Exogenous shock models in high dimensions

CIRM Workshop, February 22, 2016

J.-F. Mai, S. Schenk, M. Scherer
Technische Universität München

Research questions

(a) For which $g_1, \ldots, g_d$ is the following a multivariate distribution function

$$(x_1, \ldots, x_d) \mapsto \prod_{k=1}^{d} g_k(x_{[k]}),$$

where $x_{[1]} \leq x_{[2]} \leq \ldots \leq x_{[d]}$ is the ordered list of $x_1, \ldots, x_d \in \mathbb{R}$?

(b) What about stochastic representations?

(c) Are there interesting examples / applications?

Analytical characterizations

- First observation:

$$(X_1, \dots, X_d) \sim \prod_{k=1}^{d} g_k(x_{[k]}) \quad \Rightarrow \quad g_1(x) = \mathbb{P}(X_k \le x), \quad \forall k = 1, \dots, d.$$

- W.l.o.g. consider a copula framework, i.e. $x_1, \dots, x_d \in [0,1]$ and $g_1(x) = x$.

- For which $g_2, \dots, g_d$ is the following a copula?

$$C(x_1, \dots, x_d) := x_{[1]} \cdot \prod_{k=2}^{d} g_k(x_{[k]})$$

- If C is a copula and g_1 any univariate d.f. (resp. $\bar{g}_1$ any univariate s.f.),

$$C(g_1(x_1), \dots, g_1(x_d)) := g_1(x_{[1]}) \cdot \prod_{k=2}^{d} g_k \circ g_1(x_{[k]}) \quad \Big(\text{resp. } C(\bar{g}_1(x_1), \dots, \bar{g}_1(x_d))\Big)$$

is a multivariate distribution function (resp. survival function).

Some examples

- **Independence:** $g_k \equiv \mathrm{id}_{[0,1]}$ yields

$$C(x_1, \ldots, x_d) = \prod_{k=1}^{d} x_k.$$

- **Exchangeable Marshall–Olkin copulas** have $g_k(x) := x^{a_{k-1}}$, so

$$C(x_1, \ldots, x_d) = \prod_{k=1}^{d} x_{[k]}^{a_{k-1}} \quad \text{for } d\text{-monotone } a_0 = 1, a_1, \ldots, a_{d-1} \geq 0.$$

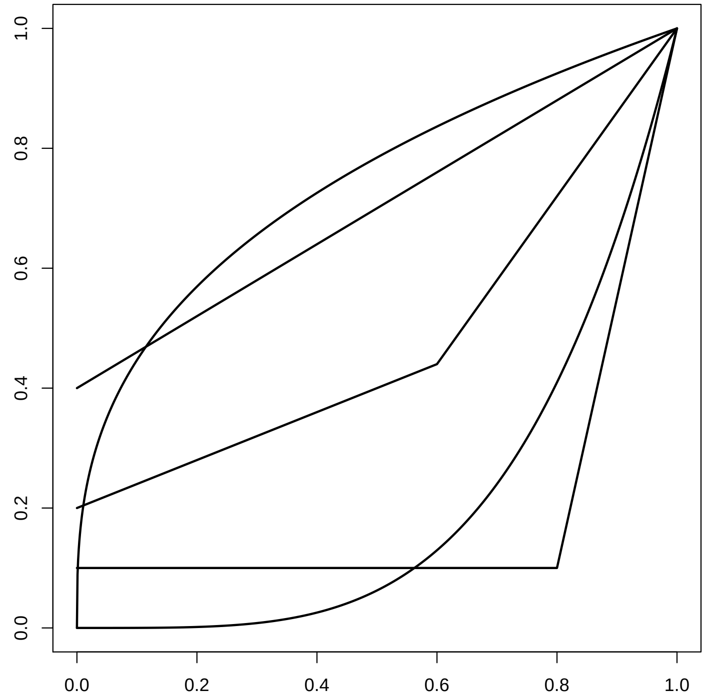
- **[Durante et al. (2007)]** study copulas with $g_k(x) := g(x), k = 2, \ldots, d$, i.e.

$$C(x_1, \ldots, x_d) = x_{[1]} \prod_{k=2}^{d} g(x_{[k]}).$$

Analytical characterizations

Set $\mathcal{D}$ of univariate d.f.'s:

$$\mathcal{D} := \{F : [0,1] \to [0,1] : \text{ continuous, non-decreasing, strictly positive on } (0,1], F(1) = 1\}$$
$$= \{\text{d.f.'s of absolutely continuous r.v.'s on } (0,1) \text{ with possibly an extra atom at } 0\}.$$



Analytical characterizations

Bivariate case:

- For $g_1, g_2 : [0,1] \to [0,1]$, $g_1(1) = g_2(1) = 1$, consider
$$C(x_1, x_2) := g_1(x_{[1]})\, g_2(x_{[2]}).$$
- **[Durante et al. (2008)]**: C is a copula if and only if $g_1 \equiv \mathrm{id}_{[0,1]}$ and

  (i) $g_2(y) - g_2(x) \geq 0, \quad x, y \in [0,1], x < y.$

  (ii) $g_1(y)\, g_2(y) - 2\, g_1(x)\, g_2(y) + g_1(x)\, g_2(x) \geq 0, \quad x, y \in [0,1], x < y.$
- **Analytical interpretation:**

  $\rightarrow$ $(i) \Leftrightarrow g_2$ is increasing.

  $\rightarrow$ $(ii) \Leftrightarrow g_2$ is continuous, strictly positive on $(0, 1]$, and g_1/g_2 is increasing.
- **Remark**: $g_1/g_2 \in \mathcal{D}$.

Analytical characterizations

Necessary condition:

- Assume that $(U_1,\dots,U_d) \sim x_{[1]} \cdot \prod_{k=2}^{d} g_k(x_{[k]})$. Then

$$g_j(x) = \frac{x \cdot \prod_{k=2}^{j} g_k(x)}{x \cdot \prod_{k=2}^{j-1} g_k(x)}$$
$$= \frac{\mathbb{P}(\max\{U_1,\dots,U_j\} \le x)}{\mathbb{P}(\max\{U_1,\dots,U_{j-1}\} \le x)} \quad \Rightarrow \quad x \cdot \prod_{i=2}^{j} g_i(x) = \mathbb{P}(\max\{U_1,\dots,U_j\} \le x).$$

- So it is necessary that $x \cdot \prod_{i=2}^{j} g_i(x) \in \mathcal{D}$ for all $j = 2,\dots,d$.

- But we need more.

Analytical characterizations

Theorem 1: The following statements are equivalent:

(i) $C(x_1,\ldots,x_d) = x_{[1]} \cdot \prod_{k=2}^{d} g_k(x_{[k]})$ is a copula.

(ii) For all $0 < x < y \leq 1$, $(k,j) \in \mathbb{N}_0 \times \mathbb{N}$ with $k + j \leq d$ we have

$$\sum_{i=0}^{j} \binom{j}{i} (-1)^i \prod_{\ell=1}^{i} g_{\ell+k}(x) \prod_{\ell=i+1}^{j} g_{\ell+k}(y) \geq 0.$$

(iii) For $m = 1,\ldots,d$ we have $G_m \in \mathcal{D}$, where

$$G_m(x) := \prod_{i=0}^{m-1} g_{d-m+1+i}^{(-1)^i \binom{m-1}{i}}(x), \quad x \in [0,1] \quad \text{(with } x = 0 \text{ as limit and } g_1(x) = x\text{)}.$$

Research questions

(a) For which $g_1, \ldots, g_d$ is the following a multivariate distribution function

$$(x_1, \ldots, x_d) \mapsto \prod_{k=1}^{d} g_k(x_{[k]}),$$

where $x_{[1]} \leq x_{[2]} \leq \ldots \leq x_{[d]}$ is the ordered list of $x_1, \ldots, x_d \in \mathbb{R}$?

(b) What about stochastic representations?

(c) Are there interesting examples / applications?

Stochastic representations
Exogenous shock model representation

Theorem 2:

Let (i)–(iii) of Theorem 1 be valid. For each $E \subset \{1, \ldots, d\}$ consider a r.v. Z_E s.t.

- the distribution function of Z_E equals $G_{|E|}$,
- all Z_E are independent.

With an arbitrary univariate survival function $\bar{g}_1$, define the random vector $(X_1, \ldots, X_d)$ by the following "exogenous shock model":

$$X_k := \min\{\bar{g}_1^{-1}(Z_E) \,:\, k \in E\}, \quad k = 1, \ldots, d.$$

$\Rightarrow$ The survival function of $(X_1, \ldots, X_d)$ is given by $C(\bar{g}_1(x_1), \ldots, \bar{g}_1(x_d))$.

Remark: Conversely, any choice for the laws $G_1, \ldots, G_d \in \mathcal{D}$ of the Z_E uniquely determines associated functions $g_2, \ldots, g_d$.

