The Empirical Econometrics and Quantitative Economics Letters
ISSN 2286 – 7147
Volume 3, Number 2 (June 2014), pp. 1 - 14.

Value at risk of biofuel portfolio based on extreme value theory and copulas

Xiao Deng*, Kanchana Chokethaworn and Chukiat Chaiboonsri

**Faculty of Economics, Chiang Mai University*
E-mail: toby898@gmail.com

ABSTRACT

Value at Risk (VaR) is a risk measurement defined as the worst loss of a portfolio over a given confidence level. Assessing the extreme events is crucial in financial risk management. This paper estimates portfolio VaR using an approach combining Copula functions, Extreme Value Theory (EVT) and GARCH model. We apply this approach to a portfolio consisting of biofuel stock sub-indices: WTI crude oil, corn, soybean and wheat. To estimate the VaR of this portfolio, wefirst use an asymmetric GARCH model and an EVT method to model the marginal distributions of each log returns series. Then we use Copula function to link the marginal distributions together into a multivariate distribution. Last, we use Monte Carlo Simulation approach to find the estimates of the portfolio VaR and backtest. The major result demonstrates that this approach performs better than traditional methods in risk management of extreme events especially for high volatility period.

Keywords: Value at risk, risk management, extreme value theory, copulas, biofuel portfolio

JEL Classification: C58, G01, G11

1. Introduction

Human face several challenges related to environmental destruction, poverty, hunger and shortage of natural energy. Rapidly growing consumption of fossil energy in the transportation sector in the last century caused problems such as increasing global warming, growing non-regeneration energy dependency, inflation and supply insecurity. One useful approach to solve these problems is to increase and apply biofuels. Biofuels have recently been used heavily in the U.S, European Union and Brazil. In Europe, the EU set clear targets for biofuel use in the transportation sector; the sector is responsible for 57.7% of global fossil fuel use. These biofuel initiatives are due to environmental and energy security or energy independence concerns.

The new development on global food commodity markets is attributed to different sources. There were some known causal factors that caused the first food price spike in 2008. They were low harvests due to unfavorable weather conditions, the exchange rate of the dollar, high oil prices and increased use of biofuel energy. The view that biofuels have created a new link between oil and food products is supported by various authors; oil price are transferred into food prices. Equity investors concern that the interconnections of agriculture and energy markets have increased through the rise in the new biofuel agribusinesses and the oil–ethanol–corn linkages. This phenomenon has led to an increasing interest in creating an optimal dynamic biofuel portfolio that consists of crude oil, corn, soybean and wheat in both the future and equity markets. This paper aims at helping petroleum companies or other index-based investors for their forecasting as well as risk management.

2. Literature Reviews

Value-at-Risk (VaR) is an approach widely used to quantifying market risk. It yields an estimate of the likely losses which can rise from price changes over a certain horizon at a given confidence level. Its advantages are that it summarizes risk in a single and easy to understand number. It does not depend on a specific kind of distribution. Therefore, theoretically, it can be applied to any kind of financial asset. Inaccurate portfolio VaR estimates may lead firms to maintain insufficient risk capital reserves so that they have an inadequate capital cushion to absorb large financial shocks. For example, several major financial institutions crashed not long after the breakout of recent financial crises, e.g. East Asian financial crisis in 1997. Some of these failures were associated with substantial portfolio VaR estimation errors.

In this paper, we restrict to Elliptical Copulas, as only the Gauss and the t-copula turn out to be tractable copula models for multidimensional portfolios consisting of more than three assets (Nelsen, 2006; Patton, 2009). The main question that this paper will try to answer is finding the VaR of a portfolio based on new methods like EVT and copulas. Its focus will be Dow Jones UBS equity market.

Currently, most of current research papers on VaR estimation focus on the univariate case making it undesirable for portfolio risk management. Moreover, most of the significant research contributions to the literature on portfolio VaR are limited to estimators of marginal VaR, component VaR, and incremental VaR instead of portfolio

VaR itself (Hallerbach, 2002). This study employs new framework for portfolio VaR estimations, which integrates asymmetric GJR-GARCH models for time-varying return distribution of individual assets, extreme value theory (Embrechts et al, 1997) for tail distributions, and copula functions (Nelsen, 1999) for the dependency structure on all assets of a portfolio.

In the field of biofuel portfolio, Ting-Huan Chang and Hsin-Mei Su (2010) study the Value-at-risk estimation with the optimal dynamic biofuel portfolio. The authors expand their analysis to encompass renewable sources, such as corn and soybeans, under the current low-carbon biofuel obligations. The paper employs GARCH (1, 1) and ARJI models to estimate the one-day-ahead Value-at-Risk (VaR) of the optimal dynamic biofuel portfolio which consists of crude oil, corn and soybeans. The results of out-of-sample forecasts are also represents that their models play important roles in VaR estimation and risk management for biofuel portfolio. The authors therefore suggest that the petroleum companies should simultaneously pay attention to avoid the risk by hedging material costs according to the prices of energy-related crops.

