Extreme Value Theory and Copula Theory:

A Risk Management Application with Energy Futures

By

Jia Liu
B.A., Xiamen University, 2001
M.A., University of Victoria, 2004

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Economics

Jia Liu, 2011
University of Victoria

SUPERVISORY COMMITTEE

Extreme Value Theory and Copula Theory:

A Risk Management Application with Energy Futures

By

Jia Liu
B.A., Xiamen University, 2001
M.A., University of Victoria, 2004

Supervisory Committee

Dr. David E.A. Giles, Department of Economics
Supervisor

Dr. Judith A. Clarke, Department of Economics
Departmental Member

Dr. Merwan H. Engineer, Department of Economics
Departmental Member

Dr. William J. Reed, Department of Mathematics and Statistics
Outside Member

ABSTRACT

Supervisory Committee

Dr. David E.A. Giles, Department of Economics
Supervisor

Dr. Judith A. Clarke, Department of Economics
Departmental Member

Dr. Merwan H. Engineer, Department of Economics
Departmental Member

Dr. William J. Reed, Department of Mathematics and Statistics
Outside Member

Deregulation of the energy market and surging trading activities have made the energy markets even more volatile in recent years. Under such circumstances, it becomes increasingly important to assess the probability of rare and extreme price movement in the risk management of energy futures. Similar to other financial time series, energy futures exhibit time varying volatility and fat tails. An appropriate risk measurement of energy futures should be able to capture these two features of the returns. In the first portion of this dissertation, we use the conditional Extreme Value Theory model to estimate Value-at-Risk (VaR) and Expected Shortfall (ES) for long and short trading positions in the energy markets. The statistical tests on the backtests show that this approach provides a significant improvement over the widely used Normal distribution based VaR and ES models.

In the second portion of this dissertation, we extend our analysis from a single security to a portfolio of energy futures. In recent years, commodity futures have gained tremendous

popularity as many investors believe they provide much needed diversification to their portfolios. In order to properly account for any diversification benefits, we employ a time-varying conditional bivariate copula approach to model the dependence structure between energy futures. In contrast to previous studies on the same subject, we introduce fundamental supply and demand factors into the copula models to study the dependence structure between energy futures. We find that energy futures are more likely to move together during down markets than up markets.

In the third part of this dissertation, we extend our study of bivariate copula models to multivariate copula theory. We employ a pair-copula approach to estimate VaR and ES of a portfolio consisting of energy futures, the S&P 500 index and the US Dollar index. Our empirical results show that although the pair copula approach does not offer any added advantage in VaR and ES estimation over a long backtest horizon, it provides much more accurate estimates of risk during the period of high co-dependence among assets after the recent financial crisis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

This dissertation could not have been completed without the help and support of many people through my Ph.D. journey. I would like to gratefully acknowledge all of them here.

First and foremost I would like to express my deepest appreciation to my advisor, David E. Giles for his expert guidance, patience and advice throughout my graduate study years. I am also grateful to the dissertation committee, Judith A. Clarke, Merwan H. Engineer and William J. Reed for their insightful comments and suggestions.

Being a Ph.D. candidate while pursuing a full-time career in the finance would not have been possible without the help of my colleagues and friends. I would especially like to thank my previous boss Stephen J. Calderwood for his support and encouragement during my time as a Research Associate at Raymond James Ltd. My thanks also go to my colleagues at TD Energy Trading Inc., especially Richard Merer.

But most importantly, I want to thank my boyfriend, Shary Mudassir for his unconditional love and being there for me at every step of the way.

DEDICATION

To my parents

CHAPTER ONE: GENERAL INTRODUCTION

1. Introduction

The Basel I framework developed by the Basel Committee on Banking Supervision[1] (1996) requires that financial institutions, such as banks and investment firms, set aside a minimum amount of regulatory capital to cover potential losses from their exposure to credit risk, operational risk and market risk. The preferred approach for measuring market risk is Value-at-Risk (VaR), which measures the worst expected losses in the market value over a specific time interval at a given significance level. Financial institutions are allowed to use "internal" models to capture their VaR. Given the opportunity cost of the regulatory capital reserves the banks have to put aside for market risk, it is desirable for the banks to develop an accurate internal VaR model. However, the capital requirement is designed in such a way that banks are not tempted to pursue the lowest possible VaR estimates. This is due to the fact that the capital requirement takes into account not only the magnitude of the calculated VaR but also penalizes the accessed number of violations of the VaR (*i.e.* actual losses exceeding the VaR). Therefore, it is in the best interest of banks to come up with an accurate VaR models to minimize the amount of regulatory capital reserve they have to set aside..

Aside from the regulatory consideration, other reasons that VaR has gained huge popularity are that it is conceptually simple and it summarizes the risk by using just one

[1] The Basel Committee on Banking Supervision is an institution created by the central bank Governors of the Group of Ten nations. The Basel Committee formulates broad supervisory standards and guidelines and recommends statements of best practice in banking supervision.

number which can be easily communicated to management. A key element in VaR calculation is the distribution assumed for the financial returns under study. The common practice in estimating VaR is to assume that asset returns are normally distributed. However, this fails to capture the observed skewness and kurtosis in most financial time series. The Normal distribution based VaR models tend to underestimate risk and require higher regulatory capital due to excess VaR backtest violations. As the Extreme Value Theory focuses on modeling of the tail behaviour of a distribution using only extreme values rather than the whole dataset, it can potentially provide a more accurate estimate of tail risk.

In recent years, an increasing number of research studies have analyzed the extreme events in financial markets as a result of currency crises, stock market crashes and credit crises (Longin (1996), Müller *et al.* (1998), Pictet *et al.* (1998), Bali (2003), Gençay and Selcuk (2004), *etc*.). It is important to note that the Extreme Value Theory (EVT) assumes that the data under study are independently and identically distributed, which is clearly not the case for most financial returns. In order to address the issue of stochastic volatility, this study adopts McNeil and Frey's (2000) approach to model financial returns and measure tail risk. McNeil and Frey's solution to observed volatility clustering in financial returns is to first fit a GARCH-type model to the returns data by quasi-maximum likelihood. The second stage of the approach is to apply the EVT to the GARCH residuals. The advantage of this GARCH–EVT combination lies in its ability to capture conditional heteroskedasticity in the data through the GARCH framework, while at the same time modeling the extreme tail behaviour through the EVT method. Applying

McNeil and Frey's approach in various financial markets, Bali and Neftci (2003), Byström (2005), Fernandez (2005), Chan and Gray (2006), and Bhattacharyya *et al.* (2007) demonstrate that a risk measurement based on the statistics of the extremes can measure the risk exposure more accurately than the Normal distribution based approaches.

In chapter two, we employ McNeil and Frey's two step approach to estimate VaR and Expected Shortfall using energy futures and compare this approach with conventional models. The backtest results are evaluated using statistical tests. Our results indicate that the GARCH-EVT approach outperforms the competing models in forecasting VaR and ES by a wide margin. This approach provides a significant improvement over the widely used Normal distribution based VaR and ES models, which tends to underestimate the true risk and fail to provide statistically accurate VaR estimates. Like Marimoutou, Raggad, and Trabelsi (2009), we find that the conditional Extreme Value Theory and Filtered Historical Simulation approaches outperform the traditional methods. Further, our results show that the GARCH-EVT approach is overwhelmingly better than the competing models, especially at lower significance levels.

However, for a portfolio consisting of multiple assets, knowing the best Value-at-Risk model for each component is not sufficient to capture the portfolio risk since VaR as a risk measure is not sub-additive (Artzner *et al.* (1997)). This means that the risk of a portfolio can be larger than the sum of the stand-alone risks of its components when measured by VaR. Therefore, we need to evaluate the portfolio in a multivariate setting to

account for the diversification benefits. While univariate VaR estimation has been widely studied, the multivariate case has only been investigated recently due to the complexity of joint multivariate modeling. Traditional methods for portfolio VaR estimation, such as the RiskMetrics method, often assume a multivariate Normal distribution for the portfolio returns. However, it is a stylized fact that the returns are asymmetric and exhibit tail dependence, which often leads to an underestimated VaR. To capture the tail dependence and properly estimate portfolio VaR, copula models are introduced in chapter three.

In chapter three, copula models are used to estimate portfolio measure of risk. Copulas, introduced by Sklar in 1959 are statistical functions which join together one-dimensional distributions to form multivariate distributions. Copulas have become a popular multivariate modeling tool in many fields such as actuarial science, biomedical studies, engineering and especially finance. During the past decade, we have witnessed an increasing number of financial applications of copula theory, mainly due to its flexibility in constructing a suitable joint distribution when facing non-normality in financial data. The key characteristic of copula models is the separation of the joint distribution of returns into two components, the marginal distributions and the dependence structure. The approach is designed to capture well-known stylized facts of financial returns using marginal distributions, leaving all of the information about the dependence structures to be estimated by copula models separately. Therefore, copula models allow for the construction of more flexible joint distributions than existing multivariate distributions.

The nature of the dependence structure between financial assets has very important implications in investment decision making. It provides insights into portfolio risk management, portfolio diversification, pairs trading and exotic derivatives pricing, especially when returns are non-Normal and simple linear correlation fails to capture the degree of association between assets. The early literature on the linkages between different asset returns mainly focused on using linear correlation as the measure of dependence for elliptical variables. However, there is strong evidence that the univariate distributions of many financial variables are non-Normal and significantly fat-tailed, which rules out the use of the multivariate Normal distribution. Since the pioneering work of Embrechts *et al.* (1999), copula models have attracted increasing attention due to the models' ability to capture different patterns of dependence while allowing for flexible marginal distributions to capture the skewness and kurtosis in asset returns.

A number of recent empirical studies have discovered significant asymmetric dependence in that returns are more dependent during market downturns than during market upturns. See Longin and Solnik (2001), Ang and Chen (2002), Patton (2006a), Michelis and Ning (2010), *etc*. In addition, most of these studies find that the dependence structure is not constant over time. Following Patton (2006a), we employ a time-varying copula-GARCH model to capture these two important characteristics of the dependence structure.

Most previous empirical studies mainly look at international stock markets and foreign exchange rate markets. Little attention has been paid to nonlinear dependence, especially tail dependence caused by extreme events, between energy futures markets. With the

increasing integration of financial markets, financial innovations and ease of information flow among investors, energy markets are becoming more intertwined in recent years. As Alexander (2005) points out, the dependence between crude oil and natural gas futures prices is strong and cannot be modeled correctly by a bivariate Normal distribution. Grégoire *et al.* (2008) fit various families of static bivariate copulas to crude oil and natural gas futures and conclude that the Student-*t* copula provides a much better fit based on the goodness-of-fit tests. Using energy futures, Fischer and Köck (2007) compare several construction schemes of multivariate copula modes and also confirm that the Student-*t* copula outperforms others. However, these papers mainly focus on the application of fitting static copula models to energy futures without detailed analysis on the dependence structure and its implication. In the empirical part of this chapter, we employ a time-varying conditional copula method to study the dependence structure of energy futures. In addition, we also consider the impact of supply-demand fundamentals as reflected on in inventory quantity on the dynamic dependence between energy futures. Natural gas consumption is seasonal but production is not. Natural gas inventories are built during the summer and used in the winter. The imbalance between supply and demand leads to the seasonality (higher winter prices and lower summer prices) in natural gas prices. Variation in weather from seasonal norm also affects prices, with above normal heating and cooling demand adding upward pressure to natural gas prices. Therefore, it is important to take into account the underlying supply-demand factors for the energy prices when analysing the relationship between energy prices. To the best of our knowledge, this chapter is the first study of energy futures returns dynamics with macroeconomic factors as exogenous variables.

In chapter four, we extend the bivariate copula models used in chapter three to multivariate copulas and test the accuracy of out-of-sample portfolio Value-at-Risk forecasts. Although many studies have demonstrated that copula functions can improve Value-at-Risk estimation, most of these approaches are based on bivariate copulas or very restricted multivariate copula functions mainly due to a lack of construction schemes for higher dimension copulas. From 2005 on, a handful of extensions and innovations for higher dimension copulas appeared, *e.g.* the Hierarchical Archimedean copula (HAC), see Savu and Trede (2006); the Generalized multiplicative (GMAC), see Morillas (2005); Liebscher copulas, see Liebscher (2008); Fischer and Köck copulas, see Fischer and Köck (2009); Koehler-Symanowski copulas, see Palmitesta and Provasi (2006); and Pair-copulas decomposition (also called vine copulas), see Aas *et al*. (2006). Comparative discussions of these different approaches are limited. The comprehensive reviews presented by Berg and Aas (2008) and Fischer *et al*. (2009) show that pair-copulas provide a better fit to multivariate financial data than do other multivariate copula constructions.

In this chapter, we employ a vine based pair-copula approach to estimate VaR and ES for a portfolio of equally weighted crude oil futures, natural gas futures, S&P 500 index and the US Dollar index. The major advantage of vine based copula models is their flexibility in modeling multivariate dependence. They allow for flexible specification of the dependence between different pairs of marginal distributions individually, while specifying an overall dependence between all of the marginal distributions. Both the

canonical-vine and the D-vine structures are tested in this chapter to evaluate their ability at forecasting VaR and ES at different confidence levels. Results are compared with traditional methods, such as RiskMetrics, Historical Simulation, and the conditional Extreme Value Theorem, *etc*.

Chapter Four is among a very small number of empirical studies of higher dimensional multivariate copula theory. To the best of our knowledge, this study is the first to explore the benefit of using vine copula theory in the estimation of VaR and ES for a diversified portfolio of energy futures and other assets. Our results show that the pair-copula decomposition does not provide any added advantage over the competing models in terms of forecasting VaR and ES over a long backtest horizon. However, we found that pair-copula can greatly improved the accuracy of VaR and ES forecasts in periods of high co-dependence among assets.

CHAPTER TWO: A DYNAMIC EXTREME VALUE THEORY APPROACH TO MEASURE TAIL RISK IN THE ENERGY FUTURES

1. Introduction

High volatility and the close connection between asset prices and the supply-demand fundamentals as reflected in inventory quantity have made energy futures a very popular trading instrument for large investors[2] and small speculators alike. Energy trading has always been recognized as a very risky business especially after the collapse of Enron in 2001. However, increasing volatility in the market and the record-high commodity prices prompted renewed interest from investors. Unfortunately, huge price swings and possibly improper risk management have let to enormous trading losses from energy derivatives. The widely used RiskMetrics[3] methodology assuming normality of returns tends to underestimate the probability of extreme losses. In September 2006, after huge concentrated positions in the natural gas market went wrong, the Connecticut based hedge fund Amaranth Advisors suffered a US$6.5 billion loss, the second highest trading loss ever recorded. A year later, SemGroup LP declared bankruptcy in July after a US$3.2 billion loss in oil trading sunk the formerly 12th-largest private U.S. Company. Many other energy trading desks and hedge funds suffered drastic trading losses or even went bankrupt after energy prices plunged in late 2008. Such events illustrate the

[2] Including exploration and production companies, energy consumers, financial institutions, commodity trading advisors, hedge funds, and institutional investors.

[3] The RiskMetrics model is a popular and widespread portfolio risk management approach introduced by J.P. Morgan in 1996.

increasing importance of assessing the probability of rare and extreme price movements in the risk management of energy futures. Similar to other financial time series, energy futures exhibit time varying volatility and fat tails. An appropriate risk measurement of energy futures should be able to capture these two features of the returns.

In recent years, Value-at-Risk (VaR) and Expected Shortfall (ES) have become the most common risk measures used in the finance industry. VaR measures the worst expected losses in market value over a specific time interval at a given significance level under normal market conditions. For example, if a portfolio has a daily VaR of $1 million at 5%, this means that there is only five chances in 100 that a daily loss greater than $1 million would occur. The reasons that VaR has gained huge popularity are that it is conceptually simple and it summarizes the risk by using just one number which can be easily communicated to management. The biggest drawback of the commonly used VaR is that it assumes that the asset returns are normally distributed, which fails to capture the observed skewness and kurtosis in returns. This implies that VaR as a measure of risk under the normality assumption underestimates the true risk. Therefore, VaR can be drastically improved if we can better understand the tail-behaviour of the underlying distribution. The Extreme Value Theory (EVT) approach uses information from the tails only to estimate the true underlying distribution of the returns. As we are only interested in the risk associated with the tails, using EVT to estimate tail risk measures such as VaR and ES can prove to be effective.

Extreme Value Theory has been applied successfully in many fields where extreme events may appear, such as climatology and hydrology. In recent years, an increasing number of research studies have analyzed the extreme events in financial markets as a result of currency crises, stock market crashes and credit crises (Longin (1996), Müller *et al.* (1998), Pictet *et al.* (1998), Bali (2003), Gençay and Selçuk (2004), *etc*.). It is important to note that Extreme Value Theory assumes that the data under study is independently and identically distributed, which is clearly not the case for most financial returns. In order to address the issue of stochastic volatility, this study adopts McNeil and Frey's (2000) approach to model financial returns and measure tail risk. McNeil and Frey's solution to observed volatility clustering in financial returns is to first fit a GARCH-type model to the returns by quasi-maximum likelihood. The second stage of the approach is to apply EVT to the GARCH residuals. The advantage of this GARCH–EVT combination lies in its ability to capture conditional heteroskedasticity in the data through the GARCH framework, while at the same time modeling the extreme tail behaviour through the EVT method. Applying McNeil and Frey's approach in various financial markets, Bali and Neftci (2003), Byström (2005), Fernandez (2005), Chan and Gray (2006), and Bhattacharyya *et al.* (2007) demonstrate that a risk measure based on the statistics of the extremes can capture the risk exposure more accurately than the Normal distribution based approaches.

Despite the high volatility in the energy market and the importance of energy production in national economies, there exist only a few studies on this topic. Krehbiel and Adkins (2005) applied McNeil and Frey's two step approach to model risk exposure in the

energy futures. The authors show that the conditional-EVT methodology offers a significant advantage for price risk measurement in the energy market. However, they failed to compare the conditional-EVT model to fat-tailed models such as Student-*t* based risk measures. They also did not consider alternative risk measures such as Expected Shortfall, nor did they conduct statistical tests on their backtests. Marimoutou, Raggad, and Trabelsi (2009) use the same approach to model VaR in oil futures. Using statistical tests on the backtests, the authors illustrate that the conditional-EVT framework offers a major improvement over the conventional risk methods. In this chapter, we analyze the commonly used risk measure, Value-at-Risk, as well as an alternative coherent risk measure, Expected Shortfall. Using energy futures with sufficient historical data, we employ McNeil and Frey's two-step approach to estimate risk and compare this approach with conventional models. The backtest results are evaluated using Kupiec's unconditional coverage test as well as Christoffersen's conditional coverage test.

This chapter is organized as follows: Section 2 introduces Extreme Value Theory and the risk measures used in this chapter. Section 3 applies the conditional-EVT method to the energy futures and compares the relative performance of different models in term of out-of-sample risk forecasting. Section 4 concludes this chapter.

2. Extreme Value Theory (EVT) and Risk Management

Extreme Value Theory provides a theoretical framework of analyzing rare events and it has been named the cousin of the well-known Central Limit Theorem as both theories tell us what the limiting distributions are as the sample size increases. Broadly speaking, there exist two methods of modeling extreme events, namely the block maxima models

and the peaks over threshold (POT) approach. The main difference between these two methods is how the extremes are identified and the principal distribution is used. Under the block maxima approach, the extremes are defined as the maximum data point in the successive periods. Fisher and Tippett (1928) recognized that the limiting distribution of these extremes is the generalized extreme value (GEV) distribution. The POT approach considers the observations that exceed a given threshold. Selecting only the block maxima is a waste of data if other extremes are available. The threshold methods use data more efficiently and have become a more popular method of choice in recent years.

2.1. The Block Maxima Approach

The block maxima approach considers the maximum the variable takes in successive periods. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent, identically distributed (*i.i.d.*) random variables with a common distribution function $F(x)=P(X_i \leq x)$, which does not have to be known. The block maxima approach requires grouping the series into non-overlapping successive blocks and indentifying the maximum from each block: $M_n=\max(X_1,\ldots,X_n)$. The limit law of the block maxima is given by the following theorem:

2.1.1. Fisher-Tippett Theorem

Let (X_n) be a sequence of *i.i.d.* random variables. If there exist sequences of constants $c_n>0$, $d_n \in R$ and some non-degenerate distribution function H such that

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H \qquad as \qquad n \to \infty$$

Then H belongs to one of the three standard extreme value distributions:

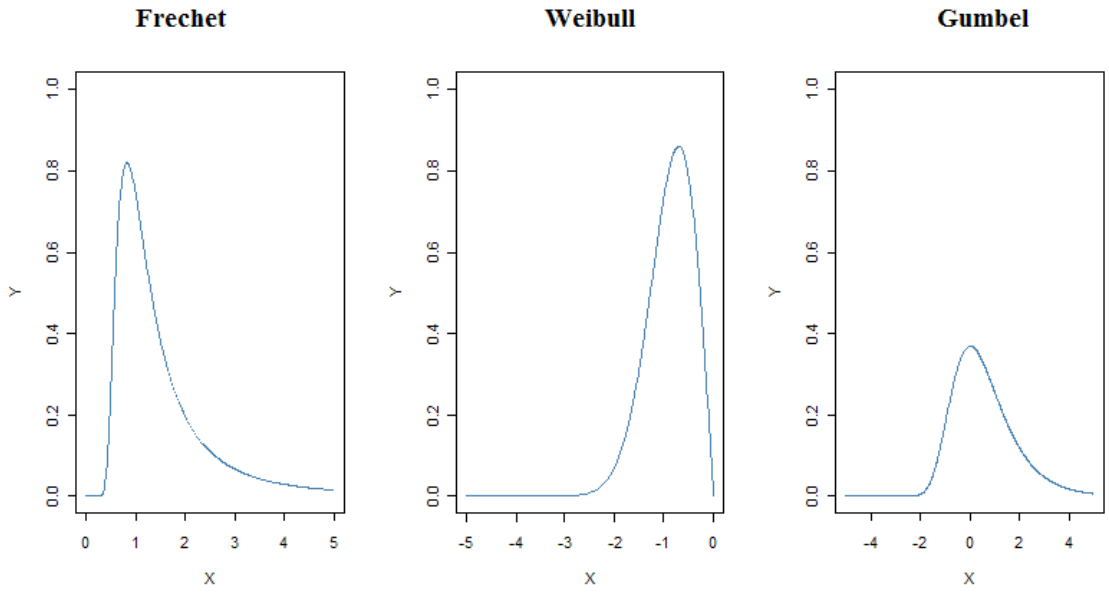
$$\textit{Fréchet}: \quad \Phi_\alpha(x) = \begin{matrix} 0, & x \leq 0 \\ e^{-(x)^{-\alpha}}, & x > 0 \end{matrix} \quad \alpha > 0,$$

$$\textit{Weibull}: \quad \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & x \le 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0,$$

$$\textit{Gumbel}: \quad \Lambda(x) = e^{-e^{-x}}, x \in R$$

Collectively, these three families of distribution are termed the extreme value distributions. Each family has a location and scale parameter, *d* and *c* respectively. Additionally, the Fréchet and Weibull families have a shape parameter α. The parameter α is the tail index, and indicates the thickness of the tail of the distribution; the thicker the tail, the smaller the tail index. The beauty of this theorem is that these three distributions are the only possible limits of the distribution of the extremes M_n, regardless of the distribution *F* for the population. In this sense, this theorem provides an extreme value version of the central limit theorem. The shape of the probability density functions for the standard Fréchet, Weibull and Gumbel distributions are shown in Figure 1. The density of *H* decays polynomially for the Fréchet distribution and therefore the Fréchet distribution suits well heavy tailed distributions such as Student's-*t* distribution.

Figure 1 Standard Extreme Value Distributions



2.1.2. The Generalized Extreme Value Distribution

Jenkinson (1955) and von Mises (1954) suggested that the three families of extreme value distributions can be generalized by a one-parameter representation:

$$H_{\xi}(x) = \begin{cases} e^{-(1+\xi x)^{-1/\xi}}, & if \quad \xi \neq 0, \\ e^{-e^{-x}}, & if \quad \xi = 0, \end{cases} \tag{1}$$

where $1 + \xi x > 0$. This representation is known as the "Generalized Extreme Value" (GEV) distribution, where the parameter $\xi = \alpha^{-1}$. This shape parameter ξ determines the type of extreme value distribution:

Fréchet distribution: $\xi = \alpha^{-1} > 0,$

Weibull distribution: $\xi = \alpha^{-1} < 0,$

Gumbel distribution: $\xi = 0$.

The biggest criticism of the block maxima approach is that it does not utilize all of the information from the extremes as it considers only the maximum points of the fixed intervals. Therefore, recent studies on the subject of extreme value analysis have concentrated on the behaviour of extreme values above a high threshold. This method is the peaks-over-threshold (POT) approach.

2.2. The Peaks over Threshold (POT) Approach

The POT approach considers the distribution of the exceedances over a certain threshold. Let (X_n) be a sequence of *i.i.d.* random variables with marginal distribution function F, which does not have to be known. We are interested in estimating the distribution

function F_u of values of X that exceed a certain threshold u. The distribution function F_u is the conditional excess distribution function and is defined as:

$$F_u(y) = P(X - u \le y | X > u), \quad 0 \le y \le x_F - u \tag{2}$$

where u is a given threshold, $y = X - u$ is termed the excess and $x_F \le \infty$ is the right endpoint of the distribution function F. The conditional excess distribution function F_u represents the probability that the value of X exceeds the threshold by at most an amount y given that X exceeds the threshold u. This conditional probability can be written as:

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

As the majority of the values of the random variable X lies between 0 and u, the estimation of F_u is not very difficult. However, due to the limited information available, the estimation of F_u is not that straightforward. The peak-over-threshold approach offers a solution to this problem. The Fisher-Tippett theorem is the basis for the theorem of peak over threshold. Based on the results of Balkema and de Haan (1974) and Pickands (1975), the distribution of the exceedances over a high threshold u can be approximated by the generalized Pareto distribution.

2.2.1. Balkema and de Haan – Pickands Theorem

It is possible to find a positive measurable function β, where β is a function of u, such that:

$$\lim_{u \to x_F} \sup_{0 \le x \le x_F - u} \left| F_u(x) - G_{\xi,\beta}(x) \right| = 0$$

if and only if $F \in MDA(H_{\xi}(x))$. That is, for a large class of underlying distributions F, as the threshold u gradually increases, the excess distribution function F_u converges to a generalized Pareto distribution.

2.2.2. The Generalized Pareto Distribution

The generalized Pareto distribution (GPD) is the limiting distribution of the peak over threshold approach and is defined as:

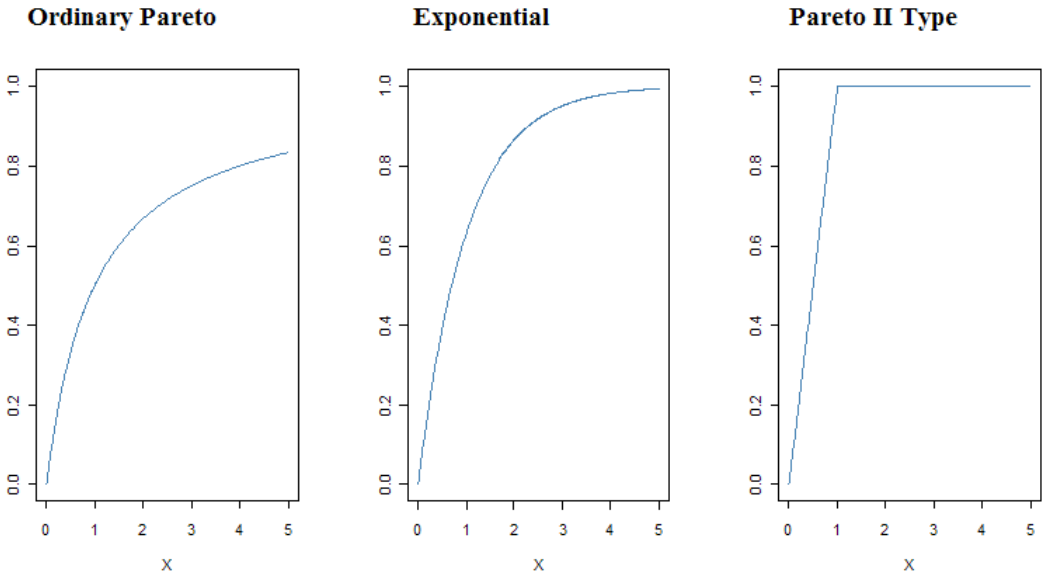
$$G_{\xi,u,\beta}(x) = \begin{cases} 1-(1+\xi(\frac{x-u}{\beta}))^{-1/\xi} & if \quad \xi \neq 0 \\ 1-e^{-(\frac{x-u}{\beta})} & if \quad \xi = 0 \end{cases} \tag{4}$$

with

$$x \in \begin{cases} [u,\infty], & if \quad \xi \geq 0 \\ [u,u-\beta/\xi], & if \quad \xi < 0 \end{cases}$$

where ξ is the shape parameter, β is the scale parameter, and u is the location parameter. The Balkema and de Haan – Pickands theorem implies that, if GEV is the limiting distribution for the block maxima, then the corresponding limiting distribution for threshold excesses is GPD. In addition, the parameters of the GPD of threshold excesses are uniquely determined by those of the associated GEV distribution of block maxima (Coles (2001)). The shape parameter ξ is equal to that of the corresponding GEV distribution and is dominant in determining the qualitative behaviour of the generalized Pareto distribution. As with the GEV distribution, the excess distribution has an upper bound if ξ<0 and has no an upper limit if ξ>0. The distribution is also unbounded if ξ=0. Similar to the GEV distribution, the GPD distribution embeds three distributions:

Figure 2 Generalized Pareto Distributions – Cumulative Probability Functions



Ordinary Pareto distribution: $\xi = \alpha^{-1} > 0,$

Exponential distribution: $\xi = \alpha^{-1} < 0,$

Pareto II type distribution: $\xi = 0$.

