Copula Based Quantile Regression

Yitong Lin

Master of Financial Mathematics, UNSW

31 May 2018

Motivating Example

"Recipe for Disaster: The Formula That Killed Wall Street" - Felix Salmon, WIRED(2009) So What Happened?

- In 2000, David Li, a PhD in Statistics working at Wall Street, devised a formula, known as the Gaussian copula, to model default correlation of securities.
- Wall Street started using this formula to price an instrument known as collateralized debt obligations (CDO). The CDO market grew from $275 billion to $4.7 trillion from 2000 to 2006.
- But, the housing bubble started to collapse, mortgage back securities default correlation became extremely high, which the Gaussian copula function failed to capture, many CDOs became worthless...

Introduction

- So why is copula powerful?
- What other applications are there?

Sklar's Theorem

Sklar's Theorem (1959):
Let $\mathbf{X} = (X_1, ..., X_d) \sim F$ with marginal distribution functions $F_1, ..., F_d$, then there exists a copula C associated with $\mathbf{X}$ such that:

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_1(x_d))$$

and the density of $\mathbf{X}$ can be expressed as:

$$f(x_1, ..., x_d) = c(F_1(x_1), ..., F_d(x_d))f_1(x_1), ..., f_d(x_d)$$

where

$$c(u_1, ..., u_d) = \frac{\partial}{\partial u_1 ... \delta u_d} C(u_1, ..., u_d)$$

is the *copula density*

Conditional distribution from copula (h-functions)

The conditional distribution can be calculated from a bivariate copula (with or without conditioned set D) as follows:

$$F(x_1|x_2, \mathbf{x}_D) = \frac{\partial}{\partial u_2|\mathbf{u_D}} C(u_1|\mathbf{u_D}, u_2|\mathbf{u_D})$$

This is also referred as the h-function for simplicity:

$$h_{X_1|X_2,\mathbf{X}_D}(x_1|x_2, \mathbf{x}_D) = \frac{\partial}{\partial u_2|\mathbf{u_D}} C(u_1|\mathbf{u_D}, u_2|\mathbf{u_D})$$

Regular vine (R-vine)

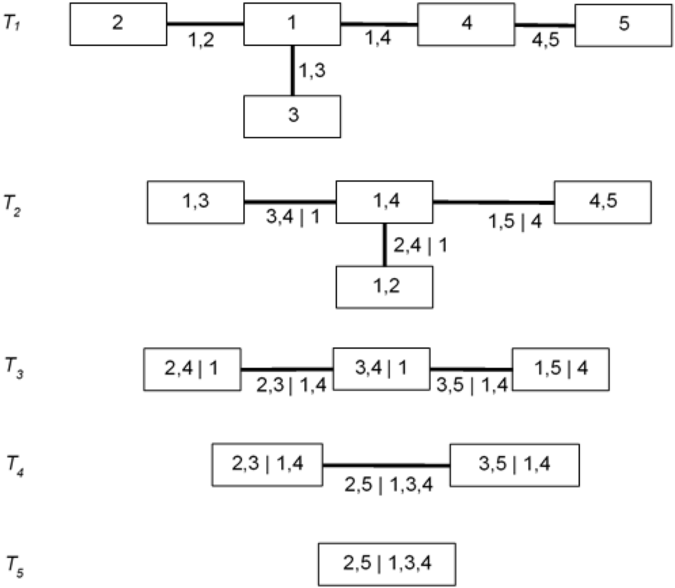
Bedford & Cooke (2002) have introduced a graphic model denoted as *Regular vines (R-vine)*. A R-vine for $\mathbf{X} = (X_1, ..., X_d)$ is a set of trees $V = (T_1, ..., T_d)$ that satisfy the following definition:

Definition (Regular Vine)

$V = (T_1, ..., T_d)$ is an R-vine on d elements if:

1. T_1 is a tree with nodes $N_1 = \{1, ..., d\}$ and set of edges denoted E_1.
2. For $i = 2, ..., d$, T_i is a tree with nodes $N_i = E_{i-1}$ and edge set E_i.
3. For $i = 2, ..., d$ and $a, b \in E_i$ with $a = a_1, a_2$ and $b = b_1, b_2$ it must hold that $\#(a \cap b) = 1$ (proximity condition), where $\#$ denotes the cardinality of a set.

