AMS517 Project

Review, replication, and additional empirical research on Chen and Fan (2006)'s estimation of copula-based semiparametric time series models

Chi-Sheng Lo & Jinglin Yang

Abstract

In this project, we summarize the contribution of Chen and Fan (2006) by discussing their estimation methodology on the copula-based semiparametric time series and making balanced critique on their strengths and weaknesses. Further, we update the literature review in order to know the latest development ever since their publication in 2006. Most importantly, we replicate the semiparametric bootstrap method suggested by Chen and Fan (2006) and finally, we conduct copula simulation and extend the research with empirical studies on interest spread with exceedence correlation and tail dependence.

Keywords: Bootstrap, copula, semiparametric estimation, nonlinear Markov models, conditional moment, conditional quantile, distributional dependence, exceedence correlation.

JEL classification: C6, C14, C22, E4.

1. Introduction

1.1 Objective

We have three main objectives. First, we aim to analyze the contributions of Chen and Fan (2006) by reviewing the methodologies and by discussing their weaknesses and strengths. Second, we desire to take a look at the development ever since their publication sixteen years ago. Third, we intend to conduct computational research with replication, simulation, and empirical study that cover the dynamics of copulas, semiparametric bootstrap, and tail dependence.

1.2 Background

Sklar (1959) pioneered the copula theory which stated that copula can capture all the scale-free dependence in the multivariate distributions. One of Skylar's theorem's strengths is that it can be extended to aid in the forecasting in time series. Darsow *et al.* (1992) set the tone for the necessary and sufficient conditions for a copula-based time series that exhibits a Markov process. Eventually, Chen and Fan (2006) rolled out the univariate copula-based semiparametric stationary Markov model which is composed by two unknown parameters. One is the copula dependence parameter α^* and another is the invariant (marginal) distribution function $G^*(\cdot)$. The former is estimated by the two-step estimator designed for bivariate copula models. The latter can be estimated by several nonparametric methods including the rescaled empirical distribution function and the kernel smoothed estimator.

Chen and Fan (2006) were the forerunner in laying the foundation for estimating the copula dependence parameter under the two-step procedure with large sample properties; in addition, they introduced estimators of moment and conditional quantile function, consistency, and $\sqrt{n}$-asymptotic normality with handily testable conditions. Finally, they concluded that the β-mixing temporal dependence measure is exclusively determined by the properties of copulas which does not depend on the invariant distributions.

Their conclusion also left room for improvement because they argued that their potential extensions should take into account three main features. First, for the serial independence, the asymptotic results can only be obtained once the true parameter value α^* is in the interior of the parameter space. To solve it, Chen and Fan (2006) suggested using a limiting distribution approach developed by Andrews (2001) who provided several asymptotic results over several testing problems with time series. Second, they should cover higher order Markov process because dimension reduction can be achieved when the process depends on nonparametric functions. Finally, they stressed that they should let dependence parameter to be time-varying in a Markov-switching manner.

1.3 Literature review of more recent development since 2006

Chen and Fan (2006) was published about sixteen years ago, thereby, we think it is necessary to update some of the latest literature in this project. Below is the summary of our collection that cover both theoretical and applied oriented literatures since 2006.

In the same year, Chen *et al.* (2006) proved that the plug-in sieve MLE of finite-dimensional copula parameters and the undisclosed marginal distributions are semi parametrically efficient. Thereafter, Chen, *et al.* (2009) estimated the empirical distribution of the GARCH residuals with a new weighted approximation which can be used to derive the asymptotic distributions of the pseudo MLE of the residual copula parameter and of the goodness-of-fit test statistic for testing the parametric specification of the residual copula.

Zimmer (2012) attempted to use other copulas instead of the Gaussian copula to gauge the risk in collaterized debt obligations in the housing market and found that the Clayton-Gumbel mixture can uncover a stronger relationship between housing prices in different geographic areas and generates a better fit to the data.

Noh *et al.* (2013) invented a new way to estimate a regression with copulas by writing the regression function in terms of a copula and marginal distributions and by using the plug-in method to create a new estimator. Dette *et al.* (2014) made a review on Noh *et al.* (2013) and they argued that the high dimensional predictors do not improve the properties of the estimator and concluded that the nonmonotonic features of the regression function cannot be reproduced by the copula-based regression estimate. Fan and Patton (2014) reviewed the applications related to the general Fréchet problems including bivariate option pricing, VaR, and distributional treatment effect and argued that copulas with bivariate models can be more advantageous than high-dimensional parametric copulas. Lin and Wu (2015) proposed a distribution free test which neither require selection of smoothing parameters nor need simulation and can be simply applied to the goodness of fit of semiparametric multivariate copula models.

More recently, Smith and Vahey (2016) found that the point and density forecasts from the copula model with asymmetric margins or the skew t margins is just as good as those from the Bayesian vector autoregression (BVAR). Wang *et al.* (2019) set forth a semiparametric approach that employs copula to account for intra-subject dependence and approximates the marginal distributions of longitudinal measurement with quantile regression.

2. Properties of copula-based Markov models

2.1 Data generating process

Y_t is a sample of a strictly stationary first order Markov process generated from $\left(F^*(\cdot), C(\cdot,\cdot;\ \alpha^*)\right)$ from which $F(\cdot)$ is the true invariant (marginal) distribution.

The copula $C(u_1,\ u_2)$ does not have the Fréchet-Hoeffding upper bound ($\min(u_1,\ u_2)$) or the lower bound ($\max(u_1 + u_2 - 1,\ 0)$). Then Y_t is almost surely a monotonic function of Y_{t-1} and the time series $\{Y_t\}$ is deterministic and stationary with upper bound $Y_t = Y_{t-1}$ and lower bound $Y_t = F^{*-1}\ (1 - F^*(Y_{t-1})$.

The parametric conditional pdf of Y_t given Y_{t-1} is:
$h^*(Y_t|Y_{t-1}) = f^*(Y_t)c(F^*(Y_{t-1}), F^*(Y_t);\ \alpha^*)$, where c is the copula density function of the copula, and f^* is the density function of the marginal distribution F^*.

