Modeling Dependence in Financial Data with Semiparametric Archimedean Copulas

José Miguel Hernández-Lobato and Alberto Suárez

Universidad Autónoma de Madrid

July 17, 2009

Outline

Outline

Copulas and Computational Finance

An important problem in Computational Finance is the modeling of the multivariate distribution of the returns generated by different financial assets. However,

Many standard univariate models do not have a direct extension to higher dimensions.

To deal with this problem we can...

Use copulas to link univariate models into a joint multidimensional model.

Definition of a copula function

Sklar's Theorem

Let $(X_1, \ldots, X_d)^\mathrm{T} \sim F$. Then there is a unique copula C such that

$$F(x_1, \ldots, x_d) = C\left[F_1(x_1), \ldots, F_d(x_d)\right] , \tag{1}$$

where $F_1, \ldots, F_d$ are the marginal distributions of F.

1. C is a distribution in the d-dimensional unit hypercube with uniform marginals.
2. C captures the dependence structure among the different univariate components.

The estimation of F can be performed by first, modeling the marginals $F_1, \ldots, F_d$ and second, by modeling the copula C.

Eliminating the marginals

Transforming the data using the marginals leads to a sample from the copula of the original distribution.

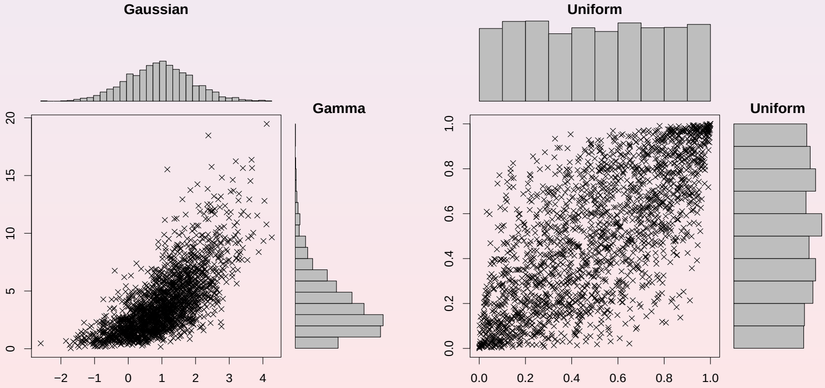


Figure: Left, sample form F. Right, sample from C.

Eliminating the marginals (continued)

Transforming the data using the marginals leads to a sample from the copula of the original distribution.

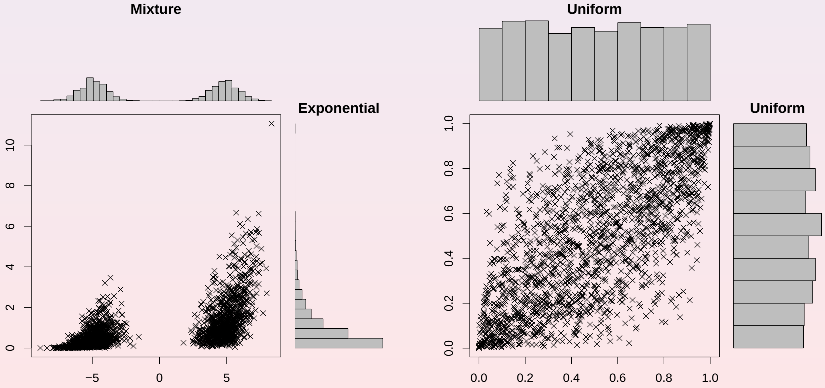
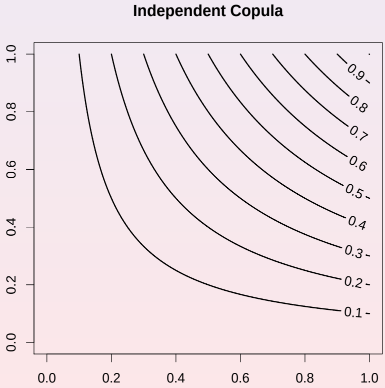
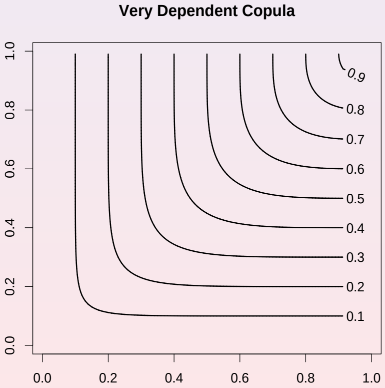


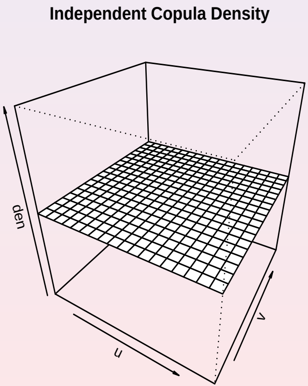
Figure: Left, sample form F. Right, sample from C.

