Model Selection of Regular Vine Copulas

Lutz F. Gruber
Joint work with Claudia Czado

Technische Universität München

December 15, 2013
Agenda

Dependence Modeling is Important

Copulas & Regular Vine Copulas

Comparison of Model Selection Procedures

Application Study: Pricing of an Exotic Financial Derivative

Conclusions & Outlook
Suppose we hold zero bonds with a Bernoulli payout distribution.

- The zero bonds are due at time $t + 1$.
- With a 90% probability, a bond repays at 100%.
- With a 10% probability, a bond defaults not repaying anything.
- The expected payout of each bond is 90%.
- Our portfolio holds 1,000 such zero bonds.

**What is the payout distribution of this portfolio?**
The Dependence Structure Determines the Portfolio Payout Distribution

Independence

Perfect Dependence

Everything in between and some more
Agenda

Dependence Modeling is Important

Copulas & Regular Vine Copulas

Comparison of Model Selection Procedures

Application Study: Pricing of an Exotic Financial Derivative

Conclusions & Outlook
Copulas Are Comprehensive Multivariate Dependence Models

- A copula $C$ is an $n$-variate probability distribution with $U(0, 1)$ margins. Any multivariate distribution $F$ can be expressed in terms of its marginal distributions $F_1, \ldots, F_n$ and a copula $C$:

$$F(x) = C(F_1(x_1), \ldots, F_n(x_n)), x \in \mathbb{R}^n.$$ 

- Examples:

- Copulas can describe dependence characteristics such as strength of association, symmetries vs. asymmetries, tail-dependencies, and any other distributional characteristic.
Copulas Separate Dependence Modeling from Marginal Modeling

- Multivariate data can be modeled with
  - Multivariate models
  - Univariate models for the margins and a copula for the dependence structure
- Major strengths of the copula approach are:
  - Well-established models for the margins can be reused
  - Highly flexible because different marginal models can be combined
  - Separation of marginal and dependence models is theoretically justified [Skl59]
- Literature includes: [Skl59, Joe01, Nel06, KC06, KJ10].
Constructing a Multivariate Copula with Regular Vine Pair Copula
Constructions Follows Lego’s Building Block Paradigm

Target Copula

Non-simplified PCC
Building blocks: complex!

Regular Vine Copula
Building blocks: tractable!

Parametric Copulas
Bad fit!
Regular Vine Copulas Are as Easy as 1-2-3! [BC01]

PCC Building Plan

PCC Building Blocks

PCC Copula Density

Regular Vine $\mathcal{V}$

$\mathcal{V} = (T_1, \ldots, T_{n-1})$

Pair Copulas $B_{\mathcal{V}}(\theta_{\mathcal{V}})$

$B_{\mathcal{V}}(\theta_{\mathcal{V}}) = (B_{T_k}(\theta_{T_k}), T_k \in \mathcal{V})$

$= (c_{i(e);j(e);\cdots;\theta_e}^{B_e}, e \in E_k, T_k \in \mathcal{V})$

$C_{12345} = C_{14} \cdot C_{15} \cdot C_{24} \cdot C_{34}$

$\cdot C_{12:4} \cdot C_{13:4} \cdot C_{45:1}$

$\cdot C_{23:14} \cdot C_{35:14}$

$\cdot C_{25:134}$
There Is a Huge Number of Regular Vine Copulas

<table>
<thead>
<tr>
<th>Dimension $n$</th>
<th>#Regular Vines $^1$</th>
<th>#Regular Vine Copulas $^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1,029</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>2,823,576</td>
</tr>
<tr>
<td>5</td>
<td>480</td>
<td>1.3559e+11</td>
</tr>
<tr>
<td>6</td>
<td>23,040</td>
<td>1.0938e+17</td>
</tr>
<tr>
<td>7</td>
<td>2,580,480</td>
<td>1.4413e+24</td>
</tr>
<tr>
<td>8</td>
<td>660,602,880</td>
<td>3.0387e+32</td>
</tr>
<tr>
<td>9</td>
<td>3.8051e+11</td>
<td>1.0090e+42</td>
</tr>
<tr>
<td>10</td>
<td>4.8705e+14</td>
<td>5.2118e+52</td>
</tr>
</tbody>
</table>

$^1$ See [MNCK09] for details.

