Journal of Applied Mathematics

volume V (2012), number III

COPULA BASED SEMIPARAMETRIC REGRESSIVE MODELS

FJODOROVS Jegors, (LV), MATVEJEVS Andrejs, (LV)

Abstract. This paper studies the estimation of copula-based semi parametric stationary Markov models. Described models allow us evaluate the parameters of copula, which has the best fit to previously selected model (simple estimators of the marginal distribution and the copula parameter are provided). These copula-based models are characterized by nonparametric marginal distributions and parametric copula functions, while the copulas capture all the scale-free temporal dependence of the processes. In our copula dependence study we used MatLab, which help to evaluate copula parameters and choose the best copula class, based on log likelihood estimation, for the selected financial market data. Also, using this MatLab we made VIX option index simulation - found the best copula fit under our condition and show the evaluation steps for copula based semi parametric autoregression.

Key words: copula, diffusion processes, time series, semi parametric regressions, VIX index.

Mathematics Subject Classification: 60J70, 62M10, 47D07

1. Introduction

The possibility of identifying nonlinear time series using nonparametric estimates of the conditional mean and conditional variance were studied in many papers (see, for example, [1], and references there). As a rule analyzing the dependence structure of stationary time series $\{x_t, t \in Z\}$ regressive models defined by invariant marginal distributions and copula functions that capture the temporal dependence of the processes. As it indicated in [1] this permits to separate out the temporal dependence (such as tail dependence) from the marginal behavior (such as fat tails) of a time series. One more advantage of this type regressive approach is a possibility to apply probabilistic limit theorems for transition from deference equations to continuous time stochastic differential equations ([2], [3]). In our paper we also study a class of copula-based semi parametric stationary Markov models in a form of scalar difference equation

$$t \in Z : X_t = X_{t-1} + \varepsilon f(X_{t-1}, \varepsilon) + \varepsilon g(X_{t-1}, \varepsilon)\xi_t \qquad (1)$$

where $\{\xi_t, t \in Z\}$ is i.i.d., N(0; 1), and ε is a small positive parameter, which will be used for diffusion approximation of (1). Regressions (1) are high-usage equations for simulation and parameter estimation of stochastic volatility models ([2]). But unfortunately defined by (1) Markov chain has incompact phase space that complicates an application of probabilistic limit theorem. Copula approach helps to simplify asymptotic analysis of (1). Let us remember that to construct a copula *C(u; v)* for pair $\{X_{t-1,} X_t\}$ from (1) one should find a marginal invariant distribution *F(x)* for X_t and to substitute this in joint distribution function $H(x, y) = P(X_{t-1} \leq x, X_t \leq y)$, that is, $C(u,v) = H(F^{-1}(u), F^{-1}(v))$ and $H(x, y) = C(F\ \ (x), F\ \ (y))$. Due to persistence of small parameter ε after a substitution $U_t = F(X_t)$ in equation (1) for a further diffusion approximation one can write a difference equation in a same form like (1):

$$t \in Z : U_t = U_{t-1} + \varepsilon f(U_{t-1}, \varepsilon) + \varepsilon g(U_{t-1}, \varepsilon)\xi_t \qquad (2)$$

But now this equation defines Markov chain on the compact [0, 1]. This makes easier formulate construction for transition probability and further estimators of functions $\hat{f}(u)$ and $\hat{f}(u)$. After diffusion approximation of (2) one can make inverse substitution and derive stochastic differential equation as diffusion approximation for (1).

In the copula approach to univariate time series modeling, the finite dimensional distributions of the time series are generated by copulas. By coupling different marginal distributions with different copula functions, copula-based time series models are able to model a wide variety of marginal behaviors (such as skewness and fat tails) and dependence properties (such as clusters, positive or negative tail dependence). (see Darsow et al. (1992) [4] and Joe (1997)[5]).

Described algorithm allow us evaluate the parameters of copula, which has the best fit to previously selected model. In our copula dependence study we used MatLab, which help to evaluate copula parameters and choose the best copula class, based on log likelihood estimation, for the selected financial market data. These copula based models are easy to simulate, and can be expressed as semi parametric regression transformation models. Also, using this MatLab we made VIX option index simulation - found the best copula fit under our condition and built semi parametric autoregression.

