Vine Copula-based Asymmetry and Tail Dependence Modeling

Jia Xu and Longbing Cao

Advanced Analytics Institute
University of Technology, Sydney
Jia.Xu-3@student.uts.edu.au, Longbing.Cao@uts.edu.au

Abstract. Financial variables such as asset returns in the massive market contain various hierarchical and horizontal relationships that form complicated dependence structures. Modeling these structures is challenging due to the stylized facts of market data. Many research results in recent decades showed that copula is an effective method to describe relations among variables. Vine structures were introduced to represent the decomposition of multivariate copula functions. However, the model construction of vine structures is still a tough problem owing to the geometrical data, conditional independent assumptions and the stylized facts. In this paper, we introduce a new bottom-up method to construct regular vine structures and applies the model to 12 currencies over 16 years as a case study to analyze the asymmetric and fat-tail features. The out-of-sample performance of our model is evaluated by Value at Risk, a widely used industrial benchmark. The experimental results show that our model and its intrinsic design significantly outperform industry baselines, and provide financially interpretable knowledge and profound insight into the dependence structures of multi-variables with complex dependencies and characteristics.

1 Introduction

Modeling complex dependence structures of financial variables is a fundamental research problem in the financial domain, useful for a wide range of applications including economics prediction and risk management. Its extreme importance has been partially demonstrated in the 2008 global financial crisis (GFC). Existing studies are usually concerned with the degree of dependencies rather than the other important respects of dependencies – the dependence structure, especially the asymmetric and tail dependence characteristics. However, as demonstrated in GFC, it is useless when all stocks tend to fall as the market falls.

Asymmetric dependencies between different markets can be easily seen from Fig. 1. Fig. 1(a) shows the correlation of daily returns between the United States comprehensive index S&P500 and the index of Eurozone stocks STOXX50E, which indicates the strong negative dependence and normal positive dependence between them. The dependence between the United Kingdom comprehensive index FTSE100 and the foreign exchange rate GBP against the USD is shown
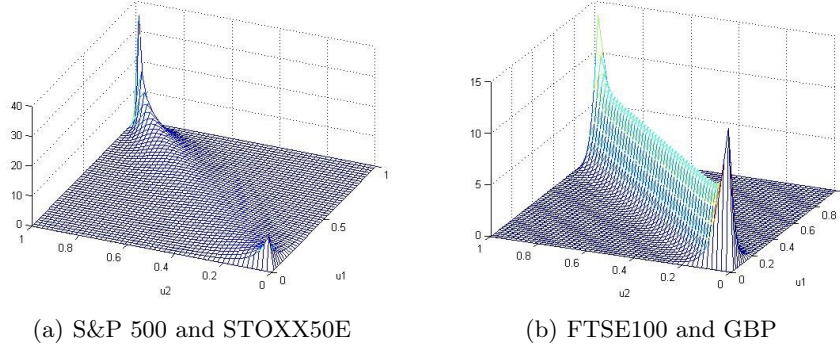
(a) S&P 500 and STOXX50E   (b) FTSE100 and GBP

Fig. 1. Asymmetric Dependencies across Markets

in Fig. 1(b), which indicates strong dependence on both sides. These examples show that financial markets are not only dependent but also asymmetric.

The challenge of modeling dependencies in financial markets lies in the three major aspects concerning us in this paper. (1) As with any complex behavioral and social system, the cross-market dependence structure is often embedded with strong and complicated coupling relationships [7, 5, 6] on high dimensionality. Flexible dependence structure without imposing any assumptions or restrictions are desired. (2) Financial variables, such as daily return, have been shown to follow non-normal distributions, which means dependence models should cover a wide range of dependencies in order to capture both positive and negative dependencies. (3) As discussed above, various lower and upper tail dependencies also need to be considered.

The dependencies across markets have been studied by different communities, including statistics and machine learning. The typical approaches in the statistical community are joint distributions with Gaussian assumption and conditional correlation. The first method has been demonstrated that Gaussian assumption is inappropriate when studying either stock markets or exchange rate markets. The second one uses conditional correlation to calculate the covariance, which is generally used in empirical studies. As the current correlation depends on the previous one, the dependence structure is not flexible. The dependence studies in the machine learning community consist of hidden Markov models and graphical probability models [7]. The hidden Markov models, however, could have a large number of hidden states when applied to a high dimensional case, which invariably leads to computational intractability in the algorithms when inferring the hidden states from observations. The graphical probability models, such as Bayesian logic program [14], impose unrealistic assumptions in constructing dependence structures.