Example: R-vine with $d = 5$



Why R-vine?

- Model the dependencies of $\mathbf{X}$ using only bivariate copula $C(u_1, u_2)$.
- Flexibility. Each bivariate copula in the R-vine can be of different families.
- Visualize dependencies of $\mathbf{X}$
- Express the joint density $f_{1,\ldots,d}(\mathbf{x})$ easily.

R-vine density

The density of **X** can be expressed as the product of all the bivariate copulas specified in the R-vine and the marginal densities.

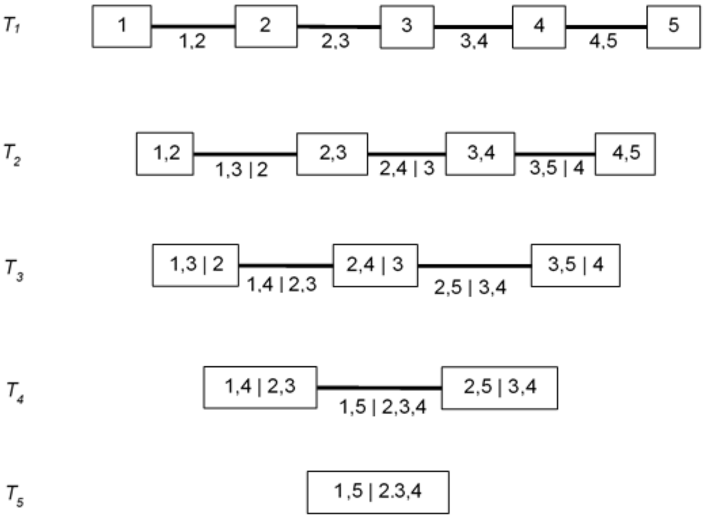
Definition (Density of R-vine)

$$f_{1,\ldots,d}(\mathbf{x}) = \prod_{k=1}^{n} f_k(x_k)$$
$$\prod_{i=1}^{n-1} \prod_{e \in E_i} c_e(.,.)$$

D-vine

- D-vine is a subclass of R-vine where each tree $T_i, i = 1, ..., d$ has a path like structure.
- This structure is useful for quantile regression which we will discuss later.

Example: D-vine with $d = 5$



Two copula based quantile regression method

Next, we investigated two copula based quantile regression method:

- Semiparametric quantile regression (based on R-vine) [Noh et al.(2014)].
- D-vine based quantile regression Kraus & Czado (2017).

Semiparametric quantile regression

Quantile regression principle

Let loss function as $\rho_\alpha(y) = y(\alpha - I_{(y<0)})$ where I is an indicator function. Koenker & Bassett Jr (1978) proves the following:

Definition (Quantile estimation)

A quantile $\alpha \in (0,1)$ can be found by minimizing the expected loss of $Y - u$ with respect to u:

$$\arg\min_u E(\rho_\alpha(Y-u)) \tag{1}$$

and the estimated sample conditional quantile can be found by solving the minimizing problem:

$$\hat{q}_\alpha(x_1, ...x_d) = \arg\min_{q\in\mathbb{R}} \sum_{i=1}^{n} \rho_\alpha(y^{(i)} - q) \tag{2}$$

Conditional density

Using the density formula for R-vine, the conditional density can be expressed as:

$$
\begin{aligned}
f(y|x_1, ..., x_d) &= \frac{f(y, x_1, ..., x_d)}{f(x_1, ..., x_d)} \\
&= \frac{c(F(y), F_1(x_1), ..., F_d(x_d))f(y)f(x_1), ..., f(x_d)}{c(F_1(x_1)...F_d(x_d))f(x_1)...f(x_d)} \\
&= f_Y(y)\frac{c(F(y), F_1(x_1), ..., F_d(x_d))}{c(F_1(x_1), ..., F_d(x_d))}
\end{aligned} \tag{3}
$$

