MODELS FOR CONSTRUCTION OF MULTIVARIATE DEPENDENCE

Daniel Berg
DnB NOR Asset Management

Kjersti Aas
Norwegian Computing Center

2nd R/Rmetrics User and Developer Workshop

June 29th – July 3rd 2008, Meielisalp, Lake Thune, Switzerland

Norwegian Computing Center (www.nr.no)

- Private, independent non-profit foundation
- Established in 1952
- 50 research scientists (mostly statisticians)
- Financed by:
  - domestic private companies
  - public sector
  - Research Council of Norway
  - EU
  - international companies
- Turnover $\sim$ 8.5 Mill Euro
- Largest client sectors:
  - petroleum
  - finance, insurance and commodity markets
  - environment, marine resources, health

Introduction

- Apart from the Gaussian- and Student copulae, the set of multivariate copulae proposed in the literature is rather limited.
- Archimedean copulae – most common mv extension extremely restrictive, allowing only one parameter
- Some attempts at constructing more flexible mv copulae
- We examine two such constructions:
  - Nested Archimedean constructions (NAC)
  - Pair-copula constructions (PCC)
- Both constructions are hierarchical in nature and model mv data using a cascade of bivariate copulae
- Differ in the modelling of the dependency structure

Outline

- ▷ Preliminaries
- ▷ Nested Archimedean constructions (NAC)
- ▷ Pair-copula constructions (PCC)
- ▷ Comparison
- ▷ Applications
- ▷ Summary

Preliminaries

Motivation

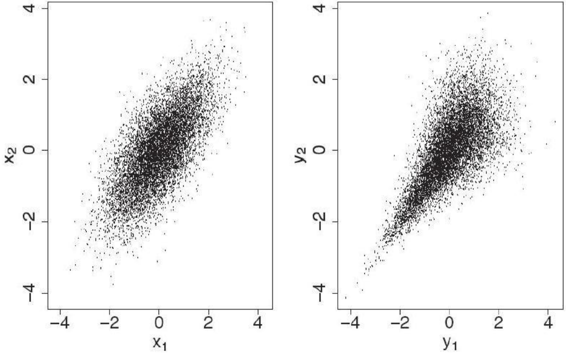


Figure: Two simulated data sets, both with standard normal margins and correlation 0.7.

Preliminaries

Motivation

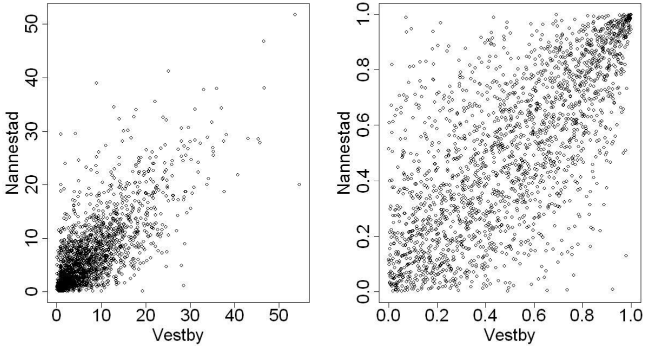


Figure: Daily nonzero precipitation values from two Norwegian cities and its copula.

Preliminaries

Copula definition & theorem

Definition (Copula)

A d-dimensional copula is a multivariate distribution function $\mathcal{C}$ with standard uniform marginal distributions.

Theorem (Sklar, 1959)

Let H be a joint distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

Preliminaries

Useful copula results

- A general d-dimensional density h can be expressed, for some copula density c, as

$$h(x_1, \dots, x_d) = c\{F_1(x_1), \dots, F_d(x_d)\} f_1(x_1) \cdots f_d(x_d).$$

- Non-parametric estimate for $F_i(x_i)$ commonly used to transform original margins into standard uniform:

$$u_{ji} = \widehat{F}_i(x_{ji}) = \frac{R_{ji}}{n+1},$$

  where R_{ji} is the rank of x_{ji} amongst $x_{1i}, \dots, x_{ni}$.