In this study, we apply this approach to four biofuel-related sub-index of Dow Jones which form w portfolio. This study started the estimation by first, performing a preliminary analysis on the four sub-index log return series to verify some typical assumed properties of log returns, namely normality and i.i.d. properties. We found that the series were heavy-tailed and moreover they were not i.i.d. To solve this we fitted an AR(1)-GJR-GARCH(1,1) model to the series and divided the obtained residuals by their corresponding volatilities to get standardized residuals which were approximately i.i.d. Then we estimate the marginal distribution of each series. By combining the Kernel Density Estimation method and the Peak over Threshold (POT) method, we can achieve this. That is, having the standardized i.i.d. residuals we estimated the interior of the empirical cumulative density functions (CDF) using the Kernel Density method and the POT method in the tails. We thus obtained semi-parametric empirical CDF for each series. After that, the most important is the modeling of dependence structure using *t* copula function by the Canonical Maximum Likelihood method. It consists of,first transform the standardized residuals to uniform variates by the semi-parametric empirical CDF and then fit the Copula to the transformed data. At last the portfolio VaR can be estimated using Monte Carlo Simulation method and back tested.

3. Extreme Value Theory

Extreme Value Theory is a robust framework which provides simple parametric models to capture the extreme tails of a distribution and to forecast risk. Mainly there are two broad approaches of applying EVT (Embrechts et al, 1997), the Block Maximum and the modern approach of peaks-over-threshold (POT). The Block Maximum's principle is dividing the time interval in equal chunks or blocks and modeling only the maximum loss for each of the blocks based on Generalized Extreme Value (GEV) distribution. This approach is considered wasteful of data because only one observation for each block is used. The POT models are modern methods for EVT. They directly model all large observations which exceed a high threshold.

3.1 Generalized Pareto Distribution (GPD)

For the marginal return distributions, separate GP models are fit to both the lower and upper distribution tails. Under the parameterization of the GP tail model, the tail distribution is represented by the complement of the GP cumulative distribution function (CDF):

$$G_{\xi,\beta}(x) = \begin{cases} 1-\left(1+\xi\frac{x}{\beta}\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ \left(1-e^{\frac{-x}{\beta}}\right) & \xi = 0 \end{cases} \tag{1}$$

Let us define the excess distribution above the threshold u as the conditional probability:

$$F_u(y) = P(X-u \leq y \mid X > u) \tag{2}$$

It is very easy to derive that in terms of the CDF of X (denoting it by F), we have

$$F_u(y) = \frac{F(y+u)-F(u)}{1-F(u)} = \frac{F(x)-F(u)}{1-F(u)} \tag{3}$$

By the theorems of Pickands (1975), for a large class of underlying distribution functions F the conditional excess distribution function $F_u(y)$, for u large, is well approximated to GPD.

$$\lim_{u\to x_0} \sup_{0\leq y\leq x_0} \left|F_u(y)-G_{\xi,\beta}(y)\right| = 0 \tag{4}$$

It means that for a large class of underlying distribution F, as the threshold u is progressively raised, the excess distribution F_u will converge to a generalized Pareto distribution. The resultant parameter estimations are functions of the selected threshold u. The choice of the threshold value u is crucial in order to obtain a good estimation in MLE. In fact, if u is too high, we have only a few exceedances data and the variance of the estimators is high. If u is too low, the estimators are biased because the relation (4) does not hold.

Setting $x = y + u$ and combining results of equation (2) and (3), our model can be written as

$$F(x) = \big(1-F(u)\big)G_{\xi,\beta}(x-u) + F(u) \qquad \text{for} \qquad x > u \tag{5}$$

Using equation (5) and the empirical estimate $\frac{n-N_u}{n}$ for $F(u)$, where n is the total number of observations and N_u the number of observations over the threshold u. Putting the empirical estimator of $F(u)$ and our estimated parameters $(\hat{\xi},\widehat{\beta})$ of the GPD together, we arrive at the tail estimator:

$$\hat{F}(x) = 1-\frac{N_u}{N}\left(1+\hat{\xi}\,\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}} \tag{6}$$

4. Copula Theory

Copulas are multivariate distribution functions that allow the decomposing of any n-dimensional joint distribution into its n marginal distributions and a copula function. In practice, a copula is often used to construct a joint distribution function by combining the marginal distributions and the dependence between the variables.

An n-dimensional copula is a multivariate cumulative function, C, with uniform distribution margins in [0, 1] (U(0, 1]) and the following properties:

$C: [0, 1]^n \rightarrow [0, 1]$

C is grounded and n-increasing;

C has margins C_i which satisfy $C_i(u) = (1, \dots, 1, u, 1, \dots 1) = u$ for all $u \in [0, 1]$.