Semiparametric quantile estimate

Hence, we can transform the expected loss from expectation with respect to $Y|\mathbf{X}$ to Y :

$$\begin{aligned} q_\alpha(x_1,...,x_d) &= \arg\min_u E(\rho_\alpha(Y-u)|x_1,...,x_d) \\ &= \arg\min_u E(\rho_\alpha(Y-u)c(F(Y),F_1(x_1),..,F_d(x_d))) \end{aligned}$$

and the conditional quantile estimate is:

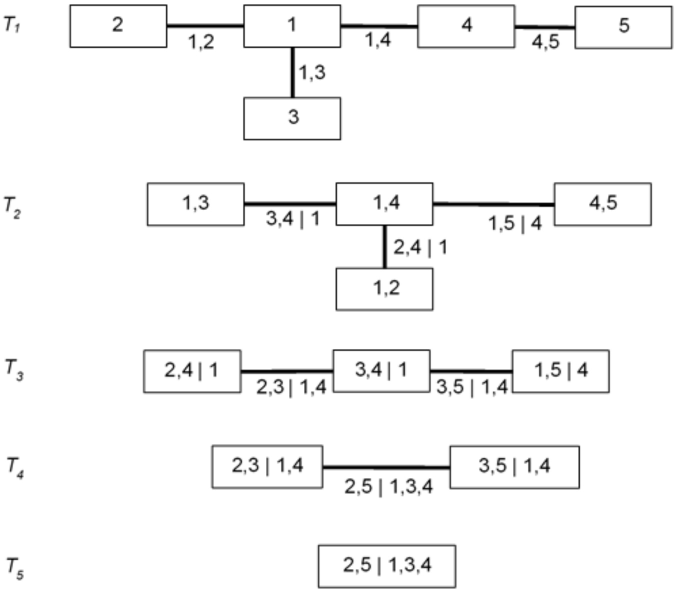
$$\hat{q}_\alpha(x_1,...,x_d) = \arg\min \sum_{i=1}^{n} \rho_\alpha(y^{(i)}-u)\widehat{c}(\widehat{F}(y^{(i)}),\widehat{F}(x_1),...,\widehat{F}(x_d))$$

R-vine selection algorithm

- To complete the quantile estimate, we need select a R-vine to obtain $\widehat{c}(\widehat{F}(y^{(i)}), \widehat{F}(x_1), ..., \widehat{F}(x_d))$.
- This selection algorithm aims to sequentially select an R-vine model with most explanatory power.
- For each tree, it aims to maximize the sum of Kendall's tau of its edges.

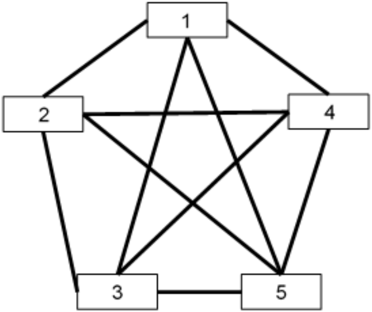
We will illustrate the intuition with an example:

Example: R-vine



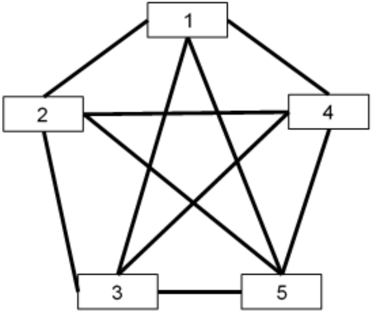
Example: R-vine selection algorithm

Tree 1



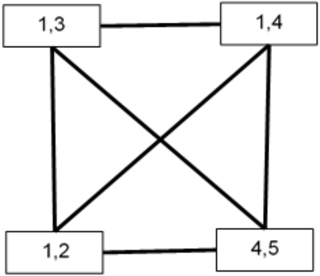
Example: R-Vine selection algorithm

Tree 1



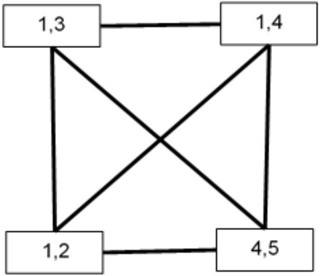
Example: R-Vine selection algorithm

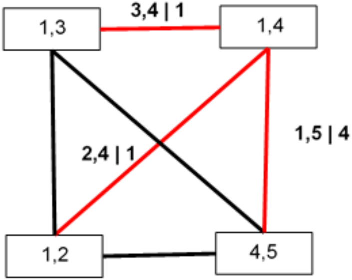
Tree 2



Example: R-Vine selection algorithm

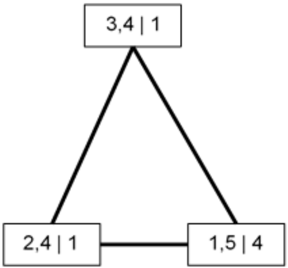
Tree 2





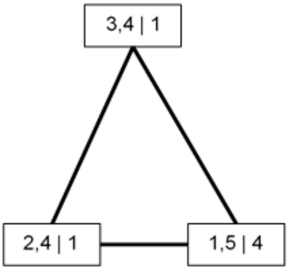
Example: R-Vine selection algorithm

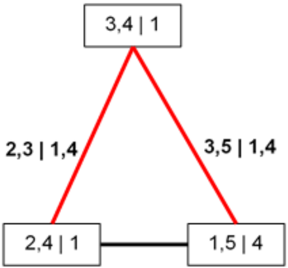
Tree 3



Example: R-Vine selection algorithm

Tree 3





D-vine quantile regression

Conditional quantile function

We need to model the relationship between Y and $\mathbf{X}$ and the *conditional quantile function* for $\alpha \in (0, 1)$:

$$q_\alpha(x_1, ..., x_d) = F^{-1}_{Y|X_1,...,X_d}(\alpha|x_1, ..., x_d)$$

Conditional quantile function cont.

First, apply probability integral transform (PIT) $V = F_Y(Y)$ and $U_j = F_j(X_j)$ with respective PIT value $v = F_Y(y)$ and $u_j = F_j(x_j)$:

$$\begin{aligned} F_{Y|X_1,\ldots,X_d} &= P(Y \leq y | X_1 = x_1, ..., X_d = x_d) \\ &= P(F_Y(Y) \leq v | F_1(X_1) = u_1, ..., F_d(X_d) = u_d) \\ &= C_{V|U_1,\ldots,U_d}(v | u_1, ..., u_d) \end{aligned} \tag{4}$$

Second, taking the inverse of (4):

$$F^{-1}_{Y|X_1,\ldots,X_d}(\alpha | x_1, ..., x_d) = F_y^{-1}\big(C^{-1}_{V|U_1,\ldots,U_d}(\alpha | u_1, ..., u_d)\big) \tag{5}$$

Conditional quantile function cont.

Finally, we have our conditional quantile estimation function expressed in terms of copula and marginal of Y

$$\hat{q}_\alpha(x_1, ..., x_d) = \hat{F}_Y^{-1}\big(\widehat{C}^{-1}_{V|U_1,...,U_d}(\alpha|u_1, ..., u_d)\big) \tag{6}$$

Estimating conditional copula

- To calculate estimated conditional quantile, we need to be able to calculate $\widehat{C}^{-1}_{V|U_1,...,U_d}(\alpha|u_1,...,u_d)$
- Kraus & Czado (2017) proposed fitting a D-vine copula to $(V, U_1, ..., U_d)$, such that V is the starting node in the first tree.
- This allows $C^{-1}_{V|U_1,...,U_d}$ to be easily calculable using recursive method.
- D-vine is fitted to maximize the conditional log likelihood.

We will illustrate the idea with an example.

