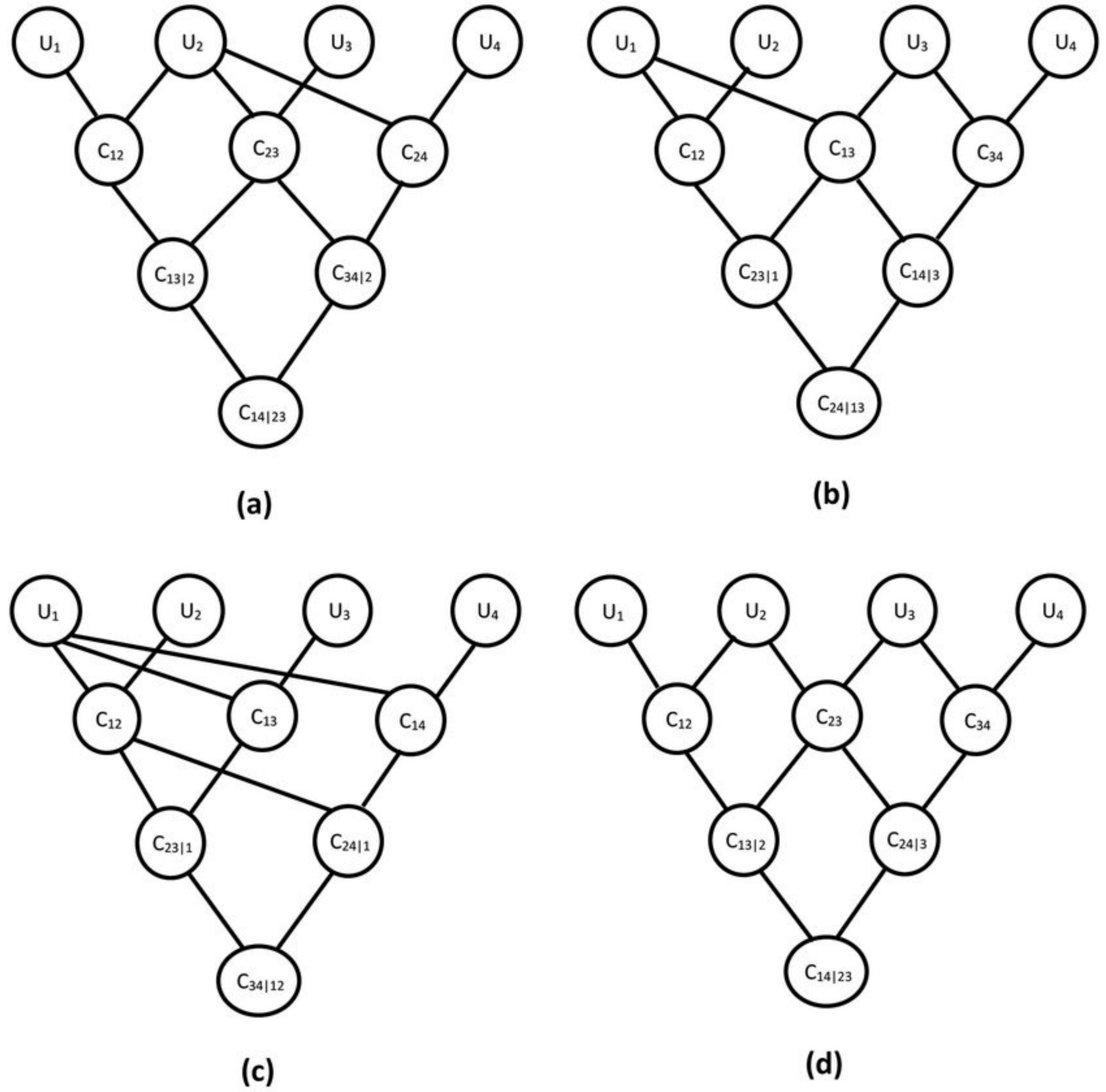
Introduction to Vine Copula for Statistical Arbitrage

HUDSON AND THAMES

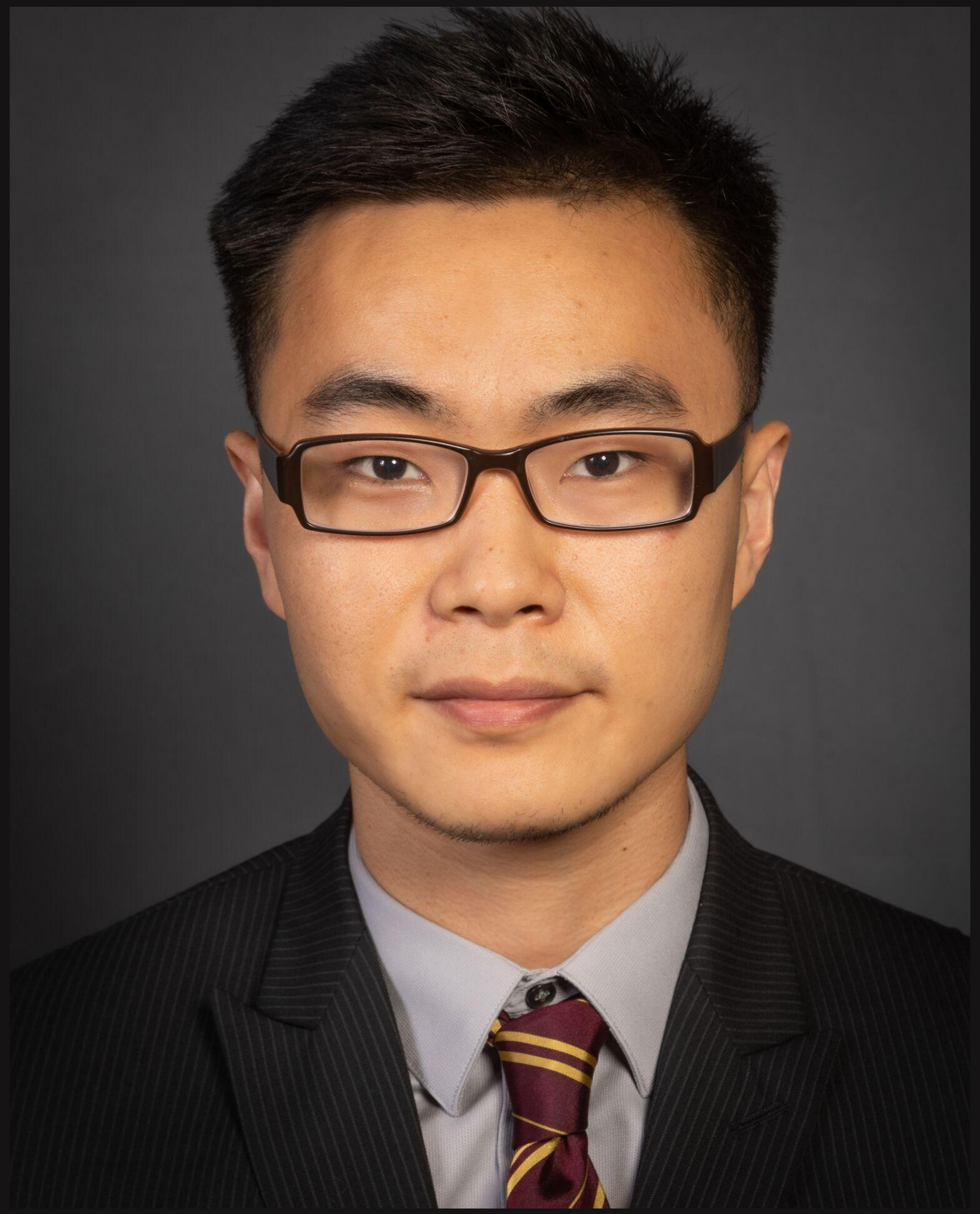
ARBITRAGELAB BY HUDSON & THAMES



Possible Vine Copula Structures, picture from [Pham et al. (2018)]

CONTENTS

ABOUT ME



- Ph.D. Candidate in Applied Mathematics at University of Delaware
- Researcher at Hudson & Thames

Research Interests:

- Stochastic Modeling
- Fokker-Planck and Broadwell Models
- Agent Based Modeling
- Numerical Methods for PDE
- Copula Modeling for Stats Arbitrage
- Applied Probability
- Stochastic Control