The paper is structured as follows. Section 2 gives a brief review of the copula functions definition. Section 3 describes our approach. In Section 4 we report our results for the VIX index data Section 5 concludes and discusses several possible avenues of future research.

2. Copula functions

Copulas became popular in the finance and insurance community in the past years, where modeling and estimating the dependence structure between several univariate time series are of great interest; see Frees and Valdez (1998) [6] and Embrechts et al. (2002) [7] for reviews.

A copula function is a multivariate distribution function with standard uniform marginals. By Sklar's (1959) [8] theorem, one can always model any multivariate distribution by modeling its marginal distributions and its copula function separately, where the copula captures all the scale-free dependence in the multivariate distribution.
The central result of this theorem, which states that any continuous *N*-dimensional cumulative distribution function *F*, evaluated at point $x = (x_1, \ldots, x_n)$ can be represented as

$$F(x) = C(F_1(x_1), \dots, F_n(x_n)),$$

where C is called a copula function and $F_i(x_i)$, $i = 1, \dots, n$ are the marginal distributions. The use of copulas therefore splits a complicated problem (finding a multivariate distribution) into two simpler tasks. The first task is to model the univariate marginal distributions and the second task is finding a copula that summarises the dependence structure between them.
It is also useful to represent of copulas as joint distribution functions of standard uniform random variables:

$$U = F(X_1) \text{ and } V = F(X_2)$$

$$C(u,v) = P(U \le u, V \le v)$$

The outcome of uniform random variables falls into the interval [0, 1], therefore the domain of a copula must be the N-dimensional unit cube. Similarly, because the mapping represents a probability, the range of the copula must also be the unit interval. Also, it is easy to determine the value of a copula on the border of its domain. When one argument equals zero, the probability of any joint event must also be zero. Similarly, when all but one of the inputs are equal to one the joint probability must be equal to the (marginal) probability of the argument that does not equal one. Finally, the function must be increasing in all its arguments.
Besides the standard distribution functions, copulas have associated densities:

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}$$

which permit the bivariate density $f(u; v)$ as the product of the copula density and the density functions of the margins

$$f(u,v) = c(F_1(u), F_2(v)) f_1(u) f_2(v)$$

This expression indicates how the simple product of two marginal distributions will fail to properly measure the joint distribution of two asset prices unless they are in fact independent and the dependence information captured by the copula density, $c(F_1(u), F_2(v))$*;* is normalized to unity.

3. Evaluation of parameters for the semi parametric regression model

Copula based semi parametric models are characterized by conditional heteroscedasticity and have been often used in modeling the variability of statistical data. In paper [1] the basic idea was to apply a local linear regression to the squared residuals for finding the unknown functions ***f*** and ***g***.

Our methodology builds on the finding conditional expectation of the first and second order.

Let $\{Y_t\}$ be a stationary Markov process of order 1 with continuous state space. Then its probabilistic properties are completely determined by the joint distribution function of $\{Y_{t-1}\}$ and $\{Y_t\}$. For the determination of the copula based model we should use Markov model in the scalar difference equation form:

$$t \in Z : X_t = X_{t-1} + \varepsilon f(X_{t-1}, \varepsilon) + \varepsilon g(X_{t-1}, \varepsilon)\xi_t$$

Where

$$E\xi_t = 0,\ DE\xi_t = 1,\ \{\xi_t, t \in Z\} \text{ is i.i.d. } N(0,1)$$

And our goal reduced to the estimation of conditional moments, which will be our base regression model parameters:

$$g(X_{t-1},\varepsilon) - ? \text{ and } f(X_{t-1},\varepsilon) - ?$$

As was mentioned above it is not easy task, especially this representation complicates an application of probabilistic limit theorem. That is why; if we have stationary distribution our suggestion is to find parameters through Markov chain using copula approach.
Firstly, let's show that copula distribution density equals Markov chain transition density:

$$P(X_{t-1} \le y) = F(y) = \int_{-\infty}^{y} p(z)dz$$

$$P(X_{t+1} \in A / X_t) = P(X, A) = \int_A p(x,y)dy$$

$$P(X_{t+1} \in A, X_t \in B) = \int_A p(z,A)dz = \iint_{AB} p(z,y)p(z)dzdy = \int P(x,A)dF(x) =$$
$$= \iint_{BA} p(x,y)p(z)dF(x)dy$$
$$C(F(y),F(x) = P(X_{t+1} \le y, X_t \le x) = \iint p(z,u)p(z)dzdu$$