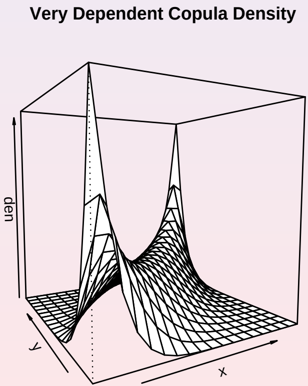
Some bivariate copula functions





Some bivariate copula densities





Parametric, non-parametric and semiparametric copulas

Modeling multivariate data with copulas requires copula functions that are flexible and robust at the same time.

- Parametric copula models are robust but they lack flexibility.
- Non-parametric copula models can represent any dependence structure but they are prone to overfitting.

Solution: semiparametric copula models

- We focus in the family of bivariate Archimedean copulas.
- These copulas are parameterized in terms of a latent unidimensional function.
- Our approach describes this latent function in a non-parametric manner.

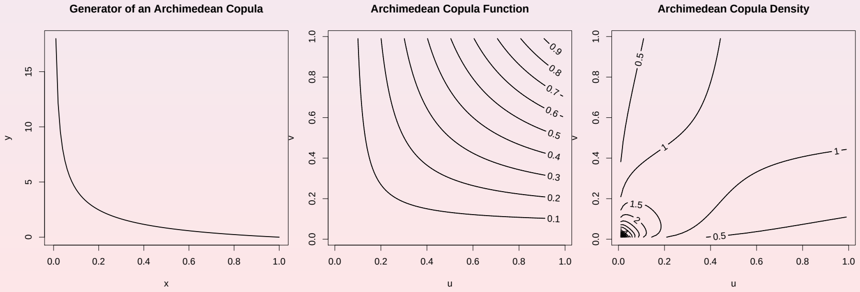
Outline

Bivariate Archimedean copulas

are defined by a generator $\phi^{-1} : [0,1] \to \mathbb{R}^{+} \cup \{+\infty\}$ that is convex, strictly decreasing and satisfies $\phi^{-1}(0) = +\infty$ and $\phi^{-1}(1) = 0$. Given ϕ^{-1}, the copula function is

$$C(u,v) = \phi\left[\phi^{-1}(u) + \phi^{-1}(v)\right], \quad u, v \in [0,1] \qquad (2)$$

where ϕ is the inverse of ϕ^{-1}.



Parameterizations of Bivariate Archimedean copulas

ϕ^{-1} is a very constrained function. For this reason, we introduce a novel latent function $g : \mathbb{R} \to \mathbb{R}$ that is in a one-to-one relationship with ϕ^{-1} and is easier to model

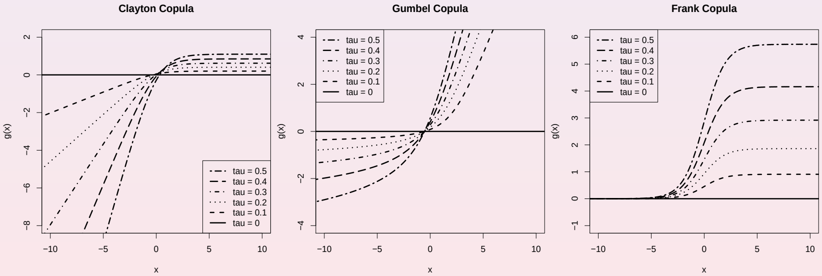
$$g(x) = \log \frac{\phi''\left\{\phi^{-1}\left[\sigma(x)\right]\right\}}{\phi'\left\{\phi^{-1}\left[\sigma(x)\right]\right\}}\,, \tag{3}$$

$$\phi^{-1}(x) = \int_x^1 \frac{1}{\int_0^y \exp\left\{g\left[\sigma^{-1}(z)\right]\right\}\,dz}\,dy\,, \tag{4}$$

where σ is the sigmoid function. Asymptotically, g behaves like a linear function.

Some plots of g for parametric Archimedean copulas

g has a central non-linear reagion.



