beta distribution is often severely wrong.

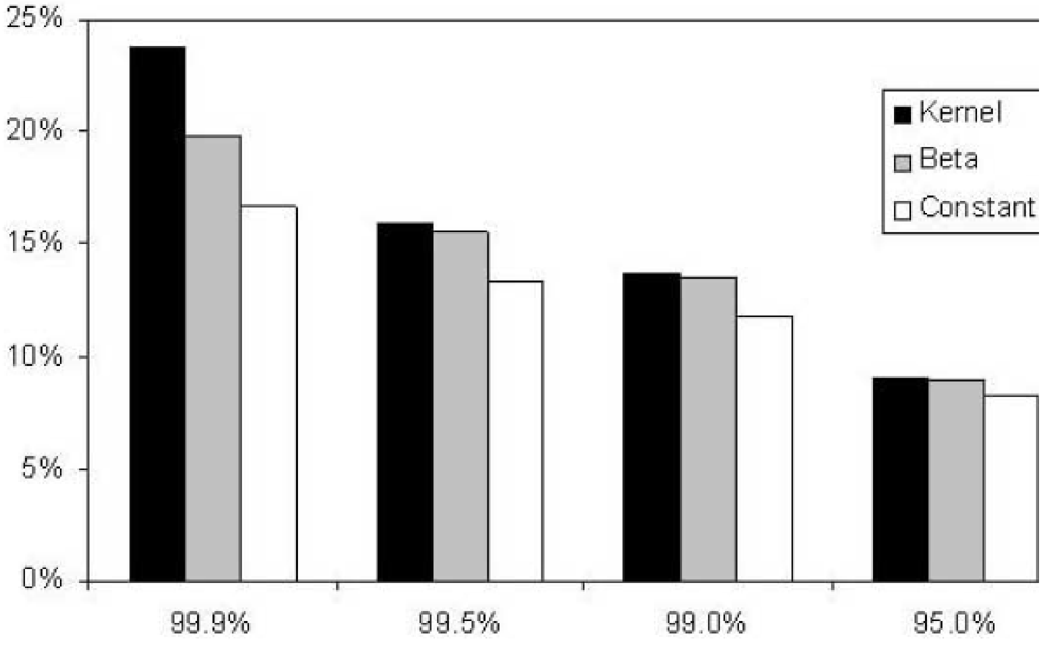
This could lead to underestimation of risk measures.



Recovery on junior bond



Impact on VaR



B.II.2. Nonparametric estimation of copulas

Use of nonparametric kernel methods to smooth empirical copulas

Let us consider a n-dimensional kernel:

$$K(x) = \prod_{j=1}^{n} K_j(x_j),$$

and its primitive function

$$\overline{K}(x) = \prod_{j=1}^{n} \int_{-\infty}^{x_j} K_j(u)du = \prod_{j=1}^{n} \overline{K}_j(x_j).$$

Let us denote

$$K(x;h) = \prod_{j=1}^{n} K_j(x_j / h_j),$$

where h_j are univariate bandwidths, and h is a diagonal matrix collecting them.

As before the pdf of Y_{jt} at y_j is estimated via

$$\hat{f}_j(y_j) = (Th_j)^{-1} \sum_{t=1}^{T} K_j\left(\frac{Y_{jt} - y_j}{h_j}\right)$$

and the pdf of Y_t at y is estimated via

$$\hat{f}(y) = (T|h|)^{-1} \sum_{t=1}^{T} K(Y_t - y; h)$$

Hence a smoothed estimator of the cdf of Y_{jt} at y_j is given by:

$$\hat{F}_j(y_j) = \int_{-\infty}^{y_j} \hat{f}_j(u)du$$

while the cdf of Y_t at y is estimated via

$$\hat{F}(y) = \int_{-\infty}^{y_1} \dots \int_{-\infty}^{y_n} \hat{f}(u)du$$

If a single Gaussian kernel $K_j(x) = \varphi(x)$ is used we get a simple form

$$\hat{F}_j(y_j) = \frac{1}{T}\sum_{t=1}^{T}\Phi((Y_{jt} - y_j)/h_j)$$

and $$\hat{F}(y) = \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{n}\Phi((Y_{jt} - y_j)/h_j),$$

where φ, Φ are the pdf, cdf of a *N(0,1)*.

In order to estimate the copula at point *u*, we can directly exploit the expression :

$$C(u_1,\ldots,u_n) = F(F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n)),$$

and use an empirical counterpart based on smoothed cdf and smoothed quantiles.