Example: Conditional copula

$j = 1$

$C_{V|U_1}(v|u_1) = h_{V|U_1}(v|u_1)$

$j = 2$

$C_{U_2|U_1}(u_2|u_1) = h_{U_2|U_1}(u_2|u_1)$

$C_{U_1|U_2}(u_1|u_2) = h_{U_1|U_2}(u_1|u_2)$

$C_{V|U_1,U_2}(v|u_1, u_2, u_3) = h_{V|U_2;U_1}(C_{V|U_1}(v|u_1)|C_{U_2|U_1}(u_2|u_1))$

$j = 3$

$C_{U_3|U_2}(u_3|u_2) = h_{U_3|U_2}(u_3|u_2)$

$C_{U_3|U_1,U_2}(u_3|u_1, u_2) = h_{U_3|U_1,U_2}(C_{U_3|U_2}(u_3|u_2)|C_{U_1|U_2}(u_1|u_2))$

$$C_{V|U_1,U_2,U_3}(v|u_1, u_2, u_3) = h_{V|U_3;U_1,U_2}(C_{V|U_1,U_2}(v|u_1, u_2)| C_{U_3|U_1,U_2}(u_3|u_1, u_2))$$

Conditional Log likelihood

The conditional log-likelihood(cll) of a D-vine given pseudo copula data u from PIT and ordering $M = (m_1, ...m_d), m_i \in \{1, ..., d\}$ is defined as:

$$cll = \sum_{i=1}^{n} ln\ c_{v|u}\big(\hat{v}^{(i)}|\hat{\mathbf{u}}^{(\mathbf{i})}\big)$$

where the conditional copula density $c_{v|\mathbf{u}}$ is defined as:

$$\begin{aligned} c_{v|\mathbf{u}}\big(\hat{v}^{(i)}|\hat{\mathbf{u}}^{(\mathbf{i})}\big) =& c_{vu_{m_1}}\big(\hat{v}^{(i)}, \hat{u}_{m_1}^{(i)}\big) \\ & \prod_{j=2}^{d} c_{vu_{m_j}|u_{m_1},...,u_{m_{j-1}}}\Big(C_{v|u_{m_1},...,u_{m_{j-1}}}\big(\hat{v}_{m_j}^{(i)}|\hat{u}_{m_1}^{(i)}, ..., \hat{u}_{m_{j-1}}^{(i)}\big), \\ & C_{u_{m_j}|u_{m_1},...,u_{m_{j-1}}}\big(\hat{u}_{m_j}^{(i)}|\hat{u}_{m_1}^{(i)}, ..., \hat{u}_{m_{j-1}}^{(i)}\big)\Big) \end{aligned}$$

Example: Inverse conditional copula

$j = 1$

$$C^{-1}_{V|U_1}(v|u_1) = h^{-1}_{V|U_1}(\alpha|u_1)$$

$j = 2$

$$C^{-1}_{V|U_1,U_2}(v|u_1, u_2) = C^{-1}_{V|U_1}(h^{-1}_{V|U_2;U_1}(\alpha|h_{U_2|U_1}(u_2|u_1))|u_1)$$

$j = 3$

$$C^{-1}_{V|U_1,U_2,U_3}(v|u_1, u_2, u_3) = C^{-1}_{V|U_1,U_2}(\alpha|h_{U_3|U_1;U_2}(h_{U_3|U_2}(u_3|u_2))|u_1, u_2)$$

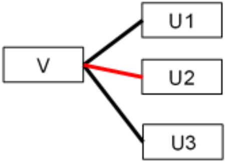
etc....

D-Vine selection algorithm

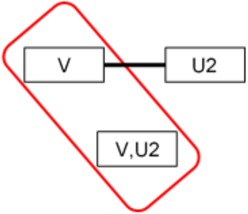
- Sequentially add predictor $\widehat{u}_{m_j}$ end of the first tree T_1 that maximizes the cll to the
- Terminate when addition of $\widehat{u}_{m_j}$ does not improve cll or all $\hat{\mathbf{u}}$ has been added to the D-vine

Example: D-Vine selection algorithm

Iteration 1:

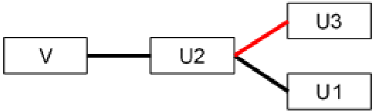


Select U that maximizes cll, which is calculated as the product of copula density within the red circle.