LinkedIn: https://www.linkedin.com/in/hansen-pei-0949691b3/

Twitter: @HansenPei

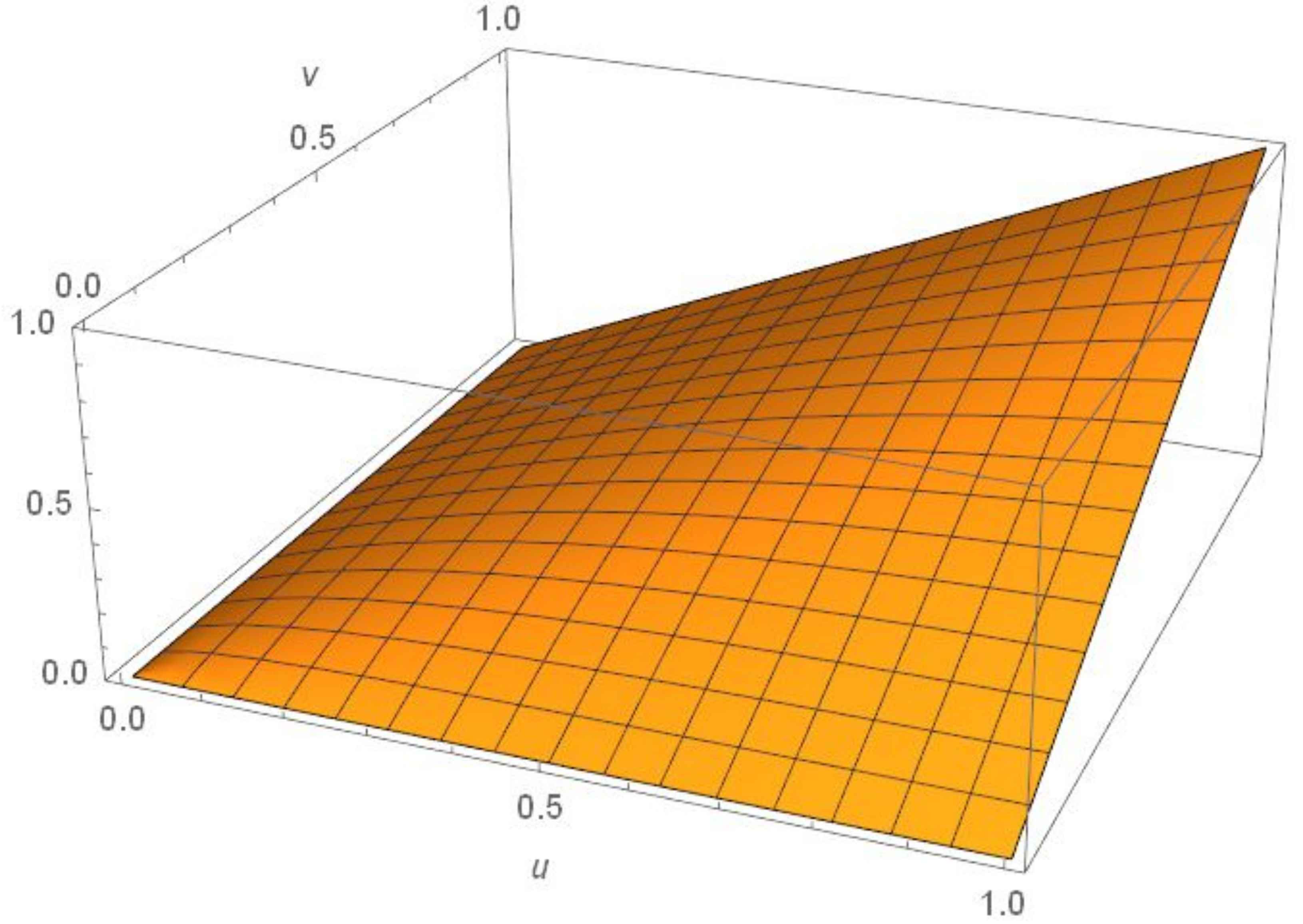
01. Key Math Concepts

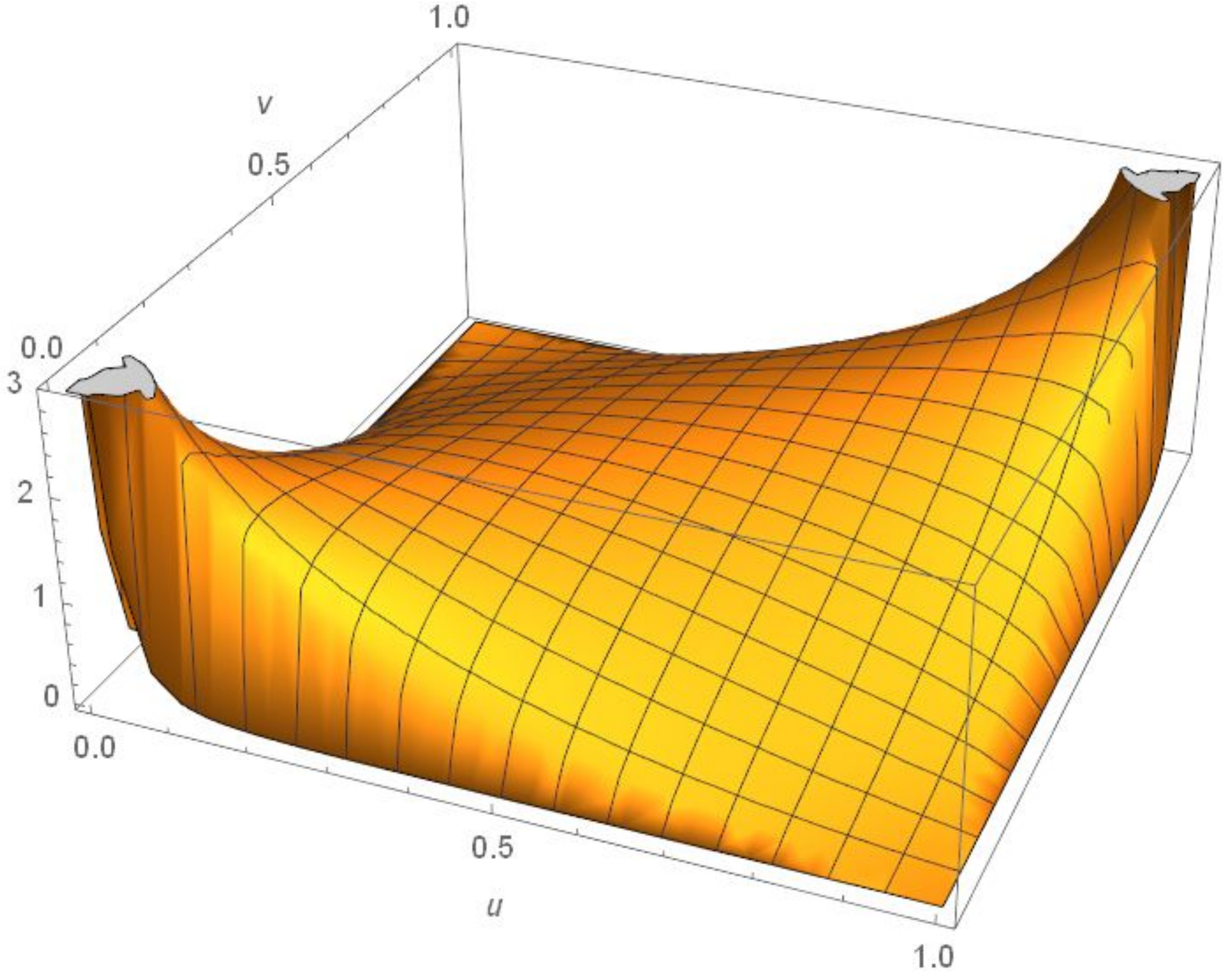
Review of Copula Concepts

1. Get data from 2 random variables, say, S_1, S_2
2. Map to their quantiles U_1, U_2 using marginal CDF and formulate

$$P(U_1 \leq u_1, U_2 \leq u_2) = C(u_1, u_2)$$

Copula: Joint Cumulative Density on Quantiles





Decompose Joint Probability Density Via Copula

Generic joint probability density:

$$p(u_1, u_2) = p(u_1|u_2)p(u_2)$$

Still the same, but different notation:

$$f(u_1, u_2) = f_{1|2}(u_1|u_2)f_2(u_2)$$

More random variables:

$$f(u_1, u_2, u_3) = f_{1|23}(u_1|u_2, u_3)f_{2|3}(u_2|u_3)f_3(u_3)$$

$$f(u_1, u_2, u_3) = f_{3|12}(u_3|u_1, u_2)f_{2|1}(u_2|u_1)f_1(u_1)$$

4 other ways

Decompose Joint Probability Density Via Copula

Say we have the joint probability density via structure

$$f(u_1, u_2, u_3) = f_{1|23}(u_1 | u_2, u_3) f_{2|3}(u_2 | u_3) f_3(u_3)$$

- Key component for statistical arbitrage
- Relative mispricing in a higher dimension