In recent decades, a number of research methods based on regular vine model capture the asymmetric dependencies in currency markets and show decent effects [8, 13, 9, 17]. A popular methodology is to combine the time series models (e.g., ARMA, GARCH) and copula to observe the joint distribution on multi-

variates. With this framework, we can simplify the problems about observing joint distribution into two parts: the marginal distribution of each variable and copula between variables. The multivariate Archimedean copula family has been studied in [2, 15]. They show that Gaussian Copula models do not have lower and upper tail dependence, while the multivariate t copula does not have flexible tail dependence as the symmetric structure of t copula. Vine structures including canonical vine and D vine copula models [1, 3] can implement a wide range of dependencies by decomposing the multivariate copula into different bivariate copulas. However, due to the structure assumptions, they do not have flexible dependence structures. Due to the assumptions imposed on dependence structures, their dependence structures may not reflect the actual dependence in the real world.

In order to model the asymmetric dependencies in multivariate data with various dependence structures, we propose a new weighted partial regular vine copula model (WPRV) with asymmetric dependencies. WPRV is more powerful, because: (1) A new partial correlation-based algorithm constructs the regular vine structure. Our WPRV can uniquely determine the correlation matrix and is algebraically independent without any strong restriction on the dependence structure. The dependence structure is more flexible, since the current tree structure is independent of the established tree structure and bivariate copulas selection. (2) The bivariate copula with different types of tail dependencies (e.g., BB1, survival BB1, BB7 and survival BB7) are implemented to capture various tail dependencies between financial variables. (3) The moving trends of lower and upper tail dependence with the multivariate data structures and also the trends of lower and upper tail dependence during the dynamic period are analyzed.

The rest of paper is structured as follows. Section 2 introduces the related definitions of copula and different tail dependencies. Section 3 discusses how to construct our weighted partial regular vine copula model, copula family selection, and the parameter estimation in partial regular vine copula and marginal distribution. The case study results are shown in Section 7. Finally, Section 8 concludes the paper.

2 Preliminaries

2.1 Vine Copula

Vine theory was introduced in [4], which is one kind of graphical models. Let V, T, E and N represent vine structure, trees, edges, nodes respectively. The regular vine and its related definitions are given below.

Definition 1. *(Regular Vine) V is a regular vine on n variables if*

(1) T_1 is a tree with nodes $N_1 = 1, ..., n$ and a set of edges denoted by E_1;
(2) For $j = 2, ..., n - 1$, T_j is a tree with nodes $N_j = E_{j-1}$ and edge set E_j;
(3) (proximity condition) For $j = 2, ..., n - 1$ and $a, b \in E_j$, $\#(a \triangle b) = 2$, where $\triangle$ denotes the symmetric difference operator and $\#$ denotes the cardinality.

Definition 2. *(Complete Union, Conditioning and Conditioned Sets of an Edge) The complete union of an edge $e_i \in E_i$ is the set $U_{e_i} = \{n_1 \in N_1 \mid \exists e_j \in E_j, j = 1, 2, \ldots, i - 1 \text{ with } n_1 \in e_1 \in e_2 \in \ldots \in e_{i-1} \in e_i\} \subset N_1$. For $e_i = \{a, b\} \in E_i, a, b \in N_i, i = 1, 2, \ldots, n - 1$, the conditioning set of an edge e_i is $D_{e_i} = U_a \cap U_b$, and the conditioned sets of an edge e_i are $C_{e_i,a} = U_a \setminus D_{e_i}, C_{e_i,b} = U_b \setminus D_{e_i}$ and $C_{e_i} = C_{e_i,a} \cup C_{e_i,b} = U_a \triangle U_b$, where $A \triangle B := (A \setminus B) \cup (B \setminus A)$ denotes the symmetric difference of two sets.*

Hence, U_{ei} is a set of all nodes in N_i that are connected by the edges e_i. By definition, $U_{ei}(1) = e_i$. The constraint set is defined below.

Definition 3. *(Constraint Set) The constraint set for V is a set:*
$$CV = \{(\{C_{e_a}, C_{e_b}\}, D_e) \mid e \in E_i, e = \{a, b\}, i = 1, \ldots, n - 1\}$$

The edge e can be written as $\{Ce \mid De\}$, or $\{C_{e(a)}, C_{e(b)} | D_e, e = \{a, b\}\}$, where the conditioning set D_e is shown to the right of "|", and the conditioned set C_e to the left. $\{U_a \setminus D_e\}$ is the set which includes all variables in the set U_a, but excludes the variables in the set D_e.