- u_{ji} commonly referred to as *pseudo-observations* and models based on non-parametric margins and parametric copulas are referred to as *semi-parametric* copulas

Nested Archimedean constructions

Outline

- Fully nested construction (FNAC)
- Partially nested construction (PNAC)
- Hierarchically nested construction (HNAC)
- Parameter estimation
- Simulation

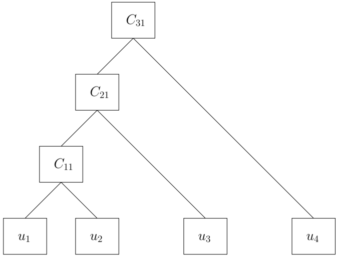
Nested Archimedean constructions

Fully nested construction (FNAC)

- Originally proposed by Joe (1997); also discussed in Embrechts et al. (2003); Whelan (2004); Savu and Trede (2006); McNeil (2008).
- Allows specification of max $d-1$ copulae while remaining unspecified copulae are implicitly given from construction
- All bivariate margins are Archimedean copulae

Nested Archimedean constructions

Fully nested construction (FNAC)



- (u_1, u_3) and (u_2, u_3) both have copula C_{21}.
- (u_1, u_4), (u_2, u_4) and (u_3, u_4) all have copula C_{31}.

$$\begin{aligned} C(u_1, u_2, u_3, u_4) &= C_{31}\{u_4, C_{21}\{u_3, C_{11}\{u_1, u_2\}\}\} \\ &= \varphi_{31}^{-1}\{\varphi_{31}(u_4) + \varphi_{13}(\varphi_{21}^{-1}\{\varphi_{21}(u_3) + \varphi_{21}(\varphi_{11}^{-1}\{\varphi_{11}(u_1) + \varphi_{11}(u_2)\})\})\}. \end{aligned}$$

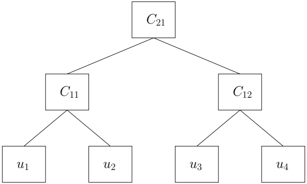
Nested Archimedean constructions

Partially nested construction (PNAC)

- Originally proposed by Joe (1997); also discussed in Whelan (2004); McNeil et al. (2006); McNeil (2008).
- Allows specification of max $d-1$ copulae while remaining unspecified copulae are implicitly given by construction
- Composite between exchangeable copula and FNAC since it is partially exchangeable

Nested Archimedean constructions

Partially nested construction (PNAC)



- (u_1, u_3), (u_1, u_4), (u_2, u_3) and (u_2, u_4) all have copula C_{21}.

$$\begin{aligned} C(u_1, u_2, u_3, u_4) &= C_{21}\{C_{11}\{u_1, u_2\}, C_{12}\{u_3, u_4\}\} \\ &= \varphi_{21}^{-1}\{\varphi_{21}(\varphi_{11}^{-1}\{\varphi_{11}(u_1) + \varphi_{11}(u_2)\} + \varphi_{12}^{-1}\{\varphi_{12}(u_3) + \varphi_{12}(u_4)\}\}. \end{aligned}$$

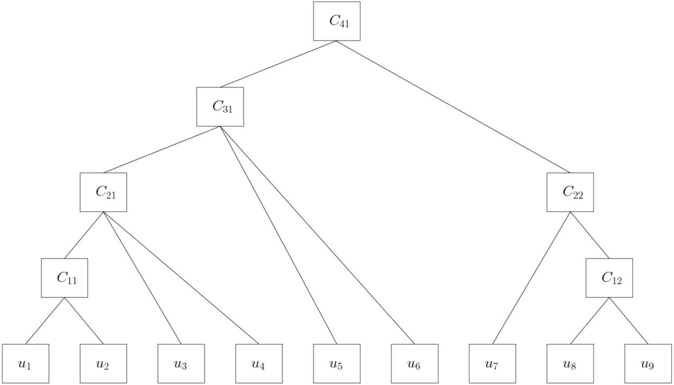
Nested Archimedean constructions

Hierarchically nested Archimedean constructions (HNAC)