Figure 2 plots the shape of these distribution functions with, for illustrative purposes, the location parameter *u* set to zero and the scale parameter β set to 1. In general, financial losses do not have an upper limit. Figure 2 suggests that distributions with shape parameter ξ>0 are more suited to model fat tailed distributions.

2.2.3. The Choice of the Threshold

An important step in applying the POT approach is to choose an appropriate threshold value *u*. In theory, *u* should be high enough so that the limiting distribution will converge to the generalized Pareto distribution. In practice, the choice of *u* should allow for enough observations to estimate the parameters. Two approaches are used in this chapter to

determine the appropriate threshold value, one is the mean excess plot and the other is maximum likelihood estimation of a parametric GPD.

2.2.3.1. The Mean Excess Plot

The mean excess plot (ME-plot) is a very useful graphical tool for selecting the threshold u. The mean excess is the expected value of the excess over a given threshold u, given that u is exceeded. The mean excess function $e(.)$ for a random variable X with right endpoint x_F is defined as:

$$e(u) = E(X - u | X > u) \text{ for } u<x_F.$$

The mean excess function is better known as the Expected Shortfall in financial risk management; see Embrechts *et al.* (1997) for a detailed discussion of the properties of this function. If the underlying distribution $X>u$ has a generalized Pareto distribution, then the corresponding mean excess is:

$$e(u) = \frac{\beta + \xi u}{1 - \xi} \tag{5}$$

where $\xi<1$ so that $e(u)$ exists. As indicated by the equation above, the mean excess function is linear in the threshold u when $X>u$ has a generalized Pareto distribution. Let n be the number of observations that exceed the threshold u. The empirical mean excess function is defined as:

$$e(u) = \frac{\sum_{i}^{n}(x_i - u)}{\sum_{i}^{n}1}, \quad where \quad x_i > u \tag{6}$$

To use the ME-plot to choose the threshold u, one has to look for a threshold u from which the plot presents approximately linear line behaviour. The mean excess plot will tend towards infinity with a positive slope for a fat-tailed distribution. When the data are

exponentially distributed, the plot is a horizontal line. For light-tailed distributions, the plot has a negative slope.

2.2.3.2. Parameter estimation

As a preliminary test, using the ME-plot to select the appropriate threshold is more of an art than a science. A further check on the preliminary conclusion is to estimate the shape parameters using the generalized Perato distribution and look for stability of the parameter estimates as the threshold is changed. By the Balkema and de Haan – Pickands theorem, if the GPD is a reasonable distribution for a threshold, then the excesses of a higher level threshold should also follow a GPD with the same shape parameter. Therefore, above a certain level of threshold, the shape parameter should be very stable.

Once a threshold level has been selected, the parameters of the GPD can be estimated using several approaches, including maximum likelihood estimation (MLE), method of moments (MOM), biased and unbiased probability weighted moments (PWMB, PWMU), *etc*. In this chapter, we use the method of MLE to estimate the shape parameter ξ and the scale parameter β. For a high enough threshold u and n excesses of the threshold (x_1-u, …, x_n-u), the likelihood function is given by:

$$L(\xi,\beta) = \begin{cases} -n\log(\beta) - \dfrac{1+\xi}{\xi}\sum_{i=1}^{n}\log\left(1+\xi\left(\dfrac{x_i-u}{\beta}\right)\right) & if \quad \xi \neq 0 \\ -n\log(\beta) - \dfrac{1}{\beta}\sum_{i=1}^{n}(x_i-u) & if \quad \xi = 0 \end{cases} \tag{7}$$

By maximizing the log likelihood function, we can obtain the estimates of the shape parameter ξ and the scale parameter β.

2.3. Measures of Extreme Risks: Value at Risk and Expected Shortfall

In recent years, Value-at-Risk (VaR) has become the most commonly used tool to measure the downside risk associated with a portfolio. The popularity of VaR started in the early 1990's when it was endorsed by the Group of Thirty (G30)[4] as the "best practices" for dealing with derivatives in its 1993 best practices report. VaR measures the worst expected losses in the market value over a specific time interval at a given significance level. VaR answers the question: "How much can I lose over a pre-set horizon with *x%* probability?" For a given probability *p*, VaR can be defined as the *p*-th quantile of the distribution *F*:

$$VaR_p = F^{-1}(1-p) \tag{8}$$

where F^{-1}, the inverse of the distribution function *F*, is the quantile function. The methods used to calculate VaR can be grouped into parametric and non-parametric approaches. The parametric approach assumes a particular model for the distribution of data; for example, the variance-covariance method and the Extreme Value VaR method. The non-parametric approach includes the historical simulation method and the Monte Carlo simulation method.

2.3.1. Variance-Covariance Method

The variance-covariance method is the simplest and the most commonly used approach among the various models used to estimate VaR. Assuming that returns r_t, *t=1,2,…,n,*

[4] The Group of Thirty, often abbreviated to G30, is an international body of leading financiers and academics which aims to deepen understanding of economic and financial issues and to examine consequences of decisions made in the public and private sectors related to these issues. The group consists of thirty members and includes the heads of major private banks and central banks, as well as members from academia and international institutions.

follow a martingale process with $r_t = \mu_t + \varepsilon_t$, where ε_t has a distribution function *F* with zero mean and variance σ_t^2, the VaR can be calculated as:

$$VaR_p = \mu_t + F^{-1}(1-p)\sigma_t \tag{9}$$

The most commonly used distribution function *F* in this case is the Normal distribution. The biggest criticism of this approach is that most financial time series exhibit the properties of asymmetry and fat tails. Therefore, the risk is often underestimated. However, this approach has been widely applied for calculating the VaR since the risk is additive when it is based on sample variance assuming normality.

In order to account for the fat tails, the standard deviation can also be estimated using a statistical model such as the family of GARCH (Bollerslev, 1986) models. The simplest GARCH (1, 1) model is as follows:

$$r_t = \sigma_t \varepsilon_t \qquad \varepsilon_t \sim i.i.d.(0,1) \tag{10}$$

$$\sigma_t^2 = \omega + \alpha\, r_{t-1}^2 + \beta\, \sigma_{t-1}^2$$

Although the conditional distribution of the GARCH process is Normal, the unconditional distribution exhibits some excess kurtosis.

2.3.2. RiskMetrics

The RiskMetrics approach is a particular, convenient case of the GARCH process. Variances are modeled using an exponentially weighted moving-average (EWMA) forecast. The forecast is a weighted average of the previous forecasts, with weight λ, and of the latest squared innovation, with weight (1- λ):

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2 \tag{11}$$

where σ_t^2 is the forecast of the volatility and r_t^2 is the squared return, which acts as the proxy for true volatility. The λ parameter, also called the decay factor, determines the relative weights places on previous observations. The EWMA model places geometrically declining weights on past observations, assigning greater importance to recent observations. Note that through backward substitution of the RiskMetrics model we arrive at the expression in Eq. (12) whereby the prediction of volatility is an exponentially weighted moving average of past squared returns.

$$\sigma_t^2 = (1-\lambda)\sum_{\tau=1}^{\infty} r_{t-\tau}^2 \tag{12}$$

Although in principle the decay factor λ, can be estimated, the RiskMetrics approach has chosen λ=0.94 for daily forecasts. A clear advantage of the RiskMetrics model is that no estimation is necessary as the decay factor has been set to 0.94. This is a huge advantage in a large portfolio. However, the disadvantage of the approach is that it is not able to capture the asymmetry and fat tails behaviour of the returns.

2.3.3. Historical Simulation

The other most commonly used method for VaR estimation is the Historical Simulation (HS). The VaR in this case is estimated by the *p*-th quantile of the sample returns. This approach is non-parametric and does not require any distributional assumptions as the HS approach essentially uses only the empirical distribution of the returns. Hence, the HS approach allows us to capture fat tails and other non-Normal characteristics without knowing the underlying distribution.

However, the HS approach assumes that the distribution of the returns is constant over the sample period. This approach relies on the selected historical database and ignores any other events that are not represented in the database. It is problematic to use the HS approach to forecast out-of-sample VaR when the distribution over the sample period does not represent the population distribution.

2.3.4. Extreme VaR

EVT focuses on the tail distribution of the returns. For that reason, it is not surprising that the extreme value based VaR is superior to the traditional variance-covariance and non-parametric methods in estimating extreme risks (Aragones *et al.*, 2000). The extreme value based VaR can be estimated by:

$$\hat{VaR} = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\frac{n}{N_u} p \right)^{-\hat{\zeta}} - 1 \qquad (13)$$

where n is the total number of observations, N_u is the number of observations above the threshold, $\hat{\beta}$ is the estimated scale parameter and $\hat{\xi}$ is the estimated shape parameter.

2.3.5. GARCH-EVT Methodology

Most financial return series exhibit stochastic volatility and fat-tailed distributions. The presence of the serial correlation in the squared returns violates the basic assumption made by the Extreme Value Theory that the series under study is independently and identically distributed. In order to address the issue of stochastic volatility, this study adopts McNeil and Frey's (2000) approach to model financial returns and measure tail risk. McNeil and Frey's solution is to first fit a GARCH-type model to the return data by quasi-maximum likelihood. The GARCH residuals have been shown to be closer to the

i.i.d. assumption than the raw returns series but continue to exhibit fat tails. The second stage of the approach is to apply the EVT to the GARCH residuals. The advantage of this GARCH–EVT combination lies in its ability to capture conditional heteroskedasticity in the data through the GARCH framework, while at the same time modeling the extreme tail behaviour through the EVT method.

We assume that the dynamics of returns can be represented by:

$$r_t = a_0 + a_1 r_{t-1} + \sigma_t Z_t \tag{14}$$

where the innovations Z_t are a strict white noise process with zero mean, unit variance and marginal distribution function $F_Z(z)$. We assume that the conditional variance σ_t^2 of the mean-adjusted series $\varepsilon_t = r_t - a_0 - a_1 r_{t-1}$ follows a GARCH (*p, q*) process:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \gamma_j \sigma_{t-j}^2 \tag{15}$$

where the coefficients α_i (i=0,…,p) and γ_j (j=0,…,q) are all assumed to be positive to ensure that the conditional variance σ_t^2 is always positive. The GARCH (*p, q*) model is fitted using a quasi-maximum-likelihood approach, which means that the likelihood for a GARCH (*p, q*) model with Normal innovations is maximized to obtain parameter estimates. The assumption of Normal innovations contradicts our belief that financial returns have fat-tailed distributions. However, the PML method has been shown to yield a consistent and asymptotically Normal estimator (see Chapter 4 of Gouriéroux (1997)). Standardized residuals can be calculated as:

$$(z_{t-n+1},\ldots,z_t) = \left(\frac{r_{t-n+1} - \hat{a}_0 - \hat{a}_1 r_{t-n}}{\hat{\sigma}_{t-n+1}},\ldots,\frac{r_t - \hat{a}_0 - \hat{a}_1 r_{t-1}}{\hat{\sigma}_t} \right)$$

where a ^ indicates estimated parameters using a PML approach. The one-step ahead forecast for the conditional variance in *t+1* is given by:

$$\hat{\sigma}_{t+1}^2 = \hat{\omega} + \sum_{i=1}^{p} \hat{\alpha}_i \hat{\varepsilon}_{t+1-i}^2 + \sum_{j=1}^{q} \hat{\gamma}_j \hat{\sigma}_{t+1-j}^2 \qquad (16)$$

where $\hat{\varepsilon}_t = r_t - \hat{a}_0 - \hat{a}_1 r_{t-1}$. For stage two of the GARCH-EVT approach, we estimate the tails of the standardized residuals computed in the stage one using EVT. The *q*th quantile of the innovations is given by:

$$\hat{VaR}(Z)_q = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\frac{n}{N_u} p \right)^{-\hat{\zeta}} - 1 \quad . \qquad (17)$$

Therefore, for a one-day horizon, an estimate of the VaR for the returns is:

$$VaR_q^{t+1} = \hat{a}_0 + \hat{a}_1 r_t + \hat{\sigma}_{t+1} \hat{VaR}(Z)_q \; . \qquad (18)$$

2.3.6. Expected Shortfall (ES)

As the most commonly used quantile-based risk measure, Value-at-Risk has been heavily criticized. First, VaR does not indicate the size of the potential loss given that this loss exceeds the VaR. Second, as Artzner *et al.* (1997, 1999) showed that the VaR is not necessarily sub-additive. That is, the total VaR of a portfolio may be greater than the sum of individual VaRs. This may cause problems if the risk management system of a financial institute is based on VaR limits of individual books. To overcome these deficiencies, Artzner *et al.* (1997) introduced Expected Shortfall (ES) as an alternative

risk measure which is not only coherent[5] but also gauges the extent of the loss when a VaR is exceeded. The Expected Shortfall is defined as:

$$ES_q = E\left(r \middle| r > VaR_q\right)$$

Therefore, the Expected Shortfall measures the expected value of loss when a VaR violation occurs. The above expression can be rewritten as:

$$ES_q = VaR_q + E\left(r - VaR_q \middle| r > VaR_q\right) \tag{19}$$

The second term of this can be interpreted as the excess distribution $F_{VaR_q}(y)$ over the threshold VaR_q. According to the Pickands-Balkema-de Haan Theorem, if the threshold VaR_q is high enough, the excess distribution is also a GPD. Therefore, the mean of the excess distribution $F_{VaR_q}(y)$ is given by:

$$(\beta + \xi(VaR_q - u)) / (1 - \xi)$$

The Expected Shortfall calculated using EVT based methods can be estimated as:

$$\hat{ES}_q = \frac{\hat{VaR}_q}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1-\hat{\xi}} \tag{20}$$

2.3.7. Backtest Risk Models

The best way to rank the competing VaR approaches is to assess the out-of-sample accuracy of the estimated VaRs in forecasting extreme returns. The simplest method is to compare the out-of-sample VaR estimates to the actual realized return in the next period. A violation occurs if the realized return is greater than the estimated one in a given day. The violation ratio is calculated by dividing the number of violations by the total number

[5] A coherent risk measure ρ is defined as one that satisfies the following four properties: (a) sub-additivity, (b) homogeneity, (c) monotonicity, and (d) translational invariance. These are described in the following equations: (a) $\rho(x) + \rho(y) \leq \rho(x + y)$, (b) $\rho(tx) = t\rho(x)$, (c) $\rho(x) \geq \rho(y)$ if $x \leq y$, and (d) $\rho(x + n) = \rho(x) + n$.

of one-step-ahead forecasts. When forecasting VaRs at a certain quantile q, we expect the realized return will be higher 100(1-q) percent of the time if the model is correct. Ideally, the violation ratio should converge to q as the sample size increases. A violation ratio higher than q implies that the model consistently underestimates the return/risk at the tail which may result in unnecessary frequent adjustments to the portfolio. On the other hand, a violation ratio less than the expected one indicates that the model consistently overestimates the return/risk which will require excessive capital commitment.

In order to determine whether the frequency of violation is in line with the expected significance level, we use the unconditional coverage test of Kupiec (1995). Assuming that the VaR estimates are accurate, the violations can be modeled as independent draws from a binomial distribution. Define $N = \sum_{t=1}^{T} I_{t+1}$ as the number of violations over T periods, where I_{t+1} is the sequence of VaR violations that can be described as:

$$\text{Right Tail: } I_{t+1} = \begin{matrix} 1, & if & r_{t+1} > VaR_{t+1}|t \\ 0, & if & r_{t+1} \le VaR_{t+1}|t \end{matrix}$$

$$\text{Left Tail: } I_{t+1} = \begin{matrix} 1, & if & r_{t+1} < VaR_{t+1}|t \\ 0, & if & r_{t+1} \ge VaR_{t+1}|t \end{matrix}$$

The null hypothesis of Kupiec's unconditional coverage test assumes that the probability of occurrence of the violations, N/T equals the expected significance level q. Let p be the expected violation rate ($p = 1-q$, where q is the significance level for the VaR). The appropriate likelihood ratio statistic LR_{uc}, the test of unconditional coverage, is:

$$LR_{uc} = 2 \log \left[\left(\frac{N}{T}\right)^{N} \left(1-\frac{N}{T}\right)^{T-N} \right] - \log\left(p^{N}(1-p)^{T-N}\right) \xrightarrow{d} \chi^{2}(1). \qquad (21)$$

Note that this is a two-sided test and a model is rejected if it generates too few or too many violations. As Christoffersen (1998) points out, the Kupiec test only provides a necessary condition to classify a VaR model as adequate. In the presence of volatility clustering or volatility persistence, the conditional accuracy of VaR estimates becomes an important issue. Christoffersen (1998) proposed a conditional coverage test that jointly investigates (1) whether the number of violations is statistically consistent with the hypothesized number, (2) whether violations are independently distributed through time. That is, the number of violations should follow an *i.i.d*. Bernoulli sequence with the targeted exceedance rate. The conditional coverage test is a joint test of two properties: correct unconditional coverage and serial independence:

$$LR_{cc} = LR_{uc} + LR_{ind}$$

which is asymptotically distributed as a chi-square variate with two degrees of freedom, $\chi^2(2)$ under the null hypothesis of independence. The statistic LR_{ind} is the likelihood ratio test statistic for the null hypothesis of serial independence against the alternative of first-order Markov dependence. The appropriate likelihood ratio statistic for the conditional coverage test is:

$$LR_{cc} = -2\log\left[(1-p)^{T-N} p^{N}\right] + 2\log\left[(1-\pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1-\pi_{11})^{n_{10}} \pi_{11}^{n_{11}}\right] \xrightarrow{d} \chi^2(2) \tag{22}$$

where n_{ij} is the number of observations with value *i* followed by *j*, for *i, j=0, 1*, and $\pi_{ij} = \dfrac{n_{ij}}{\sum_j n_{ij}}$ are the corresponding probabilities. The values *i, j = 1* indicate that a violation has been made, while *i, j =0* points to the opposite. The advantage of the

conditional coverage test is that risk managers can reject those models that generate too few or too many *clustered* violations.

Both the unconditional and conditional coverage tests only deal with the frequency of the violations. However, these tests fail to consider the severity of additional loss (excess of estimated VaR) when violations occur. A "Black Swan" event[6] could potentially wipe out all capital in a portfolio, which poses a much higher risk than many small losses. Therefore, among all models that can forecast VaR accurately, it is important to rank the competing models based on the specific concerns of the risk managers. The idea of using a loss function to address these specific concerns was first proposed by Lopez (1998). Specifically, he considered the following loss function:

$$\Psi_{t+1} = \begin{cases} 1 + \left(VaR_{t+1|t} - r_{t+1}\right)^2 & if \quad r_{t+1} > VaR_{t+1|t} \\ 0 & else \end{cases} \tag{23}$$

This loss function is defined with a negative orientation and a model which minimizes the loss is preferred over the other models. The above loss function adds a score of one whenever a violation occurs to penalize a high number of violations. Also, the penalty increases when the magnitude of tail losses $\left(VaR_{t+1|t} - r_{t+1}\right)^2$ increases. However, the individual portfolio manager's loss function is not necessarily the same as the above loss function. For example, one may want to incorporate the opportunity cost of the capital requirement imposed by the VaR models. Nevertheless, this approach provides an evaluation framework that can be adjusted to address an individual portfolio manager's objective function.

[6] A "Black Swan" event is a high-impact, hard-to-predict, and rare event beyond the realm of normal expectations.

Lopez's loss function suffers from the disadvantage that if the competing VaR models are not filtered by the unconditional and conditional coverage tests, a model that does not generate any violation is considered the optimal choice as $\Psi_{t+1} = 0$. In addition, the magnitude of the tail losses could be better measured using the ES estimates instead of the VaR estimates which does not account for the size of the expected loss. Therefore, inspired by Sarma *et al.* (2003) and Angelidis and Degiannakis (2006), backtests of the risk models in this chapter will be conducted using a two-stage approach. In stage one, the competing VaR models are tested using the unconditional and conditional coverage tests to ensure that the frequency of the violation is equal to the expected significance level and the occurrence of the violations is independently distributed. The second stage is designed to incorporate penalties for the magnitude of the tail losses and the opportunity costs of the capital requirement on the other days. Therefore, the loss function can be defined as:

$$\Psi_{t+1} = \begin{matrix} \left(r_{t+1} - ES_{t+1|t}\right)^2 & if \quad r_{t+1} > VaR_{t+1|t} \\ \theta_c ES_{t+1|t} & else \end{matrix} \tag{24}$$

where θ_c measures the opportunity cost of capital.

3. Modeling Energy Futures using A Dynamic GARCH-EVT Framework

3.1. Data Description and Preliminary Tests

As the purpose of this chapter is to study the tail behaviour of energy futures, we select all energy futures contracts that are liquid and widely held. End-of-day prices of the front month futures of West Texas Intermediate (CL), Brent Crude Oil (CO), Heating Oil (HO), and Natural Gas (NG) have been obtained from the Energy Information Administration (EIA)[7] and Bloomberg[8]. Table 1 below shows the start date and end date of the data. Gasoline futures are not included in this analysis because the current benchmark gasoline contract (reformulated gasoline blendstock for oxygen blending, or RBOB) only started trading on the New York Mercantile Exchange (NYMEX) in October 2005.

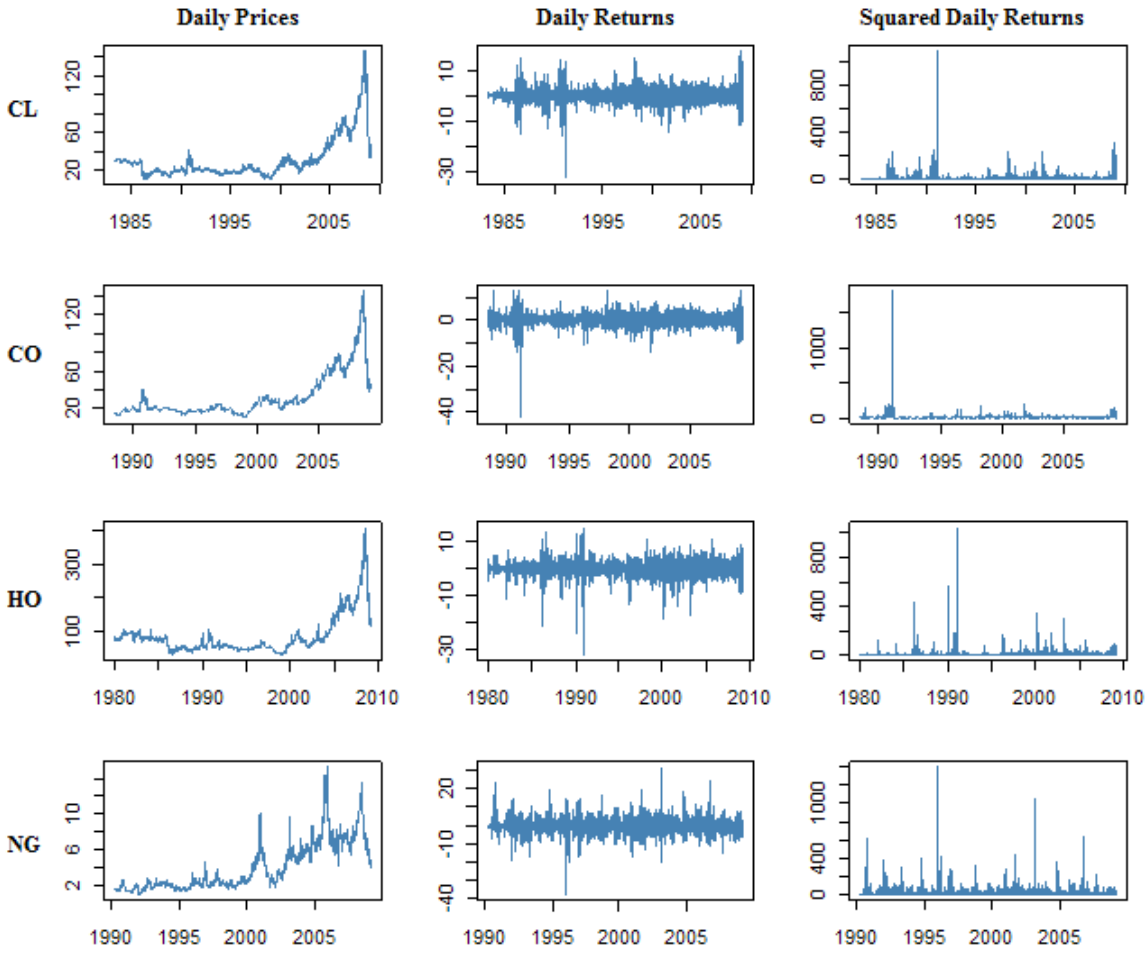
Table 1 Data Analyzed

Commodity Futures	Ticker	Unit	Start	End	Observations
WTI	CL	US$/bbl	4-Apr-83	3-Mar-09	6,496
Brent	CO	US$/bbl	23-Jun-88	3-Mar-09	5,399
Heating Oil	HO	USCent/gal.	2-Jan-80	3-Mar-09	6,527
Natural Gas	NG	USD/MMBtu	3-Apr-90	3-Mar-09	4,732

[7] http://www.eia.doe.gov/

[8] http://www.bloomberg.com

Figure 3 Daily Prices, Returns and Squared Returns



** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Since we focus on the unleveraged returns in this chapter, the daily returns are defined as:

$$r_t = \log \frac{P_t}{P_{t-1}} \times 100 . \qquad (25)$$

Figure 3 plots daily prices, returns and squared returns for each series analyzed. Each plot of daily returns confirms the typical empirical time series properties. All series of returns appear to be mean reverting and exhibit periods of low volatility followed by periods of extreme volatility. The squared daily returns exhibit evidence of volatility clustering, *i.e.* large changes tend to be followed by large changes. For risk management purposes, any measure of risk exposure should be conditional on the current volatility regime.

Basic statistics for the front month futures are summarized in Table 2. As is the case with most financial time series, the front month energy futures exhibit evidence of fat tails, with the kurtosis of all series far larger than 3. The Jarque-Bera (1980) normality test[9] also confirms that all returns are not normally distributed. The Augmented Dickey-Fuller (1979) test rejects the null hypothesis of a unit root for all series. Among all energy futures, natural gas is more volatile than others. The standard deviation of natural gas front month futures returns is 50% higher than the standard deviation of petroleum products returns. Unlike the petroleum products, front month natural gas futures returns exhibit positive skewness.

---

[9] The Jarque-Bera test statistic JB is defined as $JB = \frac{n}{6} S^2 + \frac{1}{4} K^2$, where n is the number of observations, S is the sample skewness and K is the sample kurtosis. The statistic JB has an asymptotic chi-square distribution with two degrees of freedom and can be used to test the null hypothesis that the data are from a Normal distribution. The null hypothesis is a joint hypothesis of the skewness being zero and the excess kurtosis being 0,

Table 2 Summary Descriptive Statistics

	CL	**CO**	**HO**	**NG**
Start Date	4-Apr-83	23-Jun-88	2-Jan-80	3-Apr-90
End Date	3-Mar-09	11-Sep-07	3-Mar-09	3-Mar-09
Sample Size	6,496	856	7,313	4,732
Mean	0.0053	0.0190	0.0050	0.0203
Median	0.0377	0.0275	0.0444	0.0000
Max	16.4097	13.1506	13.9942	32.4354
Min	-40.0478	-42.7223	-39.0942	-37.5749
Std. dev.	2.4522	2.2851	2.3464	3.6033
Skewness	-0.8978	-1.3889	-1.5298	0.0549
Kurtosis	19.4922	29.3368	23.1176	10.6783
Jarque-Bera Test	74481.0	157773.2	126156.2	11624.2
p-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*
ADF Unit-Root Test**	-39.0616	-73.6550	-86.6185	-69.8695
p-value	*0.0000*	*0.0001*	*0.0001*	*0.0001*
Auto Corr – r***				
Lag 1	-0.008	-0.003	-0.013	-0.016
Lag 5	-0.060	-0.037	-0.009	-0.031
Lag 10	-0.019	-0.031	-0.035	-0.024
Lag 20	0.025	0.004	-0.006	0.012
Ljung-Box (20)	88.46	48.828	77.858	36.999
p-value	*0.0000*	*0.0000*	*0.0000*	*0.0120*
Auto Corr – r^2***				
Lag 1	0.144	0.075	0.067	0.086
Lag 5	0.106	0.044	0.071	0.107
Lag 10	0.184	0.13	0.122	0.153
Lag 20	0.099	0.027	0.059	0.053
Ljung-Box (20)	938.65	264.75	550.28	425.75
p-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*

** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*
*** An intercept term (no trend) is included in the integrating regression for the ADF Unit-Root test.*
**** The Lag orders are selected to examine a range of possible autocorrelations.*

The squared daily returns shown in Figure 3 exhibit volatility clustering, *i.e.* large changes tend to follow by large changes, in "bursts". The Ljung-Box (1978) Q-statistics[10] reported in Table 2 also reject the null hypothesis of no autocorrelation through 20-lags at a 1% significance level for all series. Given these features, the next step is to test the (G)ARCH effect using an ARCH-LM test on the residuals from a simple Ordinary Least Squares regression of each returns series on an intercept and it own lagged values. The ARCH-LM test is a Lagrange multiplier (LM) test for autoregressive conditional heteroskedasticity (ARCH) in the residuals (Engle (1982)). The ARCH LM test statistic is computed from an auxiliary test regression. To test the null hypothesis that there is no ARCH up to order q in the residuals, we estimate the model:

$$e_t^2 = \lambda_0 + \sum_{i=1}^{q} \lambda_i e_{t-i}^2 + v_t$$

where e is the residuals obtained from the above-mentioned OLS regression. The *F*-statistic is an omitted variable test for the joint significance of all lagged squared residuals. The LM test statistic is asymptotically distributed $\chi^2(q)$ under quite general conditions. The test statistics shown in table 3 suggest that the null hypothesis of no ARCH effect is rejected for all series and the null hypothesis of no GARCH effect is also rejected for all series. This result is not sensitive to the number of lags included in the ARCH-LM and GARCH tests.