Conditional (cumulative) density

$$P(U_1 \leq u_1 | U_2 = u_2, U_3 = u_3)$$

$$= h(u_1 | u_2, u_3)$$

$$= \int_0^{u_1} f(u, u_2, u_3) du \,/\, \int_0^1 f(u, u_2, u_3) du$$

Decompose Joint Probability Density Via Copula

Copula Density:

$$c_{23}(u_2, u_3) = \frac{\partial^2 C_{23}(u_2, u_3)}{\partial u_2 \partial u_3}$$

Joint Density:

$$\frac{\partial^2 C_{23}(F_2(x_2), F_3(x_3))}{\partial x_2 \partial x_3}$$

$$f_{23}(x_2, x_3) = \boxed{c_{23}(F_2(x_2), F_3(x_3))} \cdot f_2(x_2) \cdot f_3(x_3)$$

Conditional Density:

$$f_{2|3}(x_2|x_3) = \frac{f_{23}(x_2, x_3)}{f_3(x_3)} = c_{23}(u_2, u_3) \cdot f_2(x_2)$$

02. Understand Vine Copula

Vine Copula: An Example

Let’s get slightly more complicated...

$$f(x_1, x_2, x_3) = f_{1|23}(x_1|x_2, x_3) \cdot f_{2|3}(x_2|x_3) \cdot f_3(x_3)$$

Vine Copula: An Example

Let’s get slightly more complicated...

$$f(x_1, x_2, x_3) = f_{1|23}(x_1|x_2, x_3) \cdot f_{2|3}(x_2|x_3) \cdot f_3(x_3)$$

$$f_{2|3}(x_2|x_3) = c_{23}(F_2(x_2), F_3(x_3)) \cdot f_2(x_2)$$

$$f_{1|23}(x_1|x_2, x_3) = c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)) \cdot f_{1|3}(x_1|x_3)$$

$$f_{1|3}(x_1|x_3) = c_{13}(F_1(x_1), F_3(x_3)) \cdot f_1(x_1)$$

Vine Copula: An Example

Let's get slightly more complicated...

$$f(x_1, x_2, x_3) = f_{1|23}(x_1|x_2, x_3) \cdot f_{2|3}(x_2|x_3) \cdot f_3(x_3)$$

$$f_{2|3}(x_2|x_3) = c_{23}(F_2(x_2), F_3(x_3)) \cdot f_2(x_2)$$

$$f_{1|23}(x_1|x_2, x_3) = c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)) \cdot f_{1|3}(x_1|x_3)$$

$$f_{1|3}(x_1|x_3) = c_{13}(F_1(x_1), F_3(x_3)) \cdot f_1(x_1)$$

Bivar Copulas

Be careful:

$$F_{1|3}(x_1|x_3) = P(X_1 \leq x_1|X_3 = x_3)$$

Vine Copula: An Example

Let’s get slightly more complicated...

$$f(x_1, x_2, x_3) = f_1(x_1) f_2(x_2) f_3(x_3)$$

$$\times c_{23}(F_2(x_2), F_3(x_3)) \cdot c_{13}(F_1(x_1), F_3(x_3))$$

$$\times c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3))$$

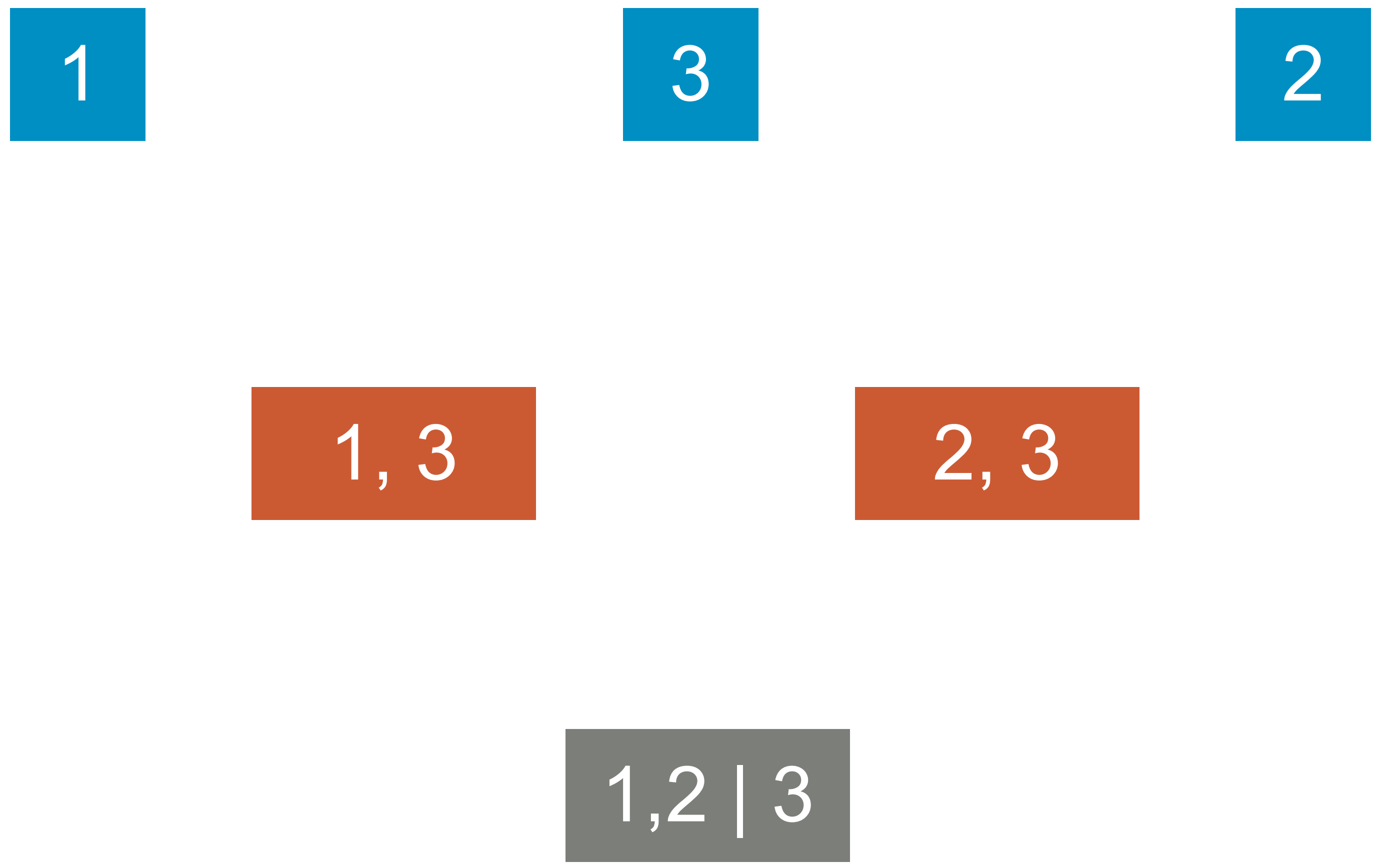
Vine Copula: An Example

Let’s get slightly more complicated...

$$f(x_1, x_2, x_3) = f_{\boxed{1}}(x_1) f_{\boxed{2}}(x_2) f_{\boxed{3}}(x_3)$$

$$\times c_{\boxed{23}}(F_2(x_2), F_3(x_3)) \cdot c_{\boxed{13}}(F_1(x_1), F_3(x_3))$$

$$\times c_{\boxed{12|3}}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3))$$