$$p(x,y) = \frac{\partial^2 C(F(y),F(x))}{\partial x \partial y} p(y)$$

As the result we see that Markov transition density is copula density.
Secondly, we should expand our semi parametric regression into Taylor series:

$$F(X_t) = F(X_{t-1}) + F'(X_t)\varepsilon f(X_{t-1},\varepsilon) + F'(X_t)\varepsilon g(X_{t-1},\varepsilon)\xi_t \ldots$$

And due to persistence of the small parameter ε, we can rewrite our expression is:

$$t \in Z : U_t = U_{t-1} + \varepsilon f(U_{t-1},\varepsilon) + \varepsilon g(U_{t-1},\varepsilon)\xi_t$$
$$f(U_{t-1},\varepsilon) = E(U_t \mid U_{t-1} = u) - ? \qquad (3)$$
$$g(U_{t-1},\varepsilon) = E((U_{t-1} - f(U_{t-1},\varepsilon))^2 \mid U_{t-1} = u) - ? \qquad (4)$$

After conditional expectations of (3) and (4) evaluation one can make inverse substitution and derive stochastic differential equation as diffusion approximation for the base semi parametric model (1). Of course, our algorithm works only if inverse function exists. For example, Gamble copula, which don't have standard inverse function.
Now we derived a tool for model (1) parameters evaluation. For describing our idea briefly, let's take a look in the next section how works our algorithm with the true market data.

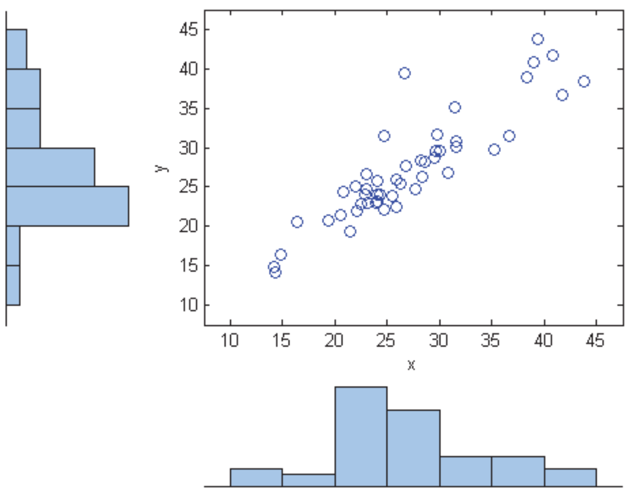
4. Practical approach of the proposed algorithm

We'll analyze the VIX - **Market Volatility Index –** daily data from 31.03.2007. to 10.12.2011. The VIX is a market mechanism that measures the 30-day forward implied volatility of the underlying

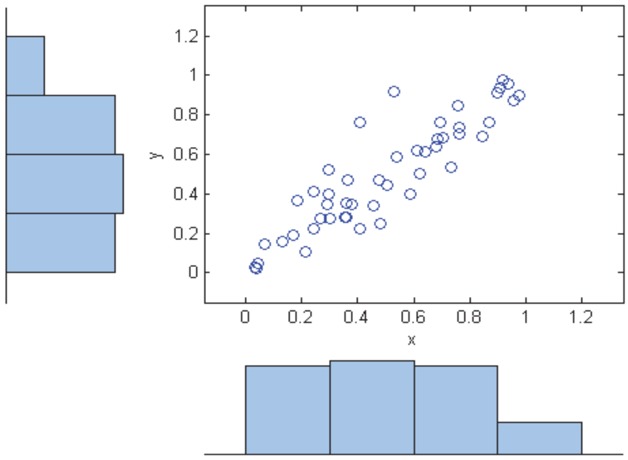
index, the S&P 500. Being able to meaningfully interpret movements in the VIX and its reaction to market events can give investors an edge in managing the risk and profitability of their trading book and in designing portfolio strategies using VIX derivatives to capitalize on the dynamic and time-varying correlation of the VIX with its underlying S&P 500 Index. Let's built for this option index semi parametric copula based model, using AIC and BIC criteria.
An easiest way of parameters estimating of the semi regressive model for the VIX index would be to hold the algorithm:

- Simulate U_t points which is R[0,1] (uniform) or transform the existing sample into R[0,1];
- Build scatter plot for (U_{t-1}, U_t);
- Make several statistical tests to find the suited distribution of data;
- Taking into account scatter plot and distribution of data try to choose copula from existing class or build your own copula, if you know marginal distributions;
- Test copula consistency to data (for example AIB and BIC);
- Find regression parameters.