---

[10] The Ljung-Box test tests whether any of a group of autocorrelations of a time series are different from zero. The test statistics can be calculated as $Q = n(n+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$, where n is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag k, and h is the number of lags being tested. For significance level α, the critical region for rejection of the hypothesis of randomness is $Q > \chi^2_{1-\alpha,h}$ .where $\chi^2_{1-\alpha,h}$ is the α-quantile of the chi-square distribution with h degrees of freedom.

Table 3 (G)ARCH Effect Test

Testing for ARCH(1)	**CL**	**CO**	**HO**	**NG**
F-statistic	143.2	31.2	32.4	35.2
p-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Testing for GARCH(1,1)	**CL**	**CO**	**HO**	**NG**
nR^2	154.1	36.8	77.0	40.0
p-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*

** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Given the above test results, the GARCH effects need to be filtered out from the data before the extreme value analysis can be applied. In order to find the appropriate GARCH specification, I estimate various GARCH (*p, q*) models (from 1 to 2 for both *p* and *q*). Based on the Schwarz (1978)[11] criterion, the selected GARCH (*p, q*) models are presented in Table 4 below along with the ARCH effect test statistics on the residuals. As the ARCH-LM test confirms that there are no ARCH effects left in the residuals, the analysis in the rest of this chapter will focus on the residuals of the GARCH (*p, q*) models fitted to the front month energy futures.

[11] Schwarz criterion (SC) is defined as: $SC(M) = -2\ln(\hat{L}) + \frac{M \ln(n)}{n}$, where *M* is the number of parameters being estimated and $\ln(\hat{L})$ is the maximized log likelihood value. Smaller SC values indicate better fit.

Table 4 Selected GARCH (p,q) Models and ARCH Effect Tests on the Residuals

	CL	**CO**	**HO**	**NG**
Selected GARCH(p,q) Model	GARCH (2,2)	GARCH (1,1)	GARCH (1,1)	GARCH (2,2)
Residual ARCH-LM Test				
F-statistic	0.1465	0.0166	0.0002	0.8817
p-value	*0.7019*	*0.8974*	*0.9882*	*0.3478*

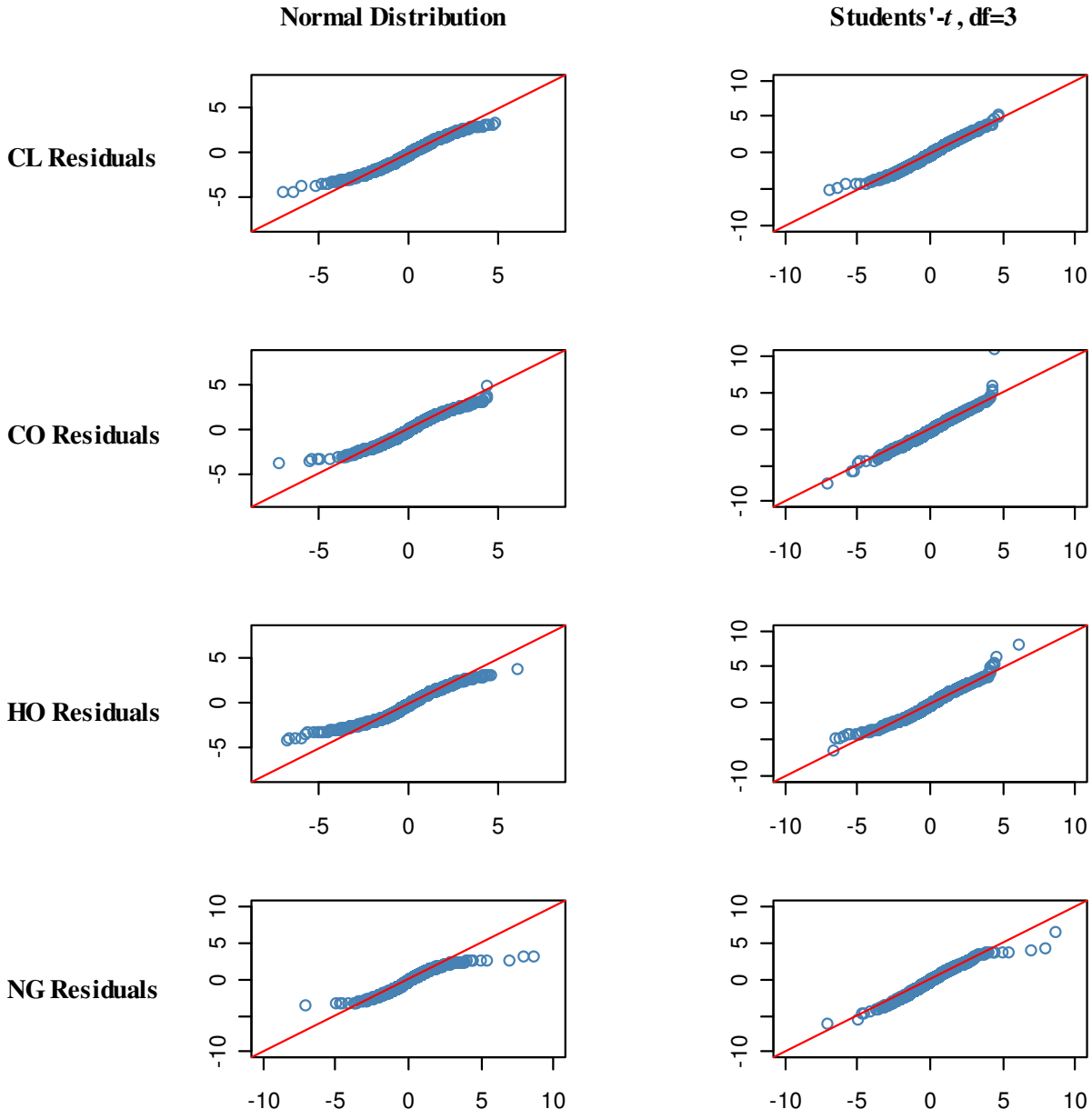
** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

3.2. Exploratory Analysis

One important exploratory tool used in extreme value analysis is the quantile-quantile plot (QQ-plot). The QQ-plot is a graphical way to determine whether two data sets come from populations with a common distribution. The graph of the quantiles makes it easy to informally assess the goodness of fit of the data to the assumed underlying distribution. Let $X_{n,n} < X_{n-1,n} < \ldots < X_{1,n}$ be the ordered data and let F be the empirical distribution. The graph of quantiles is defined by the set of the points:

$$X_{k,n}, F^{-1}\ \frac{n-k+1}{n+1}\ , k = 1,\ldots,n\ \ .$$

Figure 4 QQ-Plot against the Normal and the Student's-t Distribution



** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Basically, the QQ-plot compares the quantiles of the dataset with the quantiles of the reference distribution *F*. A perfect fit shows as a 45^{o} linear line. If the plot curves to the right, the data have a right tail which is heavier than the reference distribution *F* and if it curves up to the left the data have a heavier left tail. This is the case when comparing the residuals to the Normal distribution function, where we see that the data have heavier tails than expected from a Normal distribution. The second (right) QQ-plot compares the residuals to the Student's-*t* distribution. The Student's-*t* distribution fits the data much better than the Normal, as expected because it allows for fatter tails distribution. However, there is some evidence that the distribution of the residuals has heavier tails than the Student's-*t* distribution with three degrees of freedom.

In order to explore the potentially asymmetric behaviour between the risk and return, all residual series have been split into right tails and left tails. In the left tail case, we change the sign of the residuals to positive values. Table 5 below lists the number of observations included in each case. The following empirical analysis will use the split samples.

Table 5 Number of Observations Included in Each Tail

	CL	**CO**	**HO**	**NG**
Name	CL_Right Tail	CO_Right Tail	HO_Right Tail	NG_Right Tail
# of Obs.	3,268	2,701	3,692	2,369
Name	CL_Left Tail	CO_Left Tail	HO_Left Tail	NG_Left Tail
# of Obs.	3,226	2,697	3,619	2,361

** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

3.3. Determination of Thresholds

As discussed above, two approaches are utilized in this chapter to determine the appropriate threshold value. First, the mean excess plot will be used to select the threshold range graphically. A complementary step is to estimate the shape parameters and look for the stability of the estimates. The mean residual life plot and the maximum likelihood estimates of the shape parameters are plotted side-by-side for each tail under study. In order to closely investigate the choice of the threshold, the plots below only include the observations between the 5% quantile and the 25% quantile. The 95% asymptotic confidence band is also shown in the plots.

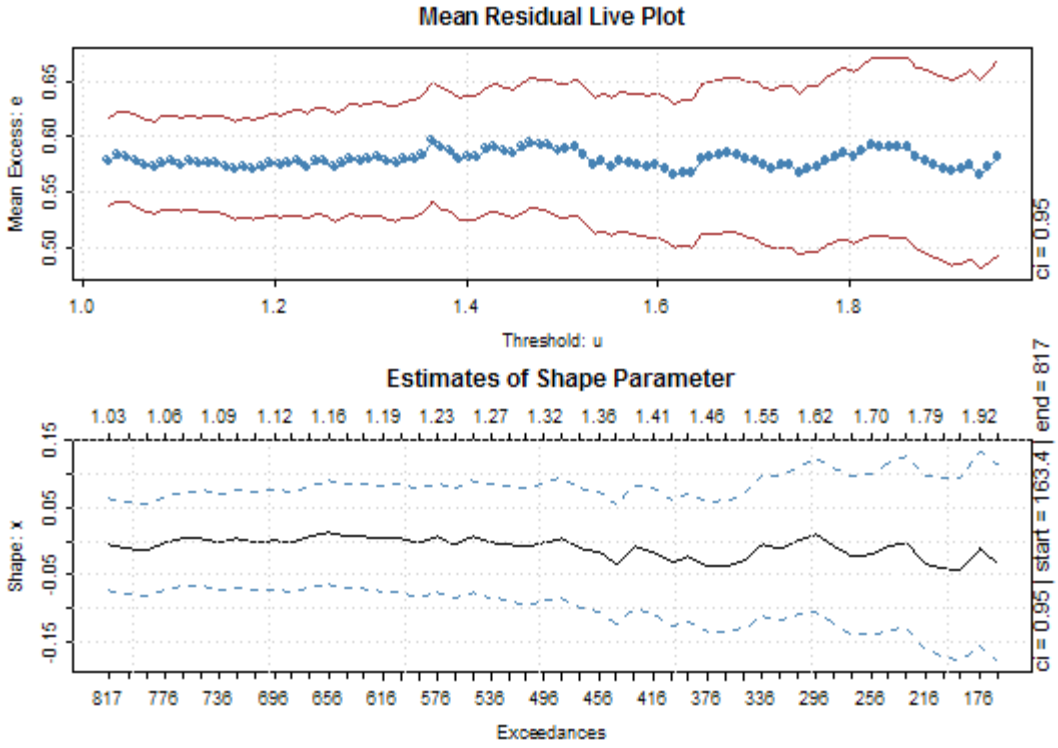
Table 6 shows the selected threshold values, corresponding sample quantiles and the number of exceedances. The estimated shape and scale parameters along with their asymptotic standard errors under different thresholds are also presented in Table 6. The estimated tail index values range between -0.029 (Brent, right tail) to 0.16 (WTI, left tail). The estimated tail index values are not overwhelmingly significant for some tails. However, we have decided not to impose the null hypothesis as the diagnostic plots in Figure 9 show excellent fit for the estimated GDP distribution to the empirical distribution (this decision is also consistent with Coles (2001)). Higher values of the estimated tail index are observed for the left tail except for the Natural Gas market. This is consistent with the fact that only the returns for the Natural Gas market exhibited a positive skew. The fact that the right tail index is less skewed than the left tail index for the petroleum product markets is an indicator that extreme losses are more likely than extreme rewards in these markets. On the other hand, the Natural Gas market has a higher

right tail index than left tail index. Therefore, high positive returns are more likely than large losses in this market.

According to the Extreme Value Theory, the excess distribution above a suitable threshold should follow a generalized Pareto distribution. In Figure 9, we compare the empirical distribution of exceedances with the cumulative distribution simulated from a GPD to determine how the GPD fits the tails of the standardized residuals. In general, the empirical excess distribution function follows the corresponding GPD closely, which suggests that the GPD likely models the exceedances well. Figure 9 also plots the survival functions for all of the tails. The survival function is simply one minus the cumulative density function, 1-$F(x)$. It is the probability of observing a value from the series x at least as large as some specified value u. Under different thresholds for both right tails and left tails, the plotted survival functions confirm that the GPD fits the residuals well.

Figure 5 Mean Residual Life & Shape Parameter Estimates Plot – WTI

Right Tail



Left Tail

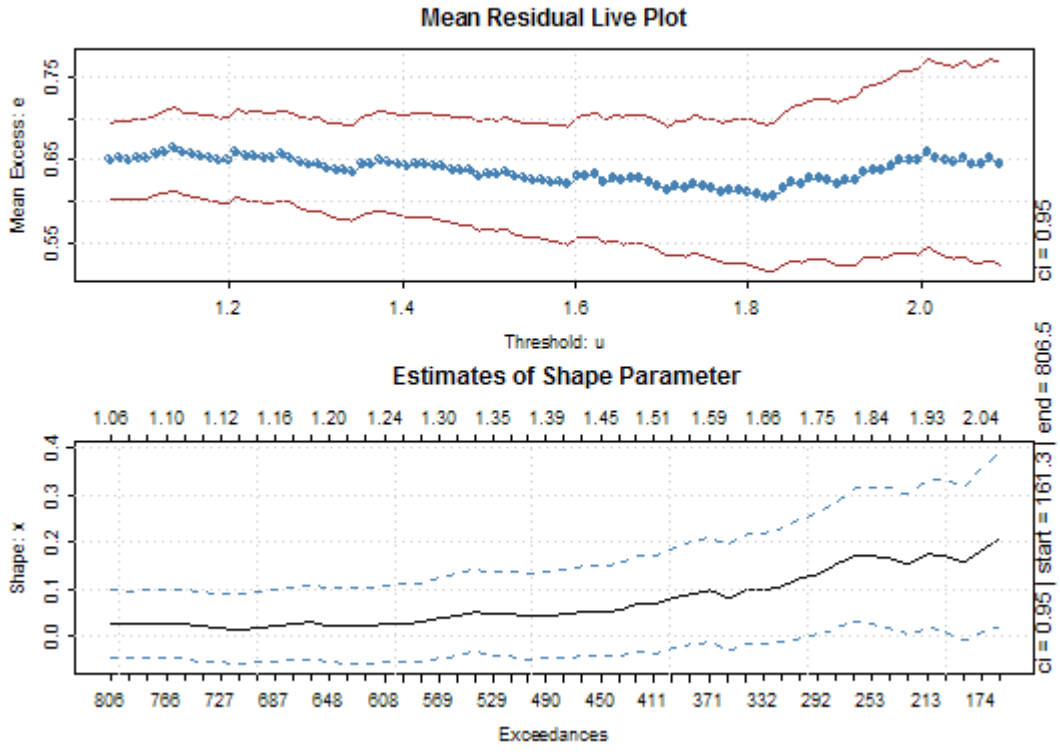
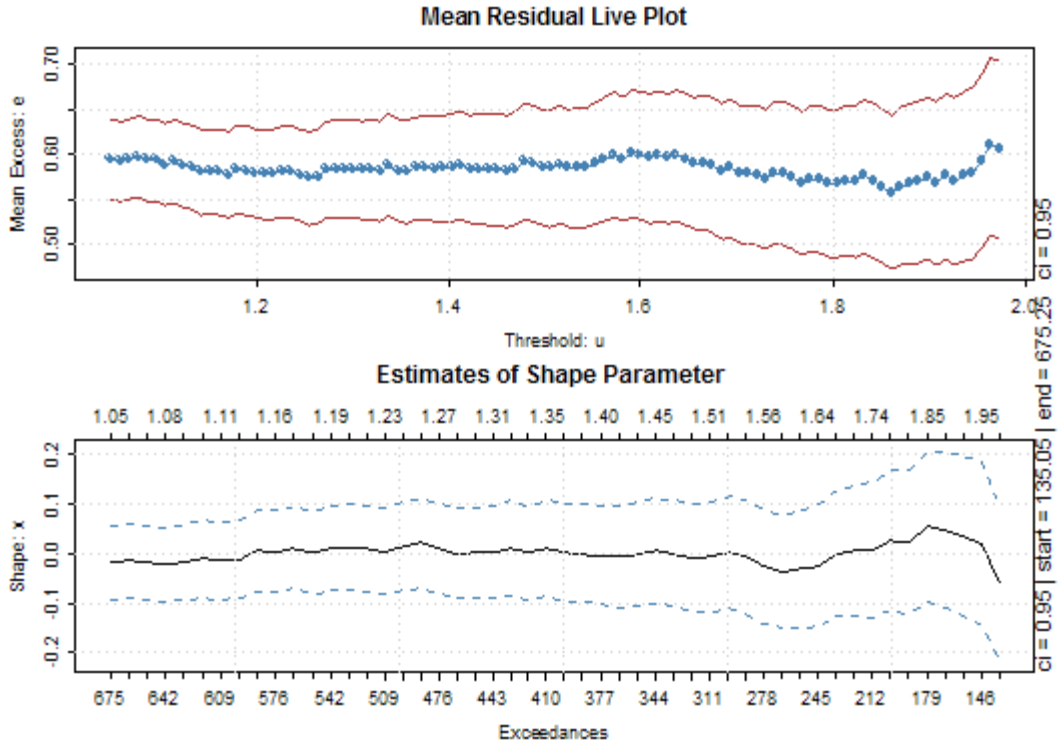


Figure 6 Mean Residual Life & Shape Parameter Estimates Plot - Brent

Right Tail



Left Tail

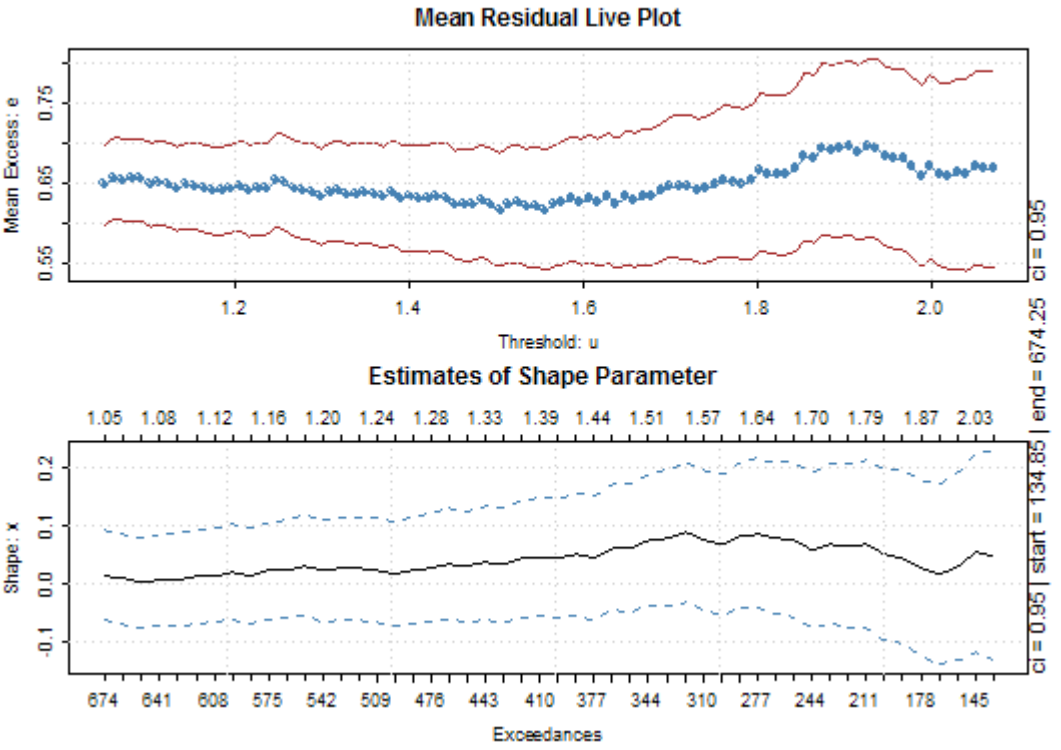
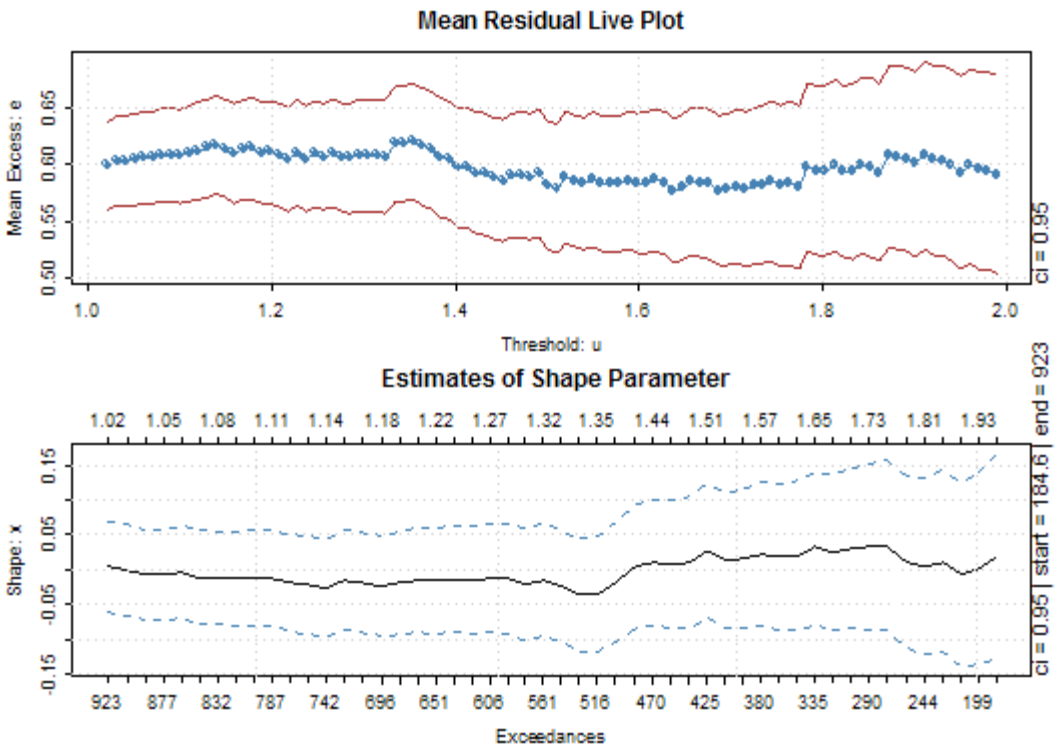


Figure 7 Mean Residual Life & Shape Parameter Estimates Plot – Heating Oil

Right Tail



Left Tail

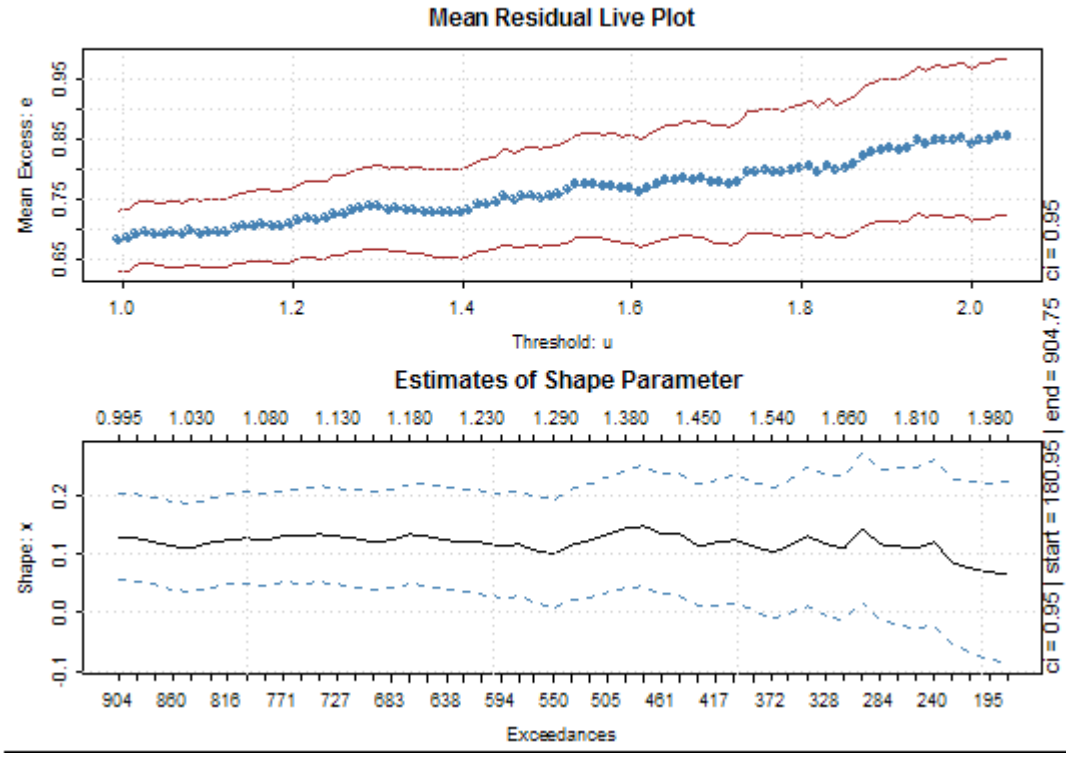
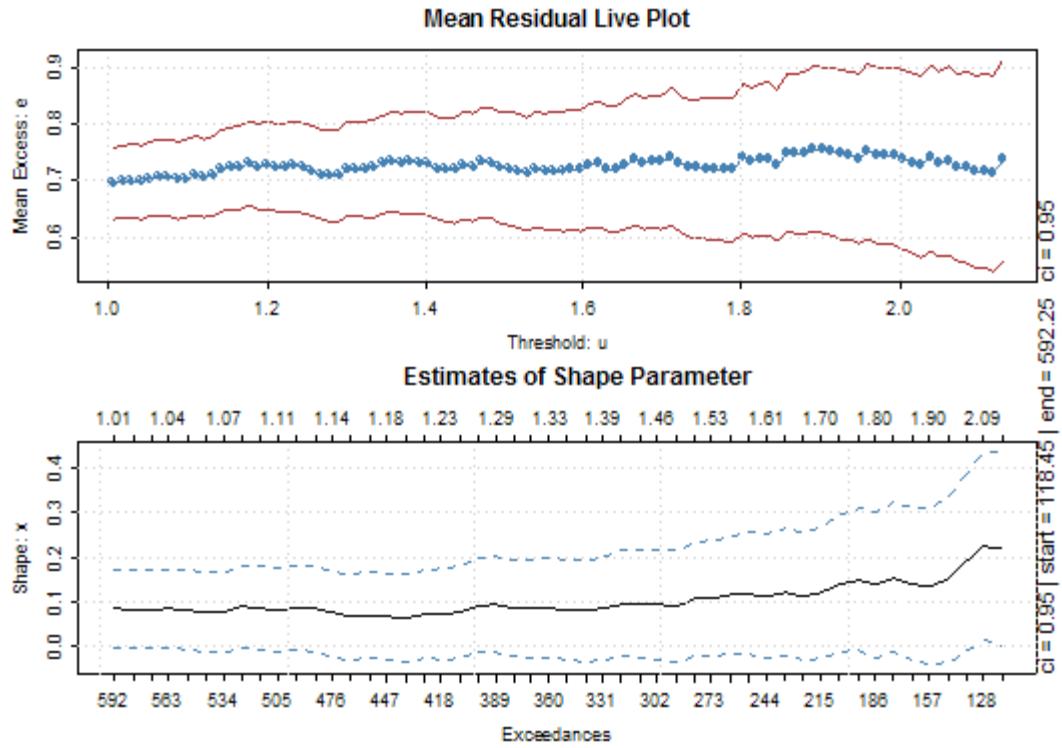