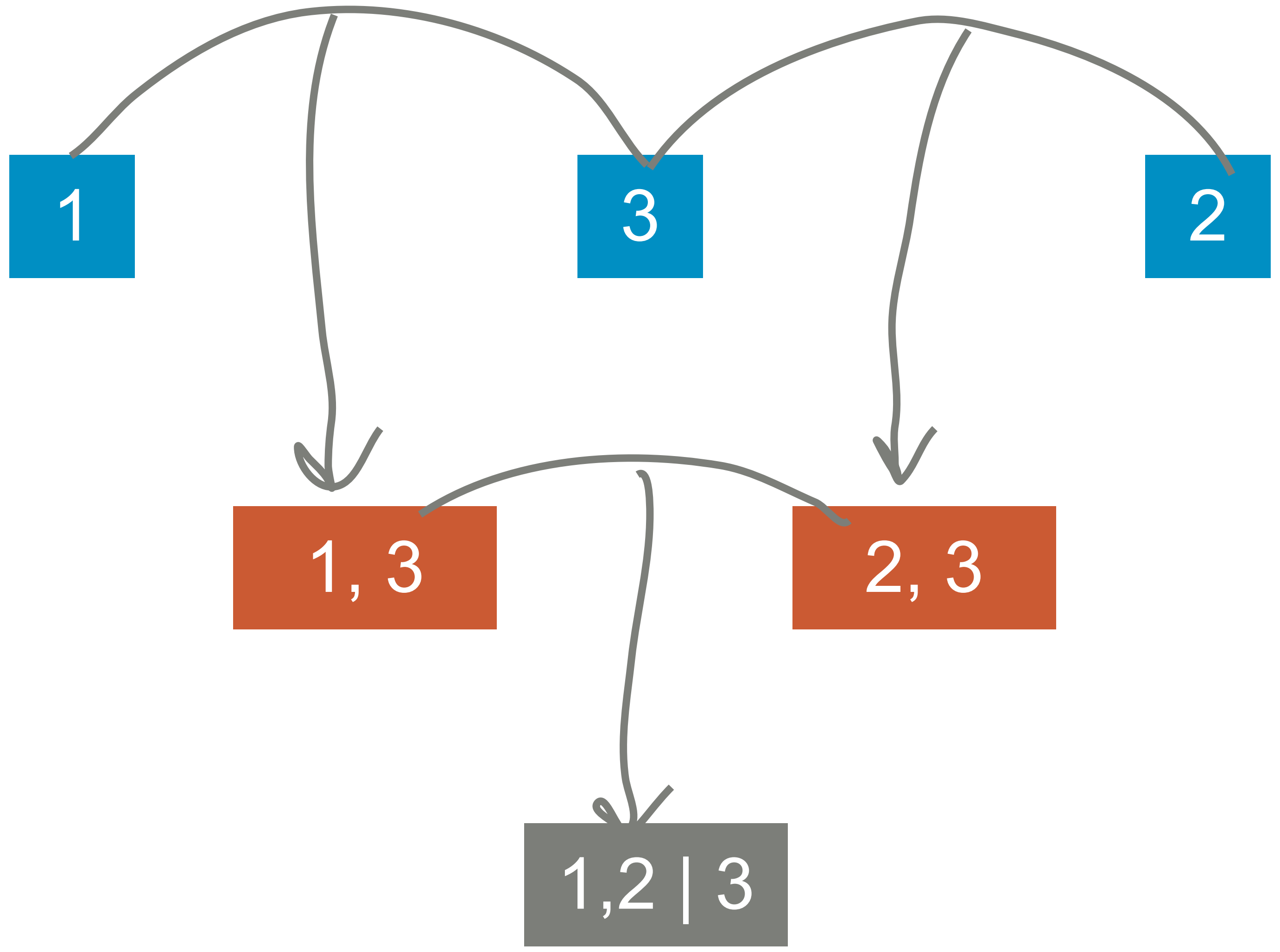
Vine Copula: An Example

Let’s get slightly more complicated...

$$f(x_1, x_2, x_3) = f_{\boxed{1}}(x_1) f_{\boxed{2}}(x_2) f_{\boxed{3}}(x_3)$$

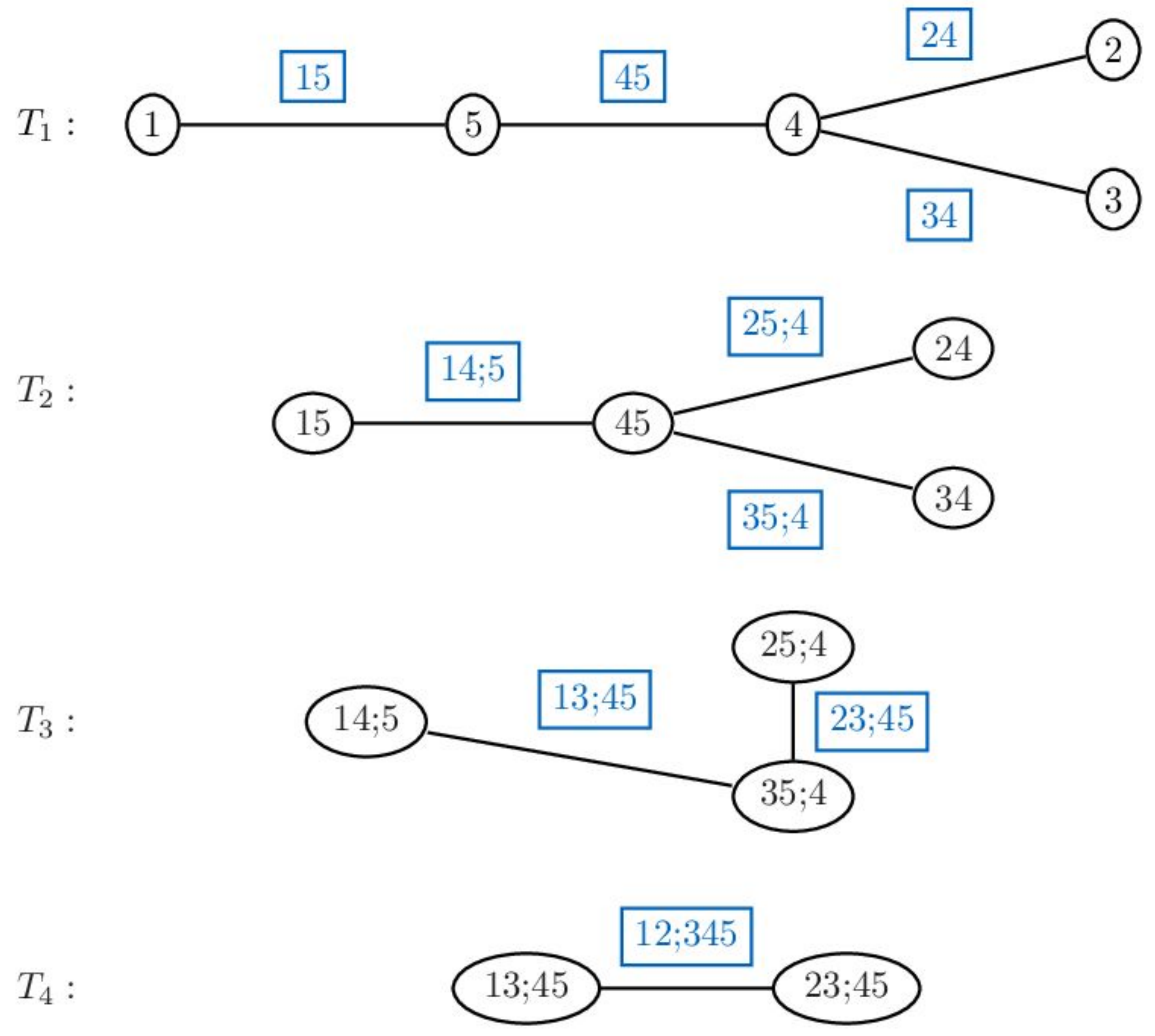
$$\times c_{\boxed{23}}(F_2(x_2), F_3(x_3)) \cdot c_{\boxed{13}}(F_1(x_1), F_3(x_3))$$

$$\times c_{\boxed{12|3}}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3))$$



Vine Copula: Decompose higher dim dependency by bivar copulas and graphical structure

Model Advantages



Picture from [Killiches et al. (2016)]