Using Matlab program we have built scatter plots for VIX index transformed into uniform distribution (R[0,1]) and non transformed data.



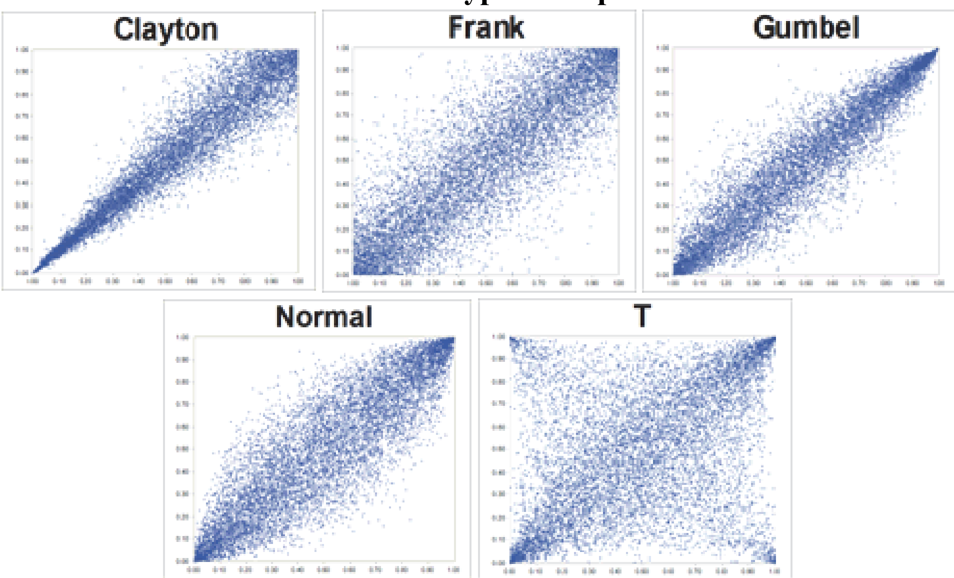
Graph1. Scatter plot for non transformed into R[0,1] VIX index data



Graph 2. Scatter plot for transformed into R[0,1] VIX index data

An important issue faced by an applied researcher interested in using the class of semi parametric copula-based time series models is the choice of an appropriate parametric copula. In different papers Chen et al. (2003) [9] propose two simple tests for the correct specification of a parametric copula in the context of modeling the contemporaneous dependence between several univariate time series and of the innovations of univariate GARCH models used to filter each univariate time series; (2) Chen and Fan (2004b) [10] establish pseudo-likelihood ratio tests for selection of parametric copula models for multivariate i.i.d. observations under copula misspecification [1]. But our suggestion is simpler – we can choose the best copula fit using AIC and BIC criteria or using χ^2 test for data distribution. We take for different copula comparisons
AIC and BIC (see Table 1).

Most common types of copula in finance



Graph 3. Scatter plot for transformed into R[0,1] VIX index data

For the first copula choosing step it is reasonable to compare graphical parametric copulas with VIX data scatter plot (Graph 2). As we can see the most suitable copulas for our data are Gumbel, Frank and Normal. For this sample of copulas is useful to calculate AIC and BIC criteria.