2.2   Tail Dependencies

One important copula-based dependence measurement is *tail dependence coefficient*, which indicates the dependencies between extreme events. The extreme dependence of a multivariate distribution F can be described by various tail dependence parameters of its copula C. Suppose that random vector $(U_1, ..., U_n) := (F_1(x_1), ..., F_n(X_n))$ with standard uniform marginal distribution. The lower and upper tail dependence coefficients are defined as follows.

$$
\begin{aligned}
\lambda_L &= \lim_{u \to 0} Pr\{U_1 \le u, ..., U_n \le u \mid U_n \le u\} \\
&= \lim_{u \to 0} \frac{C(u, ..., u)}{u} \\
\lambda_U &= \lim_{u \to 0} Pr\{U_1 > 1 - u, ..., U_n > 1 - u \mid U_n > 1 - u\} \\
&= \lim_{u \to 0} \frac{\overline{C}(1 - u, ..., 1 - u)}{u}
\end{aligned}
\tag{1}
$$

where $\overline{C}$ is the survival function of C. If λ_U exists and $\lambda_U \in (0, 1]$, then copula C has an upper tail dependence coefficient, but there is no upper tail dependence coefficient when $\lambda_U = 0$. Similarly, if λ_L exists and $\lambda_L \in (0, 1]$, then copula C has an upper tail dependence coefficient, but no upper tail dependence coefficient when $\lambda_L = 0$.

Frahm et al. [10] proposed a non-parametric method to obtain the nonparametric estimator of lower and upper tail dependence by using *Pickand's dependence function* [16]. One simple nonparametric estimator of tail dependence is the log estimator, which is denoted by :

$$
\begin{aligned}
\hat{\lambda}_L &= 2 - \lim_{u^* \to 0} \frac{log\left(1 - 2(1 - u^*) + T_{-1} \sum_{t=1}^{T} \mathbf{1}\{U_1 \le 1 - u^*, ..., U_n \le 1 - u^*\}\right)}{log(1 - u^*)} \\
\hat{\lambda}_U &= 2 - \lim_{u^* \to 0} \frac{log\left(T_{-1} \sum_{t=1}^{T} \mathbf{1}\{U_1 \le 1 - u^*, ..., U_n \le 1 - u^*\}\right)}{log(1 - u^*)}
\end{aligned}
\tag{2}
$$

In this work, the above nonparametric method is implemented for roughly analyzing the tail dependence coefficient before the regular vine model construction.

3 Our Weighted Partial Regular Vine Model

As highlighted in the introduction, our WPRV is centered on the bottom-to-top regular vine structure. According to the method proposed by Bedford and Cooke [4] for building vine structure by using partial correlation, for elliptical distributions, partial correlation is equal to the corresponding conditional correlation. We can thus use partial correlation instead of conditional correlation to measure the correlations on each node and every tree does not depend on the structure of the previous tree, which is more flexible.

3.1 Partial Regular Vine Construction

The regular vine on n variables shares several important properties (see details in [11]):

(1) There are $(j-1)$ and $(j+1)$ variables in the conditioning sets and constraint sets of an edge of the jth tree respectively;

(2) If two or more nodes have the same constraint sets, they are the same node;

(3) If variable i is a member of the conditioned set of an edge e in a regular vine, then i is a member of the conditioned set of exactly one of the m-child of e, and the conditioning set of an m-child is a subset of De.

(4) If two or more nodes have the same constraint sets, they are the same node;

(5) If variable i is a member of the conditioned set of an edge e in a regular vine, then i is a member of the conditioned set of exactly one of the m-child of e, and the conditioning set of an m-child is a subset of D_e.

According to the above properties, we derive two lemmas, which are important for constructing the partial regular vine tree structure. The two lemmas are given as follows.

Lemma 1. *Let $I \in \{1, ..., n\}$, $x_1, x_2, y_1, y_2 \in I$ and $x_1 \neq x_2$, the nodes of T_j be $N_1 = \{x_1, y_1 ; I\backslash\{x_1, x_2, y_1\}\}$ and $N_2 = \{x_2, y_2 ; I\backslash\{x_1, x_2, y_2\}\}$. For a regular vine on n variables, nodes N_1 and N_2 have a common m-child. If $y_1 \neq y_2$, the common m-child is $\{y_1, y_2 ; I\backslash\{x_1, x_2, y_1, y_2\}\}$.*

Proof. According to Definition 1, each node has two m-children. For N_1, the constraint set CV_{x_1} of its m-children are $\{x_1, I\backslash\{x_1, x_2, y_1\}\}$ and $\{y_1, I\backslash\{x_1, x_2, y_1\}\}$. For N_2, the constraint set CV_{x_2} of its m-children are $\{x_2, I\backslash\{x_1, x_2, y_2\}\}$ and $\{y_2, I\backslash\{x_1, x_2, y_2\}\}$. We can see that $\{y_1, I\backslash\{x_1, x_2, y_1\}\}$ and $\{y_2, I\backslash\{x_1, x_2, y_2\}\}$ are equal, but indexed by different variables in a conditioned set. According to Property (4), N_1 and N_2 have a common m-child. If $y_1 \neq y_2$, y_1 and y_2 should be in the conditioned set of the m-child.