Stochastic representations
Exogenous shock model representation

With $\bar{g}_1$ a continuous survival function on $(0, \infty)$ consider

$$X_k := \min\{\bar{g}_1^{-1}(Z_E) \,:\, k \in E\}, \quad k = 1, \ldots, d.$$

- The Z_E are abs. continuous on $(0, 1)$ with potential extra atom at 0.
- The $\bar{g}_1^{-1}(Z_E)$ are abs. continuous on $(0, \infty)$ with possible extra atom at ∞.
- $\bar{g}_1^{-1}(Z_E)$ = arrival time point of exogenous shock killing all components in E.
- X_k = first time point when a shock kills component k.
- **Example:** $\bar{g}_1(x) = \exp(-x)$ and $G_m(x) = x^{\lambda_m}$ for some $\lambda_m > 0$
  $\Rightarrow \bar{g}_1^{-1}(Z_E)$ exponential with rate $\lambda_{|E|}$
  $\Rightarrow g_k(x) = x^{a_k}$ for special sequences $(a_2, \ldots, a_d)$.

Stochastic representations

Exogenous shock model representation

Schematic overview: dimension = 2

g_1 g_2 $\sim$ $Z_E, |E| = 1$

g_1/g_2 $\sim$ $Z_E, |E| = 2$

Stochastic representations

Exogenous shock model representation

Schematic overview: dimension = 3

g_1	g_2	g_3	$\sim$	$Z_E, \lvert E\rvert = 1$
g_1/g_2	g_2/g_3	$\sim$	$Z_E, \lvert E\rvert = 2$	
$g_1\,g_3/g_2^2$	$\sim$	$Z_E, \lvert E\rvert = 3$		

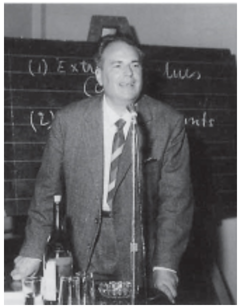
Stochastic representations
Exogenous shock model representation

Schematic overview: dimension = d

g_1	g_2	g_3	$\dots$	g_d	$\lvert E\rvert = 1$
g_1/g_2	g_2/g_3	$\dots$	g_{d-1}/g_d	$\lvert E\rvert = 2$	
$g_1\,g_3/g_2^2$	$\dots$	$g_{d-2}\,g_d/g_{d-1}^2$	$\lvert E\rvert = 3$		
$\vdots$					
$\prod_{i=0}^{d-1} g_{i+1}^{\binom{d-1}{i}(-1)^i}$	$\lvert E\rvert = d$				

Challenge: Dimensionality reduction?

Stochastic representations
de Finetti representation



Theorem 3:

Let $H = \{H_t\}_{t \in [t_0, t_1]}$ be an increasing additive process with $H_{t_0} = 0$ and $H_{t_1} = \infty$, i.e. $\Psi_t(x) := -\log\big(\mathbb{E}[\exp(-x\,H_t)]\big)$ defines a family of Bernstein functions $\{\Psi_t\}_{t \geq 0}$.

- Draw one sample $\mathcal{F}(\omega)$ from the random d.f. $\mathcal{F} := \left\{1 - e^{-H_t}\right\}_{t \in [t_0, t_1]}$.
- Let $X_1, X_2, \ldots \in (t_0, t_1)$ be i.i.d. random variables drawn from $\mathcal{F}(\omega)$.

⇒ The univariate survival function of X_k is given by

$$\bar{g}_1(x) := \mathbb{P}(X_k > x) = \mathbb{E}\left[e^{-H_x}\right] = \exp\big(-\Psi_x(1)\big), \quad x \geq 0.$$

⇒ The survival copula of $(X_1, \ldots, X_d)$ has form C with $g_2, \ldots, g_d$ given by

$$g_k(x) := \exp\Big(-\Psi_{\bar{g}_1^{-1}(x)}(k) + \Psi_{\bar{g}_1^{-1}(u)}(k-1)\Big), \quad k = 2, \ldots, d.$$

Research questions

(a) For which $g_1, \ldots, g_d$ is the following a multivariate distribution function

$$(x_1, \ldots, x_d) \mapsto \prod_{k=1}^{d} g_k(x_{[k]}),$$

where $x_{[1]} \leq x_{[2]} \leq \ldots \leq x_{[d]}$ is the ordered list of $x_1, \ldots, x_d \in \mathbb{R}$?

(b) What about stochastic representations?

(c) Are there interesting examples / applications?

Examples
Marshall–Olkin copulas

- Let $H = \{H_t\}_{t \geq 0}$ be a Lévy subordinator. This means

  $\rightarrow$ $\Psi_t = t \cdot \Psi_1$ for some fixed Bernstein function Ψ_1.

  $\rightarrow$ $g_k(x) = x^{a_k}$ for a sequence $(1, a_2, \dots, a_d)$ being d-monotone.

