Copulas Applications in Estimating Value-at-Risk (VaR): Iranian Crude Oil Prices

Faramarz Kashanchi and Pranesh Kumar

Abstract: Crude oil being the primary source of energy is been unquestioningly the main driving engine of every country in this world whether it is the oil producer economy and/or oil consumer economy. Crude oil, one of the key strategic products in the global market, may influence the economy of the exporting and importing countries. Iran is one of the major crude oil exporting partners of the Organization of the Petroleum Exporting Countries (OPEC). Analysis of the risk measures associated with the Iranian oil price data is of strategic importance to the Iranian government and policy makers in particular for the short-and long-term planning for setting up the oil production targets. Oil-price risk-management focuses mainly on when and how an organization can best prevent the costly exposure to the price risk. Value-at-Risk (VaR) is the commonly accepted instrument of risk-measure and is evaluated by analysing the negative tail of the probability distributions of the returns/profit and loss. Among several approaches for calculating VaR, the most common approaches are variance-covariance approach, historical simulation and Monte-Carlo simulation. Recently, copula functions have emerged as a powerful tool to model and simulate multivariate probability distributions. Copula applications have been noted predominantly in the areas of finance, actuary, economics and health and clinical studies. In addition, copulas are useful devices to deal with the non-normality and non-linearity issues which are frequently observed in cases of financial time series data. In this paper we shall apply copulas namely; Frank copula, Clayton copula and Gumbel copula to analyse the time series crude oil price data of Iran in respect of OPEC prices. Data considered are; i. Monthly average prices for a barrel of Iranian and OPEC crude oil, from January 1997 to December 2008, ii. Seasonal number of barrels of Iran’s crude oil export, from January 1997 to December 2008. The results will demonstrate copula simulated data are providing higher and lower relative change values on the upper and lower tails respectively in comparison to the original data.

Key words: Value-at-Risk (VaR), OPEC crude oil prices, Quantile distributions, Copula Functions, Monte Carlo simulations

* Corresponding author.
1. Introduction

Risk measurement for regulating the financial markets has been a key issue for market players since the systematic developments in present financial history. The portfolio theory [15, 17, 18, 23] placed financial risk measurement in practice which assumes that all risk could be partitioned into systematic, market risk and the residual, company-specific or idiosyncratic risk. The Capital Asset Pricing Model (CAPM) model which emerged practically as an easy risk measure model hypothesized that since market risk is relevant for securities pricing; only the market risk measurement $\beta$ (beta) is necessary. The CAPM model provided a readily measurable risk estimate that could be applied practically in real time market conditions. The only problem was that $\beta$ metrics proved to have only a tenuous connection to the actual security returns and thus raising questions on the appropriateness of $\beta$’s as the true risk measure [6,10,21].

With doubts on the appropriateness of $\beta$’s as the correct risk measure, practitioners searched for alternative risk measures which were both accurate and relatively inexpensive to estimate. Despite many other risk measures and models considered, Value-at-Risk (VaR) has been widely adopted. Since accurate risk measurement is essential to the financial institutions for proper risk management, many of such internal models were developed in-house by the financial institutions. The VaR model is complementary to many other internal risk measures such as RAROC developed by Bankers Trust in the 1970s [W1]. However, market forces during the late 1990s led to the evolution of VaR as a dominant risk measurement tool for financial firms. Concerned about the impact of the increasing risk environment on the safety and soundness of the banking systems, bank regulators in 1992 instituted the risk-adjusted capital requirements that levied a capital charge for both on-and off-balance sheet credit risk exposures. The primary reason for the widespread adoption of VaR was the decision of JP Morgan to develop an open architecture methodology known as RiskMetrics [W2]. RiskMetrics was supported by a publicly available database containing the critical inputs required to estimate the model. Another reason for the VaR adoption and popularity was the introduction in 1998 by the Bank for International Settlements (BIS) of international bank capital requirements that allowed relatively sophisticated banks to calculate their capital requirements based on their own internal models such as VaR. RiskMetrics quickly became the industry benchmark in risk measurement. Bank regulators worldwide allowed commercial banks to measure their market risk exposures using internal models that were often VaR-based. The market risk amendments to the Basel accord made in-house risk measurement models a mainstay in the financial sector [W4, W5].

In what follows, we present in section 2 the Value-at-Risk as a measure of market risk and VaR calculation methods. Assumptions and limitations of VaR are discussed in section 3. Section 4 presents a primer on copula functions, simulation algorithm and some commonly used copulas. The exploratory data analysis (EDA) of the Iran’s crude oil prices and export is described in section 5. In section 6, we present the results of the risk analysis from the data and copula simulated data on the Iran’s light and heavy crude oil prices. We provide concluding
2. Risk-Measure: Value-at-Risk (VaR)

The Chairman Weatherstone of JP Morgan in 1990’s asked a basic question: “How much can we lose with α% probability on our trading portfolio over a pre-set horizon?” [22] This was a question only about the risk metric and later became the basis for Value-at-Risk (VaR). The VaR, as a metric to answer this question of how much could be lost on a bad day, is based on a statistical/probability approach. VaR measures the worst expected loss under normal market conditions over a specific time period at a given confidence level (quantile). In statistical terms, a bad day is defined such that there is only α% probability that daily losses will exceed this amount for a given distribution of all possible daily returns over some recent past period. Thus, a bad day is defined such that there is only a α% probability of an even worse day. The 1% VaR denoted as $V_{aR1\%}$ indicates a daily loss that will be equaled or exceeded only 1 percent of the time. Alternatively, $V_{aR1\%}$ means that there is a 99% chance that tomorrow’s daily portfolio value will exceed today’s value less $V_{aR1\%}$ [W2].