Algorithm 1 Tree Structure Construction via A Bottom-to-Top Strategy

---

Require: Observations of n input variables
 1: Calculate all values of partial correlation, and then allocate the smallest absolute value of partial correlation to the node in T_{n-1} (T_{n-1} is the bottom tree).
 2: **for** $k = 1, \ldots, n-2$ **do**
 3:     **for** $i = n-1, \ldots, \lceil \frac{n}{2} \rceil$ **do**
 4:         **if** $T_i > T_k$ **then**
 5:             Find variable combinations for nodes on both sides in tree T_i which can minimize the function $|\rho_{c:d}|$, where T_i indicates the ith tree and T_k is tree inverse level tree;
 6:         **else**
 7:             Find variable combinations for nodes on both sides in tree T_i which can minimize the function of $\sum ln(1 - \rho_{c:d}^2)$
 8:         **end if**
 9:     **end for**
10: **end for**
11: There will be $n-2$ regular vines as $k = 1, \ldots, n-2$.
12: **return** A group of partial regular vine dependence structure candidatures.

Lemma 2. *For a regular vine on n variables, $j = 2, ..., n-1$, the edge e in T_j has only two constraint sets of m-children in T_{j-1}, which are indexed by different variables in a conditioned set.*

Proof. Suppose there are three identical constraint sets indexed by different variables in a conditioned set, according to Property (4), nodes with the same constraint sets should be the same node. Based on Property (5), the variables in the conditioned set will still be in the conditioned set of its m-children. This means that the node will have three variables in its conditioned set, which violates Property (3) and the proximity condition in the regular vine definition. Therefore, one edge has only two constraint sets which are indexed by different variables in a conditioned set.

According to the above properties of regular vine and the two lemmas, we construct the partial regular vine by using Algorithm 1.

3.2   Vine Structure Selection

After building the $\lceil \frac{n}{2} \rceil$ candidate regular vines, the next step is to find the 'Best' regular vine among these candidates. In order to remove the bias by only selecting the strongest correlation on the top, giving a weight to each tree can enhance the influence of the trees on the top or at the bottom, and a balanced structure can be selected. We assume each level is a unit height and the tree inverse level k is the zero potential energy level. Hence, the weight of each level will increase from level k to level $n-1$ and level 1. Since the parameters m, g, K and T are constant in a given environment, Equation (3) can be simplified as:

$$W = e^{-m_0 h} \tag{3}$$

where m_0 is the parameter and h is the distance from the level of each tree to the tree inverse level. To restrict the value of weight for each level in interval $[0, 1]$, we standardize the weight:

$$W(h) = \begin{cases} 0.5 \times \dfrac{e^{-m_0(k-h)}}{\sum_{i=1}^{k} e^{-m_0(k-i)}}, & h \in [1, k]; \\[4mm] 0.5 \times \dfrac{e^{-m_0(h-k)}}{\sum_{i=k+1}^{N-1} e^{-m_0(i-k)}}, & h \in (k, N-1]. \end{cases} \tag{4}$$

where N is the number of variables, k is the tree inverse level, h is the level of a tree and m_0 is a parameter which falls in interval $[0, 1]$.

The 'Best' regular vine structure maximizes the value of function $-ln(D)$, where D is the weighted determinant which is calculated as:

$$D = \prod_{i,j} (1 - W_i \rho_{i,j;d(i,j)}^2) \tag{5}$$

where W_i is the corresponding weight and $d(i, j)$ is the conditioning set excluding variables i and j. The corresponding conditioned set is i and j.

3.3   Bivariate Copula Selection

Once the partial regular vine tree structure is identified, the next step is to select bivariate copulas for each edge in all trees. As discussed above, the partial correlation is equal to its corresponding conditional correlation for the elliptical family. This means our partial regular vine tree structure is built based on an elliptical copula family (i.e., Gaussian or t copulas). However, according to the following theorem, the limitation of partial correlation can be removed by mapping the partial regular vine tree structure to typical regular vine via conditional correlation.

Theorem 1. *For any regular vine on n variables, there is one-to-one correspondence between the set of $n \times n$ positive definite correlation matrices and the set of partial correlation specification of the vine.*

The proof of Theorem 1 can be referred to [4], which is omitted here. It shows that there is a one-to-one relationship between the partial regular vine specification and the correlation matrix, which ensures that we can map our partial regular vine tree structure to the typical conditional correlation based-regular vine tree structure. We can then choose bivariate copulas from a large number of copula family candidates, rather than the elliptical copula family. Hence, the limitation can be removed while selecting the bivariate copulas.