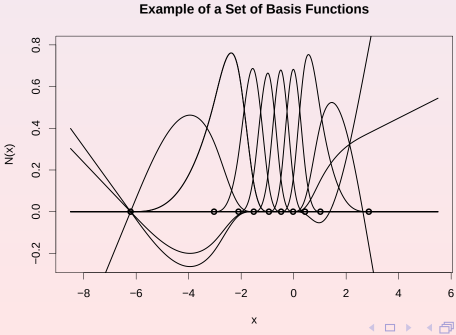
τ is a measure of non-linear dependence.

Modeling g

We model g by means of a natural cubic spline basis:

$$g_{\boldsymbol{\theta}}(x) = \sum_{i}^{K} \theta_i N_i(x) \,. \qquad (5)$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$.



Estimation of g

Given $\mathcal{D} = \{U_i, V_i\}_{i=1}^{N}$ where $U_i, V_i \sim U(0,1)$, we estimate g as the maximizer of

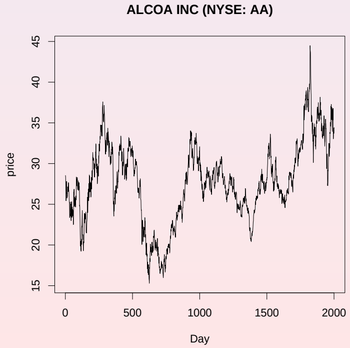
$$\mathrm{PLL}(\mathcal{D}|g_{\boldsymbol{\theta}}, \beta) = \log \mathcal{L}(\mathcal{D}|g_{\boldsymbol{\theta}}) - \beta \int \left\{ g''_{\boldsymbol{\theta}}(x) \right\}^2 \, dx \tag{6}$$

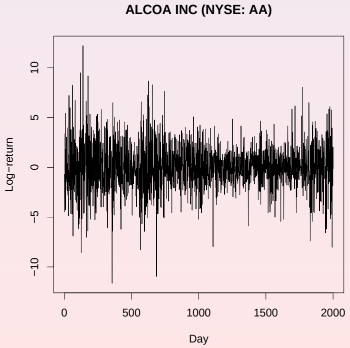
where β is a smoothing parameter fixed by a 10-fold cross validation grid search.

Outline

The data

- 64 components of the Dow Jones Composite Index.
- Daily log-returns From April 13th, 2000 to March 31st, 2008.
- We obtain 64 time series with 2000 consecutive log-returns.





Modeling the marginal distributions

We use an asymmetric GARCH process with an autoregressive component and innovations that follow an unspecified density.

$$X_t = \phi_0 + \phi_1 X_{t-1} + \sigma_t \varepsilon_t \,, \tag{7}$$

$$\sigma_t = \kappa + \alpha(|\sigma_{t-1}\varepsilon_{t-1}| - \gamma\sigma_{t-1}\varepsilon_{t-1}) + \beta\sigma_{t-1} \,, \tag{8}$$

where $\kappa > 0$, $\alpha, \beta \geq 0$, $-1 < \gamma, \phi_1 < 1$, $\varepsilon_t \sim f$ and f has zero mean and unit standard deviation.

Once we have a marginal model for each financial asset, we map each return to [0,1] using the probability integral transform.

Benchmark copula estimation methods

- **SPAC** The method that is described here.
- **LAM** A flexible Archimedean copula [Lambert, 2007].
- **DIM** A flexible Archimedean copula [Dimitrova et al., 2008].
- **GK** A non-parametric copula based on Gaussian kernels.
- **BMG** A method based on a Bayesian mixture of Gaussians.
- **ST** The Student's t copula model.
- **GC** The Gaussian copula model.
- **SST** The skewed Student's t copula model.

Experimental protocol

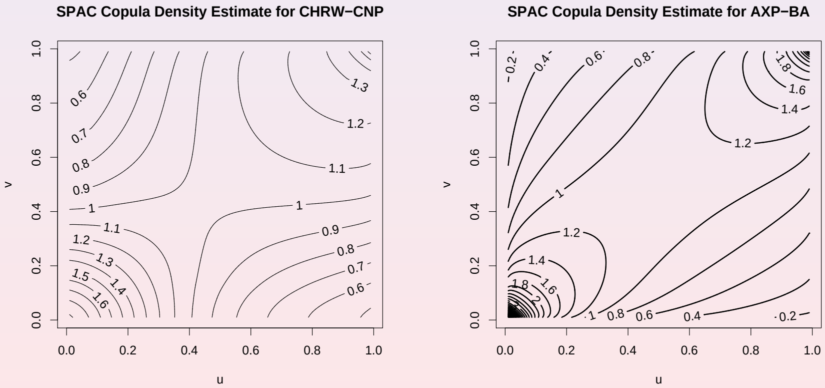
1. We form 32 pairs of financial assets and obtain 32 samples of size 2000 from the corresponding bivariate copulas.
2. Each copula sample is randomly split in 100 pairs of independent train and test sets with 1333 and 667 instances, respectively.
3. The copula estimation method is applied to each train set and its log-likelihood is evaluated on the corresponding test set.
4. For each of the 32 pairs of financial assets, we compute the average test log-likelihood of the copula estimate.