For example, suppose a portfolio consists of a single asset with its current value being 100. Further, assume returns of the asset follow a normal probability distribution with annual mean return ($\mu$) as 10% and annual standard deviation ($\sigma$) as 20%. The portfolio manager generally will be interested in answers for questions like: What is the distribution of the end-of-year portfolio value? What is the maximum loss at the end of the year with say 1% probability (This is the VaR at 1%)? What is the probability that a loss of more than 20 will occur by the year-end (What is the probability that the end-of-year portfolio value will be less than 80)? Under the assumed scenario, the probability that the end-of-year portfolio value is less than 80 is 0.02275. With a probability of 1% the end-of-year portfolio value will be less than 76.74, i.e., the VaR of the distribution is (100-76.74)=23.26. This is the maximum loss that the portfolio may sustain over one year for 99% of the time. Increasing confidence level will result in increase in VaR.

With several assets in the portfolio, the lognormal distribution is a more reasonable assumption about the distribution for the prices than the normal distribution because lognormal cannot take negative values. Suppose that the natural logarithm of the portfolio value is normally distributed with annual mean $\mu$ and annual standard deviation $\sigma$. Denoting the portfolio value at time T by $V_T$, 

$$ln(V_T) \sim Normal[ln(V_0) + (\mu - \frac{\sigma^2}{2})T, \sigma T]$$  

(2.1)

It may be noted that the term $\frac{\sigma^2T}{2}$ is due to Ito’s lemma.

With reference to the portfolio considered above, we have $V = 100$, $\mu = 10\%$, $\sigma = 20\%$. Thus, the log of the portfolio value at year-end will be
\[ \ln(V_T) \sim \text{Normal}[\ln(100) + (0.1 - \frac{0.2^2}{2})(1), 0.2(1)] = \text{Normal}(4.6852, 0.2). \] 

(2.2)

Thus, the probability that the end-of-year portfolio value is less than 80 = 0.0648. With a probability of 1\% the end-of-year portfolio value will be less than 68.03, i.e., the VaR of the distribution is (100-68.03)=31.97.

In a two-asset portfolio, VaR can be calculated using

\[
\text{VaR} = \sqrt{\omega_1^2 \sigma_1^2 + \omega_2^2 \sigma_2^2 + 2\rho \omega_1 \omega_2 \sigma_1 \sigma_2},
\]  

(2.3)

where \( \omega \)'s are the weights of assets, \( \sigma \)'s are the standard deviations of the assets and \( \rho \) is the correlation coefficient between two assets. Thus, the undiversified VaR is the weighted average of the individual standard deviations while the diversified VaR which takes into account the correlation between the assets is the square root of the variance of the portfolio. The diversified VaR measure is used to set trading limits while the larger undiversified VaR measure is used to gauge an idea of the risk exposure in the event of a significant correction or market crash. Therefore, VaR is essentially measured in terms of the standard deviation of a portfolio which is used as an indicator of the volatility of that portfolio. A portfolio exhibiting high volatility will have a high VaR value concluding that the portfolio has a higher chance of making losses [1, 2, 3, 4, 6, 22].

2.1 VaR: Definition

Let random variable \( X \) be the portfolio value over the \( d \)-period horizon and \( f(x) \) be its probability density function. Then for a confidence level \( \alpha \% \),

\[
\alpha = Pr(\text{Value} > 100 - \text{VaR}) = \int_{100-\text{VaR}}^{\infty} f(x)dx. \]  

(2.4)

Alternatively if \( g(x) \) is the probability density function of losses, then for a confidence level \( \alpha \% \),

\[
\alpha = Pr(\text{Loss} > -\text{VaR}) = \int_{-\text{VaR}}^{\infty} g(x)dx. \]  

(2.5)

The time period \( (d) \) and confidence level \( (\alpha) \) are two main parameters to be selected appropriately for the risk measure. In mathematical terms, Value-at-Risk (VaR) is the \((1 - \alpha)\)th quantile of the distribution of the \( d \)-period return value of a given portfolio \( P \). Thus,

\[
\text{VaR}_{\alpha,d}(P) = -F_{P_d}^{-1}(1-\alpha)\text{PV}(P), \]  

(2.6)

where \( \alpha \) is the confidence level, \( P_d \) is the change in value of the portfolio \( P \) over time period \( d \), i.e., the \( d \)-period return, \( F_{P_d} \) is the cumulative probability distribution of \( P_d \) and \( \text{PV}(P) \) is the present worth of the portfolio [13, 23].
2.2 VaR: Calculations

Among several approaches for calculating VaR, the most common approaches are variance-covariance approach (VCA), historical simulation, Monte-Carlo simulation (MCS).

2.2.1 Variance-Covariance Approach (VCA)

In VCA, assumptions are: (i) Observed returns are normally distributed, (ii) Correlations between risk factors are constant and (iii) Delta (price sensitivity to changes in a risk factor) of each portfolio component is constant. The volatility of each risk factor measured by its standard deviation ($\sigma$) is estimated from the historical data. The potential effect of each component of the portfolio on the overall portfolio value is then worked out from the component’s delta (with respect to a particular risk factor) and risk factor’s volatility.

The d-period VaR of a portfolio with single asset $X$ at $\alpha$% confidence level is defined by

$$ VaR_{\alpha,d}(X) = -\sqrt{d}N_{0,1}^{-1}(1 - \alpha)\sigma PV(X), $$

where $N_{0,1}^{-1}$ is the inverse of the standard normal distribution function and $PV(X)$ is the present value of the asset $X$.

VaR calculated for a portfolio with multiple $k$-assets is

$$ VaR_{\alpha,d}(X) = -\sqrt{d}N_{0,1}^{-1}(1 - \alpha)\sqrt{Y\Sigma Y^T}, $$

where $\Sigma$ is the $n \times n$ variance-covariance matrix of the $k$-assets, $Y = PV(X)$ is a vector of length $k$ and $Y$ is the vector of amounts invested in $k$-assets. It may be noted in cases such as options or bonds that the non-linearity among the risk factors, if present, is assumed to be sufficiently limited so that it may be ignored. For very short holding periods, $VaR$ of an option can be obtained using the Delta of the option (First-order Taylor series approximation of the change in value of the option).