There are a huge of copula families, which have various tail dependencies. The details of tail dependence of copula families are listed in Table 1. To capture the asymmetric characteristics, the BB1, S.BB1, BB7 and S.BB7 copulas are the best choice since they have various lower and upper tail dependencies, which can vary independently from 0 to 1.

Table 1. The Tail Dependence of Copula Family

	Lower Tail Dependence	Upper Tail Dependence
Gaussian	-	-
t	$2t_{\nu+1}(\mu(\nu,\phi))$	$2t_{\nu+1}(\mu(\nu,\phi))$
Gumbel	-	$2^{-1/\phi}$
Frank	-	-
Clayton	$2^{-1/\phi}$	-
Joe	-	$2 - 2^{1/\phi}$
BB1	$2^{-1/(\phi\delta)}$	$2 - 2^{1/\delta}$
S.BB1	$2 - 2^{1/\delta}$	$2^{-1/(\phi\delta)}$
BB6	-	$2 - 2^{1/(\delta\phi)}$
BB7	$2 - 2^{1/\delta}$	$2 - 2^{1/\phi}$
S.BB7	$2 - 2^{1/\phi}$	$2 - 2^{1/\delta}$
BB8	-	$2^{-1/\phi}$ when $\delta = 1$

S.BB1 and S.BB7 are survival BB1 and BB7 copula respectively. ϕ and δ are parameters of the corresponding copula family. For t copula, $\mu(\nu,\phi) = \left(-\sqrt{\nu+1}\sqrt{\frac{1-\phi}{1+\phi}}\right)$.

3.4   Marginal Distribution Specification and Parameter Estimation

For the financial applications of the partial regular vine copula model, we use volatility models (i.e., $ARMA$-$GARCH$ models) as the margins. Typically, let $X_t(t = 0, 1, ..., \mathfrak{T})$ be a time series of the prices of a financial asset, such as the stock market index. The return of financial asset can be defined as $r_t = log(X_t/X_{t-1})$. If there are n assets with returns $r_{t,1}, ..., r_{t,n}$, we first select the appropriate marginal distribution of individual variables (i.e., returns of financial variables), which is a univariate distribution. Due to the characteristics of financial assets, such as volatility cluster, a common choice is $ARMA(1,1)$-$GARCH(1,1)$ with skewed student t innovations, which is defined as follows.

$$
\begin{aligned}
r_{t,j} &= c_j + \Phi_j r_{j,t-1} + \Theta_j \varepsilon_{j,t-1} + \varepsilon_{j,t}, \\
\varepsilon_{j,t} &= \sigma_{j,t} \cdot Z_{j,t} \\
\sigma_{j,t}^2 &= \omega_j + \alpha_j \varepsilon_{j,t-1}^2 + \beta_j \sigma_{j,t-1}^2
\end{aligned}
\tag{6}
$$

where $j = 1, ..., n, t = 1, ..., \mathfrak{T}$ and $Z_{j,t}$ is the innovations which follow skewed student t distribution.

Let $\theta_j^m = (c_j, \Phi_j, \Theta_j, \omega_j, \alpha_j, \beta_j)$ be the parameter set of marginal distribution, θ^c be the parameters of multivariate copula functions, the multivariate joint log-likelihood is given by:

$$
\begin{aligned}
L(\theta_1^m, ..., \theta_n^m, \theta^c) &= \sum_{t=1}^{\mathfrak{T}} log f(r_{1,t}, ..., r_{n,t}; \theta_1^m, ..., \theta_n^m, \theta^c) \\
&= \sum_{t=1}^{\mathfrak{T}} \log c(F_1(r_{1,t}), ..., F_n(r_{n,t}); \theta^c) + \sum_{t=1}^{\mathfrak{T}} \sum_{j=1}^{n} \log f_j(r_{j,t}; \theta_j^m)
\end{aligned}
\tag{7}
$$

where the multivariate $c(\cdot; \theta^c)$ is denoted as the regular vine model.

4   Case Study

4.1   Data and Marginal Distribution Specification

To evaluate the performance of our model, we use real-world data, involving 12 currency exchange rates against USD. These trading currencies are EUR, GBP, CHF, SEK, CAD, BRL, AUD, NZD, JPY, HKD, SGD, and INR, which are sequentially numbered from $v1$ to $v12$. They represent major currencies in the global market and can be arranged into portfolios. The training data set uses observations from 04/01/1999 to 27/08/2004, a total of 1298 daily returns. Observations from 6/09/2004 to 21/06/2013, a total 1912 daily returns are used for out-of-sample testing. All the data was downloaded from Yahoo Finance (http://finance.yahoo.com/).