- Proposed by Joe (1997); mentioned in Whelan (2004), worked out in full generality by Savu and Trede (2006).
- Allows specification of max $d-1$ copulae while remaining unspecified copulae are implicitly given by construction
- Extension of PNAC – copulae involved do not need to be bivariate. Both FNAC and PNAC are special cases.
- A number of conditions must be satisfied to ensure that this construction yields a valid Archimedean copula:
    - All inverse generator functions must be completely montone.
    - Degree of dependence must decrease with level of nesting
    - All bivariate copulae must in general belong to the same family of Archimedean copulae (although some few exceptions exist)

Nested Archimedean constructions

Hierarchically nested Archimedean constructions (HNAC)



Nested Archimedean constructions

Parameter estimation

- For all NACs, parameters may be estimated by maximum likelihood.
- Not straightforward to derive density – resort to computer algebra system (R/D, Mathematica, ...).
- Density often obtained by recursive approach – no. of comp. steps needed to evaluate density increases rapidly with complexity of copula.

Nested Archimedean constructions

Simulation

- Simulation from high-dim NACs is not straightforward in general.
- Most algorithms involve higher order derivatives of generator, inverse generator or copula functions. These are usually extremely complex for high dimensions.
- Some exceptions exist:
  - McNeil (2008) uses Laplace-transform method for the FNAC (only for Gumbel and Clayton).
  - McNeil (2008) uses Laplace-transform method for 4-dim PNAC but does not extend the algorithm to higher-dim PNACs although this is possible (again for now only for Gumbel and Clayton).

Pair-copula constructions (PCC)

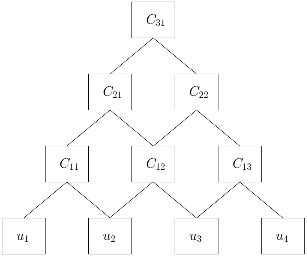
Outline

- ▷ Pair-copula construction (PCC)
- ▷ Parameter estimation
- ▷ Simulation

Pair-copula constructions (PCC)

- Originally proposed by Joe (1997) and later discussed in detail by Bedford and Cooke (2001, 2002); Kurowicka and Cooke (2006) (simulation) and Aas et al. (2008) (inference).
- Allows for the specification of $d(d-1)/2$ bivariate copulae of which the first $d-1$ are unconditional and the rest conditional.
- The bivariate copulae involved do not have to belong to the same class.

Pair-copula constructions (PCC)



- C_{21} is the copula of $F(u_1|u_2)$ and $F(u_3|u_2)$.
- C_{22} is the copula of $F(u_2|u_3)$ and $F(u_4|u_3)$.
- C_{31} is the copula of $F(u_1|u_2, u_3)$ and $F(u_4|u_2, u_3)$.
- Bedford and Cooke (2001) introduced *vines* as tree structures to help organize the many different constructions.

Pair-copula constructions (PCC)

- The density corresponding to the 4-dim figure is

$$\begin{aligned} c(u_1, u_2, u_3, u_4) = c_{11}(u_1, u_2) \cdot c_{12}(u_2, u_3) \cdot c_{13}(u_3, u_4) \\ \cdot c_{21}(F(u_1|u_2), F(u_3|u_2)) \cdot c_{22}(F(u_2|u_3), F(u_4|u_3)) \\ \cdot c_{31}(F(u_1|u_2, u_3), F(u_4|u_2, u_3)) \end{aligned}$$

- where

$$\begin{aligned} F(u_1|u_2) &= \partial C_{11}(u_1, u_2)/\partial u_2 \\ F(u_3|u_2) &= \partial C_{12}(u_2, u_3)/\partial u_2 \\ F(u_2|u_3) &= \partial C_{12}(u_2, u_3)/\partial u_3 \\ F(u_4|u_3) &= \partial C_{13}(u_3, u_4)/\partial u_3 \\ F(u_1|u_2, u_3) &= \partial C_{21}(F(u_1|u_2), F(u_3|u_2))/\partial F(u_3|u_2) \\ F(u_4|u_2, u_3) &= \partial C_{22}(F(u_4|u_3), F(u_2|u_3))/\partial F(u_2|u_3). \end{aligned}$$

Pair-copula constructions (PCC)

Parameter estimation

- Parameters of a PCC can be estimated by maximum likelihood.
- Since the density is given explicitly, the procedure is simpler than the one for the NACs.
- Likelihood must be maximized numerically – time consuming in higher dimensions.