There are different methods of calculating the relevant risk factor volatilities and correlations. In the simple historic volatility method, effects of a large one-off market move can significantly distort volatilities over the required forecasting period. For example if using 30-day historic volatility, a market shock will stay in the volatility cycle for 30 days until it falls out of the sample range and correspondingly causes a sharp drop in historic volatility 30 days after the event. This is because each past observation is equally weighted in the volatility calculation. Alternatively, past observations are assigned unequal weights. Higher weights are assigned to the recent observations so that large jumps in volatility are not caused by events that occurred some time ago. RiskMetrics is based on one such method which uses exponentially weighted moving averages (EWMA) of the square of the price returns ($R$) to estimate the variance-covariance matrix. The variances are calculated by
and covariances by
\[
\sigma^2_{d,i} = \lambda \sigma^2_{d,i-1} + \left(1 - \lambda \right) R^2_i,
\]
(2.9)

and covariances by
\[
\sigma^2_{12d,i} = \lambda \sigma^2_{12d,i-1} + \left(1 - \lambda \right) R^2_{1i} R^2_{2i},
\]
(2.10)

2.2.2 Historical Simulation

The historical simulation approach to calculate VaR avoids some of the pitfalls of the correlation based methods because three main assumptions: (i) normally distributed returns, (ii) constant correlations and (iii) constant Deltas are not required. The model calculates potential losses using actual historical returns in the risk factors and thus captures the non-normal patterns of risk factor returns. This means rare events and crashes can be included in the analysis. Since the risk factor returns used for revaluing the portfolio are actual past movements, the correlations in the calculation are also actual past correlations. They capture the dynamic nature of correlation as well as the scenarios when the usual correlation breaks down [5, 15, 26].

Historical simulation is based on the order statistics. Let \( j \)-th be the number that denotes the \( \alpha \)-th quantile of the order statistics with \( n \) observations. Then \( x_j \) is the \( \alpha \)-th quantile value of the ordered data series \( X \) which consists of \( n \) observation such that \( \alpha = \frac{j}{n} \). This approach of estimating the \( \alpha \)-th quantile by order statistics is a generalization of the median (50th quantile). The robustness of the estimator of the \( \alpha \)-th quantile depends on the quantile and number of observations. The 100(1 − \( \alpha \))% confidence interval for the \( \alpha \)-th quantile of an order statistics is approximated by \([x_r, x_s]\) where \( r \) and \( s \) are respectively the next higher natural numbers of
\[
r' = n\alpha - z_{1-c/2} \sqrt{n\alpha(1-\alpha)},
\]
(2.11)
\[
s' = n\alpha + z_{1-c/2} \sqrt{n\alpha(1-\alpha)},
\]
(2.12)
and \( z_c \) denotes the \( c \)-th quantile of the standard normal distribution [11]. This approximation can be used if \( n\alpha(1 - \alpha) > 9 \) or \( \alpha \leq 0.01 \) for \( n > 910 \). These confidence intervals are not symmetric implying that the errors in quantile estimation are not symmetric and not normally distributed. However these errors are asymptotically normally distributed with variance of \( x_j \) as
\[
\sigma^2_{x_j} = \frac{\alpha(1-\alpha)}{n} [f(x_j)]^2,
\]
(2.13)
where \( f \) is the density function of \( X \) which must be strictly greater than zero [30].

2.2.3 Monte-Carlo Simulation (MCS)
In MCS, a data series is generated using the random numbers, expected to be the representation of the future scenario. It allows the risk manager to use actual historical distributions for risk factor returns rather than assuming normal distributions for returns. A large number of randomly generated simulations are run forward in time using volatility and correlation estimates. Each simulation will be different but in total the simulations will aggregate to the chosen statistical parameters, that is, historical distributions, volatility and correlation estimates. MCS is more realistic than the previous two models VCA and Historical simulations and is likely to estimate VaR more accurately [9, 12].

The model risk is evaluated usually by doing Backtesting [W5]. Backtesting compares the d-period VaR with the actually observed profit/loss over the next d-period. If the actually observed profit/loss exceeds the VaR estimates too often, model is not appropriate. For example the 99% VaR should be exceeded by the actually observed profit/loss on average 10 times if the given data series consists of 1000 observations.

3. VaR: Assumptions and Limitations

In VCA, there are several assumptions that must be made in order to make VaR calculations and analyses tractable. The first most important assumption is the distributional assumption, i.e., daily returns follow a normal distribution. A distribution is described as normal if there is a high probability that any sampled observation will have a value that is close to the population mean and a low probability of having a value that is far from mean.

In the simple equity portfolio example, let daily return fluctuations in the S&P 500 index follow a normal distribution with mean zero and standard deviation of say 100 bp (basis points). The assumption of a zero mean is debatable, since at the very least we know that equity prices have a positive expected return the risk free rate plus a market risk premium. To calibrate the numbers for this non-zero mean return case, let us assume a mean risk free rate of 4 percent p.a. and a risk premium of 6 percent p.a. A total expected return of 10% p.a. then translates into a mean return of approximately four basis points per day (i.e., 1000 bp/250 days = 4 bp/day). Hence, an alternative assumption could have been that asset returns are normally distributed with a mean return of four basis points per day rather than zero basis points per day. Similarly, the assumption of a 100 bp daily standard deviation can be questioned. Linking daily volatility to the annual volatility using the square root rule is equivalent to assuming an annualized standard deviation of 15.8 percent p.a. for the S&P 500 index. The square root rule states that under standard assumptions the J-period volatility is equal to the one period volatility inflated by the square root of J. For example, assuming daily volatility 1% per day and 250 trading days in a year, it results in an annual volatility of 1% per day \( \sqrt{250} = 15.8 \) percent p.a. Historically, this is approximately the observed order of magnitude for the volatility of well-diversified equity portfolio.