4.1 The Sklar's theorem: Let F be a joint distribution function with margins $F_1, \dots, F_n$. Then there exists an n-copula C such that

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (7)$$

If margins F_i are all continuous, then C is unique. Conversely, if C is an n-copula

And F_i are distribution functions, the function F defined above is an n-dimensional distribution function with margins F_i. According to the theorem above, a joint distribution can be separated into two parts. One is univariate marginal distributions, and the other is copula function.

4.2 Common types of Copula functions

The most common types of copula function in financial research are Elliptical copula, including the Gaussian copula and Student-t copula, and Archimedean copula. We now introduce the two copulas in this paper.

(A) The Gaussian copula is defined by:

$$C(u_1, u_2; \rho) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right) dx_1 dx_2 \quad (8)$$

where Φ^{-1} is the inverse of the standard Normal CDF, and is the usual linear correlation coefficient of the corresponding bivariate normal distribution. Gaussian copulas have neither upper tail dependence nor lower tail dependence for -1<ρ<1.

(B) The Student-t copula is defined by:

$$C(u_1, u_2; \rho, v) = \int_{-\infty}^{T_v^{-1}(u_1)} \int_{-\infty}^{T_v^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{(x_1^2 - 2\rho x_1 x_2 + x_2^2)}{v(1-\rho^2)}\right)^{-\frac{v+2}{2}} dx_1 dx_2 \quad (9)$$

where T^{-1} is the inverse of Student-t CDF with degree-of-freedom parameter v and correlation ρ, -1<ρ<1.

Both of the two copula functions above are symmetric. But they have different feature in tail dependence that is the Gaussian copula is independent in tail, while the Student-t copula has tail dependency. Longin and Solnik found that the correlation between markets increased in the period of higher volatility. Thus the dependence between stock market returns is inclined to be higher under extreme conditions; the Student-t copula tends to be a better alternative to capture the dependence between market returns. We will illustrate it in the Portfolio of sub-indices part.

5. Empirical Analysis

5.1 Data and descriptive statistics

This study uses the data set comprising the major biofuel related sub-indices of Dow Jones UBS which are WTI Crude Oil Sub-index (CL), Corn Sub-index (CN), Soybean Sub-index (SY), and Wheat Sub-index (WH). These data are collected from Google Finance. It covered ten years' daily data over the period of January 2nd 2003 through September 27th, 2013. The total numbers of observations are 2,717. Figure 1 is a representation of the 4 sub-indices price development for the last 10 years. These sub-indices are then transformed into daily returns. Figure 2 shows the daily sub-indices returns. It's obvious that they are highly correlated.

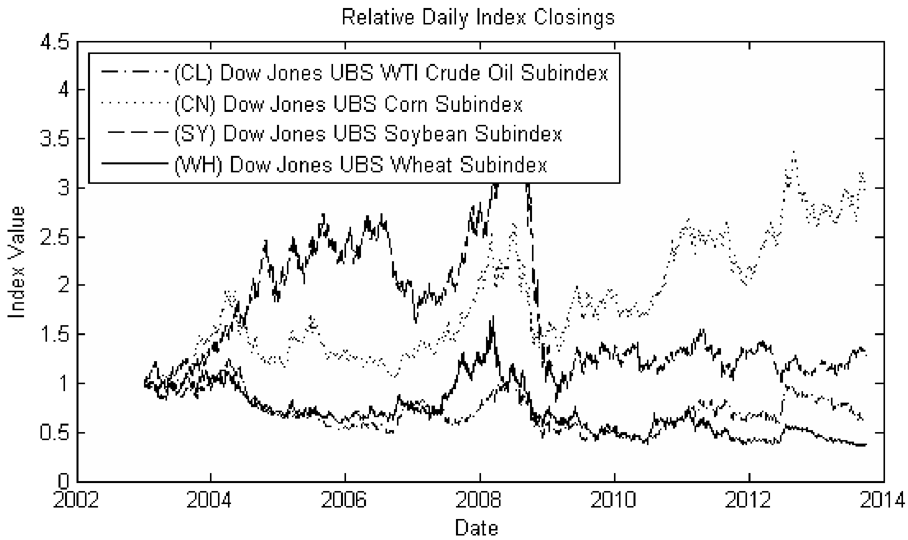

Figure 1. Daily closing prices of 4 sub-indices

TABLE 1. Summary statistics

	CL	CN	SY	WH
Mean	19.80859	152.2590	339.2966	27.55677
Std. Dev.	5.064452	48.30816	130.7388	9.322231
Skewness	0.500750	0.541964	1.147077	0.801543
Kurtosis	2.668607	2.216221	3.843920	3.196179
Minimum	10.60560	78.52200	162.2760	14.21410
Maximum	34.83900	279.6945	831.1600	65.39800