Figure 8 Mean Residual Life & Shape Parameter Estimates Plot – Natural Gas

Right Tail



Left Tail

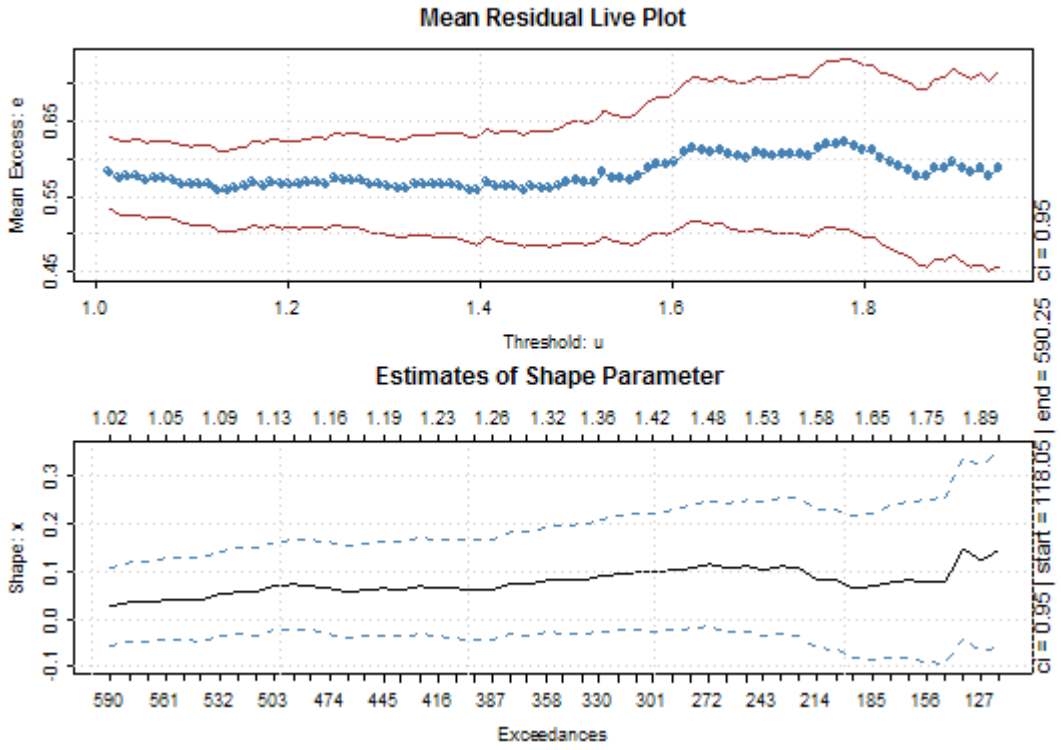


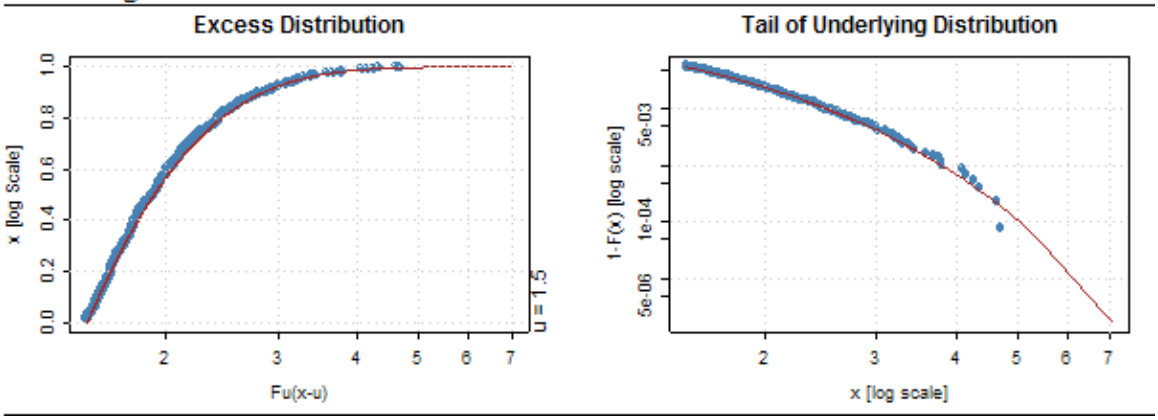
Table 6 Maximum Likelihood Parameter Estimation

	CL	CO	HO	NG
Right Tails				
# of Obs.	3,268	2,701	3,692	2,369
u	1.50	1.60	1.80	1.90
Exceedances	357	262	252	159
% of Exceedances	10.9	9.7	6.8	6.7
$\hat{\xi}$	-0.022	-0.029	0.015	0.140
(s.e.)	*0.055*	*0.068*	*0.065*	*0.078*
$\hat{\beta}$	0.599	0.614	0.585	0.643
(s.e.)	*0.046*	*0.056*	*0.053*	*0.071*
Left Tails				
# of Obs.	3,226	2,697	3,619	2,361
u	1.85	1.85	1.65	1.62
Exceedances	244	184	320	193
% of Exceedances	7.6	6.8	8.8	8.2
$\hat{\xi}$	0.160	0.032	0.116	0.063
(s.e.)	*0.074*	*0.065*	*0.066*	*0.065*
$\hat{\beta}$	0.523	0.658	0.690	0.576
(s.e.)	*0.051*	*0.064*	*0.060*	*0.056*

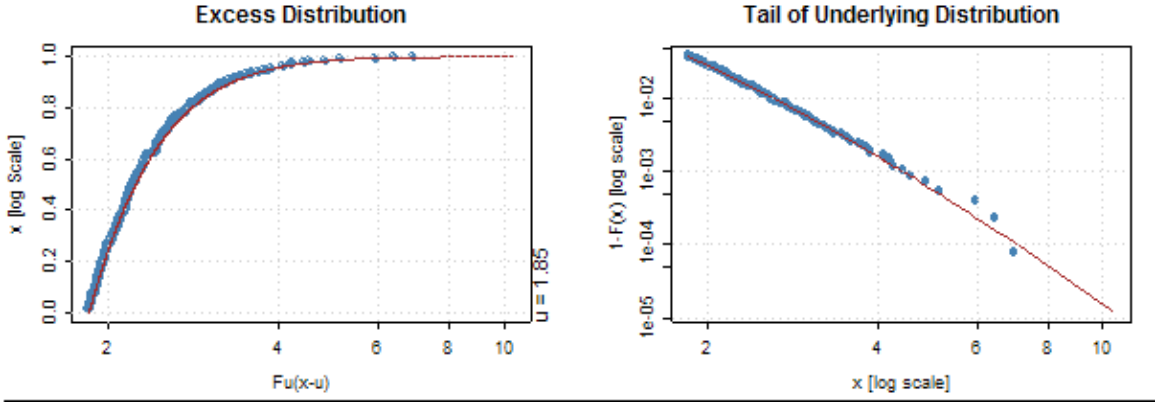
** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Figure 9 Excess Distribution Functions & Survival Functions

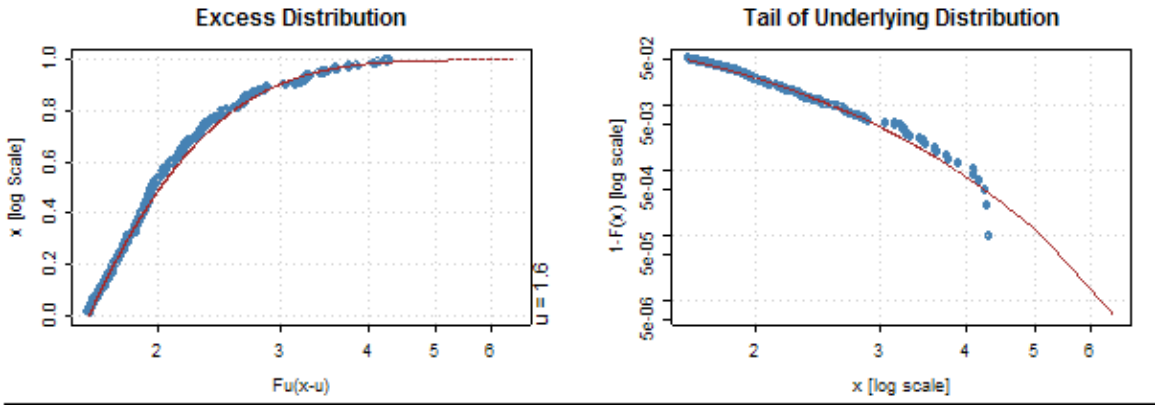
WTI - Right Tail



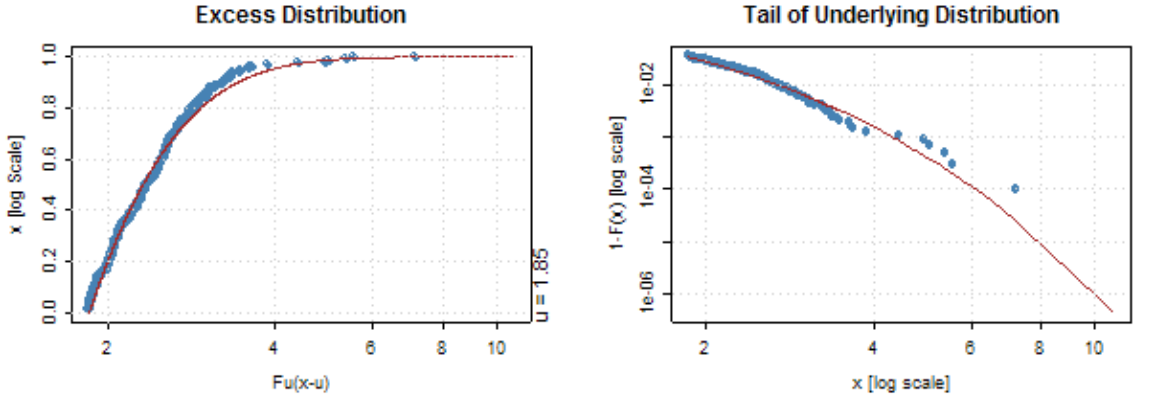
WTI - Left Tail



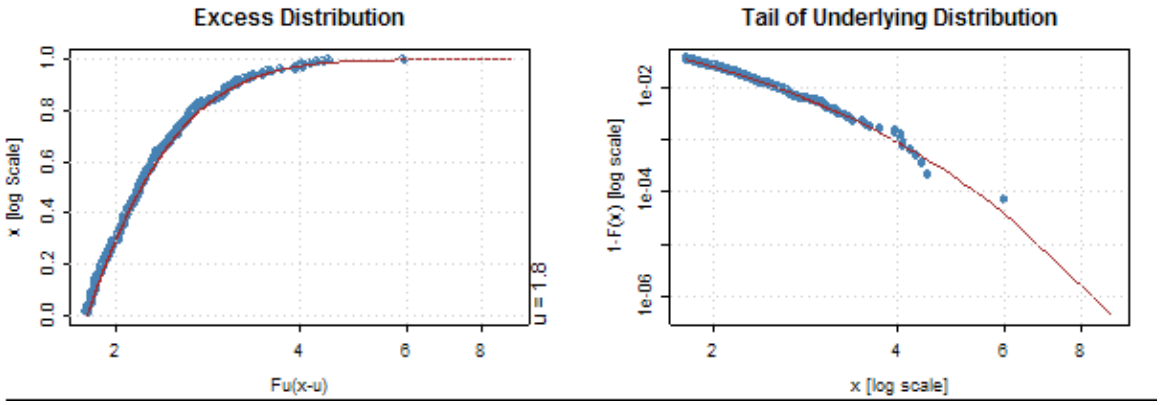
Brent - Right Tail



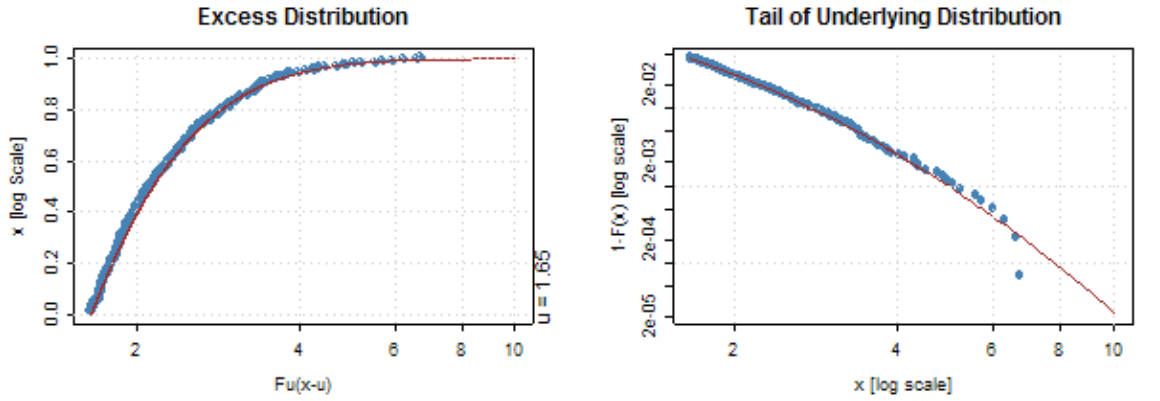
Brent - Left Tail



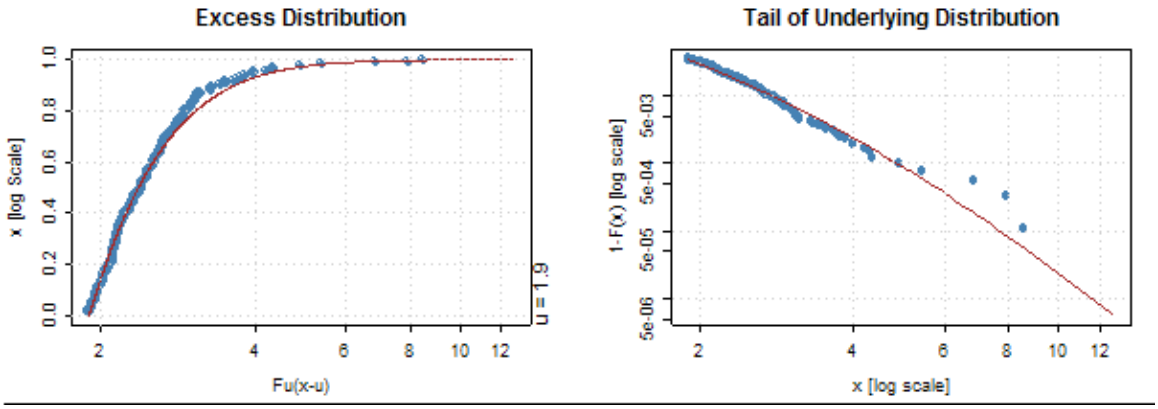
Heating Oil - Right Tail



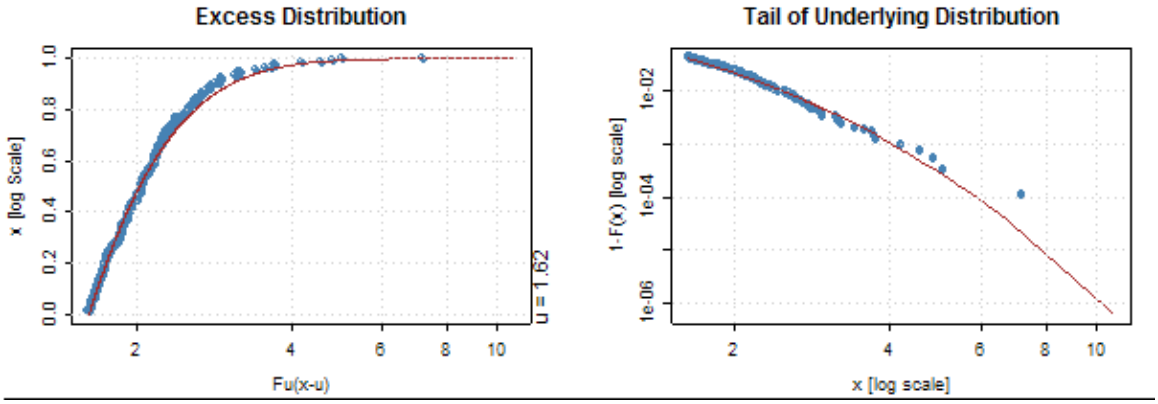
Heating Oil - Left Tail



Natural Gas - Right Tail



Natural Gas - Left Tail



3.4. Dynamic Backtest

In order to assess the accuracy of the conditional EVT approach and alternative methods of computing VaR, we backtested the methods on the four energy futures return series using the following procedure. The competing risk models included in this study are: the Normal model, the Student-*t* model with three degree of freedom, the model based on the historical simulation, the unconditional Extreme Value model, the RiskMetrics model, the conditional Normal model, the conditional Student-*t* model and the conditional Extreme Value model. For all models, we utilize a rolling sample of 1,000 observations to forecast VaR and ES for the significance level of 95%, 99%, 99.5% and 99.9%. The dynamic nature of this process allows us to capture the time-varying characteristics of the data in different time periods.

One question that arises during the process is the order of p and q of the conditional volatility model. Ideally, different p and q should be chosen for each backtest sample period based on an in-sample model selection criterion. However, we decided to use a GARCH (*1, 1*) model for the following two reasons. First, the optimization algorithm used in the R-project failed to converge at higher values of p and q. Therefore, computational constraints prevented us from specifying GARCH lags in each backtest sample period. Second, as Degiannakis and Xekalaki (2007) showed, the best-performing volatility forecasting model cannot be selected using any in-sample model selection criterion. Similarly, So and Yu (2006) demonstrated that the best fitted model according to Akaike's (1974) Information Criterion (AIC) and Schwarz's (1978) Bayesian Criterion (SBC) does not necessarily produce a more accurate VaR estimate. Therefore, instead of

selecting a higher order GARCH model using a model selection criterion, we use the simple GARCH(*1,1*) model.

Another important consideration during the backtest procedure is to choose an appropriate threshold value for the Extreme Value modeling. In such a long backtest period, it is not feasible to visually examine the mean excess plot for each backtest sample to determine the threshold value for each day and choose a different value for *u*. Therefore, in this chapter, we set the number of exceedances equal to 100 (10% of each backtest sample) as supported by the simulation study conducted by McNeil and Frey (2000) and Marimoutou, Raggad and Trabelsi (2008). As a result, observations above the 90^{th} percentile for the right tail and observations below the 10^{th} percentile for the left tail are estimated using Extreme Value theory. Table 7 summarizes the Value-at-Risk violation ratios at the significance levels of 5%, 1%, 0.5% and 0.1%. The competing models are compared and ranked by the absolute deviation from the expected violation ratios. The unconditional and conditional coverage tests results are shown in Table 8 and 9 respectively. The null hypothesis of Kupiec's (1995) unconditional coverage test assumes that the probability of occurrence of the violations equals the expected significance level. A *p*-value above 5% for the unconditional coverage test indicates that the number of violations is statistically equal to the expected one. The null hypothesis of the conditional coverage test says that the probability of occurrence of the violations equals the expected significance level and the violations are independently distributed through time. A *p*-value below 5% for the conditional coverage test shows that the model generates too few or too many clustered violations.

Table 7 VaR Violation Ratios & Model Ranking

WTI								
α	**5.00%**		**1.00%**		**0.50%**		**0.10%**	
Right Tails								
Normal	4.46%	7	1.60%	8	1.16%	8	0.78%	8
Student-*t*	5.42%	4	1.07%	2	0.66%	2	0.15%	1
HS	5.30%	3	1.13%	3	0.67%	4	0.18%	3
EVT	5.28%	2	1.20%	4	0.73%	5	0.20%	4
Risk Metrics	5.11%	1	1.57%	7	0.95%	7	0.44%	7
Conditional Normal	4.51%	5	1.40%	6	0.82%	6	0.40%	6
Conditional *t*	5.59%	8	0.64%	5	0.33%	3	0.04%	2
Conditional EVT	5.53%	6	1.02%	1	0.56%	1	0.20%	4
Left Tails								
Normal	5.13%	1	2.02%	8	1.49%	8	1.02%	8
Student-*t*	6.13%	7	1.36%	3	0.86%	5	0.20%	3
HS	5.59%	4	1.46%	5	0.78%	3	0.20%	3
EVT	5.84%	5	1.38%	4	0.84%	4	0.25%	5
Risk Metrics	6.01%	6	1.98%	7	1.22%	7	0.60%	7
Conditional Normal	5.46%	2	1.60%	6	1.06%	6	0.49%	6
Conditional *t*	6.66%	8	0.80%	2	0.47%	1	0.05%	1
Conditional EVT	5.46%	2	0.86%	1	0.47%	1	0.16%	2
Brent								
α	**5.00%**		**1.00%**		**0.50%**		**0.10%**	
Right Tails								
Normal	4.57%	5	1.43%	7	1.05%	7	0.66%	8
Student-*t*	5.89%	8	0.98%	1	0.55%	1	0.05%	2
HS	5.43%	6	1.21%	4	0.73%	4	0.18%	4
EVT	5.53%	7	1.11%	3	0.70%	3	0.23%	5
Risk Metrics	5.41%	4	1.46%	8	1.05%	7	0.55%	7
Conditional Normal	4.64%	3	1.23%	5	0.75%	5	0.36%	6
Conditional *t*	5.23%	2	0.64%	6	0.20%	6	0.02%	3
Conditional EVT	5.09%	1	0.93%	2	0.43%	2	0.09%	1
Left Tails								
Normal	5.62%	3	2.16%	7	1.61%	8	0.82%	8
Student-*t*	6.75%	8	1.46%	5	0.73%	4	0.14%	2
HS	5.84%	4	1.43%	4	0.82%	5	0.18%	4
EVT	6.00%	5	1.34%	3	0.68%	3	0.23%	5
Risk Metrics	6.00%	5	2.18%	8	1.52%	7	0.57%	7
Conditional Normal	5.39%	2	1.82%	6	1.09%	6	0.43%	6
Conditional *t*	6.43%	7	0.89%	1	0.36%	2	0.07%	1
Conditional EVT	5.32%	1	0.80%	2	0.50%	1	0.16%	3

Heating Oil

α	**5.00%**		**1.00%**		**0.50%**		**0.10%**	
Right Tails								
Normal	4.85%	1	1.57%	7	1.09%	7	0.62%	8
Student-*t*	5.83%	6	1.05%	1	0.48%	1	0.11%	1
HS	5.97%	7	1.31%	4	0.67%	3	0.24%	4
EVT	6.04%	8	1.39%	5	0.71%	4	0.29%	5
Risk Metrics	5.48%	4	1.81%	8	1.20%	8	0.48%	7
Conditional Normal	4.61%	3	1.41%	6	0.87%	6	0.41%	6
Conditional *t*	5.72%	5	0.71%	3	0.24%	5	0.02%	2
Conditional EVT	5.29%	2	1.14%	2	0.65%	2	0.22%	3
Left Tails								
Normal	4.56%	4	2.06%	8	1.58%	8	1.06%	8
Student-*t*	5.56%	5	1.57%	5	0.92%	5	0.29%	5
HS	5.74%	6	1.27%	4	0.79%	4	0.19%	3
EVT	5.75%	7	1.24%	2	0.73%	3	0.27%	4
Risk Metrics	4.90%	1	1.93%	7	1.49%	7	0.84%	7
Conditional Normal	4.74%	2	1.82%	6	1.36%	6	0.73%	6
Conditional *t*	5.94%	8	1.25%	3	0.62%	1	0.14%	1
Conditional EVT	5.28%	3	1.06%	1	0.63%	2	0.14%	1

Natural Gas

α	**5.00%**		**1.00%**		**0.50%**		**0.10%**	
Right Tails								
Normal	4.32%	6	1.77%	7	1.18%	7	0.72%	7
Student-*t*	5.52%	5	1.07%	4	0.64%	3	0.11%	1
HS	5.15%	1	1.10%	5	0.67%	5	0.11%	1
EVT	5.28%	3	0.99%	1	0.64%	3	0.13%	5
Risk Metrics	5.79%	7	2.25%	8	1.61%	8	0.75%	8
Conditional Normal	5.17%	2	1.69%	6	1.15%	6	0.46%	6
Conditional *t*	6.33%	8	0.99%	1	0.40%	2	0.08%	4
Conditional EVT	5.33%	4	1.02%	3	0.43%	1	0.11%	1
Left Tails								
Normal	4.32%	7	1.34%	6	0.99%	8	0.59%	8
Student-*t*	5.49%	6	0.88%	3	0.54%	1	0.05%	2
HS	5.31%	3	0.91%	1	0.54%	1	0.16%	3
EVT	5.36%	4	0.88%	3	0.46%	3	0.16%	3
Risk Metrics	4.88%	1	1.69%	8	0.86%	7	0.27%	6
Conditional Normal	4.21%	8	1.29%	5	0.72%	5	0.32%	7
Conditional *t*	5.36%	4	0.62%	7	0.21%	6	0.03%	5
Conditional EVT	5.15%	2	1.10%	2	0.59%	4	0.08%	1

** All models are ranked based on the deviation from the expected significance levels. The best models are highlighted by rectangles.*

Table 8 Unconditional Coverage Test

WTI				
α	**5%**	**1%**	**0.5%**	**0.1%**
Right Tails				
Normal	0.061*	0.000	0.000	0.000
Student-*t*	0.155*	0.587*	0.120*	0.317*
HS	0.319*	0.349*	0.084*	0.085*
EVT	0.349*	0.146*	0.025	0.039
Risk Metrics	0.700*	0.000	0.000	0.000
Conditional Normal	0.093*	0.005	0.002	0.000
Conditional *t*	0.050*	0.004	0.053*	0.086*
Conditional EVT	0.075*	0.887*	0.509*	0.039*
Left Tails				
Normal	0.655*	0.000	0.000	0.000
Student-*t*	0.000	0.010*	0.001	0.039
HS	0.050*	0.001	0.006	0.039
EVT	0.005	0.007	0.001	0.002
Risk Metrics	0.001	0.000	0.000	0.000
Conditional Normal	0.123*	0.000	0.000	0.000
Conditional *t*	0.000	0.124*	0.776*	0.244*
Conditional EVT	0.123*	0.269*	0.776*	0.171*
Brent				
α	**5%**	**1%**	**0.5%**	**0.1%**
Right Tails				
Normal	0.185*	0.007	0.000	0.000
Student-*t*	0.008	0.881*	0.672*	0.200*
HS	0.192*	0.185*	0.045	0.124*
EVT	0.116*	0.455*	0.070*	0.022
Risk Metrics	0.216*	0.004	0.000	0.000
Conditional Normal	0.266*	0.143*	0.028	0.000
Conditional *t*	0.488*	0.009	0.002	0.050*
Conditional EVT	0.777*	0.648*	0.513*	0.847*
Left Tails				
Normal	0.066*	0.000	0.000	0.000
Student-*t*	0.000	0.004	0.045	0.469*
HS	0.012	0.007	0.006	0.124*
EVT	0.003	0.031	0.105*	0.022
Risk Metrics	0.003	0.000	0.000	0.000
Conditional Normal	0.242*	0.000	0.000	0.000
Conditional *t*	0.000	0.442*	0.178*	0.479*
Conditional EVT	0.334*	0.158*	0.998*	0.253*

Heating Oil				
α	**5%**	**1%**	**0.5%**	**0.1%**
Right Tails				
Normal	0.577*	0.000	0.000	0.000
Student-*t*	0.003	0.718*	0.779*	0.788*
HS	0.001	0.016	0.076*	0.003
EVT	0.000	0.003	0.024	0.000
Risk Metrics	0.084*	0.000	0.000	0.000
Conditional Normal	0.150*	0.002	0.000	0.000
Conditional *t*	0.010	0.016	0.001	0.008
Conditional EVT	0.292*	0.272*	0.107*	0.008
Left Tails				
Normal	0.106*	0.000	0.000	0.000
Student-*t*	0.044	0.000	0.000	0.000
HS	0.009	0.040	0.002	0.044
EVT	0.007	0.069*	0.016	0.000
Risk Metrics	0.702*	0.000	0.000	0.000
Conditional Normal	0.334*	0.000	0.000	0.000
Conditional *t*	0.001	0.053*	0.200*	0.315*
Conditional EVT	0.319*	0.627*	0.148*	0.315*
Natural Gas				
α	**5%**	**1%**	**0.5%**	**0.1%**
Right Tails				
Normal	0.050*	0.000	0.000	0.000
Student-*t*	0.150*	0.662*	0.235*	0.890*
HS	0.684*	0.550*	0.162*	0.890*
EVT	0.436*	0.959*	0.235*	0.532*
Risk Metrics	0.031	0.000	0.000	0.000
Conditional Normal	0.630*	0.000	0.000	0.000
Conditional *t*	0.000	0.959*	0.380*	0.695*
Conditional EVT	0.355*	0.910*	0.528*	0.890*
Left Tails				
Normal	0.050*	0.047	0.000	0.000
Student-*t*	0.172*	0.469*	0.758*	0.325*
HS	0.394*	0.580*	0.758*	0.281*
EVT	0.318*	0.469*	0.697*	0.281*
Risk Metrics	0.732*	0.000	0.005	0.007
Conditional Normal	0.023*	0.092*	0.070*	0.001
Conditional *t*	0.318*	0.011	0.005	0.092*
Conditional EVT	0.684*	0.550*	0.450*	0.695*