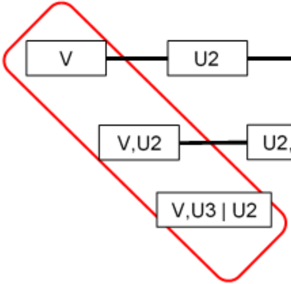


Example: D-Vine selection algorithm

Iteration 2:



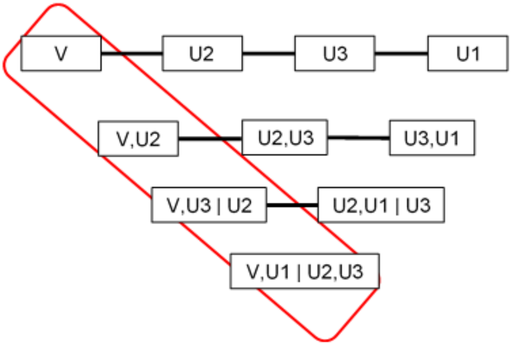
Select U that maximizes cll, which is calculated as the product of copula density within the red circle.



Example: D-Vine selection algorithm

Iteration 3:

Select U that maximizes cll, which is calculated as the product of copula density within the red circle.



Estimating marginal

So far, we haven't discussed how marginals are estimated.
We choose continuous kernel smoothing estimator, which for a given sample $(x^{(i)})_{i=1,\dots,n}$ is defined as:

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} K(\frac{x - x^{(i)}}{h}), \quad x \in \mathbb{R}$$

Thus marginals are estimated non parametrically while copula are estimated parametrically, making the two copula based quantile regression semiparametric.

Simulation

Compare four methods of quantile regression:

- Semiparametric quantile regression (SPQR) [Noh et al. (2014)]
- D-vine quantile regression (DVQR) Kraus & Czado (2017)
- Linear quantile regression (LQR) Koenker & Bassett Jr (1978)
- Boosting additive quantile regression Koenker (2011)

Linear quantile regression

Linear quantile regression is first introduced by Koenker & Bassett Jr (1978). The αth quantile function is defined as:

$$\hat{q}_\alpha(x_1, ..., x_d) = \hat{\beta}_0 + \sum_{j=1}^{d} \hat{\beta}_j x_j$$

where the parameter $\hat{\beta}_j, j = 0, ..., d$ is the solution to the linear programming problem

$$\min_{\beta_j \in \mathbb{R}, j=1,..,d} \left[\alpha \sum_{i=1}^{n} \left(y^{(i)} - \hat{q}_\alpha(\mathbf{x}^{(i)}) \right)^+ + (1-\alpha) \sum_{i=1}^{n} \left(\hat{q}_\alpha(\mathbf{x}^{(i)}) - y^{(i)} \right)^+ \right]$$

Boosting additive model

This method relaxes the linearity assumption in LQR by utilizing additive models for quantile regression. The quantile regression method proposed by Koenker (2011) is defined as:

$$\hat{q}_\alpha(x_1^{(i)}, ..., x_d^{(i)}, z_1^{(i)}, ..., z_J^{(i)}) = \hat{\beta}_0 + \sum_{j=1}^{d} \hat{\beta}_j x_j^{(i)} + \sum_{j=1}^{J} \hat{g}_j(z_j^{(i)})$$

where $\hat{g}_j$ is a smooth function on the continuous variable $z_j^{(i)}, j = 1, ..., J$

Simulation scenarios

1. **CL3** : (Y, X_1, X_2) follows a three-dimensional clayton copula with parameter δ_1 and δ_2 and margin set M_1 or M_2
2. **t5** : $(Y, X_1, ..., X_2)$ follows a five-dimensional t-copula with 3 degrees of freedom, correlation matrix S_1 or S_2 and margin set M_1 or M_2.
3. **N5** : $X \sim N(0, \Sigma)$, with $\Sigma_{ij} = 0.5^{|i-j|}$ and $Y = \sqrt{3X_1 + 0.5X_2 + 2} + (-0.7X_3 + 2)(0.2X_4^3) + \sigma\epsilon$, $\epsilon \sim N(0, 1), \sigma \in \{0.5, 1\}$

Simulation scenario cont.