2.2 Tail dependence, temporal dependence

Joe (1997) spearheaded all three copula-based measures discussed in this section. First, the invariant dependence measures under increasing transformation can be expressed in terms of the Kendal's tau (Overall dependence). The Kendal's tau can be described as:

(a) $\tau = 4\int\int_{[0,1]^2} C(u_1, u_2)\mathrm{d}C(u_1, u_2) - 1.$

In addition, the Spearman's rho is the correlation between the margins. We also say that it is the proportional to the probability of concordance minus the probability of discordance for the two vectors. The Spearman's rho can be expressed as:

(b) $\rho_S = 12\int\int_{[0,1]^2}(C(u_1, u_2) - u_1 u_2)\mathrm{d}u_1 u_2.$

Proof:
Given a copula C and margins $F,\ G$. Let (X, Y), $(X',\ Y')$ be i.i.d random vectors from the distribution $H = C(F,\ G)$.

(a)

$$\begin{aligned}\tau &= E(\mathrm{sgn}\ ((X - X')(Y - Y')))\\ &= P((X - X')(Y - Y') > 0) - P((X - X')(Y - Y') < 0)\\ &= 2P((X - X')(Y - Y') > 0) - 1\\ &= 4P(X < X',\ Y < Y') - 1 \qquad \text{Since the two vectors are interchangeable}\end{aligned}$$

$$= 4\int\int_{R^2} P(X < x, Y < y)dH(x,y) - 1$$

$$= 4\int\int_{[0,1]^2} P(X < F^{-1}(u), Y < G^{-1}(v))dC(u,v) - 1 \qquad \text{Change of variables}$$

$$= 4\int\int_{[0,1]^2} C(u,v)\mathrm{d}C(u,v) - 1.$$

(b)

Let $U \coloneqq F(X)$, $U \coloneqq G(Y)$. U and V are uniformly distributed on [0, 1] by probability integral transform. $E(U) = E(V) = \frac{1}{2}$, Var (U) = Var (V) = $\frac{1}{12}$.

$$\rho_S(X,Y) = \rho(F(X), G(Y)) = \rho(U,V)$$

$$= \frac{Cov(U,V)}{\sqrt{Var(U)}\sqrt{Var(V)}}$$

$$= 12\int\int_{[0,1]^2}(x - \tfrac{1}{2})(y - \tfrac{1}{2})\ dC(x,\ y)$$

$$= 12\int\int_{[0,1]^2} xydC(x,y) - 12(-\tfrac{1}{4} - \tfrac{1}{4} - \tfrac{1}{4})$$

$$= 12\int\int_{[0,1]^2}\left(\int\int_{[0,x]\times[0,y]} 1dudv\ \right)dC(x,y) - 3$$

$$= 12\int\int_{[0,1]^2}\left(\int\int_{[u,1]\times[v,1]} 1dC(x,y)\ \right)dudv - 3$$

$$= 12\int\int_{[0,1]^2} C(u,v)\ dudv - 3.$$

Furthermore, we can also use the tail dependence to estimate the dependence in the upper right quadrant or the lower left quadrant tail.

The lower tail dependence coefficients:

$$\lambda_L = \lim_{u\to 0^+} P(U_2 \le u | U_1 \le u) = \lim_{u\to 0^+} \frac{C(u,u)}{u}.$$

The upper tail dependence coefficients:

$$\lambda_U = \lim_{u\to 1^-} P(U_2 \ge u | U_1 \ge u) = \lim_{u\to 0^+} \frac{C(u,u)}{u}.$$

2.3 Geometric $\boldsymbol{\beta}$-mixing

The empirical processes results for strictly stationary geometrically ergodic Markov process are required to analyze the asymptotic properties of semiparametric estimators of (F^*, α^*).

The followed β -mixing introduced by Davydov (1973) can be expressed as: $\beta_t = \int sup|E[\emptyset(Y_{t+1})|Y_1 = y] - E[\emptyset(Y_{t+1})]|\mathrm{d}F^*(y)$.

The process $\{Y_t\}$ is β-mixing if $\lim_{t\to\infty} \beta_t = 0$. Moreover, based upon the foundation of Bradley (1986/2005), Chen & Fan (2006) suggested that the process can be β-mixing with exponential decay (meaning $\beta_t \leq \text{constant} \times e^{-at}$ for some $a > 0$) and polynomial decay rate ($\beta_t(1+t)^{\frac{\lambda}{1-\lambda}} \to 0$ as $t \to \infty$).

3. The model of Chen and Fan (2006)

A semiparametric copula-based time series model is completely determined by (F^*, α^*) which requires the estimation describe in the following sub-sections.

3.1.1 Estimation of the unknown true marginal distribution (Step 1)

Chen and Fan (2006) suggested that we should estimate the unknown true marginal distribution F_n(y) by the empirical distribution function $\frac{n+1}{n}F_n(y)$ where $F_n(y) = \frac{1}{n+1}\sum_{t=1}^{n} I\{Y_t \leq y\}$.

Note: The empirical CDF is obtained from the $\left[\frac{\text{Number of elements in the sample} \leq \text{y}}{\text{sample size}}\right]$.