Source: Computation

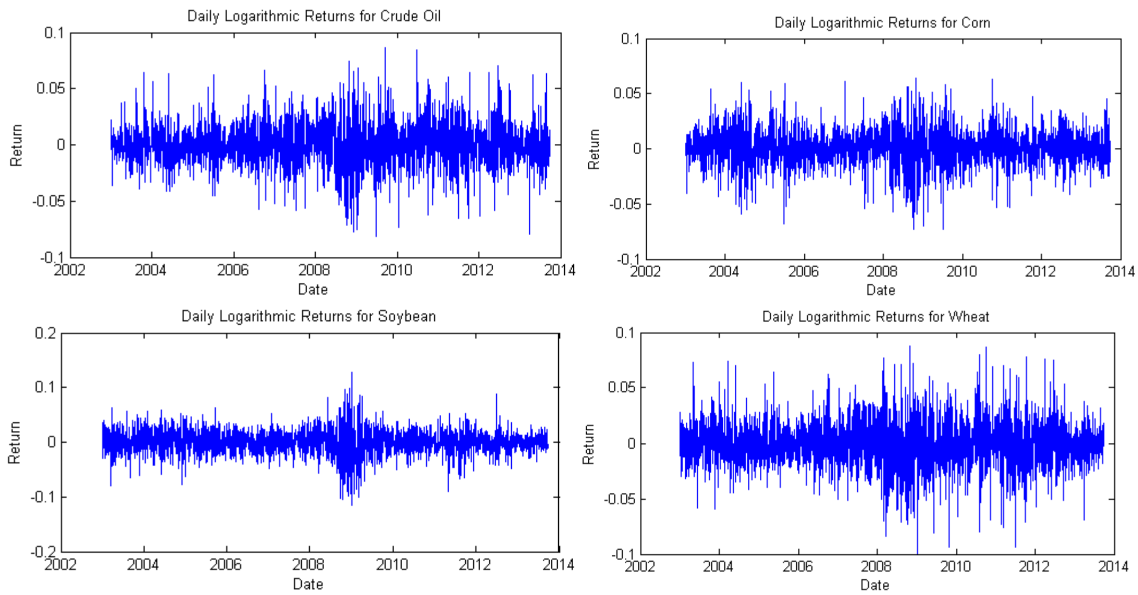

Figure 2. Daily returns of 4 sub-indices

Modeling the tails of a distribution with a GPD requires the observations to be approximately independent and identically distributed (i.i.d.). However, most financial return series exhibit some degree of autocorrelation and, more importantly, heteroskedasticity. Figure 3 shows sample ACF (autocorrelation function) of returns and sample ACF of squared returns for the four sub-indices. The ACF of returns reveals some mild serial correlation. However, the sample ACF of the squared returns illustrates significant degree of persistence in variance, which implies that we need a GARCH model to condition the data for the subsequent tail estimation process.