We use a plug-in method :

$$\hat{C}(u) = \hat{F}(\hat{\varsigma})$$

where $\hat{\varsigma} = (\hat{\varsigma}_1, \ldots, \hat{\varsigma}_n)'$ and $\hat{\varsigma}_j = \inf\{y : \hat{F}_j(y) \geq u_j\}$.

Here $\hat{\varsigma}_j$ corresponds to a kernel estimate of the quantile of Y_{jt} with probability level u_j.

Empirical illustrations :

1,700 observations of daily returns on pairs (CAC40,DAX35) and pairs (S&P500,DJI) from 01/01/1994 to 07/07/2000,

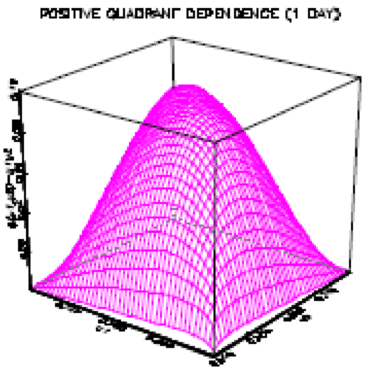
Two holding horizons: 1 day and 10 days (cf Basel Committee on Banking Supervision).

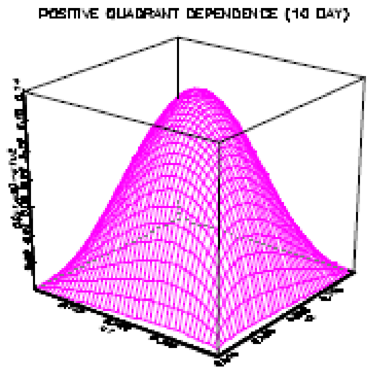
(CAC40,DAX35): comparison with

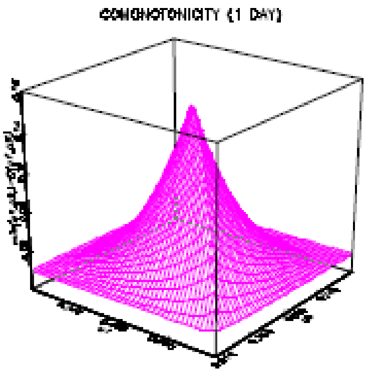
independent: $C(u_1, u_2) - u_1 u_2$ cf PQD ≥ 0

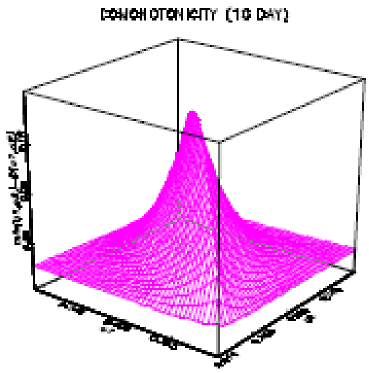
comonotonic: $\min(u_1, u_2) - C(u_1, u_2)$,

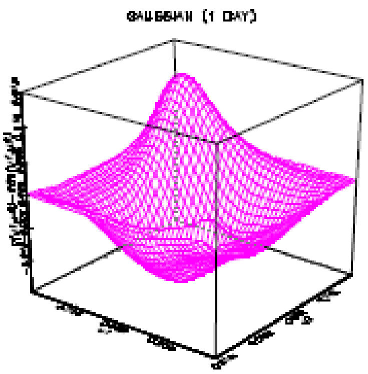
Gaussian: $C(u_1, u_2) - C_{Gauss}(u_1, u_2; \rho^*)$