3.1.2 Estimation of the copula dependence parameter (Step 2)

Chen *et al.* (2009) reiterated Chen and Fan (2006)'s estimation with the expression below:

$$\hat{\alpha}_n^{2sp} \equiv \underset{\alpha\in A}{argmax} \frac{1}{n}\sum_{t=2}^{n} logc(F_n(Y_{t-1}), F_n(Y_t);\ \alpha)$$

$$\sqrt{n}(\hat{\alpha}_n^{2sp} - \alpha_0) \xrightarrow{d} N\left(0, \sigma_{2sp}^2\right) \text{ where: } \sigma_{2sp}^2 \equiv B_0^{-1}\Sigma_{2sp}B_0^{-1} \text{ ,where } B_0 \equiv -E_0\left(\frac{\partial^2 logc(U_{t-1}, U_t;\ \alpha_0)}{\partial\alpha\partial\alpha}\right). \quad \textit{(Assured by Prop 4.3)}$$

$$\Sigma_{2sp} \equiv \lim_{n\to\infty} Var_0\left\{\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\left[\frac{\partial logc(U_{t-1}, U_t;\ \alpha_0)}{\partial\alpha} + W_1(U_{t-1}) + W_2(U_t)\right]\right\} < \infty.$$

$$W_1(U_{t-1}) = \int_0^1\int_0^1 [1\{U_{t-1} \leq v_1\} - v_1]\frac{\partial^2 logc(v_1, v_2;\alpha^*)}{\partial\alpha\partial u_1} c(v_1, v_2;\ \alpha^*)dv_1 v_2.$$

$$W_2(U_t) = \int_0^1\int_0^1 [1\{U_t \leq v_2\} - v_2]\frac{\partial^2 logc(v_1, v_2;\alpha^*)}{\partial\alpha\partial u_2}\ c(v_1, v_2;\ \alpha^*)dv_1 v_2.$$

3.2 Estimation of conditional moment and conditional quantile functions

Chen and Fan (2006) provided the following approach to compute the conditional kth moment of Y_t given Y_{t-1}:
$E\left(Y_t^k \middle| Y_{t-1} = y\right) = \int z^k h^*(z|y)dz = \int z^k c(F^*(y), F^*(z); \alpha^*)dF^*(z).$

As we know, to estimate a vector of conditional moment functions $E(\phi(Y_t)|Y_{t-1})$, where ϕ is a vector of known measurable functions of Y_t. The expression is as follows:
$E(\phi(Y_t)|Y_{t-1} = y) = \int \phi(z)c(F^*(y), F^*(z); \alpha^*)dF^*(z).$

However, it can be estimated by plugging in estimators:
$E(\phi(Y_t)|Y_{t-1} = y) = \int \phi(z)c\left(F_n(y), F_n(z); \hat{\alpha}_n^{2sp}\right)dF^*(z).$

Another crucial characteristic of the condition distribution of Y_t given Y_{t-1} is the conditional VaR (CVaR) of Y_t or conditional quantile function. Thus, for $\{Y_t\}$ satisfying assumption 1, we can estimate the qth conditional quantile of Y_t given Y_{t-1} by: $Q_q^Y(Y_{t-1}) = {F^*}^{-1}(Q_q(F^*(Y_{t-1}); \alpha^*))$ in which $Q_q(u; \alpha^*)$ is the conditional quantile of U_t given $U_{t-1} = u$.

$Q_q(u; \alpha^*) = C_{2|1}^{-1}(q|u; \alpha^*)$

Where $C_{2|1}(\cdot\,|u; \alpha^*) = P(V \leq v|U = u) = \lim_{h\to 0} \frac{C(u+h,v)-C(u,v)}{h} = \frac{\partial C(u,v;\alpha^*)}{\partial u}$.

As a result, an estimator of the conditional quantile $Q_q(u; \alpha^*)$ of U_t given $U_{t-1} = u$ is: $\tilde{Q}_q(u) = Q_q(u; \tilde{\alpha}) = C_{2|1}^{-1}(q|u; \breve{\alpha})$. Then the estimator of the conditional quantile $Q_q^Y(y)$ of Y_t given $Y_{t-1} = y$ is:

$\tilde{Q}_q^Y(y) = F_n{}^{-}\left(\tilde{Q}_q\left(F_n(y)\right)\right) = G_n{}^{-}(C_{2|1}^{-1}(q|F_n(y); \tilde{\alpha})),$

Where $F_n{}^{-}(v) = \inf\{y: F_n(y) \geq v\}$.

As conditional quantile is derived from the conditional distribution of U_t given U_{t-1}, the conditional estimators are automatically monotonic across different quantiles.

4. Asymptotic

4.1 Large sample properties

To prevent the derivatives of score function of $\tilde{\alpha}$ from approaching infinity near the boundaries, Chen and Fan (2006) established convergence of $F_n(\cdot)$ with weighted method and then use it to formulate the consistency and asymptotic normality of $\tilde{\alpha}$. As a result, the joint asymptotic distribution of $F_n(\cdot)$ and $\tilde{\alpha}$ can put together with the Delta method to establish the asymptotic properties of the conditional moment and conditional quantile estimators.

Chen and Fan (2006) defined $\widetilde{U}_n(v) = F_n\left(F^{*-1}(v)\right)$ for $v \in (0,1)$. Let $W^*(\cdot)$ be a zero-mean tight Gaussian process in $\mathrm{D}\,[0,1]$ such that $W^*(0) = W^*(1) = 0$; moreover, $\mathrm{E}\,\{W^*(v_1)W^*(v_2)\} = \min\{v_1, v_2\} - v_1v_2 + \sum_{k=2}^{\infty}\{Cov[I(U_1 \leq v_1), I(U_2 \leq v_2)] + Cov[I(U_k \leq v_1), I(U_1 \leq v_2)]\}$.

4.2 Consistency

Chen and Fan (2006) showed how to established the consistency of $\tilde{\alpha}$ for α^* by verifying the conditions in their proposition 4.2 which requires the process $\{Y_t\}$ be β-mixing with polynomial decay rate, moment condition on the score function, and the partial derivatives of the score function be dominated by a function that has a finite first moment when weighted by a weighting function w(·).

4.3 $\sqrt{n}$-normality

An estimator is asymptotically normal if it converges in distribution. The $\sqrt{n}$ is a scaled version of asymptotic normality. For this case, to obtain the $\sqrt{n}$-asymptotic normality, the following propositions are required:

(1) $\tilde{\alpha} - \alpha^* = B^{-1}A_n^* + o_p\left(n^{-\frac{1}{2}}\right)$.

(2) $\sqrt{n}(\tilde{\alpha} - \alpha^*) \rightarrow \mathrm{N}(0, B^{-1}\sum B^{-1})$ in distribution.