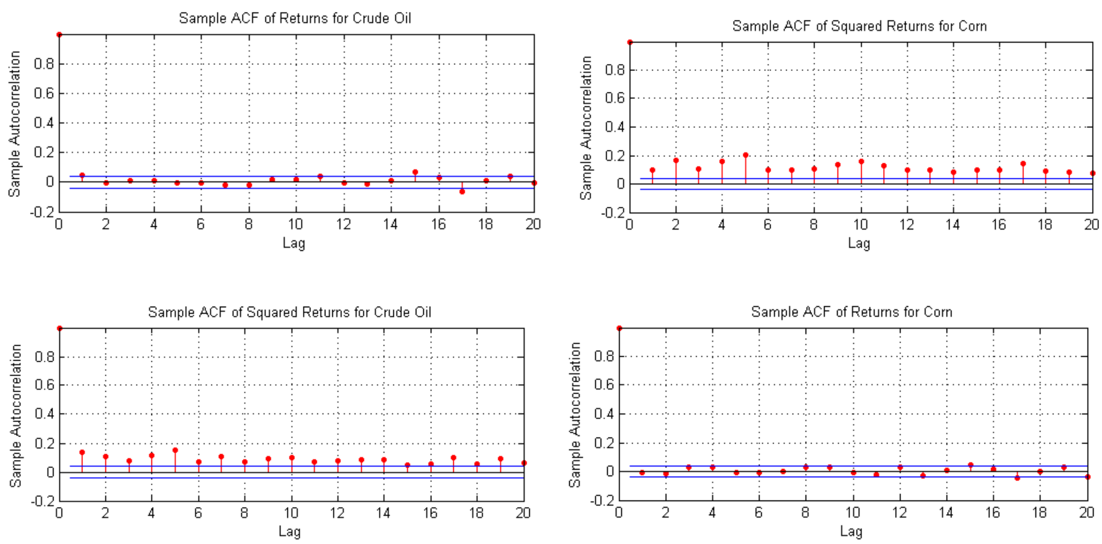

Figure 3. ACF plots of 4 sub-indices

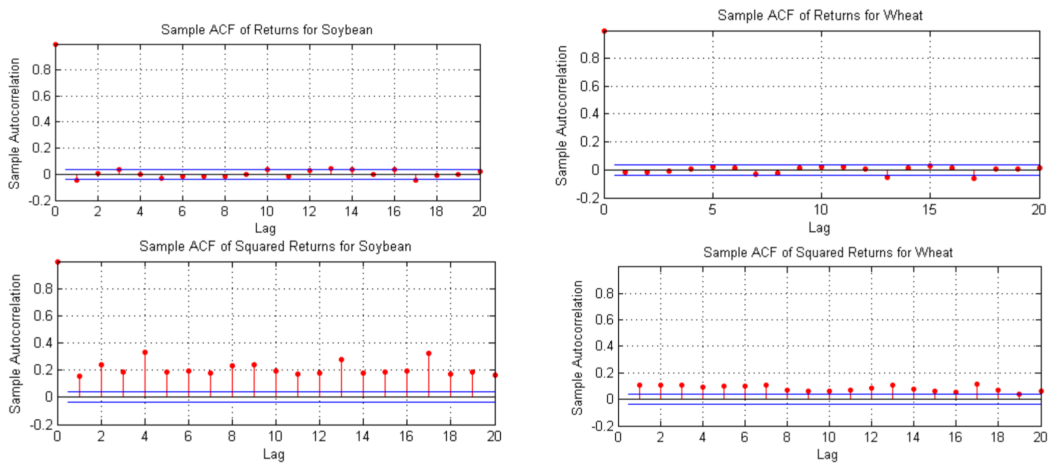

Figure 3. ACF plots of 4 sub-indices – continued

5.2 Model Estimations

In this paper, AR(1)-GJR-GARCH(1,1) model is applied to produce a series of i.i.d. observations to each index. The model is presented as follows:

$$r_t = c_0 + \theta r_{t-1} + \varepsilon_t$$

$$\varepsilon_t = \sigma_t z_t \quad z_t \sim i.i.d$$

$$\sigma_t^2 = K + \alpha\sigma_{t-1}^2 + \beta\varepsilon_{t-1}^2 + \varphi\varepsilon_{t-1}^2 I_{t-1}$$

where

$$I_{t-1} = 0 \ if \ \varepsilon_{t-1}^2 \geq 0, and \ I_{t-1} = 1 \ if \ \varepsilon_{t-1}^2 < 0.$$

In the model, r_t is the index return, and σ_t the volatility. The GJR-GARCH model could incorporate asymmetric leverage effects for volatility clustering. Additionally, the standardized residuals of each index are modeled as a standardized Student's t distribution to compensate for fat tails often associated with equity returns. The final parameters for each time series are summarized in Table 2. Figure 4 are filtered model residuals from each index. Each lower graph of Figure 4 clearly illustrates the variation in volatility (heteroskedasticity) present in the filtered residuals. Subsequently, we standardize the residuals by the corresponding conditional standard deviation. These standardized residuals represent the underlying zero-mean, unit-variance, i.i.d. series upon which the EVT estimation of the sample CDF tails is based.

TABLE 2. Parameter estimates for AR(1)-GJR-GARCH(1,1) model

	CL	CN	SY	WH
c_0	-3.5793e-04	6.6946e-04	4.7467e-04	-4.6312e-04
θ	0.041496	-0.008350	-0.030384	-0.017874
K	4.3309e-06	2.5632e-06	5.8200e-06	2.7539e-06
α	0.925521	0.946675	0.935002	0.958685
β	0.057945	0.046804	0.021240	0.049386
φ	0.016556	-0.004037	0.056993	-0.028929