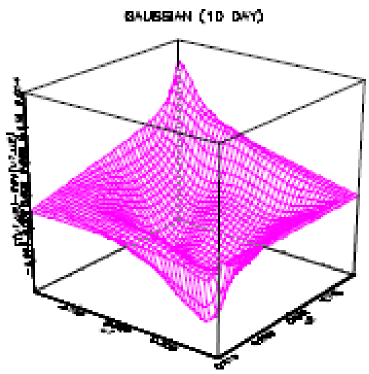




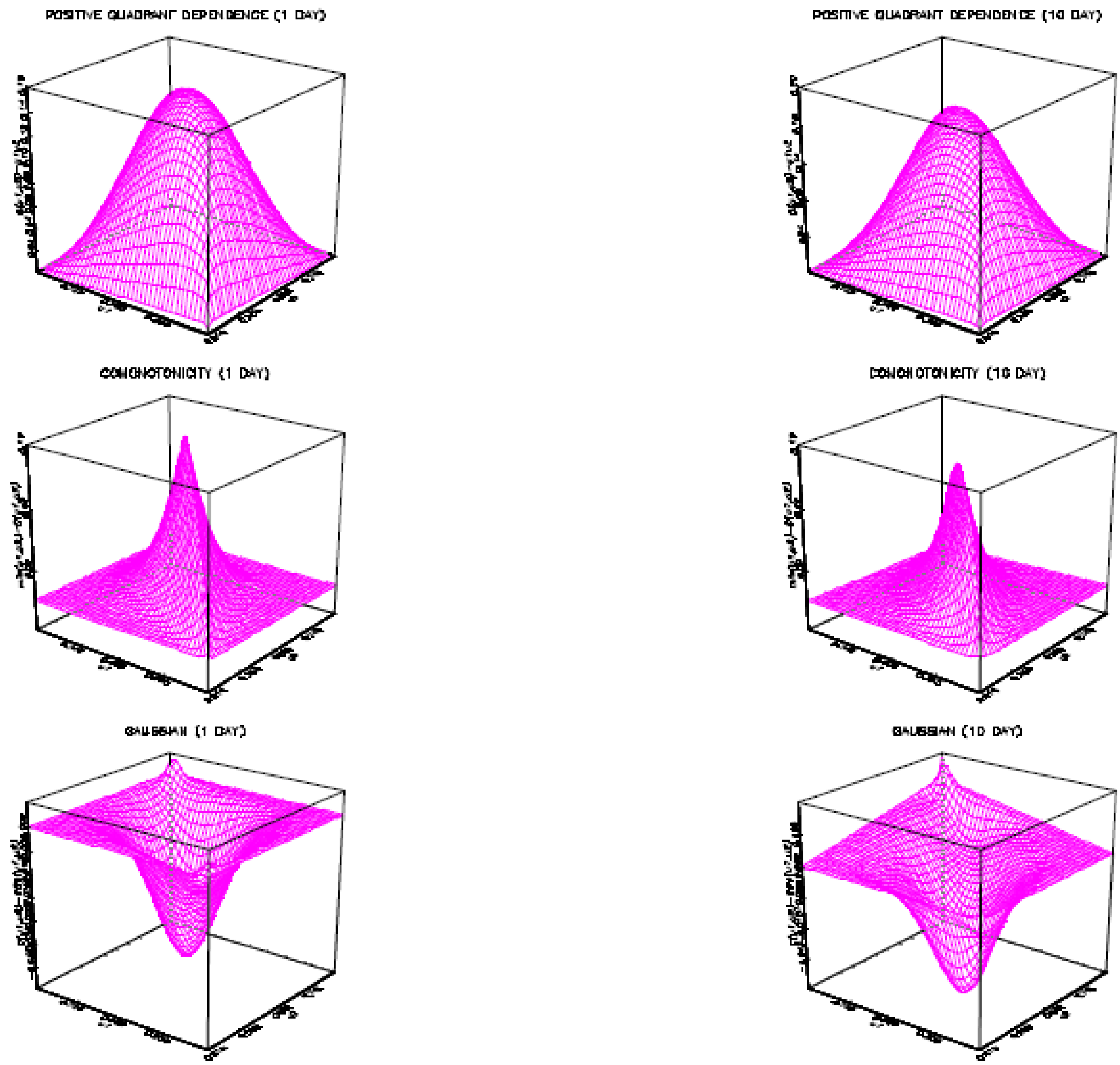








(S&P500,DJI): same comparison



simple graphical device to detect adequacy of parametric copula

Here we use the link $\rho^{*} = 2\sin(\rho\pi/6)$
where ρ is the rank correlation
to estimate the parameter value.

B.III. Parametric estimation

Method of moments:

In some cases there exist a link between Kendall's τ or Spearman's ρ and the parameter of the copula function.

Examples:

Gumbel copula:

$$C(u_1, u_2; \theta)$$
$$= \exp(-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta})$$

we get

$$\tau = 1 - 1/\theta$$

Hence $\hat{\theta} = 1/(1-\hat{\tau})$

simple estimate: $\hat{\tau} = (c - d)/(c + d)$ where c and d are the numbers of disjoint pairs which are concordant and discordant.

If $\dfrac{Y_j - Y_i}{X_j - X_i} > 0$: pair is concordant

If $\dfrac{Y_j - Y_i}{X_j - X_i} < 0$: pair is discordant

If $\dfrac{Y_j - Y_i}{X_j - X_i} = 0$: pair is a tie

(a tie add 0.5 to both counts)

If $X_j - X_i = 0$: pair is ignored

The Kendall's tau corresponds to the probability of concordance minus the probability of discordance.