5. Bootstrapping replication, simulation, and empirical research

5.1 Replication of the semiparametric bootstrap in Chen and Fan (2006)

Here we are trying to replicate a bootstrap method which is traditionally a way to estimate the distribution of an estimator or test statistic by resampling a dataset. To do so, we follow the suggested statistical inference of section 4.3 in Chen and Fan (2006) who focused on the semiparametric bootstrap method. It is a method that assumes $Y_t = G^{*-1}(U_t)$ where $\{U_t\}_{t=1}^n$ is a stationary first-order Markov process with the copula $\mathrm{C}(u_1, u_2;\, \alpha^*)$ being the joint distribution of (U_1, U_2). We implement this bootstrap on the Clayton copula. The Python code for this semiparametric bootstrap is shown in section 8.1.

The steps advised by Chen and Fan (2006) are as follows:

Given $\hat{F}_n(y) = \frac{1}{n}\sum_{t=1}^{n} K\left(\frac{y-Y_t}{a_n}\right)$,

Step 1: Generate n independent $\mathrm{U}(0,1)$ random variables $\{X_t\}_{t=1}^n$.

Step 2: Generate $U_1^b = X_1$ and $U_t^b = C_{2|1}^{-1}\left(\{X_t\}_{t=1}^n | U_{t-1}^b;\, \tilde{\alpha}\right)$ which leads to one bootstrap sample $\{U_t^b\}_{t=1}^n$.

Step 3: Let $Y_t^b = \tilde{F}_n^{-1}(U_t^b)$, where $\hat{F}_n(y)$ is the kernel estimator. Then compute the corresponding estimate using the bootstrap sample $\{Y_t^b\}_{t=1}^n$.

Step 4: Repeat steps 1-3 a large number of times and use the empirical distribution of the centered bootstrap values of the estimator to approximate its distribution.

Output:

The output is just a set of weights at random that follows the uniform distribution. Please check out our code (Bootstrap.py) in the submission folder and go to the Python for semipara bootstrap folder.

5.2 Simulation of bivariate copula-based models (Python)

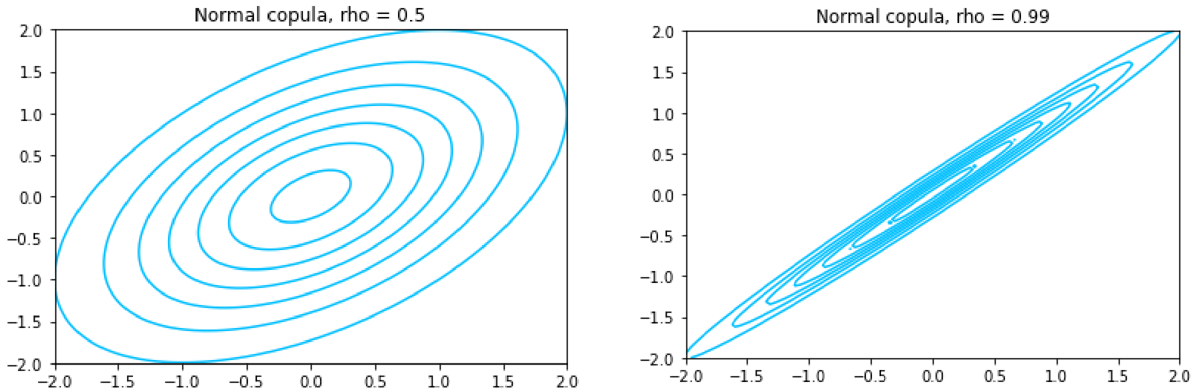
We choose normal and Clayton copulas for the simulation. The former is linear while the latter is nonlinear but both assume standard normal marginal distributions. The contour densities for the normal and the Clayton copulas are modeled and shown in figures 1 and 2, respectively. The purpose here is to compare and detect their asymmetric dependence structure. In terms of application, the asymmetric dependence structure could cause significant change in the estimation for VaR and expected shortfall. If the risk manager ignores the impact of asymmetric dependence on the left lower tail of the joint distribution, then the risk measure is subject to undervaluation. The Python code is shown in section 8.2.

The normal copula is the most popular benchmark. It allows for equal degrees of positive and negative dependence. Additionally, it includes both Fréchet bounds if within the limits. As the name implies, it does not allow for tail dependence of the underlying marginal distribution. Figure 1 shows that as its linear correlation, ρ, increases, the contour plots of normal copula become compressed yet remains symmetric. The setbacks of normal copula are quite obvious; for instances, Chen and Fan (2006) stated that the normal copula will fail to capture nonlinear asymmetric dependence and clustering phenomenon of financial market. Okimoto (2008) pointed out that it also fails to capture dependence in the tail part of joint distribution.

By contrast, the Clayton copula captures the asymmetric dependence structure and the degree of asymmetry becomes larger as the dependence parameter elevates. In general, Clayton copula tend to exhibit positive dependence and lower tail dependence. Moreover, as the dependence parameter elevates, the lower tail dependence also increases resulting in smoother time series plots. Patton (2006) stated that contour in Clayton copula has greater dependence for joint negative events than for joint positive events since it has contours which tend to be more peaked in the third quadrant.

Figure 2 reveals that as kappa (Clayton's dependence structure similar to normal's correlation; moreover, Clayton's kappa = 1 is equivalent to normal's $\rho = 0.5$) increases, the squeezed contour plot of Clayton copula is still asymmetric and remains more peaked in the negative quadrant. This result is consistent with Patton (2006).