Second assumption is the stationarity requirement. For example, a 1% fluctuation in returns is equally likely to occur at any point in time. A related assumption is the random walk assumption of inter-temporal unpredictability implying that day-to-day fluctuations in returns
are independent. Thus, a decline of say 10% on one day in the S&P 500 index has no predictive power regarding returns on the S&P 500 index on the next day. Equivalently, the random walk assumption can be represented as the assumption of an expected rate of return equal to zero. It means if the mean daily return is zero, then the best guess estimate of tomorrow’s price level is today’s price level. There is no relevant information available at time T that could help forecast price at time \( T+1 \).

Another assumption is the non-negativity requirement which stipulates that financial assets with limited liability cannot attain negative values. However, derivatives like forwards, futures and swaps can violate this assumption. The time consistency requirement states that all single period assumptions hold over the multi-period time horizon.

The most questionable assumption however is the distributional assumption because evidence shows that most securities prices are not normally distributed. Despite this, the assumption that continuously compounded returns are normally distributed is a standard assumption in finance. The very basic assumption of the Black-Scholes option pricing model is that asset returns follow a log-normal diffusion. This assumption is the key to the elegance and simplicity of the Black-Scholes option pricing formula. The instantaneous volatility of asset returns is always multiplied by the square root of time in the Black-Scholes formula. Under the model’s normality assumption, returns at any time horizon are always independent and identically normally distributed; the scale is just the square root of the volatility [19].

Many researchers have examined how best to compute VaR with assumptions other than the standardized normal. Duffie and Pan [6] refer to a comprehensive study of different distributions and the parameters that need to be estimated. Hull and White [15] suggest estimating VaR when variables are not normally distributed. They consider specifying any probability distributions for variables but requiring that transformations of the distribution still follow a multivariate normal distribution. These and other papers consider variations but have to overcome two practical problems [4, 15]. Estimating inputs for non-normal models can be very difficult especially when using historical data and the probabilities of losses and VaR are simplest to compute with the normal distribution assumption. These calculations become progressively more difficult with asymmetric and fat-tailed distributions.

Some research has been directed at the estimation techniques to yield more reliable variance and covariance estimates in the VaR calculations. Some consider improving sampling methods and data innovations that allow for better estimates of variances and covariances while others suggest improving statistical estimation to yield the better estimates from existing data. For instance, conventional estimates of VaR are based upon the assumption that the standard deviation in returns does not change over time (homoskedasticity). Engle [7] suggests that we get much better estimates by using models that explicitly allow the standard deviation to change over time (heteroskedasticity). He suggests two variants Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) that provide better forecasts of variance and better measures of Value-at-Risk.

Further, VaR is designed for portfolios where there is a linear relationship between risk and portfolio positions. Consequently, it can break down when the portfolio includes options, since
the payoffs on an option are not linear. In an attempt to deal with options and other non-linear instruments in portfolios, researchers have developed Quadratic Value at Risk measures [4]. These quadratic measures also categorized as delta-gamma models (to contrast with the more conventional linear models which are called delta-normal) allow researchers to estimate the Value at Risk for complicated portfolios that include options and option-like securities such as convertible bonds.

In the historical simulation approach, the VaR is determined by the actual price movements and thus, there are no underlying assumptions of normality driving the conclusions. Second is that each day in the time series is assigned an equal weight in measuring the VaR. Therefore a potential problem is if there is a trend in the variability, for example lower in the earlier periods and higher in the later periods. Further, this approach is based on the assumption of history repeating itself with the time period used providing a full and complete snapshot of the risks that the market is exposed to in other periods.

In the Monte Carlo simulation, the first two steps mirror the first two steps in the Variance-covariance method where we identify the markets risks that affect the asset or assets in a portfolio and convert individual assets into positions in standardized instruments [9, 26]. It is in the third step that the differences emerge. Rather than compute the variance and covariance across the market risk factors, we take the simulation route, where we specify probability distributions for each of the market risk factors and specify how these market risk factors move together (correlation parameter). While the estimation of parameters is easier if you assume normal distributions for all variables, the power of Monte Carlo simulations comes from the freedom you have to pick alternate distributions for the variables. In addition, you can bring in subjective judgments to modify these distributions. Once the distributions are specified, the simulation process starts. In each run, the market risk variables take on different outcomes and the value of the portfolio reflects the outcomes. After a repeated series of runs, numbering usually in the thousands, you will have a distribution of portfolio values that can be used to assess Value at Risk. For instance, assume that you run a series of 10,000 simulations and derive corresponding values for the portfolio. These values can be ranked from highest to lowest, and the 95% percentile Value at Risk will correspond to the 500th lowest value and the 99th percentile to the 100th lowest value.

As a caution, however, simulation is only as good as the probability distributions considered as the inputs which are used in simulation. While Monte Carlo simulations are often more sophisticated than historical simulations, many practitioners directly draw on historical data to make their distributional assumptions. Further, as the number of market risk factors increases and their co-movements become more complex, Monte Carlo simulations become more difficult to run for two reasons. First, there are many probability distributions of the market risk variables rather than just the handful in the context of analyzing a single project or asset. Second, the number of simulations that we need to run to obtain reasonable estimate of Value at Risk will have to increase substantially, like tens of thousands from the thousands. However, the strengths of Monte Carlo simulations can be seen in comparison to the other two approaches for computing Value at Risk. Unlike the variance-covariance approach, in MCS
there are no unrealistic distributional assumptions like normality in returns. In contrast to the historical simulation approach, we begin with historical data but are free to bring in both subjective judgments and other information to improve forecasted probability distributions. In addition, Monte Carlo simulations can be used to assess the Value at Risk for any type of portfolio and are flexible enough to cover options and option-like securities [27, 28, 31].