As discussed in the last section, the standardized residuals are transferred to uniform data by using the empirical probability integral transformation, which is actually the input of partial regular vine. The raw returns are fitted with univariate $ARMA(1,1) - GARCH(1,1)$ models with the skewed student-t error distribution. The Ljung-Box (LB) test [12] is introduced to remove the autocorrelation among these financial returns. In this experiment, the corresponding p values of the LB test are all greater than the significant value 0.05.

4.2   Regular Vine Copula Structure Specification and Tail Dependence Analysis

The next step is to build our weighted partial regular vine copula model. Fig. 2 shows the tree structure built by Algorithm 1. Due to space limitations, we only show the last three trees of our vine structure.

Typically, the selection of m_0 of the WPRV model is determined by the characteristics of data and domain knowledge. According to the discussion in Section 3.2, m_0 is restricted to interval $[0, 1]$. Table 2 shows the performance of the Log-likelihood of WPRV with parameter m_0 from 0.1 to 1.0. The high value of the Log-likelihood indicates good performance. According to Table 2, the WPRV model with parameter $m_0 = 0.7$ achieves the best performance.

Table 2. Log-likelihood Performance of WPRV with Parameter m_0

m_0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
LL	2767.23	2767.23	2929.91	2929.91	3031.56	3031.56	3031.56	2929.91	2929.91	2767.23

[1] LL is short for Log-likelihood.

Once the structure is identified, the next step is to choose the copula for each edge. As discussed above, the bivariate copula which can provide flexible lower and upper tail dependencies is the most appropriate to build the partial vine copula model with asymmetric dependencies. Based on Section 3.3, BB1, S.BB1, BB7 and S.BB7 copulas can provide both lower and upper tail dependencies. Therefore, the BB1, BB7, S.BB1 and S.BB7 copulas are used to build vine copula model with asymmetric dependencies to capture the asymmetric characteristics.
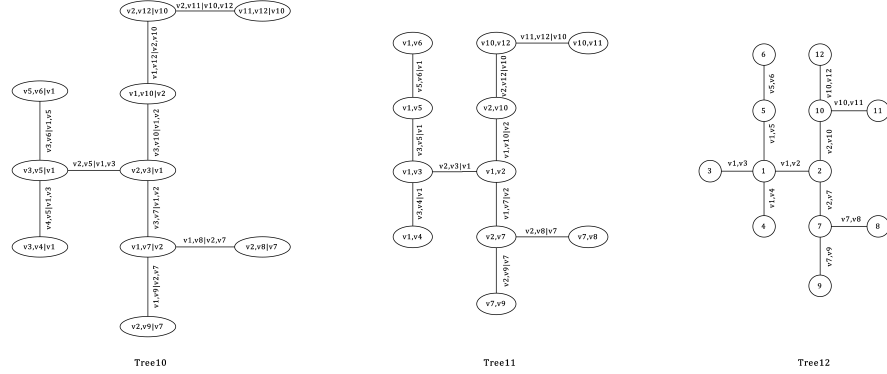
Fig. 2. Weighted Partial R vine Trees

The tail dependence in Tree 12 of Fig. 2 is shown in Table 3. The non-parametric and t copula results are listed as reference. It shows lower tail dependencies of pairs in Tree 1 are less than their corresponding upper tail dependencies. Although various bivariate copulas provide different results, similar conclusion can be drawn that lower tail dependencies are less than upper ones.