** A p-value above 5% for the unconditional coverage test indicates that the number of violations is statistically equal to the expected one.*

Table 9 Conditional Coverage Test

WTI				
α	**5%**	**1%**	**0.5%**	**0.1%**
Right Tails				
Normal	0.001	0.000	0.000	0.000
Student-*t*	0.009	0.000	0.006	0.311*
HS	0.005	0.000	0.005	0.083*
EVT	0.005	0.000	0.000	0.038
Risk Metrics	0.607*	0.000	0.000	0.000
Conditional Normal	0.045	0.001	0.001	0.000
Conditional *t*	0.041	0.003	0.049	0.086*
Conditional EVT	0.068*	0.275*	0.371*	0.038
Left Tails				
Normal	0.000	0.000	0.000	0.000
Student-*t*	0.000	0.000	0.000	0.038
HS	0.000	0.000	0.000	0.038
EVT	0.000	0.000	0.000	0.002
Risk Metrics	0.000	0.000	0.000	0.000
Conditional Normal	0.050*	0.000	0.000	0.000
Conditional *t*	0.000	0.074*	0.108*	0.243*
Conditional EVT	0.068*	0.152*	0.561*	0.167*

Brent				
α	**5%**	**1%**	**0.5%**	**0.1%**
Right Tails				
Normal	0.065*	0.004	0.000	0.000
Student-*t*	0.003	0.427*	0.501*	0.199*
HS	0.079*	0.055*	0.020	0.121*
EVT	0.041	0.345*	0.028	0.022
Risk Metrics	0.120*	0.004	0.000	0.000
Conditional Normal	0.244*	0.046	0.021	0.000
Conditional *t*	0.219*	0.008	0.002	0.050*
Conditional EVT	0.374*	0.334*	0.438*	0.830*
Left Tails				
Normal	0.000	0.000	0.000	0.000
Student-*t*	0.000	0.000	0.020	0.461*
HS	0.000	0.000	0.003	0.121*
EVT	0.000	0.001	0.039	0.022
Risk Metrics	0.001	0.000	0.000	0.000
Conditional Normal	0.169*	0.000	0.000	0.000
Conditional *t*	0.000	0.229*	0.164*	0.477*
Conditional EVT	0.207*	0.010	0.631*	0.249*

Heating Oil

α	**5%**	**1%**	**0.5%**	**0.1%**
Right Tails				
Normal	0.000	0.000	0.000	0.000
Student-*t*	0.000	0.000	0.000	0.764*
HS	0.000	0.000	0.005	0.003
EVT	0.000	0.000	0.000	0.000
Risk Metrics	0.002	0.000	0.000	0.000
Conditional Normal	0.003	0.001	0.000	0.000
Conditional *t*	0.000	0.002	0.001	0.008
Conditional EVT	0.006	0.116*	0.051*	0.008
Left Tails				
Normal	0.000	0.000	0.000	0.000
Student-*t*	0.000	0.000	0.000	0.000
HS	0.000	0.002	0.000	0.043
EVT	0.000	0.003	0.010	0.000
Risk Metrics	0.175*	0.000	0.000	0.000
Conditional Normal	0.101*	0.000	0.000	0.000
Conditional *t*	0.001	0.032*	0.144*	0.308*
Conditional EVT	0.227*	0.172*	0.106*	0.308*

Natural Gas

α	**5%**	**1%**	**0.5%**	**0.1%**
Right Tails				
Normal	0.001	0.000	0.000	0.000
Student-*t*	0.026	0.299*	0.188*	0.863*
HS	0.025	0.256*	0.129*	0.863*
EVT	0.015	0.382*	0.188*	0.524*
Risk Metrics	0.028	0.000	0.000	0.000
Conditional Normal	0.561*	0.000	0.000	0.000
Conditional *t*	0.000	0.382*	0.343*	0.689*
Conditional EVT	0.263*	0.366*	0.460*	0.863*
Left Tails				
Normal	0.013	0.016	0.000	0.000
Student-*t*	0.046	0.025	0.090*	0.324*
HS	0.031	0.251*	0.571*	0.276*
EVT	0.093	0.203*	0.573*	0.276*
Risk Metrics	0.573*	0.000	0.004	0.007
Conditional Normal	0.002	0.043	0.054*	0.001
Conditional *t*	0.281*	0.010	0.005	0.092*
Conditional EVT	0.1368	0.256*	0.359*	0.689*

** A p-value above 5% for the conditional coverage test indicates that the forecasting ability of the VaR model is adequate.*

Table 10 VaR Backtest Summary Table

WTI	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC
α												
α	**5%**			**1%**			**0.5%**			**0.1%**		
Right Tails												
Normal	7	**		8			8			8		
Student-*t*	4	**		2	**		2	**		1	**	**
HS	3	**		3	**		4	**		3	**	**
EVT	2	**		4	**		5			4		
Risk Metrics	1	**	**	7			7			7		
Conditional Normal	5	**		6			6			6		
Conditional *t*	8			5			3	**		2	**	**
Conditional EVT	6	**	**	1	**	**	1	**	**	4		
Left Tails												
Normal	1	**		8			8			8		
Student-*t*	7			3			5			3		
HS	4			5			3			3		
EVT	5			4			4			5		
Risk Metrics	6			7			7			7		
Conditional Normal	2	**	**	6			6			6		
Conditional *t*	8			2	**	**	1	**	**	1	**	**
Conditional EVT	2	**	**	1	**	**	1	**	**	2	**	**

Brent	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC
α	**5%**			**1%**			**0.5%**			**0.1%**		
Right Tails												
Normal	5	**	**	7			7			8		
Student-*t*	8			1	**	**	1	**	**	2	**	**
HS	6	**	**	4	**	**	4			4	**	**
EVT	7	**		3	**	**	3	**		5		
Risk Metrics	4	**	**	8			7			7		
Conditional Normal	3	**	**	5	**		5			6		
Conditional *t*	2	**	**	6			6			3	**	**
Conditional EVT	1	**	**	2	**	**	2	**	**	1	**	**
Left Tails												
Normal	3	**		7			8			8		
Student-*t*	8			5			4			2	**	**
HS	4			4			5			4	**	**
EVT	5			3			3	**		5		
Risk Metrics	5			8			7			7		
Conditional Normal	2	**	**	6			6			6		
Conditional *t*	7			1	**	**	2	**	**	1	**	**
Conditional EVT	1	**	**	2	**		1	**	**	3	**	**

Heating Oil	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC
α	**5%**			**1%**			**0.5%**			**0.1%**		
Right Tails												
Normal	1	**		7			7			8		
Student-*t*	6			1	**		1	**		1	**	**
HS	7			4			3	**		4		
EVT	8			5			4			5		
Risk Metrics	4	**		8			8			7		
Conditional Normal	3	**		6			6			6		
Conditional *t*	5			3			5			2		
Conditional EVT	2	**		2	**	**	2	**	**	3		
Left Tails												
Normal	4	**		8			8			8		
Student-*t*	5			5			5			5		
HS	6			4			4			3		
EVT	7			2	**		3			4		
Risk Metrics	1	**	**	7			7			7		
Conditional Normal	2	**	**	6			6			6		
Conditional *t*	8			3	**		1	**	**	1	**	**
Conditional EVT	3	**	**	1	**	**	2	**	**	1	**	**

Natural Gas	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC
α	**5%**			**1%**			**0.5%**			**0.1%**		
Right Tails												
Normal	6			7			7			7		
Student-*t*	5	**		4	**	**	3	**	**	1	**	**
HS	1	**		5	**	**	5	**	**	1	**	**
EVT	3	**		1	**	**	3	**	**	5	**	**
Risk Metrics	7			8			8			8		
Conditional Normal	2	**	**	6			6			6		
Conditional *t*	8			1	**	**	2	**	**	4	**	**
Conditional EVT	4	**	**	3	**	**	1	**	**	1	**	**
Left Tails												
Normal	7			6			8			8		
Student-*t*	6	**		3	**		1	**	**	2	**	**
HS	3	**		1	**	**	1	**	**	3	**	**
EVT	4	**	**	3	**	**	3	**	**	3	**	**
Risk Metrics	1	**	**	8			7			6		
Conditional Normal	8			5	**		5	**	**	7		
Conditional *t*	4	**	**	7			6			5	**	**
Conditional EVT	2	**	**	2	**	**	4	**	**	1	**	**

** Kupiec denotes the unconditional coverage test*

*** CC denotes the conditional coverage test*

**** All models are ranked based on the deviation from the expected significance levels. The best models are highlighted by rectangles.*

Table 10 summarizes the ranking of the violation ratios along with the two statistical tests for a total of 32 cases (four energy futures, four significance levels and two tails). The relative performances of the competing models is judged by a success rate, which is defined as the probability that a model is ranked among the top two in each case and passes both unconditional and conditional coverage tests. The results can be summarized as follows:

i. The GARCH-EVT model performs the best among the eight competing models. The GARCH-EVT model is ranked among the top two and passes two statistical tests 75% of the time. This success rate is much higher than the second highest rate of 38% based on the conditional Student-*t* model and the unconditional Student-*t* model. In addition, the GARCH-EVT model performs better for a long position (left tail) than a short position. The success rate for the left tail is 88% as compared to the success rate of 63% for the right tail. This result is encouraging as most portfolios have a long bias.

ii. The performance of the fail-tailed Student-*t* models (both conditional and unconditional) is such that they are ranked second with a success rate of 38%. However, the conditional Student-*t* model is relatively better at modeling the left tail (success rate is 50%) and the unconditional Student-*t* model is relatively superior at forecasting VaR in the right tail (success rate is 44%).

iii. The Normal distribution based models, conditional Normal, RiskMetrics and unconditional Normal, perform very poorly in terms of forecasting accurate VaR for both tails. The success rate is zero for the unconditional Normal model and 9% (1 out of 32 cases) for the RiskMetrics model and 13% for the conditional Normal

model. This result confirms our belief that the widely used Normal distribution based VaR models tend to underestimate the true risk (also supported by Table 7).

iv. Both the unconditional EVT models and the historical simulation models perform very poorly in forecasting VaR at various significance levels.

As discussed above, Expected Shortfall (ES) is a superior measurement of the true risk. We utilize a two stage approach in this chapter to backtest the ES estimates. The first stage selects models that pass both the unconditional and conditional coverage tests. In the second stage, the value of the loss function as defined in Eq. 24 for the selected models is calculated and presented in Table 11. All competing models are ranked by minimizing the loss function. The results are similar to what we observed above. The GARCH-EVT approach is ranked among the two best models 25 out of the 32 cases (78% probability), followed by the conditional Student-*t* model (31% probability). In general, the conditional models tend to pass both unconditional and conditional coverage tests more often and produce smaller values of the loss function.

Based on the backtests conducted, we conclude that the Conditional EVT model outperforms the competing models by a wild margin. Table 12 provides more details on the calculated VaR and ES based on the GARCH-EVT model using 99% significance level as an example. The descriptive statistics of the last 3,000 out-of-sample VaR and ES forecasts are presented in the table. For example, a $100 short position in WTI could lose up to $6.4 99% percent of the time. In general, average VaR and ES is higher for a long

Table 11 Backtest Expected Shortfall using Loss Function Approach

WTI

	5%		1%		0.5%		0.1%	
Right Tails								
Normal	-		-		-		-	
Student-t	-		-		-		7,186	3
HS	-		-		-		4,980	1
EVT	-		-		-		-	
Risk Metrics	1,931	1	-		-		-	
Conditional Normal	-		-		-		-	
Conditional t	-		-		-		6,886	2
Conditional EVT	2,029	2	2,994	1	3,383	1	-	
Left Tails								
Normal	-		-		-		-	
Student-t	-		-		-		-	
HS	-		-		-		-	
EVT	-		-		-		-	
Risk Metrics	-		-		-		-	
Conditional Normal	1,946	1	-		-		-	
Conditional t	-		3,647	2	4,437	2	6,832	2
Conditional EVT	2,184	2	3,332	1	3,819	1	4,999	1

Brent

	5%		1%		0.5%		0.1%	
Right Tails								
Normal	1,496	3	-		-		-	
Student-t	-		2,756	4	3,349	2	5,146	4
HS	1,541	5	2,427	2	-		3,512	2
EVT	-		2,433	3	-		-	
Risk Metrics	1,382	1	-		-		-	
Conditional Normal	1,444	2	-		-		-	
Conditional t	1,605	6	-		-		5,056	3
Conditional EVT	1,503	4	2,236	1	2,544	1	3,270	1
Left Tails								
Normal	-		-		-		-	
Student-t	-		-		-		5,121	3
HS	-		-		-		5,544	4
EVT	-		-		-		-	
Risk Metrics	-		-		-		-	
Conditional Normal	1,402	1	-		-		-	
Conditional t	-		2,664	1	3,247	2	5,002	2
Conditional EVT	1,564	2	-		2,709	1	3,478	1

Heating Oil

	5%		1%		0.5%		0.1%	
Right Tails								
Normal	-		-		-		-	
Student-t	-		-		-		8,123	1
HS	-		-		-		-	
EVT	-		-		-		-	
Risk Metrics	-		-		-		-	
Conditional Normal	-		-		-		-	
Conditional t	-		-		-		-	
Conditional EVT	-		3,381	1	3,792	1	-	
Left Tails								
Normal	-		-		-		-	
Student-t	-		-		-		-	
HS	-		-		-		-	
EVT	-		-		-		-	
Risk Metrics	2,227	1	-		-		-	
Conditional Normal	2,267	2	-		-		-	
Conditional t	-		-		5,050	1	7,781	2
Conditional EVT	2,571	3	4,248	1	5,081	2	7,448	1

Natural Gas

	5%		1%		0.5%		0.1%	
Right Tails								
Normal	-		-		-		-	
Student-t	2,413	2	4,075	4	4,951	4	7,609	5
HS	2,461	4	4,097	5	4,906	3	7,068	3
EVT	2,468	5	4,073	3	4,832	2	6,863	2
Risk Metrics	-		-		-		-	
Conditional Normal	2,175	1	-		-		-	
Conditional t	-		3,968	2	-		7,402	4
Conditional EVT	2,440	3	3,780	1	4,369	1	5,804	1
Left Tails								
Normal	-		-		-		-	
Student-t	2,391	4	4,061	4	4,935	4	7,592	5
HS	2,397	5	3,996	3	4,755	3	6,902	2
EVT	2,407	6	4,077	5	4,950	5	7,583	4
Risk Metrics	2,067	1	-		-		-	
Conditional Normal	-		2,840	1	3,100	1	-	
Conditional t	2,314	3	-		-		7,372	3
Conditional EVT	2,165	2	3,235	2	3,705	2	4,874	1

** All models are ranked based on the deviation from the expected significance levels. The best models are highlighted by rectangles.*

Table 12 Descriptive Statistics of 99% VaR and ES using the Conditional EVT approach

Right Tail	WTI	Brent	Heating Oil	Natural Gas
99% VaR				
Mean	5.9	5.7	6.0	10.0
Std. dev.	2.0	1.5	1.5	3.1
Max	18.8	14.1	18.0	34.3
Min	3.4	3.0	3.0	4.7
99% ES				
Mean	7.2	7.0	7.1	12.5
Std. dev.	2.6	2.0	1.9	4.1
Max	22.5	16.0	21.7	44.8
Min	4.4	3.6	3.6	5.9
Left Tail	**WTI**	**Brent**	**Heating Oil**	**Natural Gas**
99% VaR				
Mean	-6.4	-6.0	-6.5	-8.5
Std. dev.	1.9	1.6	1.9	2.8
Max	-18.5	-14.6	-23.3	-30.0
Min	-3.7	-3.3	-3.3	-3.9
99% ES				
Mean	-7.9	-7.3	-8.6	-10.3
Std. dev.	2.2	2.0	2.6	3.3
Max	-23.5	-17.2	-30.0	-35.2
Min	-4.5	-3.9	-3.9	-5.0

position than a short position. This is consistent with the negative skewness observed in Table 2. In general, Natural Gas appears to be riskier than other Energy futures with a higher VaR and ES for both tails. This result is consistent with the higher standard deviation recorded in Table 2.

4. Conclusion

We have tested the performance of eight competing models for calculating Value-at-Risk (VaR) and Expected Shortfall (ES). The models under study are the Normal model, the Student-*t* model, the model based on the historical simulation, the unconditional Extreme

Value model, the RiskMetrics model, the conditional Normal model, the conditional Student-*t* model and the conditional Extreme Value Theory model. Four energy futures series and four significance levels are considered in order to make sure the results are reliable. In addition, statistical tests on out-of-sample forecasting of VaR and ES are conducted to rank the models. Our results indicate that the GARCH-EVT approach outperforms the competing models in forecasting Value-at-Risk and Expected Shortfall by a wide margin. This approach provides a significant improvement over the widely used Normal distribution based VaR and ES models, which tend to underestimate the true risk and fail to provide statistically accurate VaR estimates. Our results are similar to Marimoutou, Raggad, and Trabelsi (2009), who illustrate that the conditional Extreme Value Theory and Filtered Historical Simulation approaches outperform the traditional methods. However, our results show that the GARCH-EVT approach is overwhelmingly better than the competing models, especially at lower significance levels (1%, 0.5%, and 0.1%).

CHAPTER THREE: MEASURING TIME-VARYING DEPENDENCE IN ENERGY FUTURES USING COPULAS

1. Introduction

Copulas, introduced by Sklar in 1959 as statistical functions which join together one-dimensional distributions to form multivariate distributions, have become a popular multivariate modeling tool in many fields such as actuarial science, biomedical studies, engineering and especially finance. During the past decade, we have witnessed an increasing number of financial applications of copula theory, mainly due to its flexibility in constructing a suitable joint distribution when facing non-normality in financial data. The key characteristic of copula models is the separation of the joint distribution of returns into two components: the marginal distributions and the dependence structure. The approach is designed to capture well-known stylized facts of financial returns using marginal distributions, leaving all of the information about the dependence structures to be estimated by copula models separately. Therefore, copula models allow for the construction of more flexible joint distributions than existing multivariate distributions.

The nature of the dependence structure between financial markets has very important implications in investment decision making process. It provides insights into portfolio risk management, portfolio diversification, pairs trading and exotic derivatives pricing, especially when returns are non-Normal and simple linear correlation models fail to capture the degree of association between assets. The early literature on the linkages between different asset returns mainly focused on using linear correlation as the measure

of dependence for elliptical variables. However, there is strong evidence that the univariate distributions of many financial variables are non-Normal and significantly fat-tailed, which rules out the use of the multivariate Normal distribution. Since the pioneering work of Embrechts *et al.* (1999), copula models have attracted increasing attention due to the models' ability to capture different patterns of dependence while allowing for flexible marginal distributions to capture the skewness and kurtosis in asset returns.

A number of recent empirical studies have discovered significant asymmetric dependence, in that returns are more dependent during market downturns than during market upturns. See Longin and Solnik (2001), Ang and Chen (2002), Patton (2006a), Michelis and Ning (2010), *etc*. In addition, most of these studies find that the dependence structure is not constant over time. Following Patton (2006a), we employ a time-varying copula-GARCH model to capture these two important characteristics of the dependence structure.

Most previous empirical studies mainly look at international stock markets and foreign exchange rate markets. Little attention has been paid to nonlinear dependence, especially tail dependence caused by extreme events, between energy futures markets. With the increasing integration of financial markets, financial innovations and ease of information flow among investors, energy markets are becoming more intertwined in recent years, suggesting the relevance of copula tools. As Alexander (2005) points out, the dependence between crude oil and natural gas futures prices is strong and cannot be modeled

correctly by a bivariate Normal distribution. Grégoire *et al.* (2008) fit various families of static bivariate copulas to crude oil and natural gas futures and conclude that the Student-*t* copula provides a much better fit based on the goodness-of-fit tests. Using energy futures, Fischer and Köck (2007) compare several construction schemes of multivariate copula models and also confirm that the Student-*t* copula outperform others. However, these papers mainly focus on the application of fitting static copula models to energy futures without a detailed analysis of the dependence structure and its implications on portfolio management. In the empirical part of this chapter, we employ a time-varying conditional copula method to study the dependence structure of energy futures. In addition, we also consider the impact of supply-demand fundamentals on the dynamic dependence between energy futures. To the best of our knowledge, this chapter is the first study of energy futures returns dynamics with macroeconomic factors as exogenous variables.

This chapter is organized as follows. Section 2 introduces copula theory and its extension to conditional time-varying copula functions. Section 3 discusses the dependence concepts under study. Section 4 presents examples of copula functions. Section 5 describes our adopted estimation procedure. Empirical results are presented and discussed in Section 6. Section 7 concludes the chapter.

2. Theory of Copula

What are copulas? According to Nelsen (2006, p.12), copulas can be explained from two points of view: “From one point a view, copulas are functions that join or ‘couple’ multivariate distribution functions to their one dimensional marginal distribution functions. Alternatively, copulas are multivariate distribution functions whose one-

dimensional margins are uniform on the interval (0, 1).” Therefore, 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.

2.1. Definitions and Basic Properties

For simplicity, we only focus on bivariate copulas in this chapter. The extension to the multivariate case is theoretically straightforward but computationally expensive. Following Nelsen (2006), we begin with a definition.

Definition of Copula *A 2-dimentional copula is a function C:* $[0, 1]^2 \rightarrow [0, 1]$ *with the following properties:*

1. *For every* $u \in [0, 1]$

$$C(0, u) = C(u, 0) = 0.$$

2. *For every* $u \in [0, 1]$

$$C(u, 1) = u \text{ and } C(1, u) = u.$$

3. *For every* $(u_1, u_2), (v_1, v_2) \in [0, 1] \times [0, 1]$ *with* $u_1 \leq v_1$ *and* $u_2 \leq v_2$*:*

$$C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0$$

Property 1 is also referred to as the grounded property of a copula. It says that the joint probability of both outcomes is zero if the marginal probability of any outcome is zero. Property 3 is the two-dimensional analogue of a non-decreasing one-dimensional function. A function with this feature is therefore called 2-increasing.

Sklar's (1959) Theorem *Let H be a 2-dimentional joint distribution function with marginal distributions F and G. There exists a copula C such that:*

$$H(x, y) = C(F(x), G(y)), \text{ for any } x, y.$$

If F and G are continuous, then C is unique. Conversely, if C is a copula and F and G are distribution functions, then the function H is a joint distribution function with marginal distributions F and G. The converse of Sklar's theorem is very useful in modeling multivariate distributions in finance. It implies that if we combine *two* different marginal distributions with any copula, we will have defined a valid bivariate distribution. This provides great flexibility when modeling a portfolio as we can use different marginal distributions for each asset class and use a copula to link them together.

Corollary *Let H be a 2-dimensional joint distribution function with continuous marginal distributions F , G and copula C, satisfying Sklar's theorem, then for* $u \in [0, 1]$ and $v \in [0, 1]$:

$$C(u, v) = H(F^{-1}(u), G^{-1}(v))$$

where F^{-1} and G^{-1} denote the inverses of F and G. This corollary provides a motivation for calling a copula a dependence structure as the copula links the quantiles of the two distributions rather than the original variables. Based on this corollary, one of the key properties of a copula is that the dependence structure is unaffected by a monotonically increasing transformation of the variables. For example, the same copula can be used for the joint distribution of (X, Y) as well as for the joint distribution of $(\ln X, \ln Y)$. The above Theorem and Corollary extend naturally from the bivariate case to the multivariate case.

2.2. Conditional Copula

Patton (2006a) extended and proved the validity of the Sklar's theorem for the conditional case defined below. Similar to the unconditional case, we have two variables X and Y. We introduce a conditioning vector W with a dimension of one. Following Patton (2006a), we also assume that the conditioning vector W is the same for both marginal distributions and the copula. The conditional bivariate distribution of $(X, Y) \mid W$ can be derived from the unconditional joint distribution of (X, Y, W) as follows:

$$F_{XY|W}(x, y \mid w) = f_W(w)^{-1} \cdot \frac{\partial F_{XYW}(x, y, w)}{\partial w} \text{ for } w \in \mathrm{W}.$$

Definition of Conditional Copula *The conditional copula of* $(X, Y) \mid W = w$, *where* $X \mid W = w \sim F_1(X|W)(\cdot|w)$ *and* $Y \mid W = w \sim F_1(Y|W)(\cdot|w)$, *is the conditional joint distribution function of* $U \equiv F_1(X|W)(X|w)$ *and* $V \equiv F_1(Y|W)(Y|w)$ *given* $W=w$.

The two variables U and V are known as the conditional "probability integral transforms" of X and Y given W. These variables have Uniform (0, 1) distributions, regardless of the original distribution of X and Y (Fisher (1932) and Rosenblatt (1952)). Patton (2002) shows that a conditional copula has all of the properties of an unconditional copula. The extension of Sklar's theorem to conditional distributions is presented below.

Sklar's Theorem for Conditional Copulas *Let* $F_1(X|W)(\cdot|w)$ *and* $F_1(Y|W)(\cdot|w)$ *be the conditional distribution of* $X \mid W = w$ *and* $Y \mid W = w$, $F_1(XY|W)(\cdot|w)$ *be the joint distribution of* $(X, Y) \mid W = w$, *and w be the support of W. Assume that* $F_1(X|W)(\cdot|w)$ *and*

$F_1(Y|W)(\cdot|w)$ *are continuous in x and u for all* $w \in W$. *Then there exists a unique conditional copula* $C(\cdot|w)$ *such that*

$$F_1(XY|W)(x, y|w) = C(F_1(X|W)(x|w), F_1(Y|W)(y|w)|w) = C(u, v)$$

$$\forall (x, y) \in \overline{\Re} \times \overline{\Re} \text{ and each } w \in W$$

where $u = F_1(X|W)(x|w)$ *and* $v = F_1(Y|W)(y|w)$ *are realizations of* $U \equiv F_1(X|W)(X|w)$ *and* $V \equiv F_1(Y|W)(Y|w)$ *given W=w. Conversely, if we let* $F_1(X|W)(\cdot|w)$ *and* $F_1(Y|W)(\cdot|w)$ *be the conditional distribution of X | W = w and Y | W = w, and* $C(\cdot|w)$ *be a family of conditional copulas that is measurable in w, then the function* $F_1(XY|W)(\cdot|w)$ *is a conditional bivariate distribution function with conditional marginal distributions* $F_1(X|W)(\cdot|w)$ *and* $F_1(Y|W)(\cdot|w)$.

3. Dependence Concepts

Correlation is probably the single most important concept in modern portfolio theory ever since Markowitz (1952) proved the importance of diversification in portfolio allocation fifty years ago. The concept of correlation has become so popular that people now use the term "correlation" and "dependence" interchangeably. However, as we discuss in detail below, correlation is not an appropriate measure of dependence when returns are not normally distributed. Formally, two random variables (X, Y) are dependent or associated if they don't satisfy the condition of probabilistic independence, i.e. $F(x, y) \neq F_1(x)\, F_2(y)$. A measure of dependence summarizes the dependence structure of two random variables in a single number. Now we consider the desired properties of a dependence measure. Let

$\delta(\cdot,\cdot)$ be a scalar measure of dependence. According to Embrechts *et al.* (2002), we desire the following properties:

1. $\delta(X,Y) = \delta(Y,X)$ (symmetry);
2. $-1 \le \delta(X,Y) \le +1$ (normalization);
3. $\delta(X,Y) = 1 \Leftrightarrow (X,Y)$ (comonotonic);

   $\delta(X,Y) = -1 \Leftrightarrow (X,Y)$ (countermonotonic);
4. For a strictly monotonic transformation $T : \Re \to \Re$ of X:

$$\delta(T(X),Y) = \begin{cases} \delta(Y,X) & T \quad INCREASING \\ -\delta(Y,X) & T \quad DECREASING \end{cases}$$

Dependence can be measured using several different concepts: linear correlation, concordance, and tail dependence.

3.1. Linear Correlation

The most familiar measure of dependence between two random variables (X, Y) is the Pearson product-moment correlation coefficient[12], or "Pearson's correlation", or as most people know it, the linear correlation coefficient. It is defined as:

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \qquad (1)$$

where $\mathrm{cov}(X,Y)$ is the covariance between X and Y, $\mathrm{cov}(X,Y) = E(X,Y) - E(X)E(Y)$ and σ_X , σ_Y denote the standard deviations of X and Y.

[12] It was developed by Karl Pearson from a similar but slightly different idea introduced by Francis Galton in the 1880s. Karl Pearson FRS (27 March 1857 – 27 April 1936) established the discipline of mathematical statistics.