Case	Copula parameter	Marginals			
			Y	X_1	X_2
C3	$\delta_1 = 0.8$, $\delta_2 = 4.67$	M_1	$N(0,1)$	$t_4(0,1)$	$N(1,4)$
		M_2	$st_4(0,1,2)$	$sN(-2,0.5,3)$	$st_3(1,2,5)$
			Y	X_1	X_2
t5	S_1, S_2	M_1	$N(0,1)$	$t_4(0,1)$	$N(1,4)$
		M_2	$st_4(0,1,2)$	$sN(-2,0.5,3)$	$st_3(1,2,5)$
			X_3	X_4	
			$t_4(0,1)$	$N(1,4)$	
			$sN(-2,0.5,3)$	$st_3(1,2,5)$	

Table 1: Copula parameters and and marginal setting for **CL3** and **t5**

Simulation scenario cont.

The correlation matrix is given by:

$$S_1 = \begin{bmatrix} 1 & 0.6 & 0.5 & 0.5 & 0.4 \\ 0.6 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.4 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 1 & 0.27 & 0.74 & 0.72 & 0.41 \\ 0.27 & 1 & 0.28 & 0.29 & 0.27 \\ 0.74 & 0.28 & 1 & 0.74 & 0.42 \\ 0.72 & 0.29 & 0.74 & 1 & 0.40 \\ 0.41 & 0.27 & 0.42 & 0.4 & 1 \end{bmatrix}$$

Simulation methods

The MISE of method m is defined as

$$MISE_m = \frac{1}{R}\sum_{r=1}^{R}\left[\frac{1}{n_{eval}}\sum_{i=1}^{n_{eval}}\{\hat{q}^{(r)}_{m,\alpha}(\mathbf{x}^{eval}_{r,i}) - q_{\alpha}(\mathbf{x}^{eval}_{r,i})\}^2\right]$$

where we simulate for $r = 1, ..., R = 100$ replications. For each replication r,the simulation is divided into two steps:

1. Simulate training data $(y^{train}_{r,i}, \mathbf{x}^{train}_{r,i}), i = 1, ..., n_train$ from the distribution $(Y, \mathbf{X})$. Estimate $\hat{q}^{(r)}_{m,\alpha}$ from the training set.
2. Simulate evaluation data set $\mathbf{x}^{train}_{r,i}, i = 1, ..., n_eval$ from the distribution of $\mathbf{X}$. Predict $\hat{q}^{(r)}_{m,\alpha}$ using evaluation set.
3. Calculate MISE for each methods

Replication are run in parallel in R

Result CL3

Marginals	Parameters	α	SPQR	DVQR	LQR	BAQR
M1	δ_1	0.5	0.0245	0.0246	0.0590	0.0339
		0.95	0.0420	0.0410	0.1146	0.0627
	δ_1	0.5	0.0209	0.0244	0.0419	0.0169
		0.95	0.0464	0.0594	0.1356	0.0492
M2	δ_1	0.5	0.0195	0.0179	0.0892	0.0592
		0.95	0.1214	0.0855	0.5056	0.2186
	δ_1	0.5	0.0264	0.0288	0.0836	0.0366
		0.95	0.2005	0.1562	0.4819	0.2736

Table 2: MISE for **CL3**

Result for CL3 cont.

- MISE for SPQR and DVQR perform better than their counterparts for all parameters and quantile level
- 50% quantile has better fit than 95% since the estimator for 50% quantile is more robust
- Copula based models are prefered methods in this scenario.