Pair-copula constructions (PCC)

Simulation

- Simulation algorithm for a D-vine (figure) is straightforward and simple to implement.
- As for the NACs, the conditional inversion method can be used.
- To determine each of the conditional distribution functions involved, only the first partial derivative of a bivariate copula is needed.
- Hence, simulation from a PCC is in general much simpler and faster than for a NAC.

Comparison

Flexibility

Construction	*Max no. of copulae freely specified*	*Parameter constraints*	*Copula mixing*
NAC	$d-1$	Strict generators $\varphi_{i+1,1} \circ \varphi_{i,1}^{-1}$ must have completely monotone derivatives for all levels i of the structure	May combine different Archimedean families but under strong restrictions
PCC	$d(d-1)/2$	None	May combine any copula families from any class

- When looking for appropriate data sets for the comparison of these structures it turned out to be quite difficult to find real-world data sets satisfying the parameter constraints of the NAC.

Comparison

Computational efficiency

Method	*Likelihood evaluation*	*Estimation*	*Simulation*
Gumbel			
NAC	0.32	34.39	0.02
PCC	0.04	5.09	7.56
Frank			
NAC	0.12	5.34	64.83
PCC	0.02	1.22	5.82

- Estimation and likelihood: 4-dim data with 2065 observations
- Simulation: 1000 observations.

Comparison

Structure

- Multivariate distribution defined through a NAC will always, by definition, be an Archimedean copula and all bivariate margins will belong to a known parametric copula family.
- For the PCCs, neither the multivariate distribution nor the unspecified bivariate margins will belong to a known parametric copula family in general.

Applications

- Precipitation data
- Equity returns
- Means of comparison – goodness-of-fit
- Goodness-of-fit tests based on the empirical copula and its distribution function. Both shown to perform very well in Genest et al. (2008); Berg (2007).

$$C_n(\boldsymbol{u}) = \frac{1}{n+1}\sum_{j=1}^{n} 1(U_{j1} \leq u_1, \ldots, U_{jd} \leq u_d), \quad \boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d,$$

$$K_n(t) = \frac{1}{n+1}\sum_{j=1}^{n} 1(C_n(\boldsymbol{U}_j) \leq t),$$

$$S_n = n\int_{[0,1]^d} \{C_n(\boldsymbol{u}) - C_{\theta_n}(\boldsymbol{u})\}^2 \mathrm{d}C_n(\boldsymbol{u}),$$

$$T_n = n\int_{[0,1]^d} \{K_n(\boldsymbol{u}) - K_{\theta_n}(\boldsymbol{u})\}^2 \mathrm{d}K_n(\boldsymbol{u}).$$

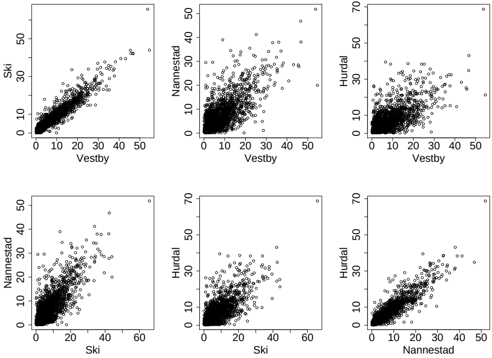
Applications

Precipitation data

- Four Norwegian weather stations.
- Daily, non-zero, rainfall data from 01.01.1990 to 21.12.2006 (2065 obs).
- We convert the observations to uniform pseudo-observations using normalized ranks before further modelling.
- Relevant for flood analysis, insurance, weather derivatives, ...