The linear correlation is a measure of linear dependence. If two variables are independent, $\rho(X,Y) = 0$ because $\text{cov}(X,Y) = 0$. In the case of perfect linear dependence, i.e. $Y = aX + b$, $\rho(X,Y) = \pm 1$. In the case of imperfect linear dependence: $-1 < \rho(X,Y) < +1$. The popularity of linear correlation can be explained for the following reasons. First, it is easy to calculate. For most bivariate distributions, it is straightforward to calculate second moments (variances and covariance) and to derive the correlation coefficient. Second, linear correlation is invariant with respect to linear transformations of the variables.

Correlation is a good measure of dependence in multivariate Normal distributions but it has several shortcomings:

1. Linear correlation is not defined for some heavy-tailed distributions whose second moments do not exist, e.g. Student's-*t* distribution with degrees of freedom less than three. This is not ideal for financial time series, which display the property of fat tails and nonexistence of higher moments.
2. Independence between two random variables implies that linear correlation is zero (uncorrelated), but the converse is true only for a multivariate Normal distribution. For example, if $X \sim N(0,1)$, and $Y = X^2$, then $\text{cov}(X,Y) = 0$, but (*X*, *Y*) are clearly dependent. Zero correlation only requires $\text{cov}(X,Y) = 0$, but zero dependence requires $\text{cov}(\phi_1(X), \phi_2(Y)) = 0$ for any functions ϕ_1 and ϕ_2.
3. Linear correlation is not invariant to strictly increasing nonlinear transformations $T : \Re \to \Re$. That is $\rho(T(X), T(Y)) \neq \rho(X,Y)$. For example, if we take the

bivariate standard Normal distribution with correlation ρ and the transformation $T(x) = \Phi(x)$ (the standard Normal distribution function), we have

$$\rho(T(X),T(Y)) = \frac{6}{\pi}\arcsin(\frac{\rho}{2}).$$

4. Moreover, there are also statistical problems with correlation; e.g., a single observation can have an arbitrarily high influence on the linear correlation. Thus linear correlation is not a robust measure.

These limitations of linear correlation motivate consideration of alternative measures of dependence such as rank correlation.

3.2. Rank Correlation

A rank correlation coefficient measures the correspondence between two rankings and assesses its significance. Two well-established measures of rank correlation are Spearman's (1904) rank correlation (Spearman's rho) and Kendall's (1938) rank correlation (Kendall's tau).

Let (X, Y) be a random vector with continuous marginal distributions (F_1, F_2) and copula $C = C_{X, Y}$. The Spearman measure is the linear correlation between $F_1(X)$ and $F_2(Y)$:

$$\rho_S(X,Y) = \rho(F_1(X), F_2(Y))$$

where ρ is the usual linear correlation. Let (X_1, Y_1) and (X_2, Y_2) be two independent vectors of random variables with identical distribution functions and copula $C = C_{X, Y}$. Kendall's measure is defined as:

$$\rho_\tau(X,Y) = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

The first term on the right, $P[(X_1 - X_2)(Y_1 - Y_2) > 0]$, is referred to as Pr[concordance], the second as Pr[discordance], therefore:

$$\rho_\tau(X,Y) = \text{Pr[concordance]} - \text{Pr[discordance]}$$

is a measure of the difference between the two. Both $\rho_S(X,Y)$ and $\rho_\tau(X,Y)$ are measures of the degree of monotonic dependence between (*X*, *Y*), whereas linear correlation measures the degree of linear dependence only. Both measures are based on the concept of concordance, which refers to the property that large values of one random variable are associated with large values of another, whereas discordance refers to large values of one being associated with small values of the other.

As proved in Embrechts *et al.* (2002), both $\rho_S(X,Y)$ and $\rho_\tau(X,Y)$ have the property of symmetry, normalization, co- and counter-monotonicity and take the value zero under independence. Further, both $\rho_S(X,Y)$ and $\rho_\tau(X,Y)$ can be expressed in terms of copulas as follows:

$$\rho_S(X,Y) = 12 \int_0^1 \int_0^1 C(u,v)\,du\,dv - 3 \qquad (2)$$

$$\rho_\tau(X,Y) = 4 \int_0^1 \int_0^1 C(u,v)\,dC(u,v) - 1. \qquad (3)$$

The rank correlation measure has the advantage over ordinary correlation because the rank correlation measure is invariant under monotonic transformations and can capture perfect dependence. Although rank correlations are not simple functions of moments and hence their computation is more complex, there are some cases when rank correlations are easier to calculate. When we are working with multivariate distributions where first

and second moments are easily determined, such as multivariate Normal or *t*-distributions then the calculation of linear correlation is easier. When we are working with a multivariate distribution which possesses a simple closed-form copula, such as the Gumbel copula, then moments may be difficult to determine and the calculation of rank correlations may be easier.

3.3. Tail Dependence

Tail dependence measures the amount of dependence in the upper-right-quadrant tail or lower-left-quadrant tail of a bivariate distribution. Such a dependence measure relates to the conditional probability that one variable exceeds some value given that another exceeds some value. For continuous marginal distributions, tail dependence is a copula property; hence it is invariant under strictly increasing transformations.

Definition of Tail Dependence Let *X* and *Y* be continuous random variables with distribution functions *F* and *G*, respectively. The upper tail dependence coefficient λ_U is the limit (if it exists) of the conditional probability that *Y* is greater than the α-th percentile of G given that *X* is greater than the α-th percentile of *F* as α approaches 1, i.e.

$$\lambda_U = \lim_{\alpha \to 1^-} P[Y > G^{-1}(\alpha) \big| X > F^{-1}(\alpha)] \tag{4}$$

If λ_U is in (0,1], *X* and *Y* are said to be asymptotically dependent in the upper tail; if λ_U = 0, *X* and *Y* are said to be asymptotically independent in the upper tail; and similarly for λ_L. When *F* and *G* are continuous distributions and the limit exists, then

$$\lambda_U = \lim_{\alpha \to 1^-} \frac{\overline{C}(\alpha, \alpha)}{1 - \alpha} \tag{5}$$

$$\lambda_L = \lim_{\alpha \to 0^+} \frac{C(\alpha, \alpha)}{\alpha} \tag{6}$$

where $\overline{C}(u,u) = 1 - 2u + C(u,u)$ denotes the survivor function of the unique copula *C* associated with (*X*, *Y*).

4. Examples of Copulas

A large number of copulas have been proposed in the financial literature to model the joint distributions of financial data. In this section, we consider a few examples of copulas which are popular in the field of finance.

4.1. Normal (Gaussian) Copula

The most popular copula is perhaps the Normal (Gaussian) copula, which has the dependence function associated with bivariate normality. It can be written as:

$$C_{Ga}(u,v;\rho) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left(-\frac{r^2 - 2\rho rs + s^2}{2(1-\rho^2)}\right) drds \tag{7}$$

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

4.2. Student-*t* copulas

The bivariate Student-*t* copula (or briefly *t* copula) with *n* degrees of freedom and correlation $\rho = Corr[t_n^{-1}(u), t_n^{-1}(v)]$:

$$C_t(u,v;n,\rho) = \int_{-\infty}^{t_n^{-1}(u)} \int_{-\infty}^{t_n^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \left(1 + \frac{r^2 - 2\rho rs + s^2}{n(1-\rho^2)}\right)^{-(n+2)/2} drds \tag{8}$$

where t_n^{-1} denotes the inverse of the cdf of the standard univariate Student-*t* distribution with *n* degrees of freedom. The parameter *n* controls the heaviness of the tails and as $n \to \infty$, $C_t(u,v;n,\rho) \to \Phi_G(u,v;\rho)$, where Φ_G is the standard bivariate normal distribution.

4.3. Archimedean Copulas

Gaussian copulas and Student-*t* copulas mentioned above belong to the family of elliptical copulas which feature symmetric dependence only. However, in many financial applications, it is more common to observe big losses together rather than big gains. Here we introduce Archimedean copulas to model such asymmetries in financial data. For the Archimedean family of copulas,

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \tag{9}$$

where $\varphi : (0,1] \to [0, \infty)$ is a strictly decreasing convex function with $\varphi(1) = 0$, and it is assumed that $\varphi^{-1}(t) = 0$ for all $t > \lim_{u \to 0} \varphi(u)$. The function φ is called the generator. The independence copula $C(u,v) = uv$ is an Archimedean copula with generator $\varphi(t) = -\log t$. We will discuss three one-parameter Archimedean copulas: Gumbel (1960a), Clayton (1978), and Frank (1979); and one two-parameter Archimedean copula, the Symmetrized Joe-Clayton (Patton 2006a) copula in this chapter.

4.3.1. Gumbel Copula

The Gumbel copula corresponds to $\varphi(t) = (-\ln(t))^\theta$ and it is defined as:

$$C_G(u,v;\theta) = \exp(-[(\ln u)^\theta + (\ln v)^\theta]^{1/\theta}) \tag{10}$$

with the dependence parameter θ restricted to the interval [1, ∞). Values of 1 and ∞ correspond to independence and the Fréchet upper bound, but this copula does not attain the Fréchet lower bound for any value of θ. In addition, the Gumbel copula does not allow negative dependence and exhibits strong right tail dependence and relatively weak left tail dependence. The Gumbel copula is appropriate for data known to be strongly correlated at high values but less correlated at low values.

4.3.2. Clayton Copula

The Clayton copula corresponds to $\varphi(t) = t^{-\theta} - 1$ and it is defined as:

$$C_C(u, v; \theta) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \tag{11}$$

with the dependence parameter θ restricted on the region (0, ∞). As θ goes to zero, the marginal distributions become independent. As θ approaches infinity, the copula attains the Fréchet upper bound. The Clayton copula cannot account for negative dependence. The Clayton copula has been widely used in the field of insurance and risk management as it exhibits strong left tail dependence and relatively weak right tail dependence.

4.3.3. Frank Copula

The Frank copula corresponds to $\varphi(t) = -\ln \frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1}$ and it is defined as:

$$C_F(u, v; \theta) = -\theta^{-1} \ln\ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \quad . \tag{12}$$

The dependence parameter θ can assume any real value in (-∞, ∞). Values of -∞, 0, and ∞ correspond to the Fréchet lower bound, independence, and Fréchet upper bound, respectively. The Frank copula is popular for several reasons. Unlike other copulas, the

Frank copula permits negative dependence between the marginal distributions. Similar to elliptical copulas, dependence is symmetric in both tails. The Frank copula is "comprehensive" in the sense that both Fréchet bounds are included in the range of permissible dependence. The Frank copula is most appropriate for data that exhibit weak tail dependence.

4.4. Symmetrized Joe-Clayton Copula (SJC)

The Symmetrized Joe-Clayton (SJC) copula proposed by Patton (2006a) is a slight modification of the original Joe-Clayton (JC) copula, proposed by Joe (1997) that is a Laplace transformation of Clayton's copula. It is defined as:

$$C_{JC}(u,v;\tau^U,\tau^L) = 1-(1-\{[1-(1-u)^{\kappa}]^{-\gamma}+[1-(1-v)^{\kappa}]^{-\gamma}-1]\}^{-1/\gamma})^{1/\kappa} \qquad (13)$$

where $\kappa = 1/\log_2(2-\tau^U)$, $\gamma = -1/\log_2(\tau^L)$ and tail dependence parameter $\tau^U \in (0,1], \tau^L \in (0,1]$. By construction, the Joe-Clayton copula always exhibits asymmetry even when the two tail dependence measures are equal. In order to overcome this drawback, Patton (2006a) proposed the "Symmetrized Joe-Clayton" copula, which has the tail dependence measures completely determining the presence or absence of asymmetry. The SJC copula is given by:

$$C_{SJC}(u,v;\tau^U,\tau^L) = 0.5(C_{JC}(u,v;\tau^U,\tau^L)+C_{JC}(1-u,1-v;\tau^U,\tau^L)+u+v-1), \qquad (14)$$

where C_{JC} represents the Joe-Clayton copula. The advantage of the Symmetrized Joe-Clayton copula is that it is symmetric when $\tau^U = \tau^L$. In practice, the fact that the Symmetrized Joe-Clayton nests symmetry as a special case makes it a more desirable specification than the Joe-Clayton copula.

4.5. Parameterizing Time Variation in the Copulas

There are many ways to capture time variation in the copulas. In this chapter, we assume that the functional form of the copula remains fixed over the sample whereas the parameters vary according to some evolution equation. An alternative way to capture time variation is to use a regime switching copula function. We follow Patton (2006a) and assume the following evolution dynamics for the SJC copula:

$$\tau_t^U = \Lambda\ \omega_U + \beta_U \cdot \tau_{t-1}^U + \alpha_U \cdot \frac{1}{10} \sum_{j=1}^{10} \left| u_{t-j} - v_{t-j} \right| \tag{15}$$

$$\tau_t^L = \Lambda\ \omega_L + \beta_L \cdot \tau_{t-1}^L + \alpha_L \cdot \frac{1}{10} \sum_{j=1}^{10} \left| u_{t-j} - v_{t-j} \right| \ , \tag{16}$$

where $\Lambda(x) = \frac{(1-e^{-x})}{(1+e^{-x})}$ is the modified logistic transformation in order to keep τ^U and τ^L within the (-1, 1) interval. The right-hand side of the model for the tail dependence evolution equation contains an autoregressive term, $\beta_U \cdot \tau_{t-1}^U$ and $\beta_L \cdot \tau_{t-1}^L$, and a forcing variable, where the autoregressive term represents the persistence effect and the forcing variable captures the variation in dependence. The mean absolute difference between u_t and v_t over the previous 10 periods is used as the forcing variable[13]; this measures how much the previous realizations of the standardized errors depart from the situation of perfect positive dependence. The expectation of this distance measure is inversely related to the concordance ordering of copulas: under perfect positive dependence it equals zero, under independence it equals 1/3, and under perfect negative dependence it equals1/2. The dynamic structure above is an ARMA(*1,10*)-type process. However, this evolution

[13] The choice of 10 periods is to follow Patton (2006a) and others to make our empirical analysis comparable.

dynamic structure can be different for different datasets. We recognize that it is difficult to identify the forcing variable for a time-varying limit probability and choose to follow Patton's (2006a) proposal to make our results comparable to previous research. We use the same evolution dynamics to capture the time variation in other copulas.

4.6. Comparison of Copulas

In order to demonstrate the flexibility of the copula functions, we now examine the isoprobability contours of very different shapes generated by various bivariate distributions, which are constructed by copula functions with standard Normal marginal distributions, each with a linear correlation coefficient of 0.5. The contour plots presented in Figure 10 show that different copulas can account for many kinds of dependence structure. The upper left panel displays the Normal copula with ρ=0.5, which has the familiar symmetric shape. The other contour plots in this figure show the dependence structures implied by other copulas, with each copula calibrated to yield ρ=0.5. The different shapes in the figure clearly demonstrate that knowing the marginal distributions and linear correlation is not sufficient to describe a joint distribution. For example, the contour of the Student's-*t* copula in the upper right panel has a quite different behaviour in the first ("positive-positive") and in the third ("negative-negative") quadrant, where the contours are more tightly clustered around the diagonal even though it is symmetric and elliptically shaped like the Normal copula. The Gumbel copula can account for returns more highly correlated in bull markets than in bear markets. On the other hand, the Clayton copula can account for returns more highly correlated in bear markets than in bull markets, which is the case for many financial time-series. Other copulas that can generate symmetric dependence are the Frank copula and the SJC which are in the

bottom two panels of Figure 1. However, the SJC copula is structured in such way that it can generate both symmetric ($\tau^U = \tau^L$) and asymmetric dependence ($\tau^U \neq \tau^L$). In addition, the SJC copula can account for returns more highly correlated in bear markets (like the Clayton copula) or in bull markets (like the Gumbel copula).

5. Copula Estimation

5.1. Statistical Inference with Copulas

Let (*X*, *Y*) be a vector of two random variables with joint distribution function *H* and marginal distribution function *F* and *G* respectively. Each marginal distribution function depends only on the parameter ϑ_i . Denote the unknown vector of parameters by $\vartheta = (\vartheta_1, \vartheta_2, \theta)$, where θ is the vector of parameters of the copula $\{C_\theta, \theta \in \Theta\}$ and C_θ is completely known except for the parameter θ . Hence by Sklar's theorem:

$$H(x, y) = C(F(x; \vartheta_1), G(y; \vartheta_2); \theta). \tag{17}$$

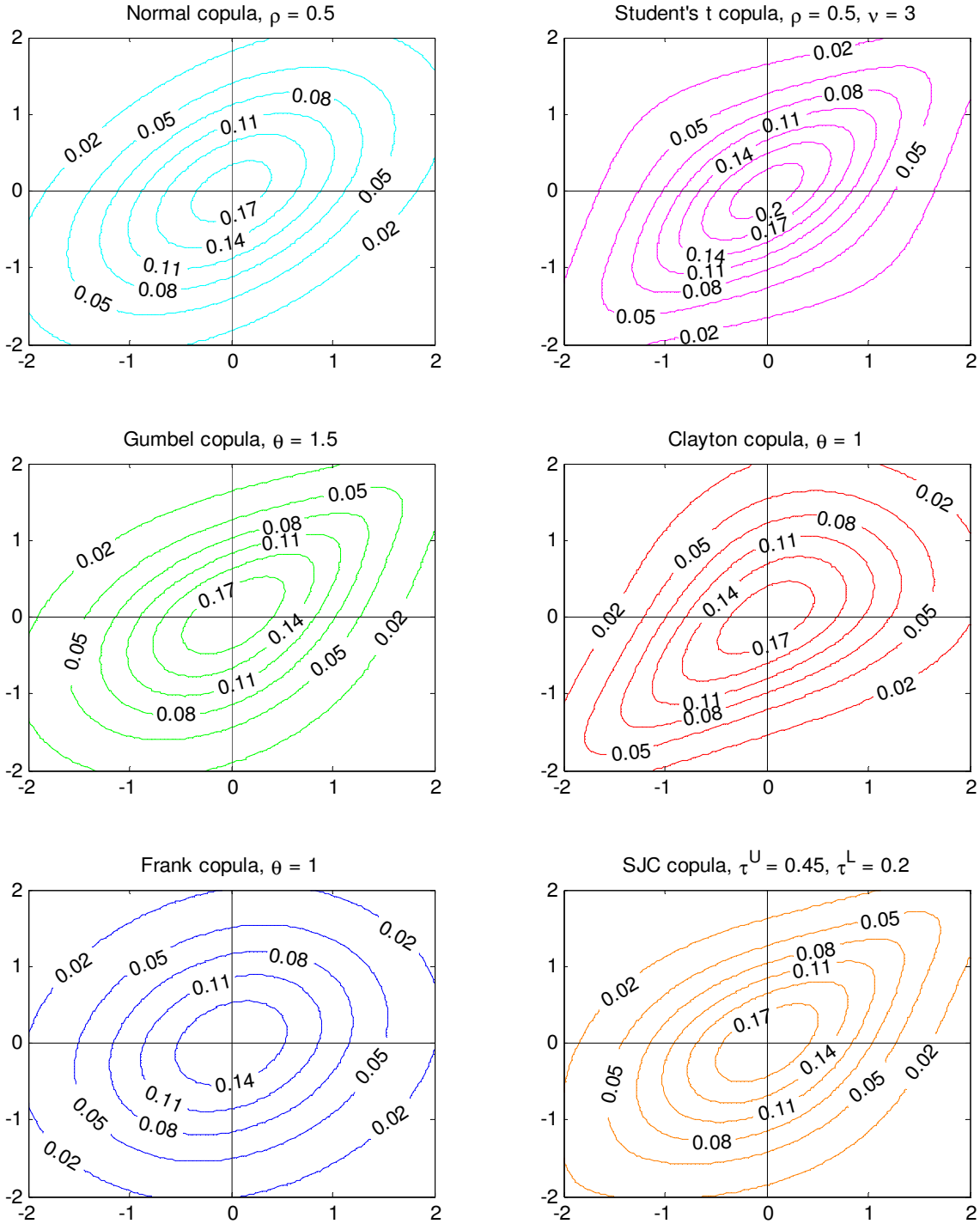
Thus the joint distribution function *H* is completely specified by the parameter vector $\vartheta = (\vartheta_1, \vartheta_2, \theta)$. Differentiating the equation above with respect to the variables yields the density function *h*:

$$h(x, y) = c(F(x), G(y)) f(x) g(y), \tag{18}$$

where f and g are the density functions associated with the marginal distributions *F* and *G* and *c* is the copula density, given by

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}. \tag{19}$$

Figure 10 Contour Plots of Various Distributions All with Standard Normal Marginal Distribution



5.2. Estimation

Based on the copula density function shown above, the log-likelihood function is the following:

$$l(\vartheta) = \sum_{t=1}^{T} \ln c(F(x,t;\vartheta_1), G(y,t;\vartheta_2), \theta) + \sum_{t=1}^{T} (\ln f(x,t;\vartheta_1) + \ln g(y,t;\vartheta_2))$$

There are two popular parametric estimation methods available for copula estimation. The first method simultaneously estimates all of the parameters of the marginal distributions and the copula using full maximum likelihood (FML). Therefore, the FML estimator $\hat{\theta}$ of the parameter vector θ is the one which maximizes the above log-likelihood function:

$$\hat{\theta} = \arg\max_{\theta} l(\theta) .$$

However, in some situations, the maximum likelihood estimation may be difficult to obtain due to the complexity of the model or just too many parameters, as both the marginal distribution and copula parameters must be estimated jointly. Especially, when we consider the complicated structure of time-varying dependence, an analytical expression for the gradient vector of the likelihood might not exist. Therefore, numerical methods have to be adopted to solve the optimization problem, which dramatically slows down computation, often leading to convergence difficulties. The second parametric estimation method, the Inference Function for Margins (IFM) method, was introduced to resolve these issues. This approach, proposed by Shih and Louis (1995) and Joe and Xu (1996), is maximum likelihood estimation of the dependence parameter given the estimated marginal distributions. In the first step, the parameters in the marginal

distributions are estimated. In the second step, the copula parameters are estimated conditioned on the previous marginal distributions estimates. This method exploits an attractive feature of copula for which the dependence structure is independent of the marginal distributions. Under certain regularity conditions, Patton (2006b) showed that the IFM estimator is consistent and asymptotically Normal.

5.3. Conditional Case

The conditional likelihood can be derived from the conditional version of Sklar's Theorem. Let F and G be the conditional distribution of $X|W$ and $Y|W$ and let H be the joint continuous conditional distribution of (X, $Y|W$), where the conditional copula function is C. Then, we have:

$$H(x, y|w) = C(F(x|w), G(y|w)|w).$$

Therefore the conditional density function is given by:

$$h(x, y|w) = c(F(x|w), G(y|w)|w) f(x|w) g(y|w),$$

where $f(\cdot|w)$ and $g(\cdot|w)$ are the conditional density functions and

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

The conditional log-likelihood function is the following:

$$l(\vartheta) = \sum_{t=1}^{T} \ln c(F(x,t;\vartheta_1|w), G(y,t;\vartheta_2), \theta|w) + \sum_{t=1}^{T} (\ln f(x,t;\vartheta_1|w) + \ln g(y,t;\vartheta_2|w)),$$

and we can use the previously mentioned estimation methods to estimate the parameters.

5.4. Copula Evaluation and Selection

With many copulas to choose from in empirical applications in finance, one needs to know how well the models fit the data, enabling us to choose an appropriate one from the many options. Like many other financial econometric models, the selection of the copula should be guided by economic theory and statistical features of the data. For example, previous research has found that financial returns are more correlated in bear markets than in bull markets. Therefore, intuitively, copulas that can capture such asymmetric dependence (such as the SJC copula) should be superior to the symmetric copulas (such as elliptical copulas).

One straight-forward approach to select a copula between non-nested parametric copulas estimated by maximum likelihood is to use a model slection criterion; e.g., the Akaike (1974) or Schwarz (1978) Bayesian information criteria are commonly adopted measures. Specifically, Akaike's information criterion (AIC) is defined as:

$$AIC(M) = -2\ln(\hat{L}) + 2M$$

where M is the number of parameters being estimated and $\ln(\hat{L})$ is the maximized log likelihood value. Smaller AIC values indicate better fit. Smith (2003) used this criterion for copula selection.

6. Empirical Results

6.1. Data Description and Preliminary Analysis

We examine the interaction between the four energy futures, selecting energy futures contracts that are liquid and widely held. End-of-day prices of the front month[14] futures of West Texas Intermediate (CL), Brent Crude Oil (CO), Heating Oil (HO), and Natural Gas (NG) have been obtained from the Energy Information Administration (EIA)[15] and Bloomberg[16] from Jan 13, 1994 to Dec 15, 2009. This time period is chosen to accommodate the longest sample period (Jan 13, 1994 was the first day that Natural Gas Futures price was available via EIA's website). Gasoline futures are not included in this analysis because the current benchmark gasoline contract (reformulated gasoline blendstock for oxygen blending, or RBOB) only started trading on the New York Mercantile Exchange (NYMEX) in October 2005. Table 13 below gives summary statistics for the energy futures returns, with each defined as:

$$r_t = \log \frac{P_t}{P_{t-1}} \times 100 .$$

Figure 11 plots daily prices, returns and squared returns for each series. Each plot of daily returns confirms the typical empirical time series properties: mean reversion, with periods of low volatility followed by periods of extreme volatility. The squared daily returns exhibit evidence of volatility clustering, *i.e.* large changes tend to be followed by large

[14] Front month is referred to the contract month with an expiration date closest to the current date. Front month contracts are generally the most liquid of futures contracts and the closest to the spot price on the underlying commodity.

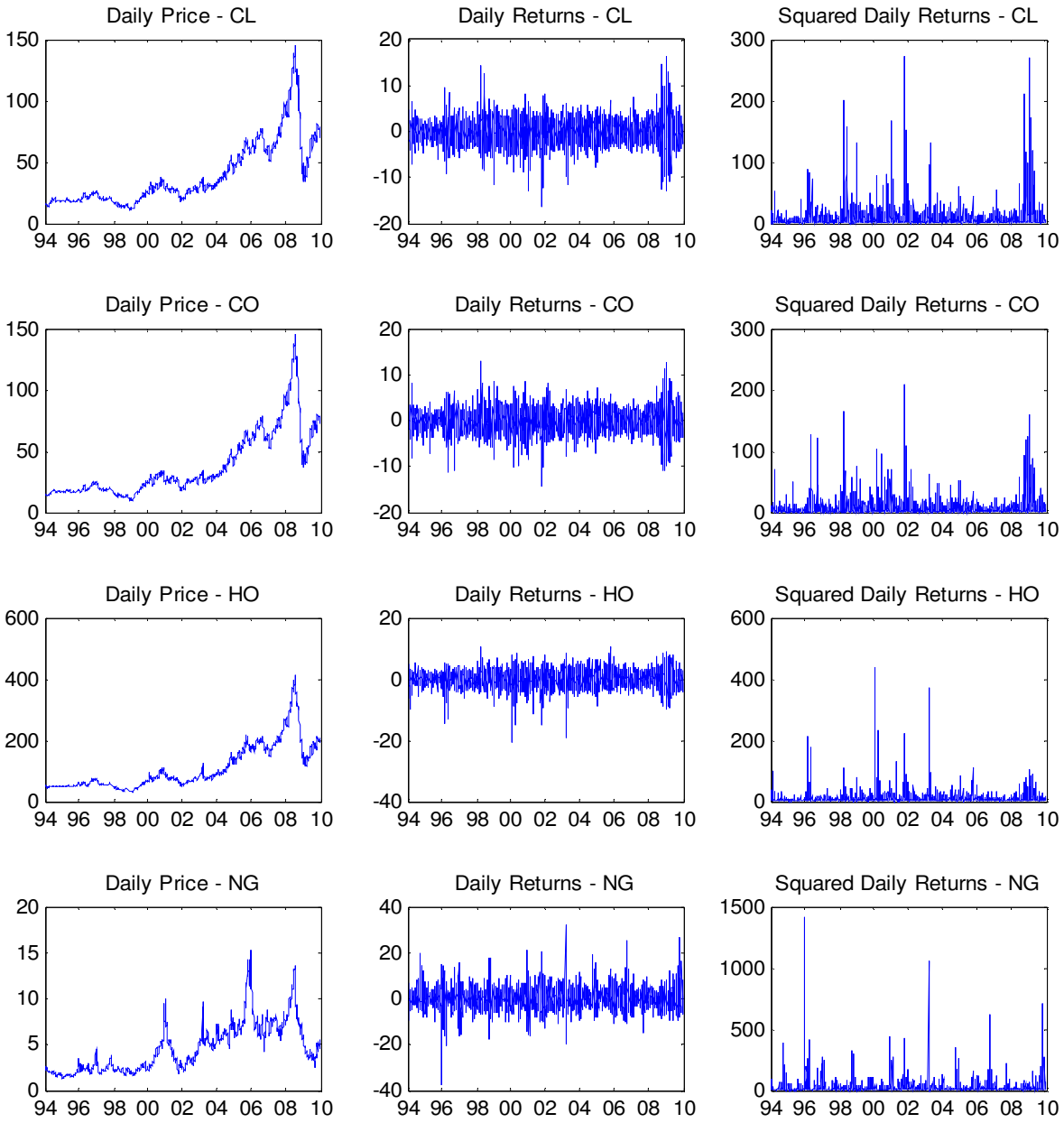
[15] http://www.eia.doe.gov/

[16] http://www.bloomberg.com

changes. For risk management purposes, any measurement of risk exposure should be conditional on the current volatility regime. Basic statistics for the front month futures are summarized in Table 13. As is the case with most financial time series, the front month energy futures exhibit evidence of fat tails, with kurtosis statistics larger than three and the Jarque-Bera (1978) test suggesting non- normality. The Augmented Dickey-Fuller test rejects the null hypothesis of a unit root for all series. Among energy futures, natural gas is more volatile than the other examined series, with the standard deviation of natural gas front month futures returns being 50% higher than the standard deviation of petroleum products returns. Unlike petroleum products, front month natural gas futures returns exhibit positive skewness, which suggests that rewards are more likely than losses.