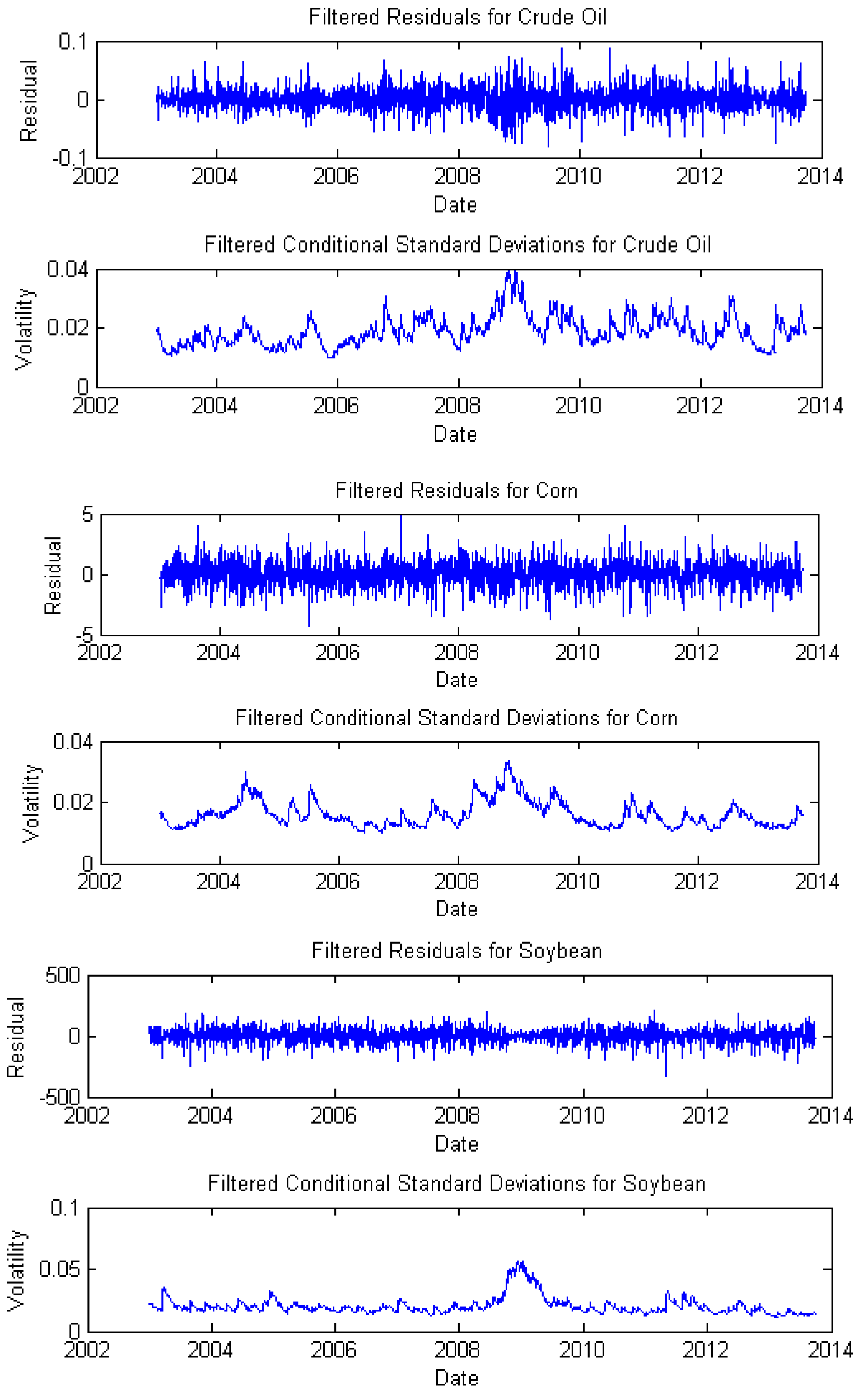

Figure 4.Filtered residuals and volatility of 4 sub-indices

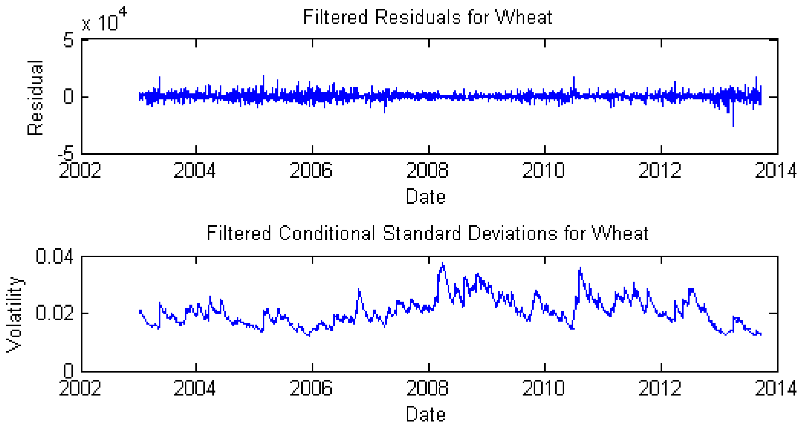

Figure 4. (continued) Filtered residuals and volatility of 4 sub-indices

5.3 VaR calculations and backtest

We transform the individual standardized residuals of AR(1)-GJR-GARCH(1,1) models to uniform variates by the semi-parametric empirical CDF, and then fit the *t* copula to the transformed data. The estimated optimal degree of freedom (*v*) of the *t* copula is 11.1248. Subsequently, this study simulates jointly dependent biofuel index returns by reversing the above steps. We simulate 10000 independent random trials of dependent standardized index residuals for a risk horizon of 1, 10 and 22 trading days. Then, using the simulated standardized residuals as the i.i.d. input noise process, reintroduce the autocorrelation and heteroskedasticity of GJR-GARCH model observed in the original index returns. Finally, given the simulated returns of each index, we form a 1/4 equally weighted index portfolio composed of the individual indices, and calculate the VaR at 99% confidence level at certain risk horizon. The estimated 90 %, 95%, and 99% VaR for *t* (11.1248) are listed in Table 3. For reference, other models of Historical Simulation and GJR-GARCH+ Gaussian are presented in Table 4.