  $\rightarrow$ C in Th. 3 is the survival copula of an **exchangeable Marshall–Olkin law**.

- Th. 2 with $\bar{g}_1(x) = \exp(-x)$ yields the [Marshall–Olkin (1967)] representation, in which arrival times of exogenous shocks are exponentially distributed.

- **Proposition:**

  $C(x_1, \dots, x_d) = x_{[1]} \cdot \prod_{k=2}^{d} g_k(x_{[k]})$ is an extreme-value copula

  $\Leftrightarrow$ it is the survival copula of an exchangeable Marshall-Olkin law.

Examples
Sato-frailty copulas

- Let $H = \{H_t\}_{t\geq 0}$ be an increasing Sato process. This means

  → $\Psi_t(x) = \Psi_{\text{sd}}(x\,t^H)$ for some $H > 0$ and some fixed, self-decomposable Bernstein function $\Psi_{\text{sd}} = \Psi_1$.

  → With $\varphi := \exp(-\Psi_{\text{sd}})$ denoting the Laplace transform of the associated self-decomposable law on $(0, \infty)$

$$C(x_1, \ldots, x_d) = C_\varphi(x_1, \ldots, x_d) = x_{[1]} \cdot \prod_{k=2}^{d} g_{k,\varphi}(x_{[k]}),$$

$$\text{with } g_{k,\varphi}(x) = \frac{\varphi(k\,\varphi^{-1}(x))}{\varphi((k-1)\,\varphi^{-1}(x))}.$$

- **Theorem 4: ("Kimberling-type" copula-characterization of SD laws)**

  C_φ is a copula for all $d \geq 2$
  $\Leftrightarrow$ φ is the Laplace transform of a self-decomposable law on $(0, \infty)$.

Examples

Sato-frailty copula $g_k(u) = \varphi(k\,\varphi^{-1}(u))/\varphi((k-1)\,\varphi^{-1}(u))$

- Laplace exponent of Gamma-distributed r.v.:

$$\Psi_{\mathsf{sd}}(x) = \beta\,\log(1 + \frac{x}{\eta}), \quad x, \beta, \eta > 0.$$

- There is a (unique) increasing Sato process $\{H_t\}_{t\geq 0}$ s.t.

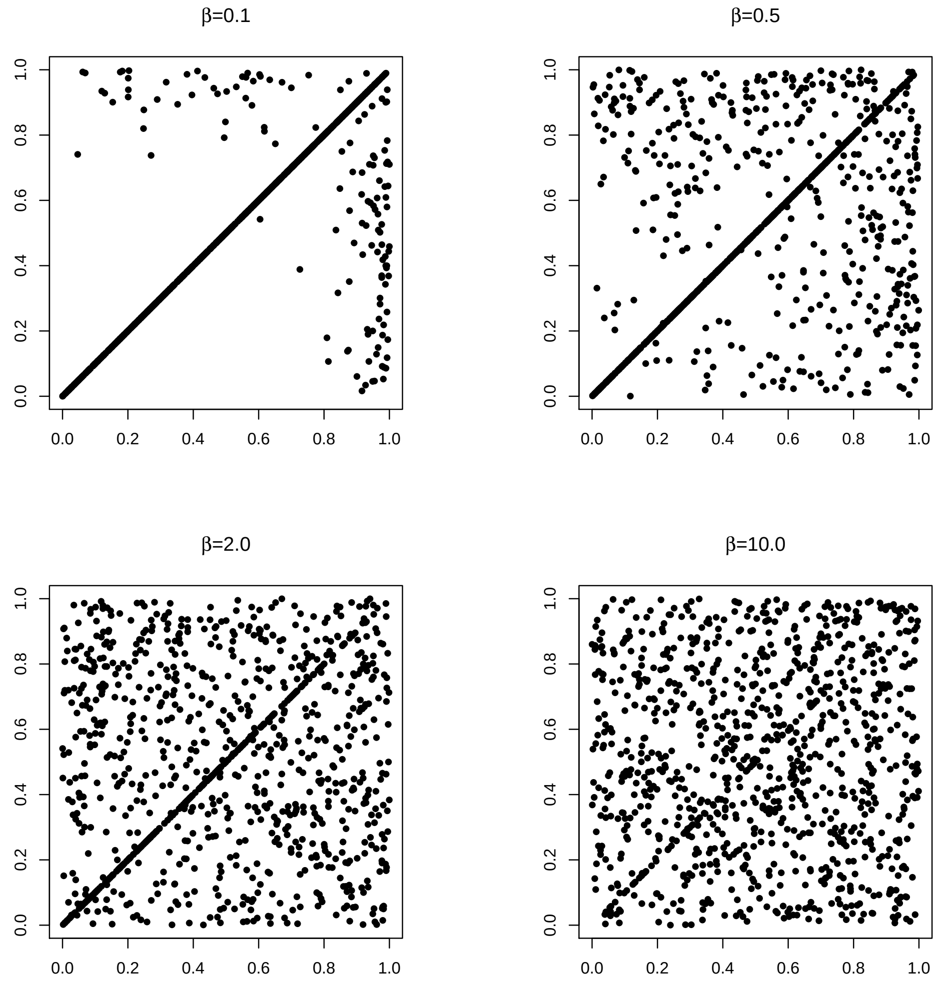
$$\varphi(x) := \mathbb{E}[\exp(-x\,H_1)] = \exp(-\Psi_{\mathsf{sd}}(x)), \quad x \geq 0.$$