Figure 1. Contour plots of Gaussian copula



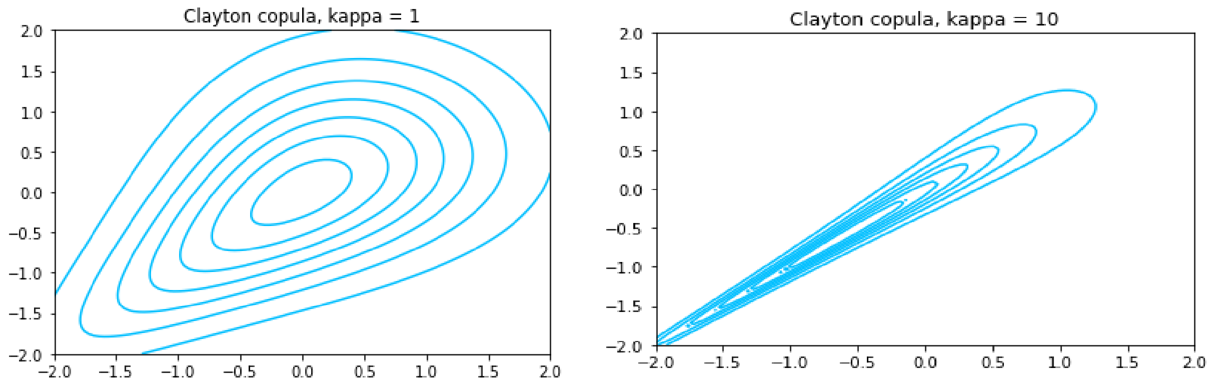
Gaussian copula

$C_{gaussian}(u_1, u_2;\ \alpha) = \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2);\ \alpha)$

$$= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\alpha^2)^{\frac{1}{2}}} \times \left\{\frac{-(s^2 - 2\alpha st + t^2)}{2(1-\alpha^2)}\right\} \text{dsdt.}$$

Where: Φ is the CDF of standard normal distribution and $\Phi_G(\cdot)$ is the standard bivariate normal distribution with correlation parameter α restricted to the interval $(-1, 1)$. The

Figure 2. Contour plots of Clayton copula



Clayton copula

$C_{clayton}(u_1, u_2;\ \alpha) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}}$ with the dependence parameter α limited to the region $(0, \infty)$. The marginals become independent as α approaches 0. The copula touches the Fréchet upper bound. The Clayton copula exhibits asymmetric dependence since the dependence is concentrated in the lower tail but does not account for negative dependence.

5.3 Empirical research on tail dependence (MATLAB)

In this section, we aim to test alternatives to copulas by using nonparametric models: exceedance correlation and quantile dependence in order to observe the distributional dependence between a pair of two assets. One main advantage of being nonparametric is that our model can avoid specification bias.

Firstly, our exceedance correlation is based on Ang and Chen (2002). This correlation at a given exceedance level is defined as the correlation between a pair of random variables when both jump or fall more than a threshold level of standard deviations distant from the means. Secondly, our quantile dependence is based on Patton (2006) and it can be simply defined as correlation coefficient within different quantiles.

The research on asymmetric dependence can be crucial for financial market practitioners in the risk management area and has been studied extensively by the academics. For instances, Ang and Chen (2002) introduced the symmetric statistics to gauge the degree of asymmetry in correlations across downside and upside markets relative to a distribution; empirically, they concluded that small beta stocks show higher correlation asymmetries. Okimoto (2008) argued that the dependence in international equity market tend to be very high especially during high volatile and bearish market.

Our empirical research focus on the asymmetric dependence between the 10-yr to 2-yr interest rate spread (T10Y2Y) and the 5-yr to federal fund rate spread (T5YFF). The times series is collected from April 2012 to December 2021. The data is extracted from the database of the Federal Reserve Economic Data (FRED) at the Federal Reserve Bank of St. Louis.

Figure 3 is the double axis graph for the time series trend of T10Y2Y and T5YFF (Note: The graph is generated in Python). Table 1 shows the regression generated in STATA. Table 2 displays the correlation tests (conducted in R) with the Pearson's correlation, Kendall's tau, and Spearman's rho. Pearson's correlation tells the linear dependence between two variables and is parametric while Kendall's tau and Spearman's rho are rank-based correlation and each of them is non-parametric. Both tables indicate that there is positive relationship between T5YFF and T10Y2Y with moderately correlation. This section is empirical in nature in which only the dataset and the statistical analysis part is totally ours, all the codes are provided in our submission folder.

Figure 3. Trend of T5YFF and T10Y2Y

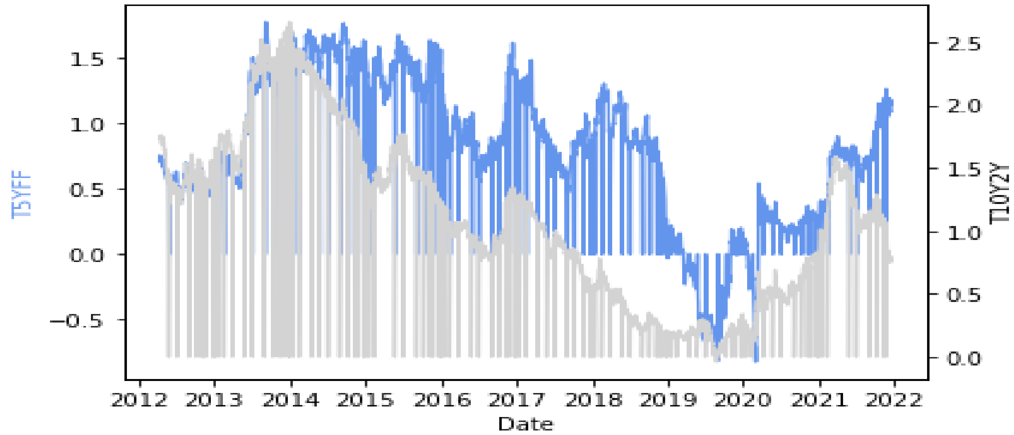


Table 1. Regression summary

```
. reg T5YFF T10Y2Y

      Source |       SS           df       MS      Number of obs   =     2,528
-------------+----------------------------------   F(1, 2526)      =   1623.88
       Model |  319.692989         1  319.692989   Prob > F        =    0.0000
    Residual |  497.292246     2,526  .196869456   R-squared       =    0.3913
-------------+----------------------------------   Adj R-squared   =    0.3911
       Total |  816.985235     2,527  .323302428   Root MSE        =     .4437

------------------------------------------------------------------------------
       T5YFF |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      T10Y2Y |   .5206789   .0129209    40.30   0.000     .4953423    .5460156
       _cons |   .2254988   .0163389    13.80   0.000     .1934598    .2575378
------------------------------------------------------------------------------
```

Table 2. Correlation test between T5YFF and T10Y2Y

	Correlation test
Pearson's correlation	0.626
Kendall's tau	0.441
Spearman's rho	0.590

Figure 4 displays the exceedence correlation which is based on the methodology of Ang and Chen (2002). The result suggests that the relationship between exceedence correlation and the quantile is nonlinear; however, the correlation tends to go down when the quantile exceeds 0.5. Moreover, this phenomenon also suggests that when there is negative shock to the financial markets, both will get hit; on the contrary, when positive shock happens, one of them will do much better than another.