TABLE 3: VaR calculated for a portfolio with equal weights at different risk horizon

Monte-Carlo Simulation and t-copula			
	1 trading day	10 trading days	22 trading days
90% VaR	1.3913%	4.6132%	6.8356%
95% VaR	1.8734%	5.9563%	8.9752%
99% VaR	2.8338%	8.7992%	13.3695%
Max Loss	6.4922%	17.6264%	21.9097%
Max Gain	7.1914%	16.9904%	22.5827%

Source: computation

After calculating the VaRs, for reliability it is necessary to back test whether the VaR model used has adequately estimated the real extreme risk or not. The failure rate is widely applied in studying the effectiveness of VaR models. The definition of failure rate is the proportion of the number of times the observations exceed the forecasted VaR to the number of all observations. Now we back test the VaR estimations at 99% over a time window of 200 days and compare the results with other traditional models. The number of VaR exceedances for each model is counted and it equals the number of time in which the effective loss is greater than the 99% VaR estimations. The principal results of this backtesing procedure are shown in Table 5. According to the results shown in Table 5, the failure rate of EVT + t copula model is nearest to 1%, 5%, 10% respectively. This finding implies our model used in this study outperforms traditional VaR model. Empirical results clearly demonstrate that the EVT and copula based method performs best. In addition, we find that the historical simulation and GJR-GARCH + Gaussian copula overestimate the VaR of portfolio. Figure 5 is the profit and loss distribution of our model based on EVT and copula. Figure 6 plot the profit and loss distributions of the model based on GJR-GARCH and Gaussian copula. Figure 6 plot the profit and loss distributions of the model based on historical simulation model.

TABLE 4: VaR of different model at 22 days risk horizon

	EVT + t copula	GARCH + Gaussian	Historical Simulation
90% VaR	6.8356%	6.7130%	6.1360%
95% VaR	8.9752%	8.9593%	8.1228%
99% VaR	13.3695%	13.6228%	11.9364%
Max Loss	21.9097%	21.9387%	23.0903%
Max Gain	22.5827%	23.8599 %	25.6832%

Source: computation

TABLE 5: Failure rate for each model

	EVT + t copula	GARCH + Gaussian	Historical Simulation
Failure Rate $\alpha = 0.1$	0.08	0.06	0
Failure Rate $\alpha = 0.05$	0.03	0.025	0
Failure Rate $\alpha = 0.01$	0.005	0	0

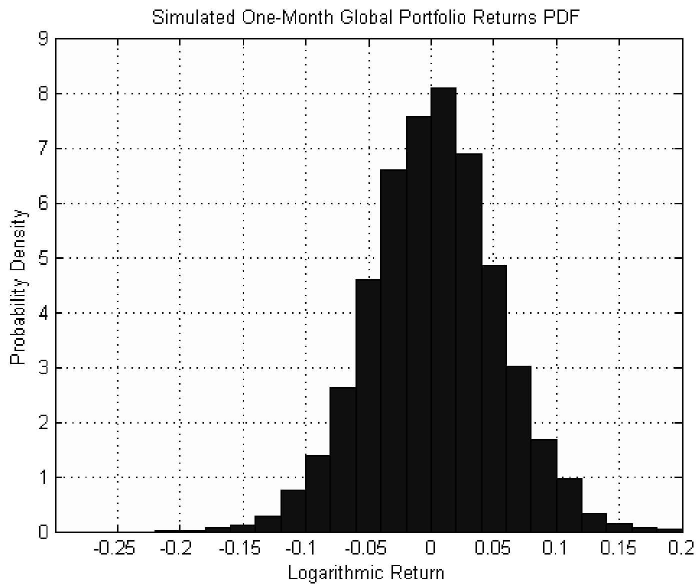

Figure 5. Portfolio profit and loss distribution (EVT + t copula)

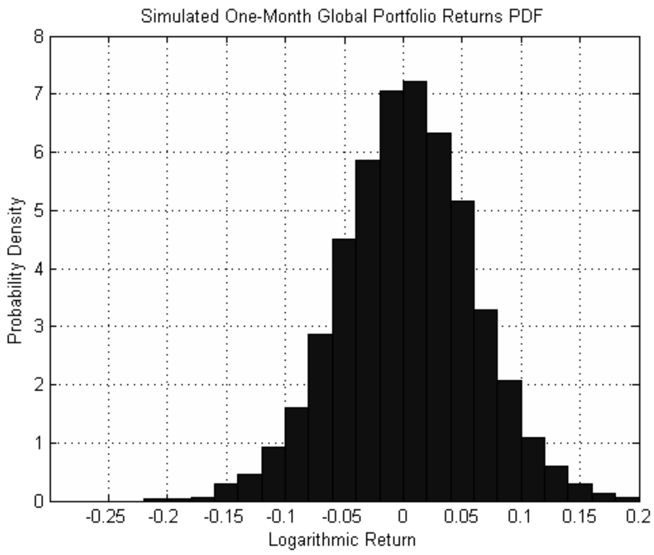

Figure 6. Portfolio profit and loss distribution (GJR GARCH + Gaussian copula)