3.1 VaR: Comparison of Approaches

In summing up, each approach to calculate VaR has its merits and demerits. The variance-covariance approach with its delta normal and delta gamma variations requires making strong assumptions about the distributions of the returns of the standardized assets. However, it is simple to compute VaR once these assumptions are made. The historical simulation approach requires no assumptions about the nature of return distributions but implicitly assumes that the data used in the simulation is a representative sample of the risks looking forward. The Monte Carlo simulation approach allows for more flexibility in terms of choosing distributions for returns and bringing in subjective judgments and external data but is the most demanding in computation efforts.

3.2 How Different Are the VaR Estimates?

The estimates of VaR from either of three approaches are functions of the inputs. VCA and the historical simulation will results in the same VaR estimates if the historical returns data is normally distributed and is used to estimate the variance-covariance matrix. Similarly, the VCA and MCS will yield approximately the same values if all of the inputs in the latter are assumed to be normally distributed with consistent values of mean and variance. The historical and MCS approaches will converge if the distributions used in MCS are entirely based upon historical data.

3.3 Which Approach is Most Reliable?

The answer to this question depends both upon what risks are being assessed and how the competing approaches are used. There are variants that have developed within each approach, aimed at improving performance [3]. Many of the comparisons across approaches are skewed by the fact that the researchers have tested variants of an approach that they have developed against alternatives. Referring to some unbiased comparative studies of the alternative approaches, evidence is mixed. Hendricks [12] compared the VaR estimates obtained using the variance-covariance and historical simulation approaches on 1000 randomly selected foreign exchange portfolios. He used nine measurement criteria, including the mean squared error (of the actual loss against the forecasted loss) and the percentage of the outcomes covered and concluded that the different approaches yield risk measures that are roughly comparable and that they all cover the risk that they are intended to cover, at least up to the 95 percent
confidence interval. He did conclude that all of the measures have trouble capturing extreme outcomes and shifts in underlying risk. Lambadrais, Papadopoulou, Skiadopoulos and Zoulis [21] computed the Value at Risk in the Greek stock and bond market with historical with Monte Carlo simulations and found that while historical simulation overstated the VaR for linear stock portfolios, results were less clear with non-linear bond portfolios.

Lastly, which VaR approach should be used can be answered by looking at the task at hand? If aim is to assess VaR for portfolios which do not include options, over very short time periods (a day or a week), the variance-covariance approach does a reasonably good job. If VaR is being computed for a risk source that is stable and where there is substantial historical data like commodity prices, historical simulations provide good estimates. In the most general case of computing VaR for nonlinear portfolios which include options over the longer time periods, where the historical data is volatile and non-stationary and the normality assumption is questionable, Monte Carlo simulations are the best choice.

4. Copulas

Copulas are functions which join uniform marginal distributions of random variables to form their multivariate distribution functions. Copulas separate joint distributions into two contributions: (i) marginal distributions of each variable and (ii) copula as a measure of dependence. Sklar’s theorem [29] states that any multivariate distribution can be expressed as the k-copula function \( C(u_1, \ldots, u_k) \) evaluated at each of the marginal distributions. Copula is not unique unless the marginal distributions are continuous. Using probability integral transform, each continuous marginal \( U_i = F_i(x_i) \) has a uniform distribution on \( I \in [0, 1] \) where \( F_i(x_i) \) is the cumulative integral of \( f_i(x_i) \) for the random variable \( X_i \in (-\infty, \infty) \). The k-dimensional probability distribution function \( F \) has a unique copula representation

\[
F(x_1, x_2, \ldots, x_k) = C(F_1(x_1), F_2(x_2), \ldots, F_k(x_k)) = C(u_1, u_2, \ldots, u_k). \tag{4.1}
\]

The joint probability density function is written as

\[
f(x_1, x_2, \ldots, x_k) = \prod_{i=1}^{k} f_i(x_i) c(F_1(x_1), F_2(x_2), \ldots, F_k(x_k)), \tag{4.2}
\]

where \( f_i(x_i) \) is each marginal density and coupling is provided by copula density

\[
c(u_1, u_2, \ldots, u_k) = \frac{\delta^k c(u_1, u_2, \ldots, u_k)}{\delta u_1 \delta u_2 \cdots \delta u_k}, \tag{4.3}
\]

if it exists. In case of independent random variables, copula density \( c(u_1, u_2, \ldots, u_k) \) is identically equal to one. The importance of the above equation \( f(x_1, x_2, \ldots, x_k) \) is that the independent portion expressed as the product of the marginals can be separated from the function \( c(u_1, u_2, \ldots, u_k) \) describing the dependence structure or shape. The dependence structure summarized by a copula is invariant under increasing and continuous transformations of the marginals.

The simplest copula is independent copula
\[ \Pi := C(u_1, u_2, \ldots, u_k) = u_1, u_2, \ldots, u_k, \]  
(4.4)

with uniform density functions for independent random variables. An empirical copula may be estimated from the N pairs of data \((x_{1,t}, x_{2,t})_{0 < t \leq N}\) by

\[ C(n/N, m/N) = \sum_{1 \leq i < n, 1 \leq j < m} 1, \]  
(4.5)

Where \(r_{1,1}\) and \(r_{1,2}\) are the rank statistics of \([x_{1,t}]_t\) and \([x_{2,t}]_t\) respectively. Frechet [9] and Hoeffding [13] lower and upper bounds for copula respectively are

\[ C(u_1, u_2, \ldots, u_k) \leq \max[1 - n + \sum_i u_i, 0], \]  
(4.6)

\[ C(u_1, u_2, \ldots, u_k) \geq \min_i [u_i]_{i \in \{1, 2, \ldots, k\}}. \]  
(4.7)

Relationships between copula and concordance Kendall’s \(\tau\), Spearman’s \(\rho\), Gini’s index \(\gamma\):

\[ \tau = 4 \int \int_{I^2} C(u_1, u_2) dC(u_1, u_2) - 1, \]  
(4.8)

\[ \rho = 12 \int \int_{I^2} u_1 u_2 dC(u_1, u_2) - 3, \]  
(4.9)

\[ \gamma = 2 \int \int_{I^3} (|u_1 + u_2 - 1| - |u_1 - u_2|) dC(u_1, u_2) \]  
(4.10)

The Pearson’s linear correlation coefficient can not be expressed in terms of copula.