- The corresponding bivariate Sato-frailty copula C_φ is

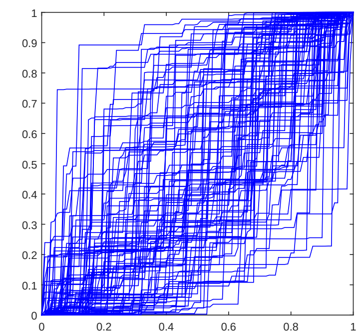
$$C_\varphi(x_1, x_2) = \frac{x_{[1]}}{(2 - x_{[2]}^{1/\beta})^\beta}.$$

Examples

Sato-frailty copula $C_{\varphi}(x_1, x_2) = x_{[1]}/(2 - x_{[2]}^{1/\beta})^{\beta}$



Examples
The Dirichlet copula



- For $c > 0$, let $H^{(c)} = \{H_t^{(c)}\}_{t\in[0,1]}$ be specified via $\{\Psi_t^{(c)}\}_{t\in(0,1)}$ as

$$\Psi_t^{(c)}(x) := \int_0^\infty (1 - e^{-x\,u})\, \frac{e^{u\,c\,(1-t)} - e^{-u\,c}}{u\,(1 - e^{-u})}\, \mathrm{d}u.$$

→ The random d.f. $\mathcal{F} = \mathcal{F}_c = \left\{1 - e^{-H_t^{(c)}}\right\}_{t\in[0,1]}$ is a **Dirichlet process**.

→ The resulting copula in Th. 3 is called **Dirichlet copula**:

$$C(x_1, \dots, x_d) = C_c(x_1, \dots, x_d) = x_{[1]} \cdot \prod_{k=2} \frac{c\,x_{[k]} + k - 1}{c + k - 1}.$$

→ Kendall's τ, Spearman's ρ_S, and tail-dependence are

$$\tau = \frac{2\,c + 3}{3(c+1)^2}, \qquad \rho_S = \frac{1}{c+1}, \qquad \text{LTD}_C = \text{UTD}_C = \frac{1}{c+1}.$$

→ **Theorem 5: (Radially symmetric exogenous shock models)**

The copula $C(x_1, \dots, x_d) = x_{[1]} \cdot \prod_{k=2}^d g_k(x_{[k]})$ is radially symmetric
$\Leftrightarrow$ $C = C_c$ is a Dirichlet copula for some $c \in [0, \infty]$.

Examples
The Dirichlet copula

- Dirichlet process is known in non-parametric Bayesian statistics.
- It is an interesting model for a distorted random number generator.
- **[Ferguson (1973)]**: Simulation $(X_1, \ldots, X_d)$ from Dirichlet copula C_c is easy:

→ Simulate $X_1 \sim \mathcal{U}(0, 1)$.

→ For $k = 2, \ldots, d$ simulate X_k as follows:

(i) Simulate discrete random variable $N \in \{1, \ldots, k\}$ with

$$\mathbb{P}(N = i) = \frac{1}{c + k - 1}, \quad i = 1, \ldots, k - 1, \quad \mathbb{P}(N = k) = \frac{c}{c + k - 1}.$$

(ii) If $N = k$ simulate $X_k \sim \mathcal{U}(0, 1)$, else set $X_k := X_N$.

References

[Mai, Schenk, Scherer (2016a)]:
"Exchangeable exogenous shock models", Bernoulli 22, 1278–1299.

[Mai, Schenk, Scherer (2016b)]:
"Two novel characterizations of self-decomposability on the half-line",
Journal of Theoretical Probability, forthcoming.

[Mai, Schenk, Scherer (2016c)]:
"Analyzing model robustness via a distortion of the stochastic root: a Dirichlet prior approach", working paper.