$^2$ This assumes $|B| = 7$ candidate pair copula families.
Agenda

Dependence Modeling is Important

Copulas & Regular Vine Copulas

Comparison of Model Selection Procedures

Application Study: Pricing of an Exotic Financial Derivative

Conclusions & Outlook
## Three Approaches to Finding Regular Vine Copulas to Observed Data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Priors</td>
<td>Frequentist level-by-level</td>
<td>Bayesian level-by-level</td>
<td>Bayesian all levels jointly</td>
</tr>
<tr>
<td></td>
<td>$T_k \sim \text{Uniform}(\cdot)$</td>
<td>$\theta_k \mid T_k, B_k \sim \text{Uniform}(\cdot)$</td>
<td>$\mathcal{V} \sim \text{Uniform}(\cdot)$</td>
</tr>
<tr>
<td>$B_k \mid T_k \sim \exp(-\lambda d_k)$</td>
<td>$\theta_v \mid \mathcal{V}, B_v \sim \text{Uniform}(\cdot)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta_v \mid \mathcal{V}, B_v \sim \text{Uniform}(\cdot)$</td>
<td>$B_v \mid \mathcal{V} \sim \exp(-\lambda d_v)$</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>Select maximum spanning trees with absolute value Kendall’s $\tau$ weights</td>
<td>Reversible Jump MCMC</td>
<td>Reversible Jump MCMC</td>
</tr>
<tr>
<td>Performance</td>
<td>Good</td>
<td>Better</td>
<td>Best</td>
</tr>
</tbody>
</table>
Get the Proposals Right—Or You Risk Rejection! [GC12, GC13]

- Use a mixture of two mutually exclusive, collectively exhaustive algorithms for the between models move:
  1. Algorithm FAM only updates the pair copula families;
  2. Algorithm TREE updates the tree structure and the pair copula families (TREE) and guarantees that the current tree is not proposed.
- Draw proposal trees from a uniform distribution over all trees allowed by the proximity condition (only TREE).
- Compute the maximum likelihood estimates of the parameters of all candidate pair copula families.
- Draw the proposal pair copulas from a discrete distribution with weights proportional to the copulas' maximum likelihoods.
- Draw the proposal parameters from a mixture of truncated normal distributions with varying variances, centered at the MLE.
We Test the Three Procedures with 10x500 Simulated Observations from Four Given Vine Copula Distributions

---

This slide shows only Model 1

Lutz F. Gruber (TUM)
Joint Model Selection Clearly Beats Sequential Selection! [GC13]

Average log likelihoods of the estimated models as percentages of the true models’ log likelihoods:

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gruber II</td>
<td>98.4</td>
<td>98.6</td>
<td>96.5</td>
<td>99.2</td>
<td>98.2</td>
</tr>
<tr>
<td>Gruber I</td>
<td>88.6</td>
<td>80.9</td>
<td>84.9</td>
<td>99.9</td>
<td>88.6</td>
</tr>
<tr>
<td>Dißmann</td>
<td>86.6</td>
<td>75.1</td>
<td>78.9</td>
<td>99.7</td>
<td>85.1</td>
</tr>
</tbody>
</table>

- Model 4 is the multivariate Gaussian copula, which can be obtained through a pair copula construction with any regular vine.
- Because the selection of the regular vine does not matter here, all model selection procedures perform uniformly well.
Joint Model Selection Clearly Beats Sequential Selection! [GC13]
Agenda

Dependence Modeling is Important

Copulas & Regular Vine Copulas

Comparison of Model Selection Procedures

Application Study: Pricing of an Exotic Financial Derivative

Conclusions & Outlook
Have Weekly Historical Prices of 9 DJIA Components (2000–2012)