Table 3. Tail Dependence Analysis by Using Various Copulas

	Non-para[1]		t		BB1		S.BB1[2]		BB7		S.BB7[2]	
	λ_L	λ_U	λ_L	λ_U	λ_L	λ_U	λ_L	λ_U	λ_L	λ_U	λ_L	λ_U
$\{v1, v2\}$	0.18	0.33	0.25	0.25	0.28	0.41	0.39	0.62	0.41	0.48	0.46	0.44
$\{v1, v3\}$	0.45	0.23	0.37	0.37	0.45	0.61	0.50	0.69	0.61	0.68	0.63	0.67
$\{v1, v4\}$	0.50	0.60	0.54	0.54	0.63	0.76	0.71	0.78	0.76	0.82	0.77	0.82
$\{v1, v5\}$	0.73	0.63	0.71	0.71	0.74	0.81	0.78	0.82	0.83	0.85	0.82	0.87
$\{v1, v6\}$	0.28	0.49	0.42	0.42	0.59	0.70	0.65	0.76	0.72	0.74	0.73	0.76
$\{v2, v7\}$	0.29	0.42	0.19	0.19	0.41	0.56	0.49	0.68	0.53	0.63	0.58	0.60
$\{v2, v10\}$	0.53	0.60	0.63	0.63	0.72	0.80	0.78	0.82	0.79	0.83	0.80	0.83
$\{v7, v8\}$	0.28	0.43	0.35	0.35	0.37	0.51	0.49	0.72	0.51	0.58	0.56	0.53
$\{v7, v9\}$	0.20	0.32	0.23	0.23	0.30	0.45	0.39	0.56	0.43	0.51	0.46	0.51
$\{v10, v11\}$	0.55	0.36	0.51	0.51	0.55	0.69	0.65	0.80	0.72	0.80	0.73	0.77
$\{v10, v12\}$	0.38	0.57	0.47	0.47	0.50	0.59	0.53	0.65	0.61	0.65	0.62	0.65

[1] Non-para means that the tail dependence coefficient is calculated via the nonparametric method;
[2] S.BB1 and S.BB7 are the survival BB1 and BB7 copula respectively.

In order to investigate the tail dependence and its movement trend, two different fixed periods (24 months and 36 months) are used as the investigation period of tail dependencies to show the relationship between the length of period and the movement trend. Then, a moving window of 620 daily observations is introduced, from 07/02/2011 to 21/06/2013. The result of pair $\{v1, v2\}$ in tree 1 is shown in Fig. 3.

The gap in a short investigation period with 24 months is larger than those in a long investigation period with 36 months. It indicates that the difference between lower and upper tail dependencies is more significant in a short inves-

tigated period than in a long one. However, the difference decreases when the length of investigation period increases.
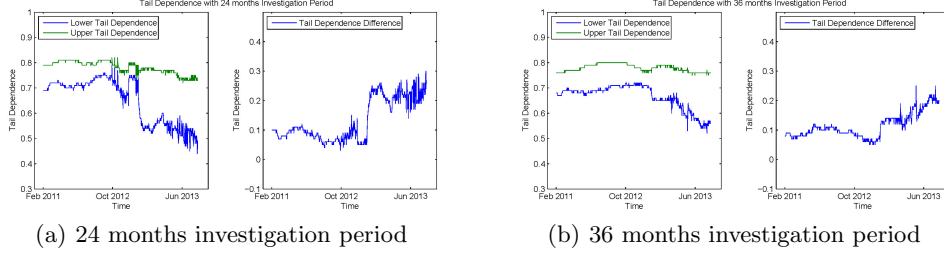

(a) 24 months investigation period   (b) 36 months investigation period

Fig. 3. Lower and Upper Tail Dependencies

4.3   Out-of-Sample Performance Analysis

The out-of-sample performance is evaluated by the Value at Risk (VaR), which a widely used industrial benchmark. Typically, backtesting methods based on Log-likelihood ratios and a null hypothesis consist of unconditional and conditional coverage tests. A large p-value indicates that the VaR forecastings are accurate and reliable. However, the p-value should at least be greater than 0.05. In this experiment, we do not use any machine learning models since they do not directly support forecasting VaR. Table 4 presents the backtesting results of WPRV, canonical vine and D vine with various copula. The bivariate copula selection is indicated in the second row. The results indicate the BB1 and S.BB1 copulas have the best performance, followed by the BB7 and S.BB7 copulas. The model with t copula is better than Clayton, Gumbel and BB6 copulas which have only one tail dependencies.

   In conclusion, the results of VaR forecasting indicate that the partial regular vine copula with asymmetric dependencies is better than those with symmetric lower and upper tail dependencies, and the models with two tail dependencies are better than those with only one tail dependencies.

5   Conclusion and Future Work

Modeling the dependencies between multivariate variables in asymmetric and tail-dependent data is very challenging in demanding applications related to big data and financial business, and existing methods cannot handle it well. This paper presents a weighted partial regular model to resolve the issue without imposing restrictions on dependence structures. The model is demonstrated through analyzing the complicated structures of portfolios in currency markets. The out-of-sample performance evaluation results highly outperform other methods from statistic and risk evaluation perspectives. Our future work will explore the performance of the regular vine with other high-dimensional time series data.