The tail dependence index of a multivariate distribution describes the amount of dependence in the upper right tail or lower left tail of the distribution and can be used to analyze the dependence among extreme random events. Joe [17] defines the upper and lower tail dependence

\[ \lambda_U := \lim_{u \to 1} \left[ 1 - 2u + C(u, u) \right] \]  
(4.11)

\[ \lambda_L := \lim_{u \to 0} C(u, u)/u. \]  
(4.12)

Copula has lower (upper) tail dependence for \(\lambda_U (\lambda_L) \in (0, 1]\) and no lower (upper) tail dependence for \(\lambda_L (\lambda_U) = 0\). This tail dependence measure is the probability that one variable is extreme given that other is extreme. Tail dependence measures are copula-based and copula is related to the full distribution via quantile transformations, i.e., \(C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2))\). Copulas can be simulated using univariate conditional distributions. The conditional distribution of \(U_i\) given first \(i - 1\) components is

\[ c(u_i|u_1, \ldots, u_{i-1}) = \frac{\delta^i C(u_1, \ldots, u_i)}{\delta u_1 \cdots \delta u_{i-1}} / \frac{\delta^{i-1} C(u_1, \ldots, u_{i-1})}{\delta u_1 \cdots \delta u_{i-1}}. \]  
(4.13)

4.1 For \(k \geq 2\), Copula simulation
i. Select a random number $u_1$ from Uniform $(0,1)$ distribution.
ii. Simulate a value $u_k$ from $c(u_k \mid u_1, \ldots, u_{k-1}), n = 2, 3, \ldots$

### 4.1 For $k \geq 2$, Copula simulation

Following are two very frequently used bivariate copulas, i.e., $k=2$, which will be used in our data analysis [24]:

**Clayton copula:**

$$C(u, v) = \left[\max(u^{-\theta} + v^{-\theta} - 1, 0)\right]^{-1/\theta}, \theta \in [-1, \infty) / [0]$$  \hspace{1cm} (4.14)

The relationship between Clayton copula parameter $\theta$ and Kendall’s rank correlation $\tau$ is:

$$\theta = \frac{2\tau}{1 - \tau}$$  \hspace{1cm} (4.15)

**Frank Copula:**

$$C(u, v) = -\frac{1}{\theta} \ln[1 + \frac{(e^{-u\theta} - 1)(e^{-v\theta} - 1)}{(e^{-\theta} - 1)}], \theta \in (-\infty, \infty) / [0]$$  \hspace{1cm} (4.16)

The relationship between Frank copula parameter $\theta$ and Kendall’s rank correlation $\tau$ is: 4

$$\tau = 1 - \frac{4}{\theta}[1 - D_1(\theta)],$$  \hspace{1cm} (4.17)

where $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} \, dt$ is the Debye function for any positive integer $k$.

**Gumbel Copula:**

$$C(u, v) = \exp[-\left(\frac{1}{\theta} ln(u) + \frac{1}{\theta} ln(v)\right)], \theta \in (1, \infty)$$  \hspace{1cm} (4.18)

The relationship between Gumbel copula parameter $\theta$ and Kendall’s rank correlation $\tau$ is:

$$\theta = \frac{1}{1 - \tau}$$  \hspace{1cm} (4.19)

### 5 Exploratory Data Analysis

In this section we intend to describe an analytical study on one of the riskiest products in the global economy, i.e., changes in crude oil prices. Iran is one of the largest crude oil exporting countries and a risk analysis of this product may affect major economic policy decisions. Iran is a member of the Organization of the Petroleum Exporting Countries (OPEC). OPEC is one of the organizations that set crude oil prices in the world and is a reliable data source. The Central Bank of Iran is another reliable data source which provides economical and statistical data and reports about Iran. We referred to OPEC and The Central Bank of Iran and compiled the following data:
i. Monthly average price for a barrel of crude oil from OPEC and Iranian light and Iranian heavy crude oil (in US dollar) from January 1997 to December 2008. [D 1]

ii. Number of barrels of crude oil exported by Iran per season from January 1997 to December 2008. [D 2] Here we divided the seasonal number of crude oil barrels exported by three in order to form a monthly dataset. (Note that this data on export was presented in Iranian calendar seasons and we have converted it to the English calendar’s seasons and months).

In Figure 1 we can see the time series plots of the data (there are 144 observations in all). The left of Figure 1 shows the oil prices of OPEC and Iran light and heavy crude oil prices which are positively correlated in the period of twelve years (January 1997-December 2008). The right of Figure 1 shows the number of barrels of crude oil (in 1000) exported by Iran and we can see a significant fluctuations in the years 2001 and 2002.

![Figure 1: Time series plots of the crude oil prices and export data used in the study.](image)

In Figure 2, the graphical summaries of the above data are presented. The distribution of crude oil prices are skewed to the right. During April 2008 and August 2008 the prices are above 97 US Dollars per barrel. We also noted suspected outliers in Figure 2’s left box plots. The right side graph of Figure 2 shows the distribution of Iran’s crude oil export that is almost symmetric with few suspected outliers on both ends. May 2001 and April 2002 with 1192 and 608 thousand barrels per day were the maximum and minimum Iran’s crude oil export respectively.
The numerical description of the data is provided in Table 1. Two interesting observations are:

i. The average light crude oil prices are higher than the heavy crude oil for the suppliers.
ii. We notice a high standard error indicating large variations in the Iran’s crude oil exports.