Figure 5 illustrates the quantile dependence pioneered by Patton (2006). The result indicates that the tail dependence tends to go down significantly at either the low and high ends of quantile. In the contrast, the dependence is close to a steady state during tranquil period.

Figure 4. Exceedence correlation

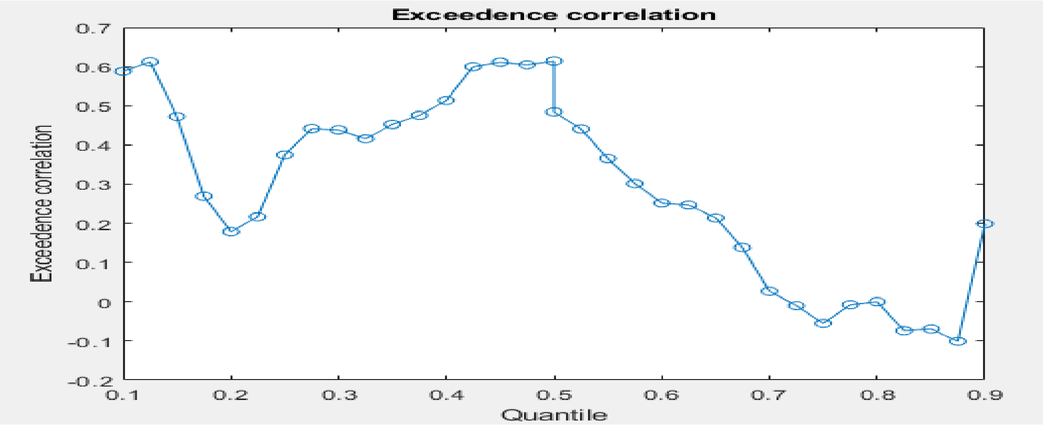
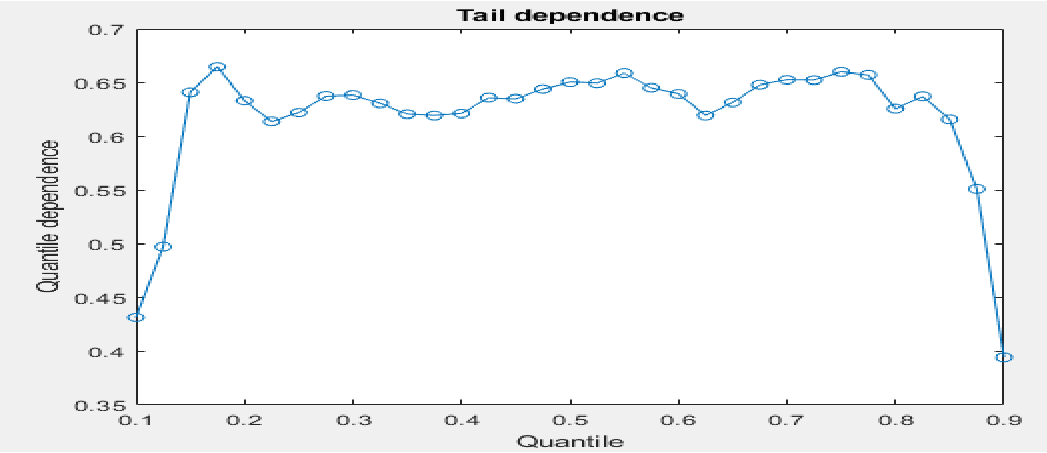


Figure 5. Quantile dependence



6. Conclusion

In this project, we discuss the estimation methodology of Chen and Fan (2006) and extend it with the latest literature review and our own replication for the semiparametric bootstrap method, simulation on copula, and empirical research with tail dependence.

In sum, Chen and Fan (2006) made a significant contribution by proposing simple estimators of the unknown marginal distribution and the copula dependence parameters with large sample properties with readily valid conditions. As a result, they enhance the calculation for CVaR since the semiparametric conditional quantile estimators can be automatically monotonic across different quantiles.

Nonetheless, as Chen and Fan (2006) mentioned, their research can be enhanced by incorporating higher order Markov process and allowing for time-varying copula dependence parameter. Also, in our opinion, Chen and Fan (2006) could have done better if they can provide computational simulation based upon their own model.

Then we conduct our own coding and/or data analysis for three major milestones: replication of semiparametric bootstrap suggested by Chen and fan (2006), simulation of copulas, and empirical research of tail dependence. First, we had replicated the semiparametric bootstrap method of Chen and Fan (2006) on the Clayton copula with Python. Second, in our own simulation model coded with Python for normal and Clayton copulas, the results of our own simulation models for the normal and Clayton iso-probability contour are consistent with the theory. Third, in our empirical research which adds the feature of multivariate time series and nonparametric models, we apply the dataset of T10Y2Y and T5YFF to examine exceedence correlation and quantile dependence. We find that tail dependence between them has sharp fall at the left and right end of distribution and they do not act consistently under both positive and negative macro or exogeneous policy shocks.

As for future topic, we can, for example, incorporate time varying feature into the copula estimation by Chen and Fan (2006) and test for asymmetric dependence between volatilities of bond market and equity market under two different regimes: expansionary and contractionary monetary policies.