The squared daily returns shown in Figure 11 exhibit volatility clustering, *i.e.* large changes tend to follow by large changes, in "bursts". The Ljung-Box Q-statistics reported in Table 14 also reject the null hypothesis of no autocorrelation through 20-lags at the 1% significance level. Therefore, the next step is to test the (G)ARCH effect using an ARCH-LM test on the residuals from simple Ordinary Least Squares regression of each returns series on an intercept and it own lagged values. The test statistics shown in Table 2 suggest that the null hypothesis of no ARCH effect and the null hypothesis of no GARCH effect are rejected for all series. This result is not sensitive to the number of lags included in the ARCH-LM and GARCH tests. Given these results, the GARCH effects need to be filtered out from the data before the extreme value analysis is applied.

Figure 11 Daily Prices, Returns and Squared Returns



** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Table 13 Summary Descriptive Statistics

	CL	CO	HO	NG
Sample Size	3,984	3,984	3,984	3,984
Mean	0.040	0.042	0.034	0.023
Median	0.106	0.103	0.040	0.000
Max	16.410	12.898	10.403	32.435
Min	-16.545	-14.437	-20.971	-37.575
Std. dev.	2.477	2.281	2.417	3.883
Skewness	-0.112	-0.219	-0.565	0.209
Kurtosis	7.075	5.925	7.681	9.758
Jarque-Bera Test	2,764*	1,452*	3,850*	7,611*
p-value	*0.000*	*0.000*	*0.000*	*0.000*
ADF Unit-Root Test	-52.030*	-72.460*	-76.300*	-74.050*
p-value	*0.000*	*0.000*	*0.000*	*0.000*
Auto Corr – *r*				
Lag 1	-0.006*	-0.011*	-0.062*	-0.032*
Lag 5	-0.037	-0.022	0.000	-0.028
Lag 10	0.001	0.026	-0.003	0.005
Lag 20	0.009	0.023	-0.003	-0.005
Ljung-Box (20)	49.3*	44.8*	82.4*	44.0*
p-value	*0.000*	*0.000*	*0.000*	*0.001*
Auto Corr - *r2*				
Lag 1	0.110*	0.074*	0.051*	0.051*
Lag 5	0.046*	0.019*	0.024*	0.084*
Lag 10	0.045*	0.04*	0.023*	0.064*
Lag 20	0.046*	0.048*	0.033*	0.037*
Ljung-Box (20)	822.0*	258.3*	264.6*	268.9*
p-value	*0.000*	*0.000*	*0.000*	*0.000*

** Indicates significant at the 5% level*
*** An intercept term (no trend) is included in the integrating regression for the ADF Unit-Root test.*
**** The Lag orders are selected to examine a range of possible autocorrelations.*
***** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Table 14 (G)ARCH Effect Test

Testing for ARCH(1)	**CL**	**CO**	**HO**	**NG**
F-statistic	61.1*	28.5*	13.3*	13.2*
p-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Testing for GARCH(1,1)	**CL**	**CO**	**HO**	**NG**
nR^2	60.4*	28.4*	13.3*	13.1*
p-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*

** Indicates significant at the 5% level*
*** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Table 15 Unconditional Correlation Measures Matrix

	Linear Correlation			**Spearman's Rho**			**Kendall's Tau**		
	CL	**CO**	**HO**	**CL**	**CO**	**HO**	**CL**	**CO**	**HO**
CO	0.82			0.85			0.70		
	(0.000)			*(0.000)*			*(0.000)*		
HO	0.73	0.74		0.76	0.73		0.60	0.57	
	(0.000)	*(0.000)*		*(0.000)*	*(0.000)*		*(0.000)*	*(0.000)*	
NG	0.22	0.21	0.26	0.26	0.25	0.28	0.18	0.17	0.20
	(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*

** Indicates significant at the 5% level; standard errors are presented in parentheses*
*** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

In order to form an initial opinion on the correlation and dependence relationship among energy futures, we present unconditional correlation measures in Table 15. As expected, the unconditional correlation measure matrices indicate that there exists a rather high dependence between energy futures. Note that although the underlying assumptions are different, both the linear correlation and Spearman statistics measure the proportion of the variation in one variable that might be considered as being associated with the variation in the other variable. On the other hand, although sharing similar underlying assumptions, the Kendall statistics measure the difference between the probability that

the observed data are on the same order[17] *versus* the probability that the observed data are not of the same order. Not surprisingly, the correlations between petroleum products are higher than those of petroleum products and natural gas. In addition, as a competitor to heating oil in the winter season for home heating, natural gas is more strongly associated with heating oil than other petroleum products.

6.2. The Models for the Marginal Distributions

We adopt the two-step estimation method in this chapter due to the large number of parameters in the time-varying models. The first step is to estimate the marginal models. The model used for the marginal distribution is AR(*1*) – GARCH(*1,1*) – *t* as described below:

$$r_t = \mu + ar_{t-1} + \varepsilon_t$$

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

$$\varepsilon_t \cdot \sqrt{\frac{\upsilon}{\sigma_t^2(\upsilon-2)}} \sim iid\ t_n$$

where r_t is the return at time *t*. The parameter estimates and standard errors for the marginal distribution models are presented in Table 16. All parameters for the variance equation are highly significant. The autoregressive effect in the volatility specification is strong as it is in the range of 0.90 (natural gas) to 0.96 (WTI). In addition, the condition for covariance-stationarity is satisfied because it appears that GARCH (1) + ARCH (1) <1 for all the series. Note that GARCH (1) + ARCH (1) is really close to one, which implies that shocks have much longer effects.

[17] The orderings of the data when ranked by each of the quantities.

Table 16 Results for the Marginal Distribution

	CL	CO	HO	NG
Constant	0.079*	0.091*	0.068*	0.016
	(0.032)	*(0.030)*	*(0.032)*	*(0.048)*
AR(1)	-0.005	-0.034*	-0.019	-0.036*
	(0.016)	*(0.016)*	*(0.016)*	*(0.016)*
GARCH constant	0.043*	0.0329*	0.059*	0.320*
	(0.014)	*(0.011)*	*(0.017)*	*(0.071)*
GARCH (1)	0.955*	0.957*	0.948*	0.899*
	(0.007)	*(0.006)*	*(0.008)*	*(0.012)*
ARCH (1)	0.037*	0.037*	0.042*	0.081*
	(0.005)	*(0.005)*	*(0.006)*	*(0.010)*
Degrees of freedom	7.532*	6.775*	7.193*	5.767*
	(0.796)	*(0.696)*	*(0.744)*	*(0.482)*
Residual ARCH Test	**CL**	**CO**	**HO**	**NG**
F-statistic	1.329	0.207	0.729	0.011
p-value	*(0.716)*	*(0.886)*	*(0.393)*	*(0.915)*

** Indicates significant at the 5% level; standard errors are presented in parentheses*

Table 16 above also presents the ARCH effect test statistics on the residuals. As the ARCH-LM test confirms that there are no ARCH effects left in the residuals, the analysis in the rest of this chapter will focus on these standardized residuals.

After fitting the marginal models and obtaining the standardized residuals, we now examine the dependence between the filtered returns[18]. Figure 12 shows the scatter plots of a pair of filtered returns. It is clear that fitting of the marginal models did not remove the dependence between each series. Not surprisingly, the dependence between the petroleum products exhibits more linearity than the relationship between natural gas and the petroleum products. However, all scatter plots appear to be clouded circles with dispersions in lower and upper tails displaying the absence of linear correlation between series. In order to further investigate the dependence relationship between energy futures, we map the standardized residuals into the unit square applying the probability-integral transformation.

[18] We use the term "filtered returns" and "standardized residuals" interchangeably.

Figure 12 Scatter Plot of Bivariate Standardized Residuals

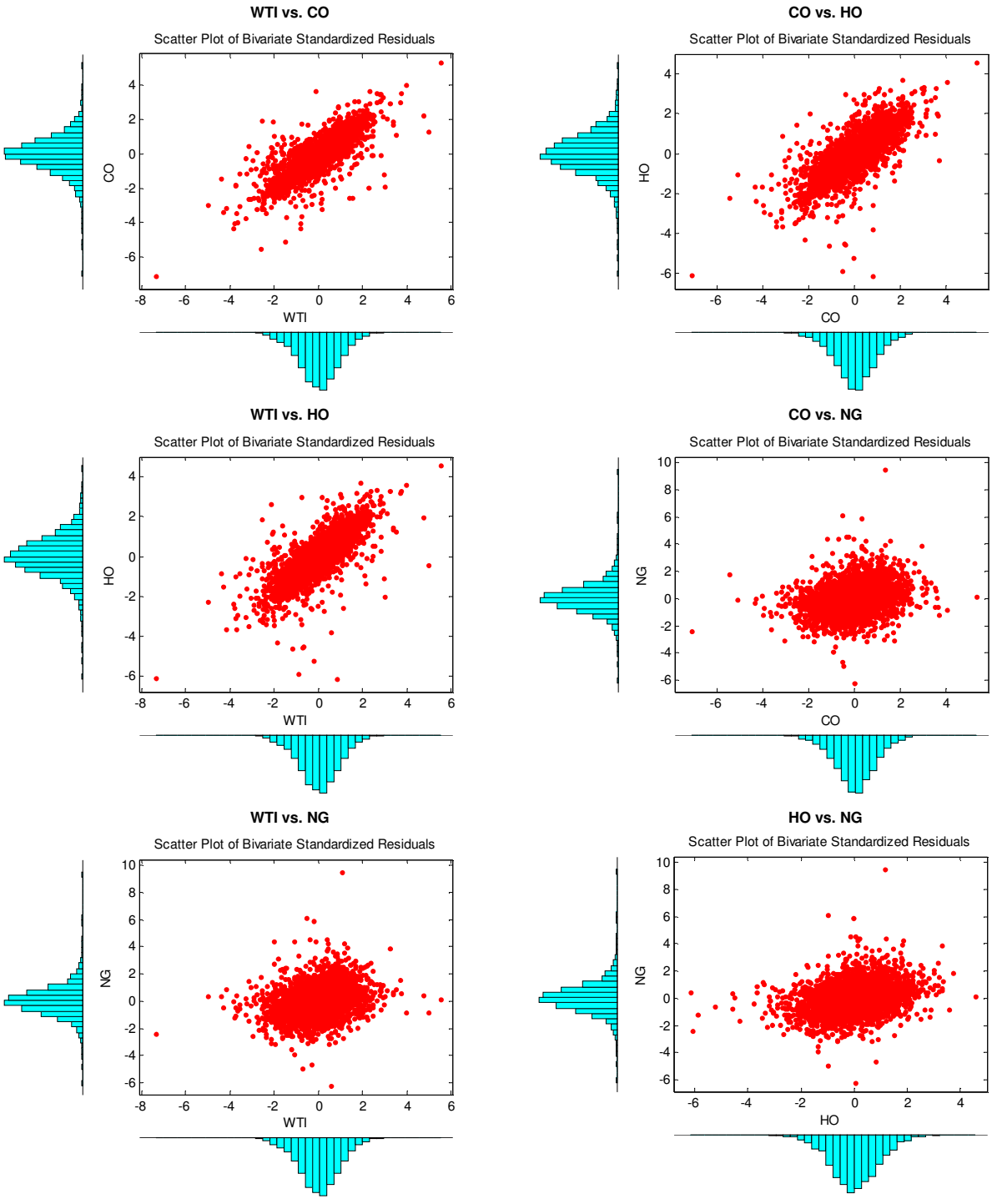


Figure 13 Dependence between Energy Futures – Scatter Plots of the Probability Transformed Standardized Residuals

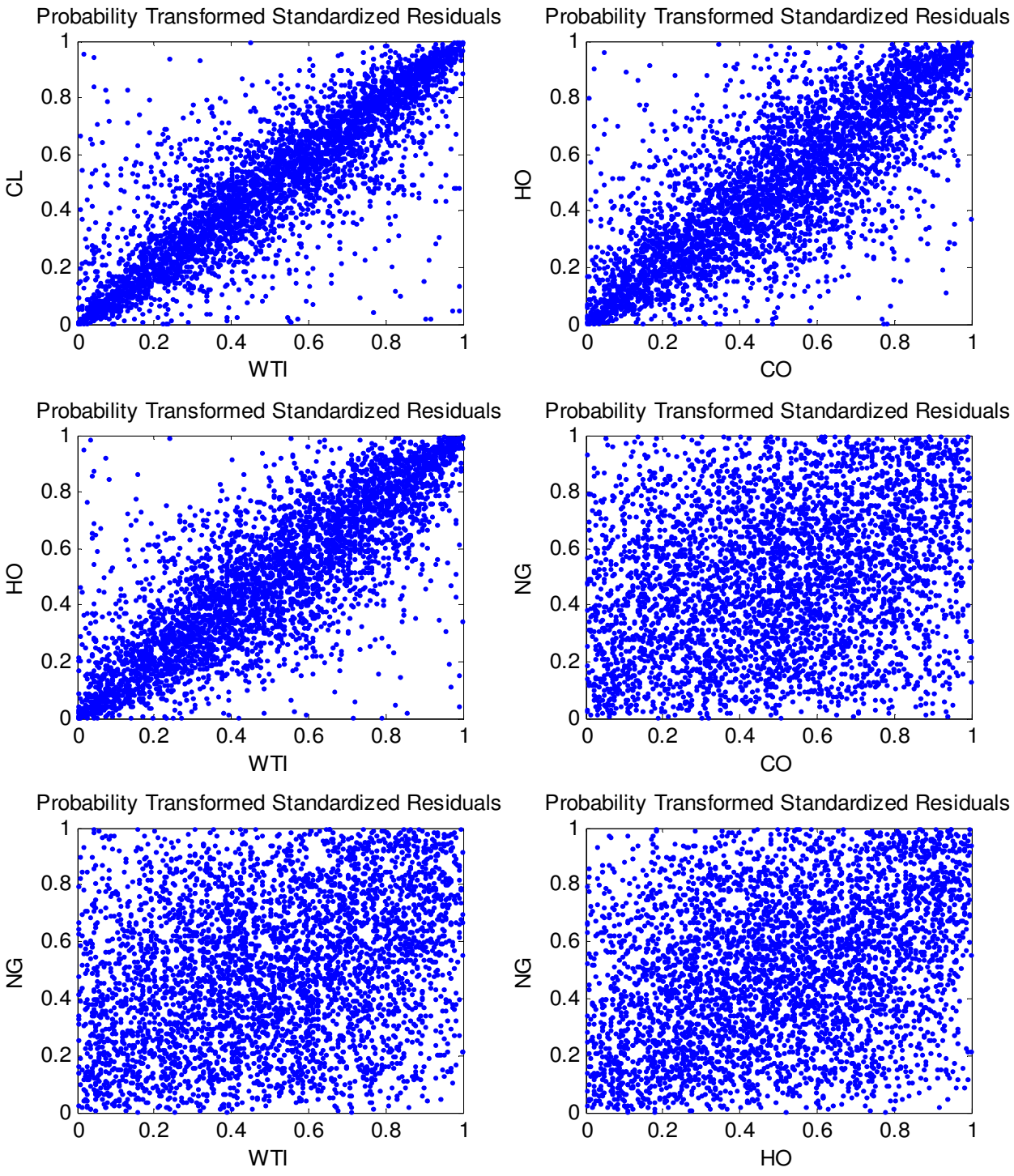


Table 17 Conditional Correlation Measure Matrix

	Linear Correlation			Spearman's Rho			Kendall's Tau		
	CL	**CO**	**HO**	**CL**	**CO**	**HO**	**CL**	**CO**	**HO**
CO	0.86			0.86			0.70		
	(0.000)			*(0.000)*			*(0.000)*		
HO	0.82	0.79		0.82	0.79		0.64	0.61	
	(0.000)	*(0.000)*		*(0.000)*	*(0.000)*		*(0.000)*	*(0.000)*	
NG	0.29	0.27	0.33	0.29	0.27	0.32	0.20	0.18	0.22
	(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*

** Indicates significant at the 5% level; standard errors are presented in parentheses*
*** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Figure 13 shows the scatter plots of the probability transformed filtered returns $\{(x_{1i}, x_{2i}) : i = 1,2,...,n\}$:

$$\{(\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i})) : i = 1,2,...,n\},$$

where $\hat{F}_1$ and $\hat{F}_2$ are the fitted *t*-distributions for the standardized residuals. Figure 13 demonstrates that the filtered returns are positively dependent and fairly symmetric around their mean. Table 17 summarizes the conditional correlation measures. Similar to the unconditional correlation measures, the conditional correlation matrices indicate that there exists a rather high dependence between energy futures. In addition, the conditional correlation measures are slightly higher than the unconditional ones.

6.3. Estimation of Copulas

Visual observation of these plots confirms that the filtered returns are associated. We now turn to finding a suitable copula to describe this dependence. The following copula families are fitted to the filtered returns: Normal, *t*, Gumbel, Clayton, Frank, and the

Symmetrized Joe-Clayton copula. Table 18 reports parameter estimates for constant dependence copulas. For the purpose of comparison, parameter estimates for time-varying dependence copulas are also presented in Table 19. The associated likelihood values and calculated AIC values are also reported in Table 18 and 19. All models are ranked in Table 20 based on minimization of AIC with "1" indicating the preferred model. The order of the ranking does not change if we use the likelihood values or other information criterion such as Bayesian Information Criterion (BIC).

The results reported in Tables 18 and 19 show that most estimated parameters are statistically significant. As expected, the estimated correlation coefficients from the Normal copula are similar to the linear pairwise correlations in Table 17. Taking fat-tails into consideration, the estimates for ρ based on the Student-*t* copula show higher correlation than using the Normal copula. The results also show that *t* copulas, especially the time-varying *t* copula, consistently outperform the other copula models. In general, Clayton copulas, both constant dependence and time-varying dependence copulas, perform very poorly and are consistently ranked on the bottom two in the 12 scenarios presented (six constant and six time-varying dependence copulas). The time-varying Frank and time-varying SJC copula are both ranked third among all 12 models. However, we prefer the time-varying SJC copula to the time-varying Frank copula as it can capture the asymmetric tail dependence while the time-varying Frank copula assumes symmetric tail behaviour.

Table 18 Results for the Copula Models – Constant Dependence

	CL/CO	CL/HO	CL/NG	CO/HO	CO/NG	HO/NG
Constant Normal Copula						
$\hat{\rho}$	0.85	0.80	0.29	0.79	0.27	0.33
	(0.003)	*(0.004)*	*(0.020)*	*(0.004)*	*(0.021)*	*(0.020)*
Copula Likelihood	2,543.1	2,074.2	170.3	1,912.6	150.0	227.8
AIC	-5,086.2	-4,148.3	-340.7	-3,825.3	-300.0	-455.7
Constant *t* Copula						
$\hat{\rho}$	0.92	0.87	0.38	0.86	0.37	0.43
	(0.067)	*(0.000)*	*(0.018)*	*(0.004)*	*(0.017)*	*(0.014)*
Dof	1.73	3.11	8.83	3.08	8.31	9.62
	(0.082)	*(0.154)*	*(0.064)*	*(0.020)*	*(0.342)*	*(0.041)*
Copula Likelihood	3,349.5	2,489.3	202.3	2,235.6	182.2	266.6
AIC	-6,698.9	-4,978.7	-404.6	-4,471.2	-364.4	-525.3
Constant Gumbel Copula						
$\hat{\theta}$	3.71	2.97	1.31	2.78	1.29	1.35
	(0.051)	*(0.040)*	*(0.019)*	*(0.038)*	*(0.019)*	*(0.019)*
Copula Likelihood	3,010.2	2,321.2	182.1	2,079.5	164.9	235.8
AIC	-6,020.3	-4,642.4	-364.2	-4,159.1	-329.7	-471.5
Constant Clayton Copula						
$\hat{\theta}$	3.48	2.60	0.49	2.44	0.47	0.58
	(0.069)	*(0.055)*	*(0.007)*	*(0.053)*	*(0.016)*	*(0.032)*
Copula Likelihood	2,295.4	1,769.4	144.3	1,646.8	130.7	190.6
AIC	-4,590.8	-3,538.9	-288.7	-3,293.7	-261.4	-381.1
Constant Frank Copula						
$\hat{\theta}$	13.49	10.50	2.34	9.56	2.21	2.66
	(0.214)	*(0.174)*	*(0.116)*	*(0.163)*	*(0.117)*	*(0.116)*
Copula Likelihood	2,998.7	2,365.5	190.7	2,106.4	165.7	245.1
AIC	-5,997.3	-4,731.0	-381.4	-4,212.9	-331.3	-490.2
Constant SJC Copula						
$\hat{\tau}^U$	0.81	0.74	0.22	0.70	0.21	0.25
	(0.000)	*(0.007)*	*(0.026)*	*(0.008)*	*(0.026)*	*(0.024)*
$\hat{\tau}^L$	0.73	0.67	0.20	0.67	0.19	0.25
	(0.000)	*(0.011)*	*(0.027)*	*(0.010)*	*(0.027)*	*(0.255)*
Copula Likelihood	2,858.3	2,215.1	192.2	2,027.5	176.0	251.9
AIC	-5,716.6	-4,430.2	-384.4	-4,055.1	-351.9	-503.8

** Indicates significant at the 5% level; standard errors are presented in parentheses*
*** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Table 19 Results for the Copula Models – Time-Varying Dependence

	CL/CO	CL/HO	CL/NG	CO/HO	CO/NG	HO/NG
			Time-varying Normal Copula			
$\hat{\omega}$	5.56	6.26	0.00	4.71	0.03	0.01
	(3.188)	*(1.019)*	*(0.009)*	*(0.914)*	*(0.024)*	*(0.011)*
$\hat{\beta}$	0.07	0.80	0.11	0.76	0.16	0.17
	(0.092)	*(0.106)*	*(0.026)*	*(0.122)*	*(0.050)*	*(0.044)*
$\hat{\alpha}$	-3.18	-5.00	2.07	-3.28	1.96	2.04
	(3.623)	*(1.218)*	*(0.037)*	*(1.134)*	*(0.086)*	*(0.049)*
Copula Likelihood	2,603.5	2,155.9	205.6	1,994.9	185.3	274.8
AIC	-5,207.0	-4,311.7	-411.1	-3,989.7	-370.5	-549.6
			Time-varying *t* Copula			
$\hat{\omega}$	8.73	6.65	2.71	5.27	2.58	2.02
	(0.576)	*(0.385)*	*(0.309)*	*(0.835)*	*(0.484)*	*(0.019)*
$\hat{\beta}$	-4.95	-2.78	-1.49	-1.46	-1.24	-0.33
	(0.634)	*(0.404)*	*(0.346)*	*(0.824)*	*(0.653)*	*(0.055)*
$\hat{\alpha}$	-11.80	-13.45	-5.38	-11.98	-5.39	-3.98
	(1.058)	*(0.919)*	*(0.951)*	*(1.368)*	*(1.183)*	*(0.066)*
Dof	1.73	3.11	8.83	3.08	8.31	9.62
Copula Likelihood	3,406.4	2,598.8	220.1	2,349.4	199.4	281.8
AIC	-6,812.8	-5,197.6	-440.1	-4,698.3	-398.8	-563.7
			Time-varying Gumbel Copula			
$\hat{\omega}$	3.03	2.63	1.59	0.81	1.10	1.28
	(0.112)	*(0.155)*	*(0.385)*	*(0.014)*	*(0.743)*	*(0.125)*
$\hat{\beta}$	-0.18	-0.16	-0.39	0.26	-0.10	-0.19
	(0.021)	*(0.035)*	*(0.237)*	*(0.002)*	*(0.439)*	*(0.080)*
$\hat{\alpha}$	-7.41	-6.26	-2.12	-1.53	-1.74	-1.78
	(0.553)	*(0.556)*	*(0.452)*	*(0.098)*	*(0.747)*	*(0.128)*
Copula Likelihood	3,122.1	2,441.1	200.0	2,210.5	180.4	253.7
AIC	-6,244.1	-4,882.2	-400.0	-4,420.9	-360.9	-507.5

Time-varying Clayton Copula						
$\hat{\omega}$	3.26	3.26	1.57	0.98	1.59	0.76
	(0.112)	*(0.055)*	*(0.138)*	*(0.002)*	*(0.258)*	*(0.266)*
$\hat{\beta}$	-0.16	-0.28	-0.46	0.27	-0.22	0.37
	(0.019)	*(0.003)*	*(0.108)*	*(0.000)*	*(0.191)*	*(0.181)*
$\hat{\alpha}$	-8.90	-7.18	-2.60	-0.58	-3.25	-0.89
	(0.702)	*(0.347)*	*(0.517)*	*(0.029)*	*(0.752)*	*(0.693)*
Copula Likelihood	2,399.1	1,907.2	158.0	1,778.9	153.6	209.9
AIC	-4,798.2	-3,814.5	-316.0	-3,557.9	-307.2	-419.9
Time-varying Frank Copula						
$\hat{\omega}$	30.54	1.28	2.24	1.44	2.46	2.64
	(1.506)	*(0.000)*	*(0.025)*	*(0.002)*	*(0.000)*	*(0.002)*
$\hat{\beta}$	-0.69	0.93	0.52	0.92	0.48	0.50
	(0.078)	*(0.003)*	*(0.006)*	*(0.004)*	*(0.000)*	*(0.014)*
$\hat{\alpha}$	-80.58	-4.53	-4.40	-5.02	-5.21	-5.31
	(7.006)	*(0.006)*	*(0.021)*	*(0.000)*	*(0.000)*	*(0.000)*
Copula Likelihood	3,070.1	2,461.8	205.7	2,223.3	179.2	265.9
AIC	-6,140.2	-4,923.5	-411.4	-4,446.5	-358.4	-531.8
Time-varying SJC Copula						
$\hat{\omega}_U$	0.48	3.91	2.30	3.86	0.27	1.33
	(0.000)	*(0.000)*	*(0.880)*	*(0.000)*	*(1.861)*	*(0.293)*
$\hat{\beta}_U$	-4.33	-3.98	-13.89	-9.91	-6.44	-7.81
	(0.000)	*(0.000)*	*(3.494)*	*(0.000)*	*(4.673)*	*(1.194)*
$\hat{\alpha}_U$	1.67	-3.28	-1.66	-2.50	-0.23	-2.36
	(0.000)	*(0.000)*	*(1.158)*	*(0.000)*	*(3.928)*	*(0.587)*
$\hat{\omega}_L$	0.64	-1.44	1.73	1.11	2.49	-0.76
	(0.000)	*(0.000)*	*(0.816)*	*(0.000)*	*(0.966)*	*(0.454)*
$\hat{\beta}_L$	-3.36	-3.86	-9.55	-9.91	-15.71	-4.38
	(0.000)	*(0.000)*	*(3.744)*	*(0.000)*	*(3.514)*	*(1.625)*
$\hat{\alpha}_L$	0.95	3.93	-4.09	1.28	-1.30	2.51
	(0.000)	*(0.000)*	*(1.524)*	*(0.000)*	*(1.337)*	*(0.616)*
Copula Likelihood	2,944.5	2,393.8	212.3	2,204.6	200.4	271.0
AIC	-5,889.0	-4,787.6	-424.6	-4,409.3	-400.8	-542.0

** Indicates significant at the 5% level; standard errors are presented in parentheses*
*** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Table 20 Ranking of the Copula Models

Ranking	CL/CO	CL/HO	CL/NG	CO/HO	CO/NG	HO/NG	Total Ranking
Constant Normal Copula	10	10	10	10	11	10	**10**
Constant *t* Copula	2	2	5	2	4	5	**2**
Constant Gumbel Copula	5	7	9	7	9	9	**9**
Constant Clayton Copula	12	12	12	12	12	12	**12**
Constant Frank Copula	6	6	8	6	8	8	**7**
Constant SJC Copula	8	8	7	8	7	7	**8**
Time-varying Normal Copula	9	9	4	9	3	2	**6**
Time-varying *t* Copula	1	1	1	1	2	1	**1**
Time-varying Gumbel Copula	3	4	6	4	5	6	**5**
Time-varying Clayton Copula	11	11	11	11	10	11	**11**
Time-varying Frank Copula	4	3	3	3	6	4	**3**
Time-varying SJC Copula	7	5	2	5	1	3	**3**