<table>
<thead>
<tr>
<th>Sector</th>
<th>Company Name</th>
<th>Ticker Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Services</td>
<td>McDonald’s Corp.</td>
<td>MCD</td>
</tr>
<tr>
<td></td>
<td>The Walt Disney Company</td>
<td>DIS</td>
</tr>
<tr>
<td></td>
<td>Wal-Mart Stores Inc.</td>
<td>WMT</td>
</tr>
<tr>
<td>Basic Materials</td>
<td>Alcoa Inc.</td>
<td>AA</td>
</tr>
<tr>
<td></td>
<td>E. I. du Pont de Nemours and Company</td>
<td>DD</td>
</tr>
<tr>
<td></td>
<td>Exxon Mobil Corporation</td>
<td>XOM</td>
</tr>
<tr>
<td>Technology</td>
<td>International Business Machines Corporation</td>
<td>IBM</td>
</tr>
<tr>
<td></td>
<td>Intel Corporation</td>
<td>INTC</td>
</tr>
<tr>
<td></td>
<td>Cisco Systems, Inc.</td>
<td>CSCO</td>
</tr>
</tbody>
</table>
Objective: To Estimate the Expected Payout of a Basket Option

\[ P_t := \left( \sum_{i \in I} 1\{u_{i,t} < B\} \right)^3 \]

- Option term: 1 week
- Underlyings: \( I = \{MCD, DIS, WMI, AA, DD, XOM, IBM, INTC, CSCO\} \)
- Normalized returns: \( u_{i,t} \) are the quantiles of the ARMA-GARCH innovations, hence approximately i.i.d. \( U(0, 1) \) [FY01]
- The payout of the option depends exclusively on the dependence characteristics of the underlying assets
Vine Copulas Capture Dependence Characteristics of Financial Data Better than the Student’s t Copula [GC12]
Agenda

Dependence Modeling is Important

Copulas & Regular Vine Copulas

Comparison of Model Selection Procedures

Application Study: Pricing of an Exotic Financial Derivative

Conclusions & Outlook
Conclusions & Outlook on Further Research

- Reversible jump Markov chain Monte Carlo-based procedures allow model selection of regular vine copulas.
- Algorithms that estimate all trees jointly produce better models than stepwise estimation procedures.
- Stepwise level-by-level procedures produce biased estimates that favor models with strong unconditional dependencies and may ignore relevant conditional dependencies.
- **TODO** Investigate the effect of prior choices.
- **TODO** Modify the algorithm to implement different acceptance/rejection mechanisms, e.g., simulated annealing.
- **TODO** Vine Mixture Copulas (VMCs): use mixture of pair copula families to smoothen family selection, cover a wider range of dependence patterns.
References I


References II


A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publications de l'Institut de Statistique de l'Université de Paris **8** (1959), 229–231.
Agenda

Details of the Regular Vine Pair Copula Construction
The Regular Vine Specifies the Construction of the Copula Density

Definition [BC01]: A sequence of trees $\mathcal{V} = (T_1, \ldots, T_{n-1})$ is a regular vine on $n$ elements, if:

1. Tree $T_1 = (N_1, E_1)$ has nodes $N_1 = \{1, \ldots, n\}$ and edges $E_1$.

2. Trees $T_k$, $k = 2, \ldots, n-1$, have nodes $N_k = E_{k-1}$ and all edges $e \in E_k$ satisfy the proximity condition: nodes $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ may only be connected by an edge, if one of the $a_i$ equals one of the $b_i$.

3. An edge $e$ in tree $T_k$, $k = 1, \ldots, n-1$, represents a pair copula conditional on $k-1$ variables $D(e)$—these are the overlapping elements of one of the $a_i$ and $b_i$:

$$
c_{1:n}(u_{1:n}; \mathcal{V}, \mathcal{B}_\mathcal{V}(\theta_\mathcal{V})) = \prod_{T_k \in \mathcal{V}} \prod_{e \in E_k} c_{i(e),j(e);D(e)}(u_{i(e)|D(e)}, u_{j(e)|D(e)}; \mathcal{B}_e(\theta_e)), \text{ where}
$$

$$
u_{i(e)|D(e)} := C_{i(e)|D(e)}(u_{i(e)}; (T_1, \ldots, T_{k-1}), (\mathcal{B}_{T_1}(\theta_{T_1}), \ldots, \mathcal{B}_{T_{k-1}}(\theta_{T_{k-1}}))|u_{D(e)}).
$$