7. References

Andrews, DWK 2001, 'Testing when a parameter is on the boundary of the maintained hypothesis', *Econometrica*, vol. 69: 3, pp. 683-734.

Ang, A & Chan, J 2002, 'Asymmetric correlations of equity portfolios', *Journal of Financial Economics*, vol. 63, pp. 443-494.

Bradley, R 2005, 'Basic properties of strong mixing conditions. A survey and some open questions', *Probability Surveys,* vol. 2, pp. 108-134.

Chan, NH, Chan, J, Chen, X, Fan, Y & Peng, L 2009, 'Statistical inference for multivariate residual copula of GARCH models', *Statistica Sinica,* vol. 9: 1, pp. 53-70.

Chen, X & Fan, Y 2006, 'Estimation of copula-based semiparametric time series models', *Journal of Econometrics*, vol. 130, pp. 307-335.

Chen, X, Fan, Y & Tsyrennikov, V 2006, 'Efficient estimation of semiparametric multivariate copula models', *Journal of the American Statistical Association*, vol. 101: 475, pp. 1228-1240.

Chen, X, Koenker, R & Ziao, Z 2009, ‘Copula-based nonlinear quantile autoregression’, *The Econometrics Journal*, vol. 12: S1, pp. S50-S67.

Chen, X, Wu, WB & Yi, Y 2009, ‘Efficient estimation of copula-based semiparametric Markov models’, *The Annals of Statistics*, vol. 37: 6B, pp. 4214-4253.

Darsow, WF, Nguyen, B & Olsen, ET 1992, ‘Copulas and Markov processes’, *Illinois Journal of Mathematics*, vol. 36, pp. 600-642.

Davydov, Y 1973, ‘Mixing conditions for Markov chains’, *Theory of Probability and Its Applications*, vol. 18: 2, pp. 312-328.

Dette, H, Hecke, RV & Volgushev, S 2014, ‘Some comments on copula-based regression’, *Journal of the American Statistical Association*, vol. 109: 507, pp. 1319-1324.

Fan, Y & Patton, AJ 2014, ‘Copulas in Econometrics’, *Annual Review of Economics*, vol. 6, pp. 179-200.

Joe, H 1997, *Multivariate models and dependence concepts*, London, UK: Chapman & Hall.

Lin, J & Wu, X 2015, ‘Smooth tests of copula specifications’, *Journal of Business & Economic Statistics*, vol. 33: 1, pp. 128-143.

Noh, H, Ghouch, AI & Bouezmarni, T 2013, ‘Copula-based regression estimation and inference’, *Journal of the American Statistical Association*, vol. 108: 502, pp. 676-688.

Okimoto, T 2008, ‘New evidence of asymmetric dependence structures in international equity markets’, *Journal of Financial and Quantitative Analysis*, vol. 43: 3, pp. 787-816.

Patton, AJ 2006, ‘Modelling asymmetric exchange rate dependence’, *International Economic Review*, vol. 47: 2, pp. 527-556.

Sklar, A 1959, ‘Fonctions de Répartition à n Dimensions et Leurs Marges’, *Publications de l’Institut Statistique de l’Université de Paris*, vol 8, pp. 229-231.

Smith, MS & Vahey, SP 2016, ‘Asymmetric forecast densities for US macroeconomic variables from a gaussian copula model of cross-sectional and serial dependence’, *Journal of Business & Economic Statistics*, vol 34: 3, pp. 416-434.

Wang, HJ, Feng, X & Dong, C 2019, 'Copula-based quantile regression for longitudinal data', *Statistica Sinica*, vol. 29: 1, pp. 245-264.

Zimmer, DM 2012, 'The role of copulas in the housing crisis', *Review of Economics and Statistics*, vol. 94: 2, pp. 607-620.

8. Appendix

8.1 Python code for the replication of semiparametric bootstrap method by Chen and Fan (2006)

```
import numpy as np
from pynverse import inversefunc
import csv

# def inverse_con_copula(x, u, alpha):# x is drawn from uniform distribution. Alpha is tail dependence.
# l1 = np.log(1/np.dot(x,np.power(u,1+alpha)))
# l2 = np.exp(np.dot((1/(1+1/alpha)),l1))
# l3 = (-np.log(l2) + 1 - np.power(u,-alpha))/alpha
# return np.exp(l3)

def bootrap(n, alpha):
    rv_x_lst = sorted(np.random.uniform(0, 1, n))
    u_lst = [rv_x_lst[0]]
    i = 1
    while i < n:
        # u_lst.append((inverse_con_copula(rv_x_lst[i], u_lst[i-1], alpha)))
        f = (lambda v: np.power(u_lst[i - 1], -alpha - 1) * np.power(
            (np.power(u_lst[i - 1], -alpha) + np.power(v, -alpha) - 1), -1 - (1 / alpha)))
        u_lst.append(inversefunc(f, y_values=rv_x_lst[i]).tolist())
        i = i + 1
    return u_lst

print(bootrap(100, 0.01))
```