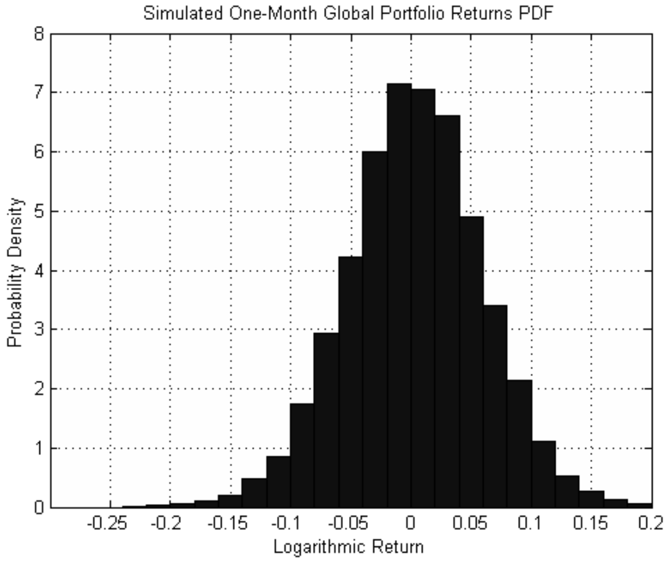

Figure 7. Portfolio profit and loss distribution (Historical Simulation)

6. Conclusions

The study incorporated a GJR-GARCH model with the copula function and EVT to model the time-varying return distribution. This approach focuses on the entire distribution rather than only the tail distribution (Byström, 2004). It estimates portfolio VaR more accurately than traditional models.

Our procedure starts with the GJR-GARCH model to estimate the conditional mean and volatility of the each asset. Then, in the second stage, the POT method of EVT is used to model the tail distribution of the residual. Finally, a four-dimensional *t* copula is fitted to the data and used to induce correlation between the simulated residuals of each asset.

In addition, our method can also be extended to make investment portfolio containing assets from many sectors. Results of this study can be used for risk management on global investments.

Our conclusion is that the Extreme Value Theory is a good tool for risk measurement of extreme events and especially for high volatility period. For high volatility samples and for 22 trading days, we got a VaR estimation of 13.36% (at 99% confidence level), 8.97% (at 95% confidence level), and 6.83% (at 90% confidence level).

We suggest that future research may consider dynamic copula in the dependence structure. Moreover, the further work needs to be done to test the sensitivity of this model based on the choice of threshold level *u*. Another point of interest may be at the sensitivity analysis based on the choice of degrees of freedom of *t* copula.

REFERENCES

Byström, H. N. E. 2004. "Managing extreme risks in tranquil and volatile markets using conditional extreme value theory," International Review of Financial Analysis 13: pp. 133–152.

Chang, T.H. and Su, H.M., 2010. "The substitutive effect of biofuels on fossil fuels in the lower and higher crude oil price periods," Energy 35: pp. 2807–2813.

Embrechts, P., F. Lindskog, and A. J. McNeil 2001, Modelling dependence with copulas and applications to risk management, ETH Zurich, preprint.

Gencay, R. and Selcuk, F. 2004. "Extreme value theory and value-at-risk: Relative performance in emerging markets," International Journal of Forecasting 20, 2: pp. 287-303.

Ho, L.C., P. Burridge, J. Cadle, and M. Theobald. 2000. "Value-at-Risk: Applying the extreme value approach to Asian markets in recent financial turmoil," Pacific-Basin Finance Journal, 8: pp. 249-275

Marimoutou, V., Raggad, B. and Trabelsi, A. 2009. "Extreme value theory and value at risk: application to oil market," Energy Economics 31: pp. 519-530

McNeil, A. and Frey, R. 2000. "Estimation of Tail Related Risk Measure for Heteroscedastic Financial Time Series: An Extreme Value Approach," Journal of Empirical Finance 7: pp. 271-300.

Miller, D. E. 2003. "The fundamentals of risk measurement," Financial Analysts Journal 59: pp. 108–109.

Nelsen, R. 2006. An Introduction to Copulas, 2nd ed. New York: Springer International

Patton A. 2009. Copula based models for financial time series. Handbook of financial time series. Hidelberg: Springer Verlag.

Sadeghi, M. and Shavvalpour, S. 2006. "Energy risk management and value at risk modeling," Energy Policy 34: pp. 3367–3373.