Flexibility

Higher dim copulas are very rigid

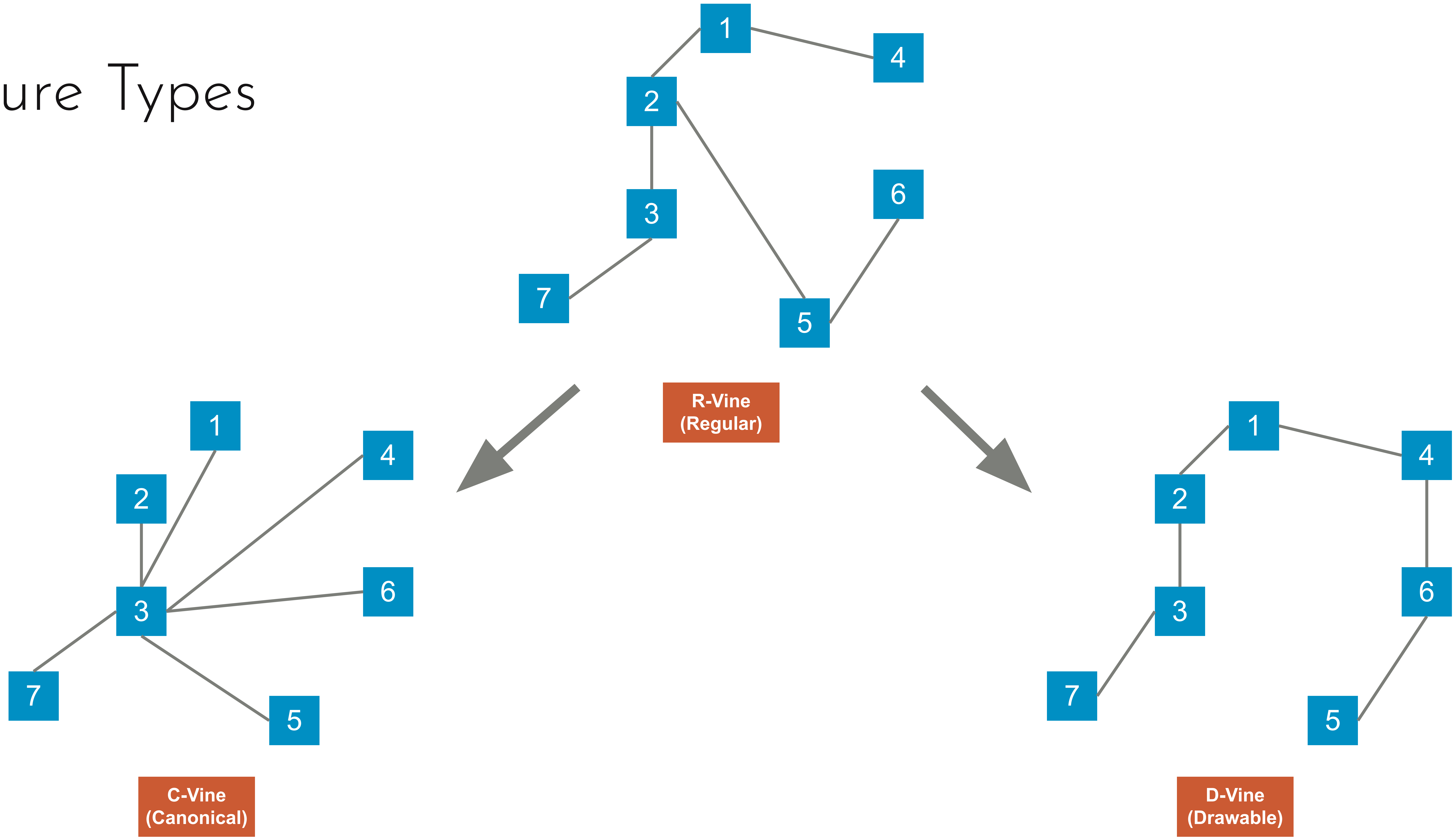
Interpretability

Visual dependence structure

Risk Control

Tail risk is limited

Structure Types

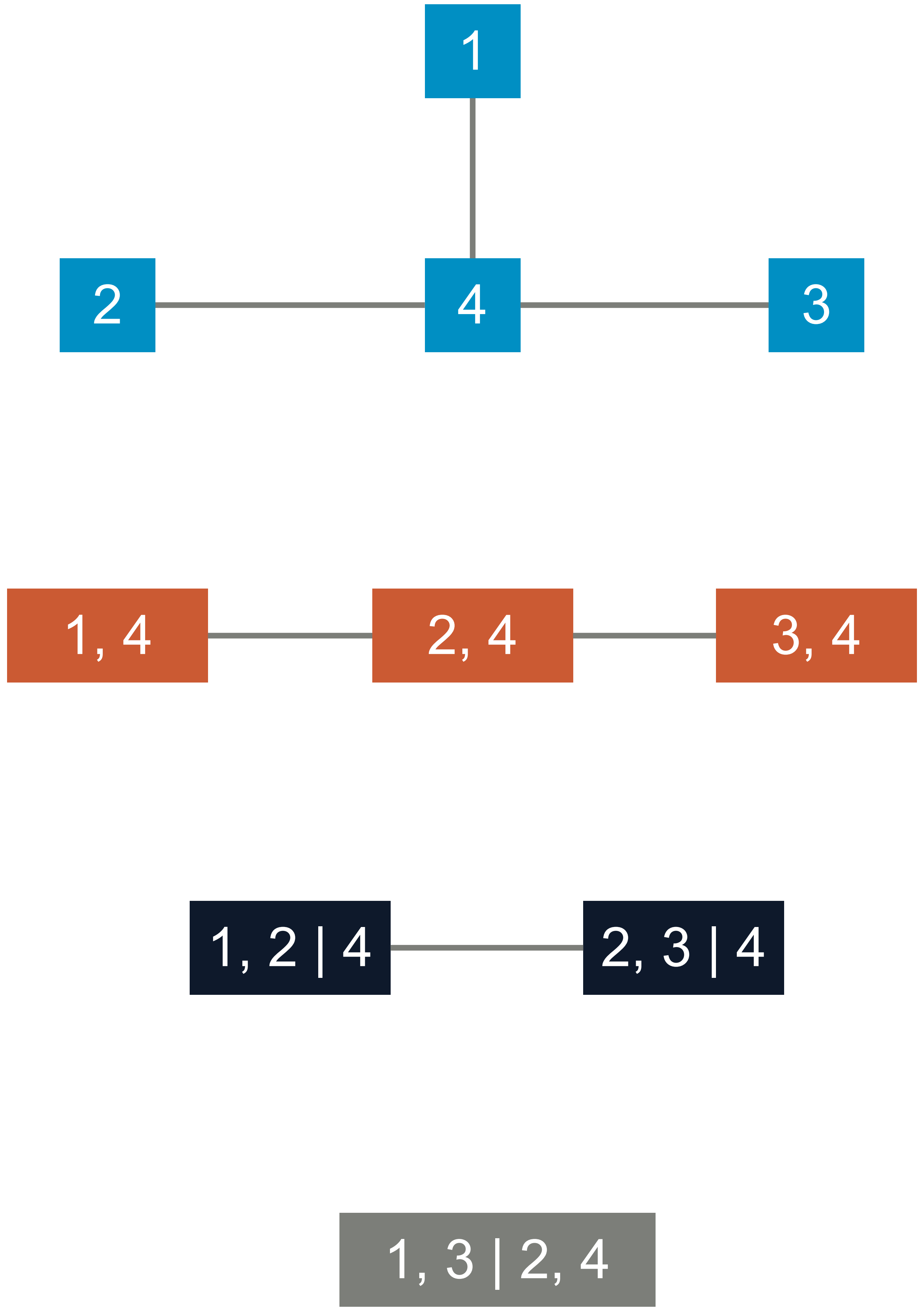


C-Vine Notation

Structure has bijection map with an ordered tuple:

(1, 3, 2, 4)