Table 4. The Backtesting Results of Value at Risk Forecasting

	$1-\alpha$	Weighted Partial Regular Vine							Canonical Vine		D Vine	
		BB1	S.BB1	BB7	S.BB7	t	Clayton	Gumbel	BB1	BB7	BB1	BB7
LR_{uc}	99%	0.002	0.002	0.100	0.100	0.048	0.354	3.74	0.048	0.248	0.100	0.100
		(0.964)	(0.964)	(0.751)	(0.751)	(0.626)	(0.552)	(0.503)	(0.826)	(0.618)	(0.751)	(0.751)
	95%	0.020	0.068	0.068	0.145	0.051	0.224	1.466	0.145	0.385	0.220	0.145
		(0.899)	(0.794)	(0.794)	(0.703)	(0.621)	(0.636)	(0.264)	(0.703)	(0.309)	(0.488)	(0.225)
	90%	0.473	0.919	0.023	0.175	1.101	1.101	0.258	0.357	0.385	0.423	0.175
		(0.492)	(0.338)	(0.880)	(0.676)	(0.294)	(0.294)	(0.212)	(0.550)	(0.535)	(0.338)	(0.676)
LR_{cc}	99%	0.282	0.282	0.341	0.341	0.369	0.859	0.934	0.369	0.569	0.641	0.541
		(0.869)	(0.869)	(0.843)	(0.843)	(0.831)	(0.651)	(0.609)	(0.831)	(0.831)	(0.467)	(0.869)
	95%	1.436	2.599	2.599	2.479	1.946	1.729	2.662	1.685	2.351	1.436	1.436
		(0.488)	(0.273)	(0.273)	(0.290)	(0.378)	(0.421)	(0.264)	(0.431)	(0.309)	(0.488)	(0.488)
	90%	1.467	2.316	0.862	0.837	1.613	1.997	1.023	1.633	1.374	2.316	1.387
		(0.480)	(0.314)	(0.650)	(0.658)	(0.446)	(0.369)	(0.312)	(0.442)	(0.503)	(0.394)	(0.363)

[1] LR_{uc} and LR_{cc} are short for the likelihood ratio of unconditional and conditional coverage respectively. The first row shows the value, while the corresponding p value is given the parenthesis in the following row. The critical value of LR_{uc} and LR_{cc} are 3.841 and 5.991.

References

1. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. Insurance: Mathematics and economics 44(2), 182–198 (2009)
2. Abdous, B., Genest, C., Rémillard, B.: Dependence properties of meta-elliptical distributions. In: Statistical modeling and analysis for complex data problems, pp. 1–15. Springer (2005)
3. Almeida, C., Czado, C., Manner, H.: Modeling high dimensional time-varying dependence using d-vine scar models. arXiv preprint arXiv:1202.2008 (2012)
4. Bedford, T., Cooke, R.M.: Vines: A new graphical model for dependent random variables. Annals of Statistics pp. 1031–1068 (2002)
5. Cao, L.: Non-iidness learning in behavioral and social data. The Computer Journal p. bxt084 (2013)
6. Cao, L.: Coupling learning of complex interactions. J. Information Processing and Management 51(2), 167–186 (2015)
7. Cao, L., Ou, Y., Yu, P.S.: Coupled behavior analysis with applications. IEEE Trans. on Knowledge and Data Engineering 24(8), 1378–1392 (2012)
8. Czado, C., Brechmann, E.C., Gruber, L.: Selection of vine copulas. In: Copulae in Mathematical and Quantitative Finance, pp. 17–37. Springer (2013)
9. Dissmann, J., Brechmann, E.C., Czado, C., Kurowicka, D.: Selecting and estimating regular vine copulae and application to financial returns. Computational Statistics & Data Analysis 59, 52–69 (2013)
10. Frahm, G., Junker, M., Schmidt, R.: Estimating the tail-dependence coefficient: Properties and pitfalls. Insurance: Mathematics and Economics 37(1), 80–100 (2005)
11. Kurowicka, D., Cooke, R.M.: Uncertainty analysis with high dimensional dependence modelling. John Wiley & Sons (2006)
12. Ljung, G.M., Box, G.E.: On a measure of lack of fit in time series models. Biometrika 65(2), 297–303 (1978)
13. de Melo Mendes, B.V., Semeraro, M.M., Leal, R.P.C.: Pair-copulas modeling in finance. Financial Markets and Portfolio Management 24(2), 193–213 (2010)
14. Neville, J., Jensen, D.: Relational dependency networks. Journal of Machine Learning Research 8, 653–692 (2007)
15. Nikoloulopoulos, A.K., Joe, H., Li, H.: Extreme value properties of multivariate t copulas. Extremes 12(2), 129–148 (2009)

16. Pickands, J.: Multivariate extreme value distributions. In: Proceedings 43rd Session International Statistical Institute. vol. 2, pp. 859–878 (1981)
17. Riccetti, L.: A copula–garch model for macro asset allocation of a portfolio with commodities. Empirical Economics 44(3), 1315–1336 (2013)