Average log-likelihood for each method on each problem

Red 1st method.
Blue 2nd method.
Green 3rd method.

Assets		SPAC	ST	SST	GS	LAM	DIM	BMG	GK
WMB	WMT	5.97	6.47	6.38	4.72	2.78	2.97	4.66	-0.81
KO	LSTR	13.90	11.96	11.08	11.90	12.26	10.82	11.69	7.52
FDX	FE	14.35	13.43	12.73	12.45	11.07	11.16	13.31	6.09
CHRW	CNP	15.63	12.95	12.82	13.09	14.19	13.28	13.33	9.04
EXC	EXPD	15.41	15.98	15.44	14.05	14.18	12.77	14.01	9.89
PEG	PFE	17.80	17.77	17.80	15.10	14.58	14.80	16.02	10.44
OSG	PCG	17.90	16.37	17.57	16.20	16.84	15.86	15.80	13.18
LUV	MCD	18.21	17.66	17.47	17.15	16.38	16.11	17.14	13.22
DIS	DUK	18.84	20.99	20.30	17.25	17.27	15.60	18.10	12.84
NI	NSC	20.66	20.43	19.50	18.70	19.52	17.69	18.67	14.99
AES	AIG	21.71	21.84	21.53	20.28	19.66	19.58	20.22	15.40
PG	R	22.89	23.46	22.80	20.24	20.14	20.10	21.76	16.76
FPL	GE	23.33	23.26	23.10	20.12	20.24	19.68	21.78	17.16
AA	AEP	23.66	23.28	23.33	22.36	21.67	21.31	22.11	16.52
SO	T	23.88	23.54	24.19	21.12	22.18	21.58	22.91	15.58
XOM	YRCW	24.83	23.53	23.24	22.36	22.41	22.28	22.44	16.05
MRK	MSFT	25.65	24.50	23.69	22.81	22.39	20.71	24.02	20.16
MMM	MO	24.93	24.90	24.10	24.57	22.57	21.57	24.04	19.81
D	DD	26.37	26.35	25.97	24.90	24.35	23.95	24.57	17.25
JNJ	JPM	27.19	29.38	29.31	23.00	24.65	24.11	28.82	24.38
ALEX	AMR	29.87	28.75	28.76	28.97	27.62	27.04	28.57	23.56
UTX	VZ	33.88	33.25	32.21	33.11	30.98	31.06	32.48	24.15
CAL	CAT	35.23	35.43	35.55	31.31	34.10	34.18	33.41	25.96
INTC	JBHT	44.22	42.90	42.77	41.09	42.58	41.11	42.00	42.06
GM	GMT	45.21	44.52	44.20	41.60	43.57	43.22	44.33	41.87
AXP	BA	52.06	50.03	51.47	47.40	50.86	50.23	49.96	46.07
HD	HON	56.84	57.17	56.13	52.55	55.30	54.36	54.69	47.07
BNI	C	61.36	60.55	60.43	58.39	60.25	58.34	58.58	55.56
CNW	CSX	80.36	80.59	80.09	75.93	79.19	77.24	77.65	71.23
UNP	UPS	80.86	80.63	79.90	75.21	79.38	78.49	78.72	74.53
HPQ	IBM	89.44	90.05	89.27	82.27	87.64	85.35	88.37	79.22
ED	EIX	93.15	90.99	93.26	86.71	91.97	89.84	93.23	88.80

Some copula density estimates



Summary

- We have proposed a novel estimator of semiparametric bivariate Archimedean copulas.
- The estimator is based on a new function g that uniquely determines the copula and is easy to model.
- A basis of natural cubic splines is used to model g in a non-parametric manner.
- Estimation is performed by maximum penalized likelihood.
- Experimental results show the improved performance of the proposed estimator with respect ot other benchmark methods.
- Accurate multivariate financial models must capture asymmetric dependence structures.

Thanks!

Thanks!