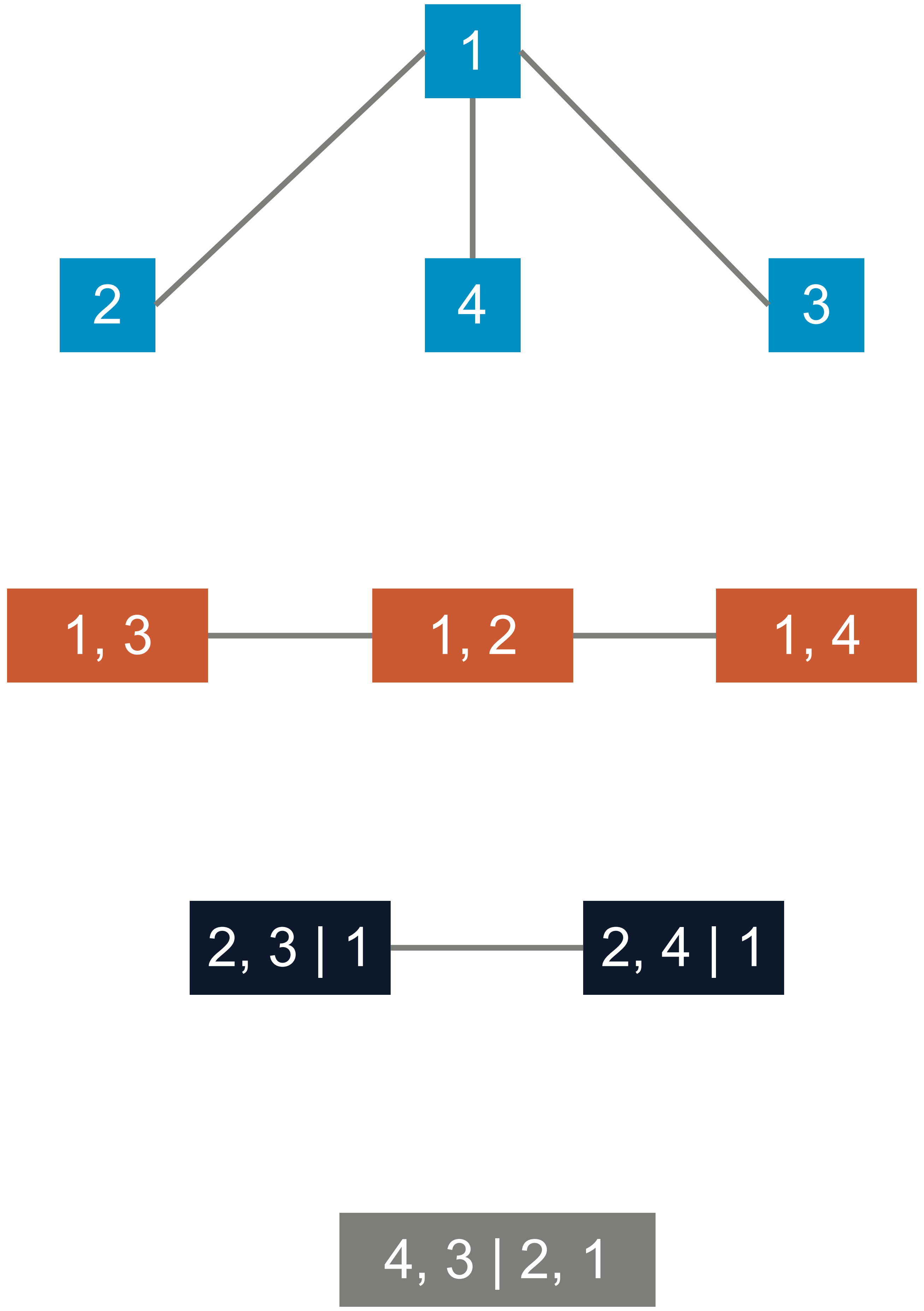
Center for each tree



R-Vine Notation

Structure has bijection map with an upper triangular matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & \\ 3 & 3 & & \\ 4 & & & \end{bmatrix}$$



Vine Copula Workflow

$$\left(\boldsymbol{X}_i, \boldsymbol{F}_i, \boldsymbol{U}_i = \boldsymbol{F}_i(\boldsymbol{X}_i)\right)_{i\,=\,1,\ldots,4}$$

Data

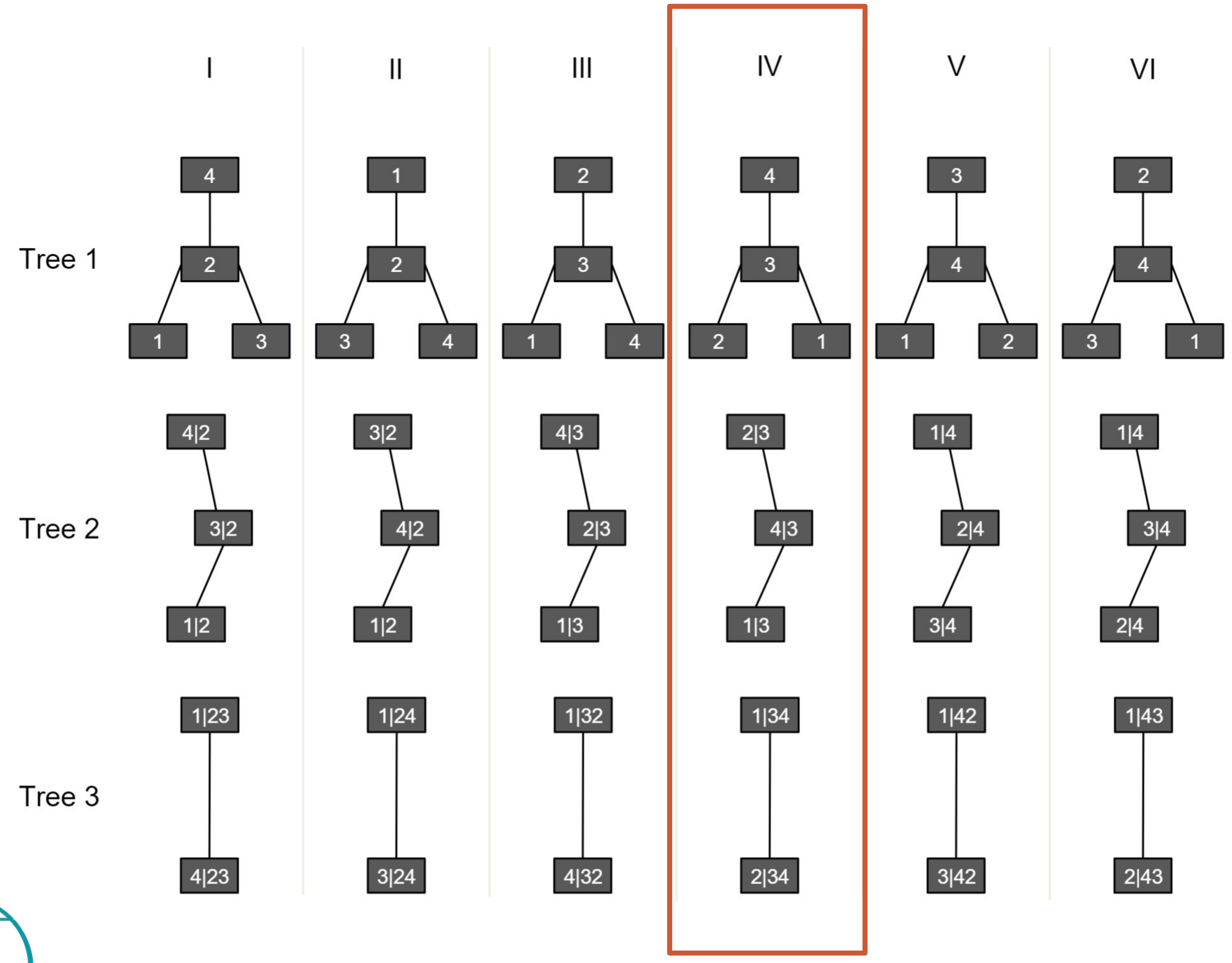
Marginal CDF

Pseudo-observation/ Quantile

1. Get data
2. Figure out the vine copula structure
3. Calculate point density
4. Calculate conditional probability
5. Generate signals

Vine Copula Workflow

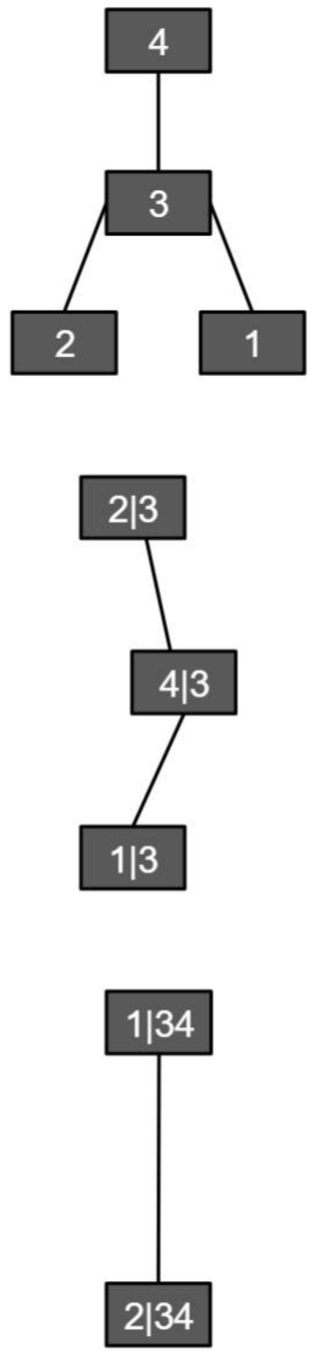
Complicated. Assume your computer can handle it for now.