Result for t5

Marginals	Parameters	α	SPQR	DVQR	LQR	BAQR
M1	R_1	0.5	0.0782	0.0577	0.0297	0.0539
		0.95	0.2811	0.2260	0.2944	0.2543
	R_2	0.5	0.0619	0.0472	0.0160	0.0389
		0.95	0.1978	0.1686	0.2002	0.1727
M2	R_1	0.5	1.9064	1.8895	1.8176	1.9393
		0.95	2.9410	1.3998	1.2937	0.9686
	R_2	0.5	0.9067	0.8850	0.9339	0.9488
		0.95	3.3226	3.2512	0.9113	1.6017

Table 3: MISE for **t5**

Result for t5 cont.

- DVQR and SPQR has similar performance compared with linear and boosting additive methods at 50% quantile level for $M1$.
- DVQR has the best result for $M1$ at the tail, since LQR can't model tail quantile effectively.
- When the marginal distributions are skewed, the MISE for SPQR and DVQR larger since the estimation of marginal distributions are more imprecise.

Result for N5

σ	Parameters	SPQR	DVQR	LQR	BAQR
σ_1	0.5	0.2689	0.2672	0.3648	0.1645
	0.95	0.5972	0.5296	0.7608	0.4339
σ_2	0.5	0.3052	0.2923	0.3757	0.2150
	0.95	0.3627	0.3598	0.4687	0.2951

Table 4: MISE for **N5**

Result for N5 cont.

- Author [Dette et al. (2014)] argues that non-monotonic relationship between the response and predictor variables cannot be modeled by a parametric copula.
- This simulation confirms the results. Both SPQR and DVQR underperformed compared with boosting additive models.
- However, it's still preferable to linear model.

Application

- We will attempt to model the interdependence in the Australian equity market.
- We want to be able to measure the median return of a stock as well as the risk of extreme returns.
- Consider two cases: How the return of a equity is affected by (1)equities outside the industry (2)equities within the industry

Methodology

- Sample 7 equity from the top 30 equity S&P/ASX 200 index
- 3032 daily observations of log returns (01/01/2006 - 01/01/2018)
- Divide data into two sets: evaluation set and validation set. A quarter of the sample to be in the evaluation set, so that $n_{train} = 2274$ and $n_{eval} = 758$.
- Performance measured with averaged tick loss

$$L_{\alpha}^{m} = \frac{1}{n_{eval}} \sum_{i=1}^{n_eval} \rho_{\alpha}(y^{(i)} - \widehat{q}_{\alpha,m}^{(i)})$$

where $\rho_{\alpha}(y) = y(\alpha - I_{(y<0)})$

Outside industry

- Response Y from the industry goods sector and X_1, X_2 from basic material sector and X_3 from energy sector.
- Copula based models outperformed their counter parts.
- Moving away from the median to the tail quantiles linear model can't effectively describe the tail behaviour of t-copula.

α	SPQR	DVQR	LQR	BAQR
0.05	1.2223	1.1886	1.4300	1.3767
0.50	0.7510	0.7405	0.7867	0.7715
0.95	0.2129	0.2122	0.3650	0.2732

Table 5: Tick loss for between industries

Within industry

- Response Y and predictor X_1, X_2 all from the real estate sector
- Copula based model performed worse than linear and boosting additive methods for all quantile level.
- Equity within the same sector will usually have the same movement, and thus their quantile dependencies will be more linear like.

α	SPQR	DVQR	LQR	BAQR
0.05	0.2622	0.2470	0.1721	0.1336
0.50	0.4660	0.4582	0.3880	0.3814
0.95	0.1714	0.1686	0.1566	0.1182

Table 6: Tick loss for within industry

The end

Thank you!

Reference

Bedford, T. & Cooke, R. M. (2002), 'Vines: A new graphical model for dependent random variables', *Annals of Statistics* pp. 1031–1068.

Koenker, R. (2011), 'Additive models for quantile regression: Model selection and confidence bandaids', *Brazilian Journal of Probability and Statistics* **25**(3), 239–262.

Koenker, R. & Bassett Jr, G. (1978), 'Regression quantiles', *Econometrica: journal of the Econometric Society* pp. 33–50.

Kraus, D. & Czado, C. (2017), 'D-vine copula based quantile regression', *Computational Statistics & Data Analysis* **110**, 1–18.