Bivariate elliptical copula:

such as Normal, Student copulas:

$$\tau = \frac{2}{\pi} \arcsin \theta$$

Normal copula : $\rho = \frac{6}{\pi} \arcsin \frac{\theta}{2}$

Hence $\hat{\theta} = \sin(\hat{\tau}\pi / 2)$,
$\hat{\theta} = 2\sin(\hat{\rho}\pi / 6)$

Maximum likelihood estimation:

The log-likelihood is given by

$$l(\theta,\beta) = \sum_{t=1}^{T} \ln c(F_1(Y_{1t};\beta_1),\ldots,F_n(Y_{nt};\beta_n);\theta) + \sum_{t=1}^{T}\sum_{j=1}^{n} \ln f_j(Y_{jt};\beta_j)$$

The parameter θ is associated with the copula function.

The parameter β_j is associated with the j-th margins.

The ML estimator of both parameters is given by $(\hat{\theta},\hat{\beta}) = \arg\max l(\theta,\beta)$

Numerical optimization procedures are required to optimize the criterion (cf. gradient methods).

It is also possible to opt for a *two step procedure:*

1) estimate the parameter β_j of each margin by optimizing the univariate log-likelihood:

$$l(\beta_j) = \sum_{t=1}^{T} \ln f_j(Y_{jt};\beta_j)$$

in order to get $\hat{\beta}_j$

2) plug the estimated parameters $\hat{\beta}_j$, and optimize the concentrated log-likelihood:

$$l(\theta) = \sum_{t=1}^{T} \ln c(F_1(Y_{1t};\hat{\beta}_1),\ldots,F_n(Y_{nt};\hat{\beta}_n);\theta)$$

Advantage:
numerically easier (stability and tractability) to optimize in parameter spaces of lower dimensions.

Inconvenient:
less efficient asymptotically, i.e. less precise estimator (larger asymptotic variance)

Semiparametric estimation:

In previous estimation method, distributions of margins were parametrically specified.

Alternative modeling: leave the margins unspecified and use a parametric specification for the copula only.

Distributions of margins are estimated by empirical cdfs.

Concentrated log-likelihood becomes

$$l(\theta) = \sum_{t=1}^{T} \ln c(\hat{F}_1(Y_{1t}),\ldots,\hat{F}_n(Y_{nt});\theta)$$

Advantage:
Avoid potential misspecification of margins

Inconvenient:
less efficient asymptotically, i.e. less precise estimator (larger asymptotic variance)

Example: Monte Carlo study concerning the impact of misspecification.

Estimation performance measured in terms of Bias and MSE (Mean Square Error)

MC experiments: 1000.
Sample size: 200,500,1000.
True model: Frank Copula, Student margins.

Parameter of copula:

$\theta = 1$ ($\rho = .1645, \tau = .1100$),

$\theta = 2$ ($\rho = .3168, \tau = .2139$)

Parameter of Student margins:

$\beta = 3$ (degrees of freedom)

Pseudo distribution: Normal margins

We assume the margins to be normal instead of student to get the misspecified model.

We get a positive bias (overestimation of the dependence in the data)

Almost no efficiency loss of semiparametric approach

TABLE 1: Bias and MSE of copula parameter estimators

	Sample size: $n = 200$			
$\theta = 1$	one-step	two-step	semi	true
Bias	0.7012	0.6094	-0.0206	-0.0164
MSE	1.5119	1.0591	0.1754	0.1797
$\theta = 2$	one-step	two-step	semi	true
Bias	1.1144	0.9292	-0.0202	-0.0124
MSE	2.2869	1.4851	0.1913	0.1928
	Sample size: $n = 500$			
$\theta = 1$	one-step	two-step	semi	true
Bias	0.7720	0.6931	-0.0081	-0.0067
MSE	1.1114	0.8184	0.0673	0.0684
$\theta = 2$	one-step	two-step	semi	true
Bias	1.2165	1.0354	-0.0067	-0.0045
MSE	2.0494	1.3977	0.0730	0.0729
	Sample size: $n = 1000$			
$\theta = 1$	one-step	two-step	semi	true
Bias	0.7904	0.7190	-0.0046	-0.0048
MSE	0.9647	0.7397	0.0360	0.0360
$\theta = 2$	one-step	two-step	semi	true
Bias	1.2553	1.0784	-0.0037	-0.0037
MSE	1.9702	1.3853	0.0387	0.0382

B.IV. Conclusions

Nonparametric tools are easy to implement and are useful graphical guides.

If one has any doubt about the correct modeling of margins there is probably little to loose but lots to gain from shifting towards a semiparametric approach.