1. Get data
2. Figure out the vine copula structure
3. Calculate point density
4. Calculate conditional probability
5. Generate signals

Vine Copula Workflow

Calculate point density via vine structure:



$$f(x_1, x_2, x_3, x_4) = f(x_1)f(x_2)f(x_3)f(x_4)$$

$$\times c_{23} \cdot c_{34} \cdot c_{13}$$

$$\times c_{24|3} \cdot c_{12|3}$$

$$\times c_{12|34}$$

1. Get data
2. Figure out the vine copula structure
3. Calculate point density
4. Calculate conditional probability
5. Generate signals

Vine Copula Workflow

$$f(x_1, x_2, x_3, x_4) \to f(u_1, u_2, u_3, u_4)$$

$$P(U_1 \le u_1 | u_2, u_3, u_4)$$
$$= \frac{\int_0^{u_1} f(u, u_2, u_3, u_4) du}{\int_0^{1} f(u, u_2, u_3, u_4) du}$$
$$= h_C(u_1 | u_2, u_3, u_4)$$

1. Get data
2. Figure out the vine copula structure
3. Calculate point density
4. Calculate conditional probability
5. Generate signals

Vine Copula Workflow

Stock 1 Overpriced

$$h_C(u_1|u_2, u_3, u_4) > 0.5$$

Stock 1 Underpriced

$$h_C(u_1|u_2, u_3, u_4) < 0.5$$

1. Get data
2. Figure out the vine copula structure
3. Calculate point density
4. Calculate conditional probability
5. Generate signals

03. Trading Strategy

Key Components

1. Pairwise Spearman’s rho
2. Generalized Spearman’s rho [Schmid and Schmidt (2007)]
3. Geometric distance to diagonal on Q-Q plot
4. Extremal approach [Mangold (2015)]

1. 4 stocks each cohort from top 20 stocks

2. C-Vine assumption

3. CMPI strategy (returns)

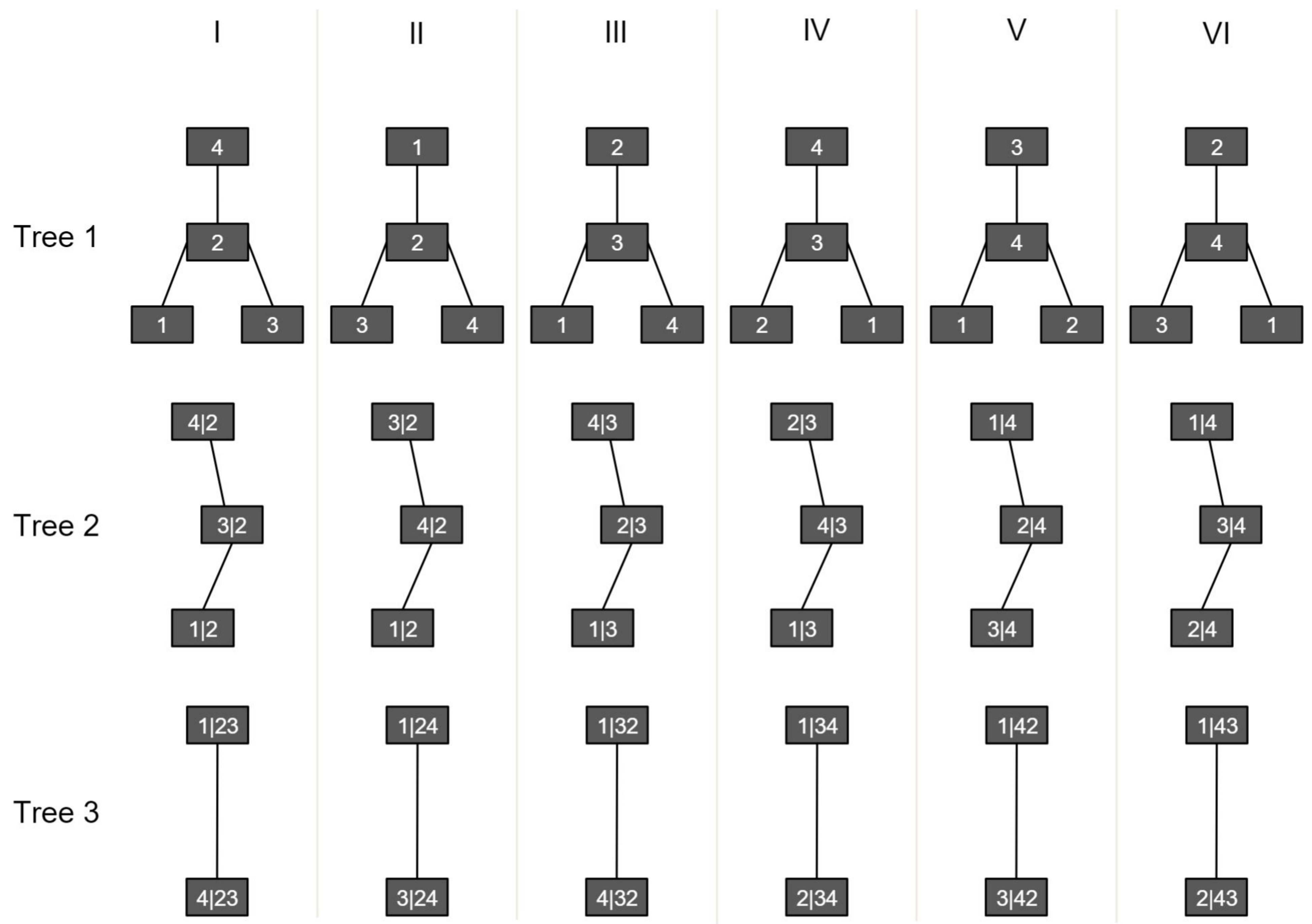
4. Bollinger Band

5. Against SPY index

6. Dollar neutral

Note: This approach is based on [Stübinger et al. 2016]