Copula	AIC	BIC
Gumbel copula	-124,1	-119,3
Frank copula	-267,4	-261,5
Normal copula	-230,3	-227,9

Table 1. **AIC and BIC criteria for VIX index data.**

Taking into account AIC and BIC criteria we should choose Frank copula for further model estimation. Let see how to derive semi parametric regression parameters using Frank copula representation:

$$C(u_{t+1},u_t) = -a^{-1}\ln\left(1+\frac{g_{u_{t+1}}g_{u_t}}{g_1}\right) \quad g_{ut} = e^{-au_t}-1 \tag{5}$$

And insert expression (5) into conditional expectation, we get our parameters:

$$E(U_{t+1}\,|\,U_t=u) = \int_0^1 u_{t+1}\,dF_{u_{t+1}|u_t}(u) = \int_0^1 u_{t+1}\,p(u_{t+1}\,|\,u_t)du_{t+1} = \int_0^1 u_{t+1}\frac{\partial C(u_{t+1},u_t)}{\partial u_{t+1}\partial u_t}du_{t+1} = \int_0^1 u_{t+1}\,c(u_{t+1},u_t)du_{t+1}$$

$$= -a\int_0^1 u_{t+1}\frac{g_1\left(1+g_{u_t+u_{t+1}}\right)}{\left(g_{u_t}g_{u_{t+1}}+g_1\right)^2}du \tag{6}$$

$$\mathrm{g}((U_{t+1}\,|\,U_t=u) = E((U_{t+1}-f(U_t)^2/U_t=u) = \int_0^1 (U_{t+1}-f(U_t)^2 c(U_{t+1},U_t)dU_{t+1} \tag{7}$$

It is impossible to solve analytically (5) and (6) expressions. But numerically it is doable for example in the Matlab. For the Frank copula we can use inverse function with the aim to return to our base equation (1). Of course, if we want use this model in practice, it is crucial to compare different class models which could be suitable for this data. This can give applied added value for this method.

But if we deal with copulas we should not skip some facts. For example, it is not easy to say which parametric copula best fits a given dataset, since some copulas may fit better near the center and other near the tails and many copulas do not have moments that are directly related to the Pearson correlation, it is difficult to compare financial models based on correlation.

Conclusions and further work

The algorithm for copula simulation and semi parametric regression coefficients finding through Markov chain have been presented. For Option VIX index data was found via MatLab the best fitted copula model, which is Frank copula. According to this copula, were shown principals of the semi parametric regression model coefficients evaluation. The next step in this research can be evaluating of the applied characteristic of the copula based semi parametric model as well as studying efficient estimation of conditional variance function in stochastic regression and to build continuous stochastic model using limit theorems.

References

[1] X.CHEN, Y.FAN. Estimation of copula-based semiparametric time series models. Journal of Econometrics, 2006.

[2] D.B. NELSON. ARCH models as di_usion approximations. Journal of Econometrics, 441 1-2: 7|38, 1990.

[3] Y. AIT-SAHALIA, R. KIMMEL. Maximum likelihood estimation of stochastic volatility models. Journal of Financial Economics, 2007.

[4] DARSOW, W., NGUYEN, B., OLSEN, E., 1992. Copulas and Markov processes. Illinois Journal of Mathematics 36, 600–642.

[5] JOE, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall/CRC;

[6] FREES, E.W., VALDEZ, E.A., 1998. Understanding relationships using copulas. North American Actuarial Journal 2, 1–25.

[7] EMBRECHTS, P., MCNEIL, A., Straumann, D., 2002. Correlation and dependence properties in risk management: properties and pitfalls. In: Dempster, M. (Ed.), Risk Management: Value at Risk and Beyond. Cambridge University Press, Cambridge, pp. 176–223.

[8] SKLAR, A., 1959. Fonctions de r'epartition 'a n dimensionset leurs marges. Publ. Inst. Statis. Univ. Paris 8,229–231.

[9] CHEN, X., HANSEN, L.P., CARRASCO, M., 1998. Nonlinearity and temporal dependence. Working Paper,University of Chicago.

[10] CHEN, X., FAN, Y., 2004b. Pseudo-likelihood ratio tests for model selection in semiparametric multivariate copula models. Canadian Journal of Statistics.

Current address

Jegors Fjodorovs, Phd student
Affiliation: Department of Probability Theory and Mathematical Statistics, Riga Technical University
Address: 1 Kalku Street, LV-1658, Riga, Latvia
e-mail: Jegors.Fjodorovs@rtu.lv

Andrejs Matvejevs, Dr.sc.ing., Professor
Affiliation: Department of Probability Theory and Mathematical Statistics, Riga Technical University
Address: 1 Kalku Street, LV-1658, Riga, Latvia
e-mail: Andrejs.Matvejevs@rtu.lv