Table 1: The descriptive statistics of the data used in this study.

<table>
<thead>
<tr>
<th>N=144</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Error</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
<td>Statistic</td>
<td>Statistic</td>
<td>Statistic</td>
<td>Statistic</td>
<td>Statistic</td>
<td>Statistic</td>
</tr>
<tr>
<td>Iran-Light ($)</td>
<td>118.39</td>
<td>9.80</td>
<td>128.19</td>
<td>38.138</td>
<td>2.167</td>
<td>26.008</td>
</tr>
<tr>
<td>Iran-Heavy ($)</td>
<td>117.24</td>
<td>9.51</td>
<td>126.75</td>
<td>36.831</td>
<td>2.097</td>
<td>25.163</td>
</tr>
<tr>
<td>OPEC-Ref ($)</td>
<td>121.53</td>
<td>9.69</td>
<td>131.22</td>
<td>38.539</td>
<td>2.146</td>
<td>25.756</td>
</tr>
<tr>
<td>Export (000)</td>
<td>584</td>
<td>608</td>
<td>1192</td>
<td>820.25</td>
<td>6.999</td>
<td>83.985</td>
</tr>
</tbody>
</table>

Table 2 presents the estimates of copula parameters, Pearson correlation and rank correlation (Kendall’s tau) between the pairs of data. There is a very high correlation between Iran’s crude oil price and the OPEC crude oil prices and a low correlation between Iran’s crude oil prices and the amount of Iran’s crude oil export. Copula parameters are calculated with respect to the Kendall’s rank correlation among the pairs of the mentioned data in the table 2. These Copula parameters are calculated using equations (4.15), (4.17) and (4.19). We have used the MATLAB functions to estimate the Kendall’s tau and the copula parameters. MATLAB codes are provided in the Appendix.
In this study, we also simulated random data for the Iran’s light and heavy crude oil prices, using Clayton, Frank, and Gumbel copulas. The Copula simulation of the random data is multivariate and requires a pair of two variables. We have considered Iran’s light and heavy crude oil prices as the first variable and OPEC prices and Iran’s crude oil export as the second variable. The following are the steps to obtain the copula simulation of random data:

i. Calculate the Kendall’s $\tau$ between Iran’s crude oil prices and the OPEC prices. And calculate the Kendall’s $\tau$ between Iran’s crude oil prices and Iran’s crude oil export amounts.

ii. Using the Kendall’s $\tau$ we can calculate the copula parameters using (4.16), (4.17) and (4.19) formulas.

iii. Monte Carlo algorithm uses the copula parameters to generate the random data. Here we generated the data with the same size as the original data (144 observations) and also for the size of 1000 observations.

Figure 3 shows an example of simulations using the Clayton copula for Iran’s light crude oil price with respect to the Iran’s crude oil export. The simulated data sizes are 144 and 1000 observations.

From the graphs in Figure 3, data on Iran light crude oil prices are slightly positively skewed. The estimated distribution of prices using Clayton copula also indicates the similar pattern. As expected, distribution is very similar to the distribution of the actual price data however it tends to be closer to the normal distribution when there are large numbers of simulated data (1000 observations).
In this section we will discuss the results of the risk analysis for the data and copula simulated data of the Iran’s heavy and light crude oil prices. Here the Value-at-Risk (VaR) is calculated in terms of the relative change of the monthly crude oil prices. We are defining risk as the unexpected rare event of observing either a very low relative change (drop in oil price) or a very high relative change (high oil price). The formula (2.11) in historical simulation section provides the calculation steps of the value-at-Risk. Figure 4 shows the upper right tail and lower left tail distributions of the VaR for the monthly relative price change of Iran’s light and heavy crude oil. We see that in all the cases the upper right tail have higher relative change values than the lower left tail. Also we can observe, extreme monthly relative price changes have lower probability of occurrence at the end of right and left tails. The copula simulated data are providing higher and lower relative change values on the upper and lower tails respectively in comparison to the original data. The left graph of the second row in Figure 4 shows that the size 1000 simulation of Frank and Gumbel copulas of the Iran’s heavy crude oil with respect to the OPEC prices are unusually high in the upper tail. The reason could be because of the values of copula parameters shown in Table 2 used in simulation or due to the insufficient numbers of iterations.

![Figure 4: Value-at-Risk for the monthly relative change of the crude oil prices and the copula simulations.](image)

Referring to the upper right tails of the graphs in Figure 4, the Frank and Gumbel copula simulations with size 144 obtained the closest values at risks to the original data values. These are presented in Table 3. The tables on the first row of Table 3 present the original data’s
monthly relative change value at risk. The preceding tables on the second and third rows are the copula simulations that obtained the closest value at risk to the original data.

Table 3: Right column is the original data of Iran’s light crude oil and the copula simulated data. Left column is the original data of Iran’s heavy crude oil and the copula simulated data.