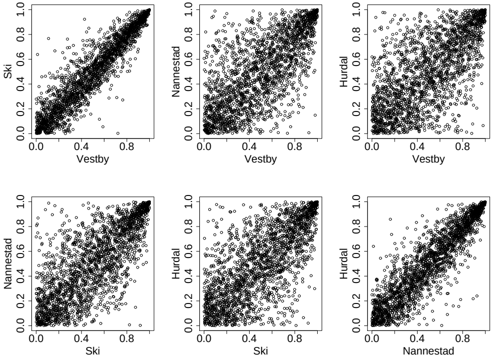
Applications

Precipitation data



Applications

Precipitation data



Applications

Precipitation data

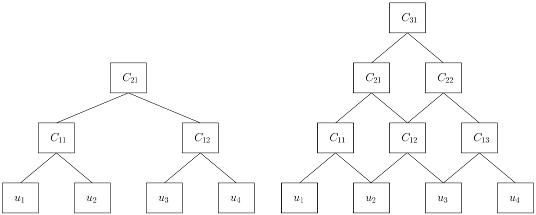
- Choose variable ordering based on pairwise Kendall taus:

Location	Ski	Nannestad	Hurdal
Vestby	0.79	0.49	0.47
Ski		0.56	0.53
Nannestad			0.71

Applications

Precipitation data

- ▷ We compare:



- ▷ Copulae at level one in both constructions are those corresponding to the largest pairwise Kendall taus.

Applications

Precipitation data

Parameter	*NAC*		*PCC*		
	Gumbel	*Frank*	*Gumbel*	*Frank*	*Student*
$\theta_{11} \setminus \nu_{11}$	4.32	16.69	4.34	16.78	0.93 \ 3.6
$\theta_{12} \setminus \nu_{12}$	3.45	13.01	2.24	7.10	0.78 \ 6.7
$\theta_{13} \setminus \nu_{13}$	-	-	3.45	12.98	0.90 \ 5.5
$\theta_{21} \setminus \nu_{21}$	1.97	5.96	1.01	0.08	0.01 \ 9.6
$\theta_{22} \setminus \nu_{22}$	-	-	1.02	0.61	0.09 \ 14.5
$\theta_{31} \setminus \nu_{31}$	-	-	1.03	0.27	0.04 \ 17.3
Log-likelihood	4741.05	4561.72	4842.25	4632.19	4643.38
p-value of S_n	0.000	0.000	0.000	0.000	0.000
p-value of T_n	0.002	0.000	0.089	0.013	0.070

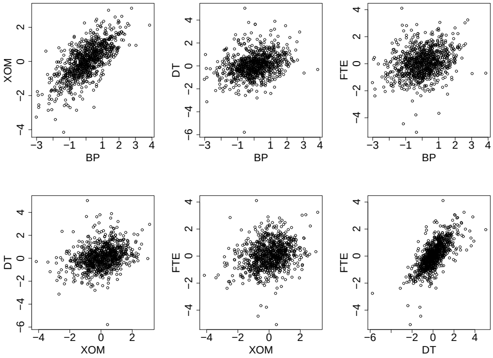
Applications

Equity returns

- Four stocks, two from oil sector and two from telecom.
- Daily data from 14.08.2003 – 29.12.2006 (852 observations).
- Log-returns are filtered by a GARCH-NIG filter and converted to uniform pseudo-observations before further modelling.

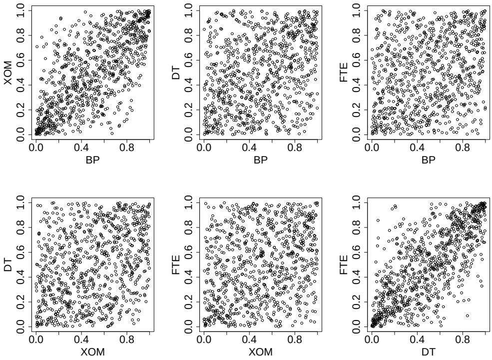
Applications

Equity returns



Applications

Equity returns



Applications

Equity returns

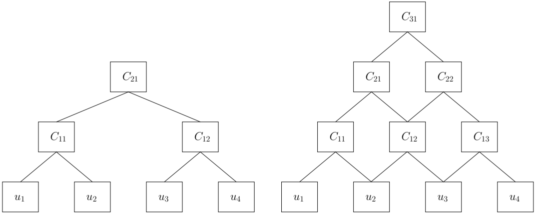
- Choose variable ordering based on pairwise Kendall taus:

Firm	XOM	DT	FTE
BP	0.45	0.19	0.20
XOM		0.23	0.17
DT			0.48

Applications

Equity returns

- We compare:



- Copulae at level one in both constructions are those corresponding to the largest pairwise Kendall taus.