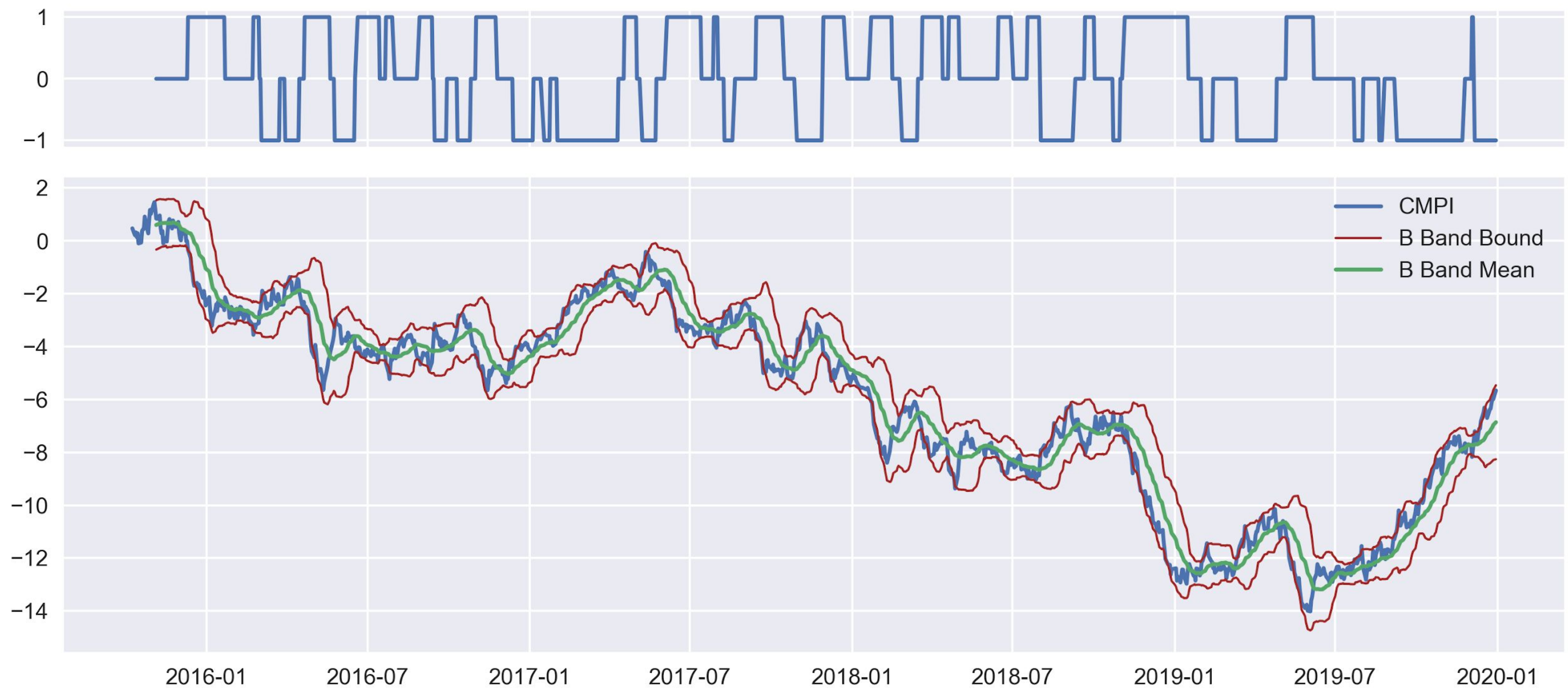
Key Components



1. 4 stocks each cohort from top 20 stocks
2. C-Vine assumption
3. CMPI strategy (returns)
4. Bollinger Band
5. Against SPY index
6. Dollar neutral

Note: This approach is based on [Stübinger et al. 2016]

Key Components



1. 4 stocks each cohort from top 20 stocks
2. C-Vine assumption
3. CMPI strategy (returns)
4. Bollinger Band
5. Against SPY index
6. Dollar neutral

Note: This approach is based on [Stübinger et al. 2016]

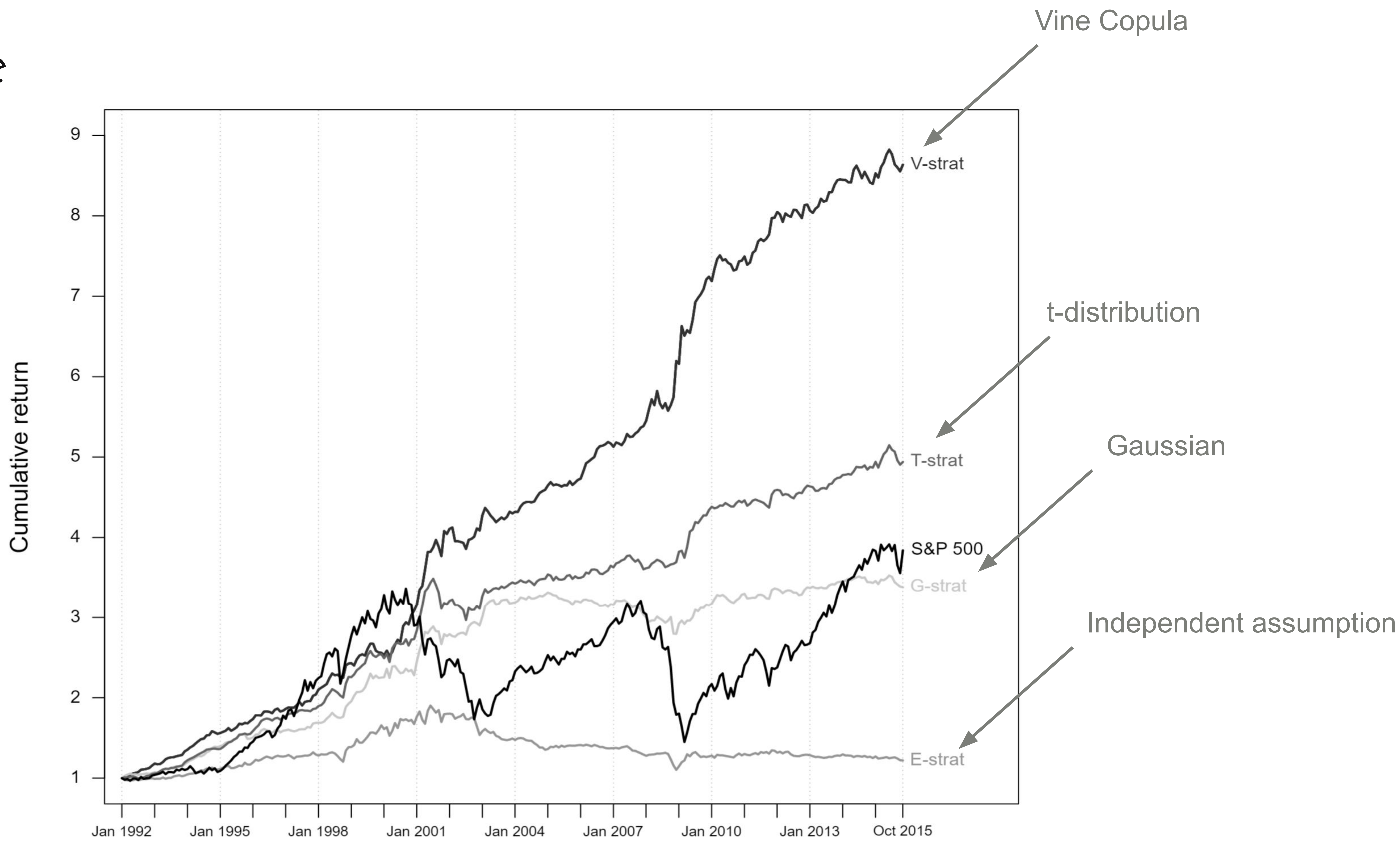
Key Components

- Long/short the target stock
- Hedge against SPY

1. 4 stocks each cohort from top 20 stocks
2. C-Vine assumption
3. CMPI strategy (returns)
4. Bollinger Band
5. Against SPY index
6. Dollar neutral

Note: This approach is based on [Stübinger et al. 2016]

Performance



Picture from [Stübinger et al. 2016]

Functionalities In Our Module

Finished:

- Automatic C-Vine Fit
- Generate positions for the target stock via Bollinger band
- Translate positions as units against an index

Working on now:

- Automatic R-Vine Fit
- Stocks selection
- Speed optimization

Note: This approach is based on [Stübinger et al. 2016]

Possible Issues

1. Exiting too early.

2. Performance seem only significant on stock groups.

3. Computation time.

Interesting Problems

HUDSON AND THAMES

- Strategies for smaller cohorts
- Stocks selection
- Term structures
- Fast computation for high dimensions
- R-Vine fit
- Optimal exit
- Higher frequency data
- Alternative data

Q&A

References

- Killiches, M., Kraus, D. and Czado, C., 2016. Using model distances to investigate the simplifying assumption, goodness-of-fit and truncation levels for vine copulas. arXiv preprint arXiv:1610.08795.
- Pham, M.T., Vernieuwe, H., Baets, B.D. and Verhoest, N.E., 2018. A coupled stochastic rainfall–evapotranspiration model for hydrological impact analysis. Hydrology and Earth System Sciences, 22(2), pp.1263-1283.
- Dissmann, J., Brechmann, E.C., Czado, C. and Kurowicka, D., 2013. Selecting and estimating regular vine copulae and application to financial returns. Computational Statistics & Data Analysis, 59, pp.52-69.
- Stübinger, J., Mangold, B. and Krauss, C., 2018. Statistical arbitrage with vine copulas. Quantitative Finance, 18(11), pp.1831-1849.
- Czado, C., 2019. Analyzing dependent data with vine copulas. Lecture Notes in Statistics, Springer.
- Vine Copula Models - Lehrstuhl für Mathematische Statistik