8.2 Python code for normal and Clayton copulas simulations and contour plotting

function.py

```
import numpy as np
```

```
import math
def theta2rho(theta):
    m = len(theta)
    k = (1+np.sqrt(1+8*m))/2
    if np.mod(k, 1) != 0:
        print('vector is not of correct length')
    else:
        k = int(k)
    out1 = np.ones((k, k))*(-999.99)
    counter = 0
    for ii in range(0, k, 1):
        for jj in range(ii, k, 1):
            if ii == jj:
                out1[ii][jj] = 1
            else:
                out1[ii][jj] = theta[counter]
                out1[jj][ii] = theta[counter]
                counter = counter+1
    return out1


def bivnormpdf(X, Y, MU, VCV):
    k = 2
    detVCV = np.linalg.det(VCV)
    invVCV = np.linalg.inv(VCV)
    T = max(X.shape[0],1)
    N = 1

    if T >= N:
        if X.shape[0] < T:
            X = X[0][0]*np.ones((T, 1))
        if 1 < T:
            Y = float(Y)*np.ones((T, 1))
            Y = np.mat(Y)
        X = np.append(X, Y, axis=1)
        out1 = np.ones((T, 1))*(-999.99)
        for tt in range(0, T, 1):
            data =(X[tt] - MU)
```

```
            out1[tt] = ((2*np.pi)**(-k/2))*(detVCV**(-0.5))*math.exp(-0.5*data*invVCV*data.T)
    else:
        if X.shape[1] < N:
            X = X[0][0]*np.ones(1, N)
            Y = Y*np.ones(1, N)
        X = np.mat(np.append(X, Y, axis=0))
        X = X.T
        out1 = (-999.99)*np.ones((1, N))
        for tt in range(0, N, 1):
            data = np.mat(X[tt] - MU)
            out1[tt] = (2 * np.pi) ** (-k / 2) * detVCV ** (-0.5) * math.exp(-0.5 * (X[tt] - MU) * invVCV * data.T)
    return out1

#def tcopulapdf(U, V, RHO, NU):

def claytonpdf(u, v,k1):
    T = max(u.shape[0], v.shape[0], 1)
    if u.shape[0] < T:
        u = np.mat(u*np.ones((T, 1)))
    if v.shape[0] < T:
        v = np.mat(float(v)*np.ones((T, 1)))
    if 1 < T:
        k1 = np.mat(k1*np.ones((T, 1)))
    uv = np.multiply(u, v)
    a = (np.zeros((T, 1)))
    b = (np.zeros((T, 1)))
    c = (np.zeros((T, 1)))
    d = (np.zeros((T, 1)))
    for i in range(0, T, 1):
        a[i] = np.array(uv[i])**np.array(-k1[i]-1)
        b[i] = np.array(u[i])**np.array(-k1[i])
        c[i] = np.array(v[i])**np.array(-k1[i])
    for i in range(0, T, 1):
        d[i] = np.array(b[i]+c[i]-1)**np.array(-2-1/k1[i])
    pdf = np.multiply(np.multiply(1+k1, a), d)
    return pdf
```

contour.py

```
import numpy as np
from scipy.stats import norm
import math
import function
import matplotlib.pyplot as plt

T = 100
xx0 = np.arange(-2, 2+4/(T-1), 4/(T-1))
xx = np.mat(xx0)
xx = xx.T
uu = np.mat(norm.cdf(xx, 0, 1))
v = np.arange(0.02, 0.2+0.03, 0.03)

# 1. Normal copula
rho = [0.5]
zz = np.mat(np.ones((T, T))*(-999.99))

for ii in range(0, T, 1):
    zz[:, ii] = function.bivnormpdf(xx, xx[ii, :], [0, 0], function.theta2rho(rho))

# plot the figure
fig1 = plt.figure(1)
plt.contour(xx0, xx0, zz, colors='deepskyblue')
plt.title('Normal copula, rho = 0.5')

# 2. Clayton copula
kappa = 1

zz = np.mat(norm.pdf(xx)*np.ones((1, T)))
for ii in range(0, T, 1):
    zz[:, ii] = np.multiply(np.multiply(zz[:, ii], norm.pdf(xx[ii])), function.claytonpdf(uu, uu[ii], kappa))

# plot the figure
fig3 = plt.figure(3)
plt.contour(xx0, xx0, zz, colors='deepskyblue')
plt.title('Clayton copula, kappa = 1')
fig4 = plt.figure(4)
```

```
print('complete')
```

8.3 Code for figures (Python), statistical results (STATA/R), and asymmetric dependence (MATLAB) in 5.2

Figure 3 Python code

```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import pyplot
import numpy as np

df=pd.read_excel(r'D:\Stony Brook AMS\AMS 2022 Spring\AMS517\AMS517 project\Paper 6 draft\DBP.xlsx')

sample_timeseries_data ={'Date':df["Date"], 'T5YFF':df["T5YFF"], 'T10Y2Y':df["T10Y2Y"]}

dataframe = pd.DataFrame(
    sample_timeseries_data,columns=[
        'Date', 'T5YFF', 'T10Y2Y'])

dataframe["Date"] = dataframe["Date"].astype("datetime64")
dataframe = dataframe.set_index("Date")
dataframe

x= df["Date"]
y1 =df["T5YFF"]
y2 = df["T10Y2Y"]

fig, ax1 = plt.subplots()
ax2 = ax1.twinx()
ax1.plot(x, y1, 'cornflowerblue')
ax2.plot(x, y2, 'lightgray')

ax1.set_xlabel('Date')
ax1.set_ylabel('T5YFF', color='cornflowerblue')
ax2.set_ylabel('T10Y2Y', color='lightgry')
plt.show()
```

Table 1 STATA code

```
import excel "D:\Stony Brook AMS\AMS 2022 Spring\AMS517\AMS517 project\Paper 6 draft\DBP.xlsx",
```

```
sheet("Sheet1") firstrow clear
reg T5YFF T10Y2Y
```

Table 2 R code

```
library(readxl)
library(ggpubr)
DBP <- read_excel("D:/Stony Brook AMS/AMS 2022 Spring/AMS517/AMS517 project/Paper 6 draft/DBP.xlsx")
View(DBP)
DF<-as.data.frame(DBP)

T5YFF<-DBP$T5YFF
T10Y2Y<-DBP$T10Y2Y

pcr <- cor.test(T10Y2Y, T5YFF,
                method = "pearson")
pcr


kcr<-cor.test(T10Y2Y, T5YFF,    method="kendall")
kcr

scr <-cor.test(T10Y2Y, T5YFF,    method = "spearman")
scr
```

Instruction for using MATLAB code for asymmetric dependence in our submission folder

Open the MATLAB for AsymDep folder, the main work is executed on Main.m which requires the functions of ang_chen1.m and quantiledep.m as well as corrcoef12.m and nines.m. All of them are developed by Professor Andrew Patton at Duke Univeristy and the required functions and the main execution file are included in the folder. The dataset is from ours and is collected from the FRED of the Federal Reserve Bank of St. Louis.