<table>
<thead>
<tr>
<th>Light Crude Oil Original Data</th>
<th>Light Crude Oil Simulated Using Gumbel Copula with respect to OPEC (144)</th>
<th>Light Crude Oil Simulated Using Frank Copula with respect to Export (144)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X&lt;z)</td>
<td>0.001 0.005 0.01 0.025 0.05 0.1</td>
<td>0.001 0.005 0.01 0.025 0.05 0.1</td>
</tr>
<tr>
<td>VaR(month)</td>
<td>-0.274 -0.271 -0.256 -0.201 -0.162 -0.120</td>
<td>-0.183 -0.170 -0.156 -0.101 -0.062 -0.020</td>
</tr>
<tr>
<td>P(X&gt;z)</td>
<td>0.001 0.005 0.01 0.025 0.05 0.1</td>
<td>0.001 0.005 0.01 0.025 0.05 0.1</td>
</tr>
<tr>
<td>VaR(Month)</td>
<td>0.386 0.384 0.369 0.315 0.275 0.233</td>
<td>0.403 0.401 0.385 0.329 0.279 0.248</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Heavy Crude Oil Original Data</th>
<th>Heavy Crude Oil Simulated Using Gumbel Copula with respect to OPEC (144)</th>
<th>Heavy Crude Oil Simulated Using Frank Copula with respect to Export (144)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X&lt;z)</td>
<td>0.001 0.005 0.01 0.025 0.05 0.1</td>
<td>0.001 0.005 0.01 0.025 0.05 0.1</td>
</tr>
<tr>
<td>VaR(month)</td>
<td>-0.287 -0.284 -0.289 -0.212 -0.163 -0.132</td>
<td>-0.179 -0.175 -0.170 -0.101 -0.061 -0.021</td>
</tr>
<tr>
<td>P(X&gt;z)</td>
<td>0.001 0.005 0.01 0.025 0.05 0.1</td>
<td>0.001 0.005 0.01 0.025 0.05 0.1</td>
</tr>
<tr>
<td>VaR(Month)</td>
<td>0.401 0.401 0.385 0.329 0.279 0.248</td>
<td>0.403 0.401 0.385 0.329 0.279 0.248</td>
</tr>
</tbody>
</table>

In the above Table 3 for the Iran’s light oil, 0.1% of the times the relative monthly price change is at most -27.4% and at least 38.6%. For the Iran’s heavy oil, 0.1% of the times the relative monthly price change is at most -28.7% and at least 40.3% per month. The copula simulations are providing the results which are far from the actual data’s results. This may be due to the fewer number of simulations and the assumptions on the marginal probability distribution functions.

The above estimates of the Value-at-Risk are calculated using the empirical probability distribution of the original data and the estimated probability distributions in the copula simulated data. We also calculated VaR based on the normal probability distribution assumption. Normal distribution resulted very close to the empirical distribution for the original data. But the copula simulated data resulted quite different for the empirical and normal distribution. This may be due to the fact that the data set can be closely approximated by a normal distribution. Figure 5 shows the empirical and normal quantiles of the monthly crude oil price relative changes for the original data and the copula simulated data. Referring to the left of the first row in figure 5, the empirical quantiles for monthly crude oil price relative change are having similar pattern for the original data and the copula simulated values with two sizes of 144 and 1000. The left graph of the second row in Figure 5 shows a significant difference between the copula simulation of size 1000 with respect to the OPEC prices and that with
respect to the Iran’s crude oil export. Also we can see in third row of Figure 5 that for the Iran’s light crude oil price relative change there is not much difference between the copula simulations of data with respect to the OPEC prices and that with respect to the Iran’s crude oil export. As mentioned in the above paragraph the reason for these observations could be because of the values of copula parameters shown in Table 2 or due to insufficient numbers of simulations.

![Figure 5: Quartiles of the relative changes based on the empirical and normal distributions. The simulated copulas are with respect to the OPEC prices and the Iran’s export (N= 144 and 1000).](image)

Accurate risk measurement in the financial market is of paramount importance to the financial managers and policy makers. Value-at-Risk (VaR) is the most commonly practised risk measure which is evaluated by analysing the negative tail of the probability distributions of the returns/profit and losses. However calculation of VaR is severely influenced by various statistical assumptions on the probability distributions of these returns/profit/losses. We think that an appropriate method of risk analysis needs to be based on the current data available, data on the historical patterns of the distribution, if any, and should avoid any assumptions to fit data to the standard statistical models. In addition, we believe that replication methods should be used for multiplicity of the given data set and the inference should be based on the distribution generated by such simulated data sets. In this paper, we have demonstrated how such an
approach can be used in risk analysis with the help of copula simulation technique and the Iran’s crude oil price, export and OPEC price data. The following summarizes the major findings of our study:

i. Clayton, Frank and Gumbel copulas resulted in similar simulation results. The Copula simulations obtained different results in comparison with the original data’s results which might be due to insufficient simulations and the inadequate estimated marginal probability distributions.

ii. Copula generation with respect to various variables may result in different simulated data. As we observed in this study the simulation of Iran’s light oil with respect to OPEC prices and that with respect to Iran’s crude oil export, generated different results.

iii. The Value-at-Risk distribution in Figure 4, shows extreme monthly relative price changes have lower probability of occurrence at the end of right and left tails.

Acknowledgement

Authors wish to thank referee for valuable comments. This study was supported by the NSERC Discovery Grant of Dr. Pranesh Kumar.

Appendix

%MATLAB CODE FOR THE COPULA SIMULATION OF THE DATA %Input: data1 and data2 %output: copula simulated values of data1 and data2

function [b1,b2,b3]=copulasim(data1,data2)
    tau = corr(iranl,opec,'type','Kendall');
    %CLAYTON COPULA SIMULATION WITH SIZE 144 rho = copulaparam('clayton',tau);
    u = copularnd('clayton',rho,144); b1 = betainv(u,2,2);
    %FRANK COPULA SIMULATION WITH SIZE 144 rho = copulaparam('frank',tau);
    u = copularnd('frank',rho,144); b2 = betainv(u,2,2);
    %GUMBEL COPULA SIMULATION WITH SIZE 144 rho = copulaparam('gumbel',tau);
    u = copularnd('gumbel',rho,144); b3 = betainv(u,2,2);
end

References


**Web-Resources**


Data Sources


Received September 14, 2013; accepted February 1, 2014.

Faramarz Kashanchi
Department of Mathematics and Statistics
University of Northern British Columbia
Prince George, BC V2N 4Z9, Canada.
kashanc@unbc.ca

Pranesh Kumar
Department of Mathematics and Statistics
University of Northern British Columbia
Prince George, BC V2N 4Z9, Canada.
pranesh.Kumar@unbc.ca
Copulas Applications in Estimating Value-at-Risk (VaR): Iranian Crude Oil Prices