Applications

Equity returns

Parameter	*NAC*	*PCC*	
	Frank	*Frank*	*Student*
$\theta_{11} \setminus \nu_{11}$	5.57	5.56	0.70 \ 13.8
$\theta_{12} \setminus \nu_{12}$	6.34	1.89	0.32 \ 134.5
$\theta_{13} \setminus \nu_{13}$	-	6.32	0.73 \ 6.4
$\theta_{21} \setminus \nu_{21}$	1.78	0.91	0.14 \ 12.0
$\theta_{22} \setminus \nu_{22}$	-	0.30	0.06 \ 20.6
$\theta_{31} \setminus \nu_{31}$	-	0.33	0.07 \ 17.8
Log-likelihood	616.45	618.63	668.49
p-value of S_n	0.006	0.008	0.410
p-value of T_n	0.385	0.385	0.697

Applications

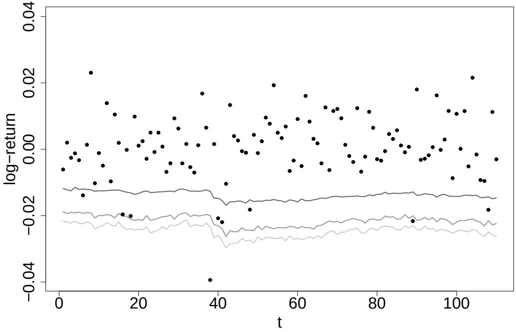
Equity returns

- Increasing complexity – risk of overfit
- To examine this we validate the GARCH-NIG PCC model out-of-sample
- We put together an equally-weighted portfolio of the four stocks considered
- The estimated model is used to forecast the 1-day VaR for each day in the period 30.12.2006–11.06.2007

Applications

Equity returns

- ▷ Out-of-sample VaR $(0.5\%, 1\%, 5\%)$:



Applications

Equity returns

- We use the likelihood ratio statistic by Kupiec (1995) to compute violation p-values:

Significance level	0.005	0.01	0.05
Observed	1	2	9
Expected	0.55	1.1	5.5
P-value	0.13	0.44	0.16

- PCC seems to work well out-of-sample and overfit is not a problem in this case.

Summary

- The NACs have two important restrictions
  - The level of dependence must decrease with the level of nesting
  - The involved copulae have to be Archimedean
- The PCCs are in general more flexible and computationally more efficient than the NACS both for simulation and in particular parameter estimation
- The likelihood and goodness-of-fit test judged in favour of PCC in our applications
- The PCC seems not to overfit our equity returns
- R-library *CopulaLib* under development – estimation, simulation, model selection and evaluation for copulae, including higher-dimensional constructions

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2008). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics 42*. In press.

Bedford, T. and R. M. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence 32*, 245–268.

Bedford, T. and R. M. Cooke (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics 30*, 1031–1068.

Berg, D. (2007). Copula goodness-of-fit testing: an overview and power comparison. Technical report, University of Oslo. Statistical research report no. 5, ISSN 0806-3842.

Embrechts, P., F. Lindskog, and A. McNeil (2003). *Modelling dependence with copulas and applications to risk management. In Handbook of heavy tailed distributions in finance*. Elsevier/North-Holland, Amsterdam.

Genest, C., B. Rémillard, and D. Beaudoin (2008). Omnibus goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics 42*. In press.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of derivatives 2*, 173–184.

Kurowicka, D. and R. M. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. New York: Wiley.

McNeil, A., R. Frey, and P. Embrechts (2006). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

McNeil, A. J. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation 78*, 567–581.

Savu, C. and M. Trede (2006). Hierarchical Archimedean Copulas. Technical report, Institute of Econometrics, University of Münster.

Whelan, N. (2004). Sampling from Archimedean copulas. *Quantitative Finance 4*, 339–352.

Contact details

Daniel Berg
E-mail: daniel@danielberg.no
Web: www.danielberg.no

Kjersti Aas
E-mail: kjersti.aas@nr.no
Web: www.nr.no/∼kjersti