6.4. Tail Dependence

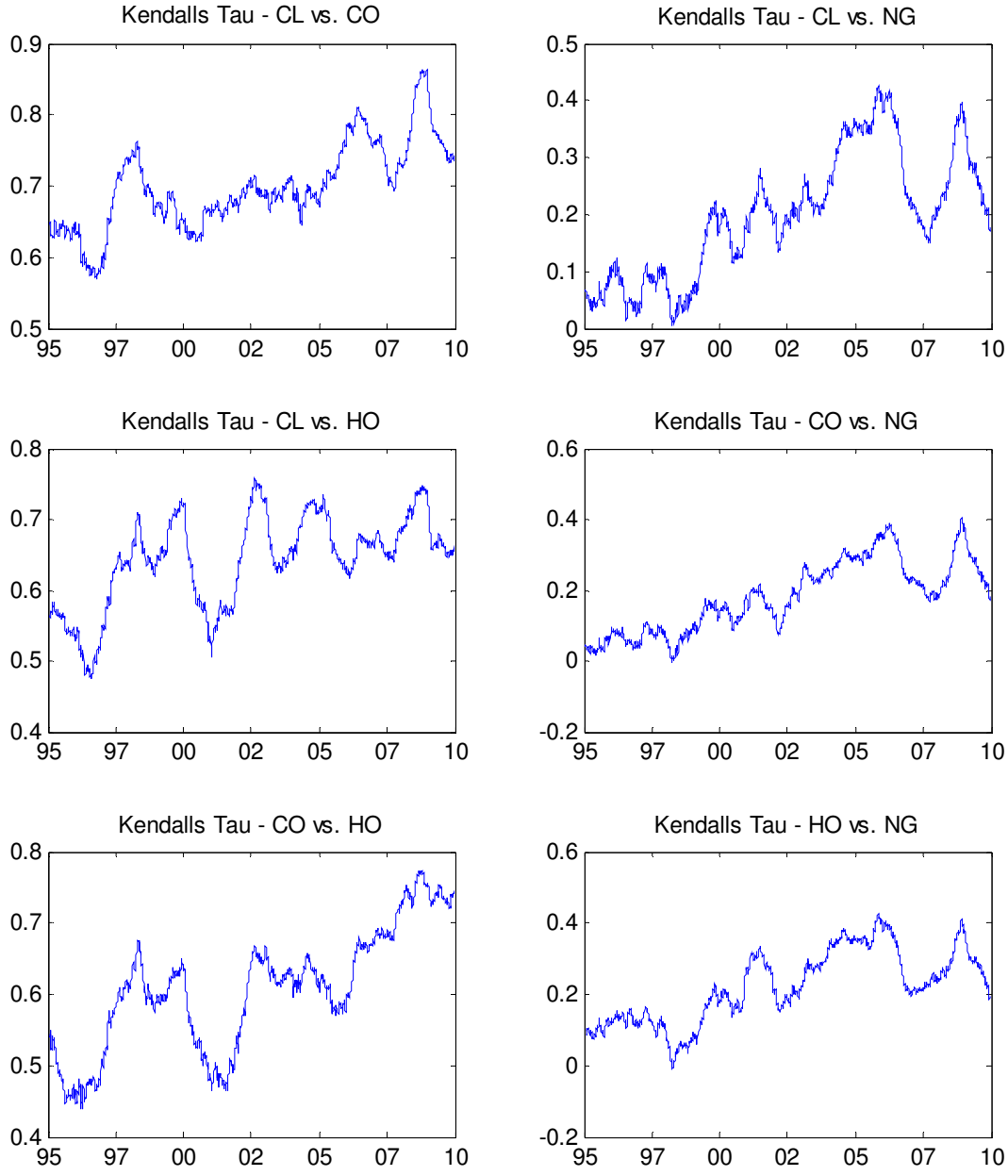
Based on the results from Tables 18 to 20, we focus on the *t*-copula and SJC copula from this point forward. The parameters of interest in this chapter are the correlation (ρ) in the *t* copula and the upper tail dependence (τ^U) and the lower tail dependence (τ^L) in the SJC copula. The contour plots of a *t*-distribution vary with the correlation (ρ) in a way similar to the Normal distribution. When ρ=0, the contours of the distribution are symmetric around the mean. As ρ increases, the contours stretch out along the 45 degree line. Results in Table 18 show that after allowing for fat tails, the standardized residuals from the marginal models show higher correlation as evidenced by higher estimated correlation coefficients in the constant *t*-coupla model than in the constant Normal copula model. Tail dependence refers to the degree of dependence in the corner of the lower-left quadrant or upper-right quadrant of a bivariate distribution. When $\tau^U = \tau^L$, the

distribution is symmetric (see Figure 24 in Appendix A). When $\tau^U > \tau^L$, the upper tail of the contour gets thinner, meaning that there is a greater probability that the energy futures will move higher jointly. Similarly, when $\tau^U < \tau^L$, the lower tail of the contour becomes thinner and the probability that the energy futures will move lower jointly is higher.

Although the *t*-copula model generates the highest log-likelihood value as well as the smallest AIC and BIC values, the symmetric nature of the distribution does not reveal more information about the dependence structure over and above the correlation coefficient. For that reason, we now investigate estimates from the SJC copula. In general, the upper tail dependence is higher than the lower tail dependence (except for the heating oil and natural gas pair), meaning that energy futures prices are more likely to move up together than go down together. Similar to previous research on the dependence structure between financial assets, our results confirm that the dependence structure is not symmetric. However, most of the previous studies show higher lower tail dependence, meaning that financial assets are more dependent during extreme downturns than upturns of the markets. In addition, the difference between the upper and lower tail dependence is significant for pairs of petroleum futures only.

In order to investigate the time-varying behaviour of the dependences between energy futures, we first take a look at 250-trading day rolling Kendall statistic between the standardized residuals. The results are plotted in Figure 14. It is clear that the dependence relationship between energy futures is not constant over time.

Figure 14 250-Trading Day Rolling Kendall's Tau



** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

For example, the dependence between WTI and natural gas was very low in late 1990's, varying between 0 to 0.1 in that period. Over the first half of the 2000's, the dependence oscillated around 0.2 before it started moving higher to around 0.4 by the end of 2006. Over the last four years, the dependence has experienced much larger ovements between 0.18 and 0.4. This preliminary analysis supports our hypothesis that the dependence is not constant over time.

The results of time-varying *t*-copula and SJC copula are presented in panel 2 and 5 of Table 19. The parameter estimates are mostly significant at the five percent level, suggesting that the correlation coefficient and tail dependence are time varying. For the *t*-copula, the estimate of β_ρ is always negative, meaning that the time-varying *t*-copulas exhibit negative correlation between energy futures pairs. The estimates of the parameter α_ρ of the forcing variable are also always negative, meaning high variation in the dependency between energy futures movements over the last ten trading days reduces correlation between them in the current period.

The top panel of Figure 15 below plots the time path of the time-varying correlation coefficients from the *t*-copula between crude oil and natural gas. For the SJC copula, the estimate of β_ρ is negative, indicating that extreme co-movements do not happen in clusters. The forcing variable estimates for the upper tail dependence are mostly negative, except for the WTI and Brent pair, and mostly positive for the lower tail dependence, except for the WTI and Natural Gas as well as Brent and Natural Gas pair. As Patton (2006a) suggested, a negative forcing variable parameter implies that a fall in the mean

absolute difference $\left|u_{t-j} - v_{t-j}\right|$ leads to an increase in tail dependence. Therefore, an unexpected shock between markets leads to less tail dependence.

For space reasons in this main chapter, the time-varying correlation coefficients and tail dependence parameters for all pairs are provided in plots in Appendix A. whereas summary statistics are presented in Table 21 below. The averages of the difference between the upper and lower tail dependence estimates provide information about dynamic asymmetries. The differences are all positive, suggesting a higher probability of joint extreme events during bull markets than during bear markets. However, they are only significantly different from zero for the pair of WTI and Brent. In addition, volatility is evident in the plots provided in Figure 15. As energy prices are strongly influenced by underlying fundamental supply demand factors, it is important to understand how correlation relationships and tail dependence structures change if we take into account of the supply demand factors in these markets.

Figure 15 Time Path of Time-Varying Correlation Coefficient and Tail Dependence – An Example (CL vs. NG)

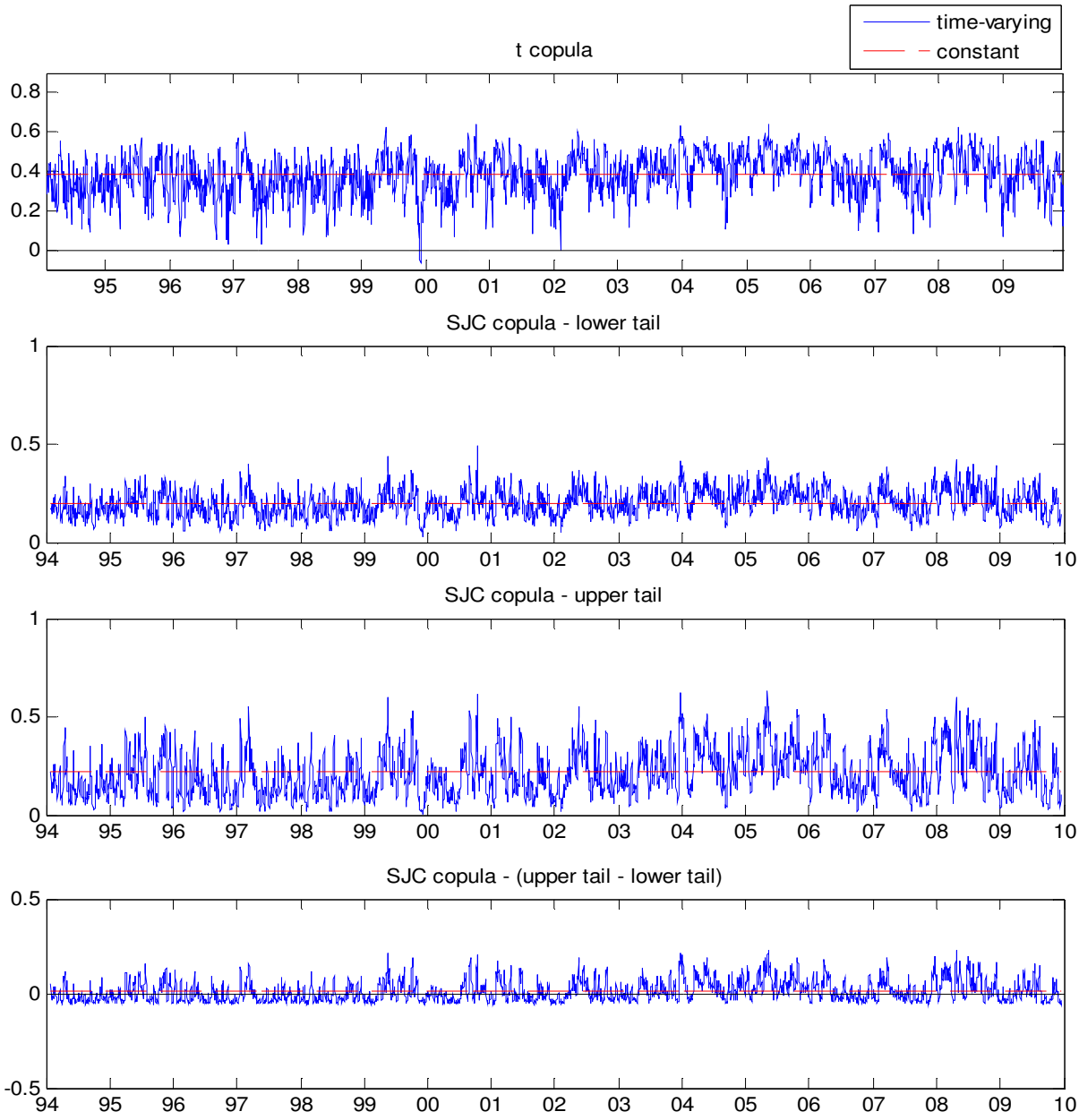


Table 21 Summary of Time-Varying Correlation Coefficient and Tail Dependence Parameters

	CL/CO	CL/HO	CO/HO	CL/NG	CO/NG	HO/NG
Correlation Coefficient (*t*-copula)						
Maximum	0.964	0.953	0.947	0.751	0.769	0.653
Minimum	0.170	0.050	0.086	-0.066	-0.046	-0.030
Mean	0.912*	0.865*	0.844*	0.380*	0.363*	0.426*
Standard Deviation	0.041	0.075	0.082	0.104	0.107	0.099
Upper Tail Dependence (SJC copula)						
Maximum	0.864	0.787	0.833	0.810	0.531	0.450
Minimum	0.452	0.580	0.283	0.006	0.054	0.053
Mean	0.805*	0.738*	0.706*	0.210	0.207*	0.250*
Standard Deviation	0.047	0.024	0.072	0.115	0.065	0.068
Lower Tail Dependence (SJC copula)						
Maximum	0.791	0.862	0.876	0.644	0.882	0.599
Minimum	0.499	0.072	0.099	0.028	0.007	0.054
Mean	0.737*	0.644*	0.670*	0.200*	0.189	0.244*
Standard Deviation	0.038	0.206	0.147	0.066	0.122	0.092
(Upper - Lower) Tail Dependence (SJC copula)						
Maximum	0.075	0.584	0.381	0.379	0.078	0.101
Minimum	-0.048	-0.094	-0.066	-0.063	-0.350	-0.186
Mean	0.068*	0.094	0.036	0.010	0.018	0.006
Standard Deviation	0.010	0.188	0.080	0.054	0.058	0.039

** Indicates significant at the 5% level*
*** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

6.5. Tail Dependence with Fundamental Factors

The last section highlighted the significant variation over time in correlation and tail dependence parameter estimates. We now introduce exogenous variables into the models to examine the robustness of the previous results in this direction. In order to avoid confusion, we call the copula models examined in the previous sections the "plain" models; these only included lagged endogenous explanatory variables. Energy prices are highly correlated with underlying fundamental factors, as energy prices adjust to imbalances of supply and demand when the market is efficient. However, as imbalances in supply and demand are unobservable in real life, investors turn to energy inventory, especially changes in energy inventory, for a better understanding of the supply demand balance. For natural gas, short-term storage changes are mostly determined by seasonal demand for heating (in the winter) and cooling (in the summer). Therefore, we use the deviation of storage changes from its five year average as the exogenous variable for natural gas and changes of inventory for petroleum products. Weekly changes in storage for both oil and gas have been downloaded from the Energy Information Administration's website[19]. As storage numbers are only recorded weekly, we have converted the weekly changes in inventory into a daily time series by assuming equal weights during the week. As discussed previously, for the conditional copula, the conditional vector W is the same for both the marginal distributions and the copula. Consequently, we include both exogenous variables in all four marginal models as well as the copula models for the six pairs.

[19] www.eia.com

Table 22 Results of Marginal Distributions with Fundamental Factors as Explanatory Variables

	CL	CO	HO	NG
Constant	0.097*	0.103*	0.067*	-0.003
	(0.040)	*(0.038)*	*(0.041)*	*(0.057)*
AR(1)	-0.024	-0.054*	-0.039	-0.050
	(0.018)	*(0.019)*	*(0.019)*	*(0.019)*
Petro_Stock_Change	-17.566*	-9.965	-29.117*	-11.539
	(8.695)	*(8.221)*	*(8.657)*	*(11.939)*
NG_Delta	-1.492	-1.474	-2.674*	-7.893*
	(1.335)	*(1.273)*	*(1.346)*	*(1.929)*
GARCH constant	0.108*	0.076*	0.131*	0.315*
	(0.028)	*(0.024)*	*(0.037)*	*(0.086)*
GARCH (1)	0.942*	0.947*	0.941*	0.914*
	(0.010)	*(0.010)*	*(0.011)*	*(0.013)*
ARCH (1)	0.040*	0.039*	0.038*	0.065*
	(0.007)	*(0.007)*	*(0.007)*	*(0.010)*
Degress of freedom	7.988*	8.678*	7.605*	6.066*
	(1.006)	*(1.191)*	*(0.909)*	*(0.601)*

** Indicates significant at the 5% level; standard errors are presented in parentheses*
*** Petro_Stock_Change captures weekly changes in petroleum stocks*
**** NG_Delta represents the deviation of weekly inventory changes from its five year average*

Table 23 Conditional Correlation Measure Matrix with Fundamental Factors

	Linear Correlation			**Spearman's Rho**			**Kendall's Tau**		
	CL	**CO**	**HO**	**CL**	**CO**	**HO**	**CL**	**CO**	**HO**
CO	0.86*			0.87*			0.72*		
	(0.000)			*(0.000)*			*(0.000)*		
HO	0.81*	0.80*		0.84*	0.81*		0.66*	0.63*	
	(0.000)	*(0.000)*		*(0.000)*	*(0.000)*		*(0.000)*	*(0.000)*	
NG	0.32*	0.31*	0.36*	0.34*	0.32*	0.37*	0.23*	0.22*	0.25*
	(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*

** Indicates significant at the 5% level; standard errors are presented in parentheses*

Table 24 Summary of Time-Varying Correlation Coefficient and Tail Dependence Parameters - with Fundamental Factors as Explanatory Variables

	CL/CO	CL/HO	CO/HO	CL/NG	CO/NG	HO/NG
Correlation Coefficient (*t*-copula)						
Maximum	0.998	0.956	0.847	0.824	0.599	0.683
Minimum	0.070	0.055	-0.459	-0.073	-0.122	-0.012
Mean	0.917*	0.879*	0.858*	0.430*	0.407*	0.465*
Standard Deviation	0.047	0.068	0.077	0.113	0.107	0.101
Upper Tail Dependence (SJC copula)						
Maximum	0.813	0.876	0.849	0.897	0.458	0.452
Minimum	0.075	0.434	0.187	0.008	0.085	0.055
Mean	0.467*	0.742*	0.704*	0.241*	0.236*	0.251*
Standard Deviation	0.066	0.061	0.067	0.123	0.067	0.068
Lower Tail Dependence (SJC copula)						
Maximum	0.905	0.893	0.878	0.866	0.690	0.555
Minimum	0.309	0.244	0.175	0.014	0.007	0.062
Mean	0.756*	0.707*	0.704*	0.236*	0.214*	0.304*
Standard Deviation	0.094	0.101	0.109	0.097	0.127	0.090
(Upper - Lower) Tail Dependence (SJC copula)						
Maximum	0.067	0.231	0.305	0.316	0.167	0.002
Minimum	-0.365	-0.022	-0.066	-0.136	-0.277	-0.135
Mean	-0.289*	0.035	0.000	0.006	0.021	-0.053*
Standard Deviation	0.037	0.043	0.046	0.048	0.070	0.026

** Indicates significant at the 5% level*
*** CL: WTI; CO: Brent; HO: Heating Oil; NG: Natural Gas*

Table 22 summarizes results for the marginal models with both fundamental factors and Table 23 shows the conditional correlation measures using the standardized residuals. The marginal effects on inventories for both crude oil and natural gas are negative, which confirms our hypothesis that returns on energy futures is lower when inventory increases. When compared to the conditional correlation measures without fundamental factors presented in Table 17, the correlation measures presented in Table 24 for the petroleum pairs are similar but the correlation is stronger between natural gas and all petroleum products. Unlike crude oil, which is a global commodity, natural gas in North American is determined by supply and demand in the continent. Therefore, natural gas prices are more likely to be influenced by its inventories. Therefore, when taking fundamental factors into consideration, we expect to see a bigger impact on natural gas.

Table 24 displays the descriptive statistics for the time-varying correlation coefficient and tail dependence parameter estimates. Compared to the outcomes reported in Table 21, the difference between the upper and lower tail dependence parameter estimates is now negative for the WTI/Brent and Heating Oil/Natural Gas pairs, with the difference being statistically significantly different than zero. The difference between tail dependence parameter estimates remains positive for all other pairs but the difference is not significantly different than zero. This result shows that it is important to consider the supply and demand balance when studying dependence structure among energy futures. Fundamentally, WTI and Brent are more dependent as both are futures contracts based on crude oil (but traded at different locations). Natural Gas and Heating Oil are substitutes for each other; therefore the relationship between these two futures should be stronger

than any other pairs involving them. After we take into account such underlying fundamental factors, WTI/Brent and Heating Oil/Natural Gas pairs are more likely to move together in a bear market than in a bull market (which is the opposite of the conclusion made before without considering the supply and demand picture but consistent with findings in the literature).

7. Conclusions

Dependence structure is a very important concept in financial time series analysis. It is widely used in pairs trading, hedging, risk management, and portfolio management. Linear correlation, due to its simplicity and ease of calculation, has become the most widely used measure to describe co-movement between two assets. Linear correlation works well in a multivariate Normal world, but it has several shortcomings and may not be robust when facing non-Normal distributions. In this chapter, we employ copula models to study the dependence structure of energy futures. In order to capture the stylized features of financial returns, we follow Patton (2006a) and employ a time-varying conditional copula method, which allows dependence parameters to vary over time. Moreover, we investigate whether the parameters are affected by supply and demand balance of oil and gas. Our results indicate that:

1. For the constant copula models which only include lagged endogenous explanatory variables, the difference between upper tail dependence and lower tail dependence is generally positive, which implies that energy futures are more likely to move together during bull markets than in bear markets. This result is opposite to most of the previous research on copulas which has found financial assets are more likely to move down together than move up together.

2. Similar to other studies utilizing Patton's (2006a) dynamic evolution structure for the time-varying copula, the dependence parameters can be very volatile over time and deviate from their constant levels frequently.
3. For the time-varying copula models which only include lagged endogenous explanatory variables, the means of the difference for all pairs are positive, suggesting a higher probability of joint extreme events during a bull market than during a bear market. However, after we introduce the fundamental supply and demand balance into the copula model, the most theoretically associated pairs, WTI/Brent and Heating Oil/Natural Gas appear to be more likely to move together during down markets than up markets. As far as we know, this is the first study that introduces supply and demand balance in modeling the dependence structure of energy futures using copulas. Our results confirm our hypothesis that it is important to consider fundamental supply-demand relationship for modeling energy futures.

CHAPTER FOUR: ESTIMATING PORTFOLIO VALUE AT RISK USING PAIR-COPULAS: AN APPLICATION ON ENERGY FUTURES

1. Introduction

Commodities have recently been advocated by many as an alternative investment vehicle as they "… offer excellent diversification benefits because they have a very low (and negative) correlation to equities."[20] Based on investors' quest for diversification and the introduction of many commodity indices and exchange traded funds (ETFs), assets under management in the commodities markets increased from approximately US$5 billion globally in 2000 to US$80 billion in 2005[21] and reached a record high of US$340 billion in October 2010[22]. However, the benefit of diversification may have been greatly reduced after the recent financial crisis.

Two by-products of the recent financial crisis are high volatility of financial asset prices and high correlation between different asset classes. The Chicago Board Options Exchange Market Volatility Index (VIX)[23], a popular measure of the implied volatility of S&P 500 index options, rose from its historical average of below 20 to as high as 80 after the Lehman Brothers bankruptcy in late 2008.

---

[20] http://www.auspicecapital.com/docs/Auspice_Diverse_1pager_1.0.pdf
[21] http://www.investmentreview.com/analysis-research/commodities-and-the-efficient-frontier-4416
[22] http://www.bloomberg.com/news/2010-11-26/commodity-assets-under-management-climb-to-record-barclays-capital-says.html
[23] http://www.cboe.com/micro/vix/introduction.aspx

Figure 16 Chicago Board Options Exchange Market Volatility Index

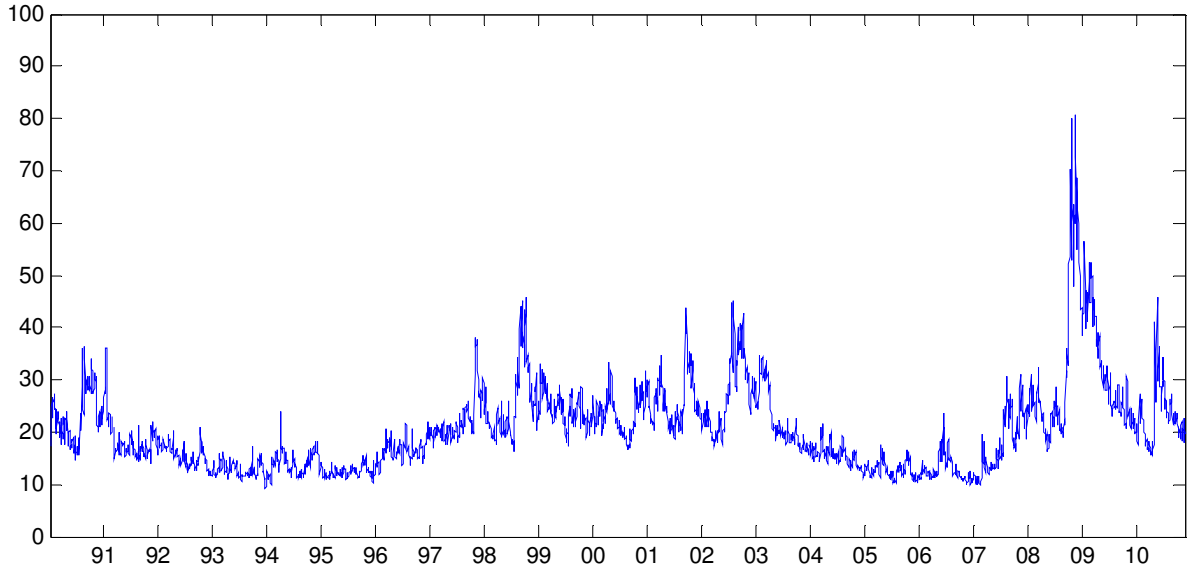
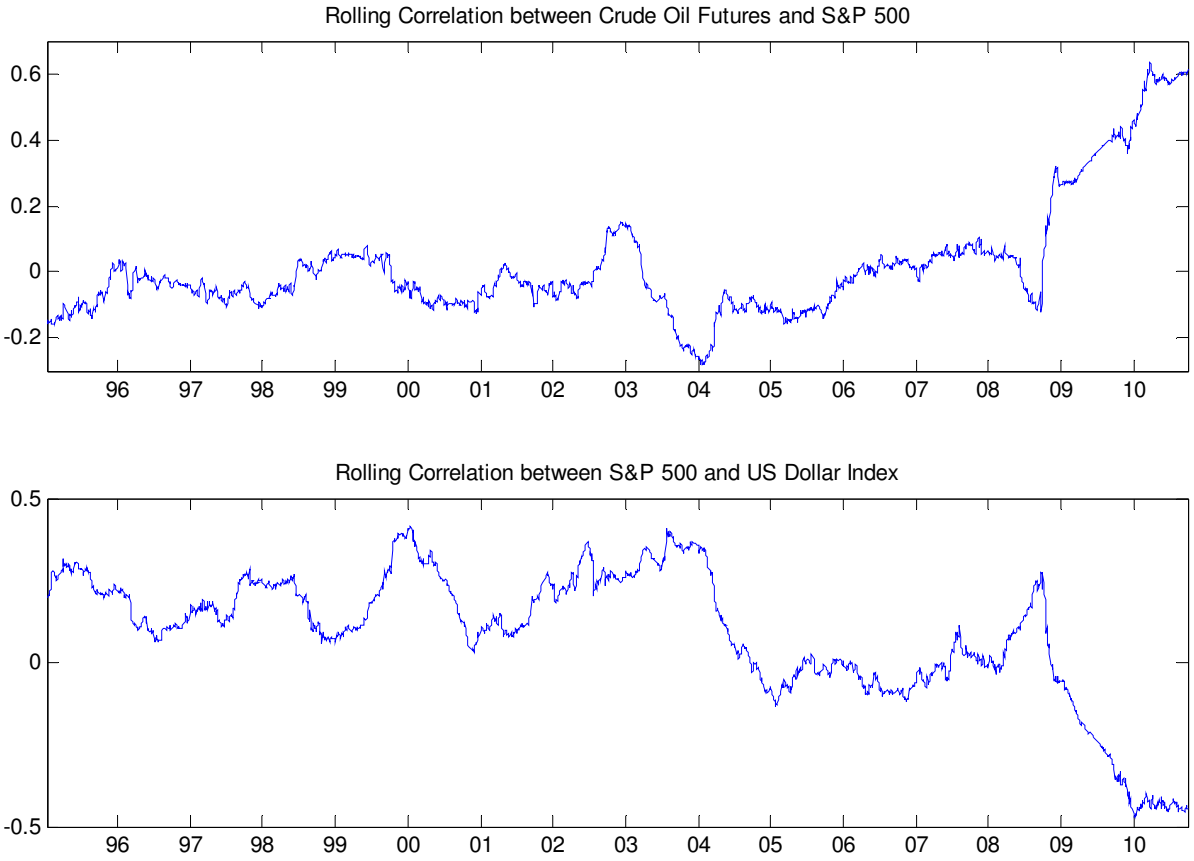


Figure 17 250-Day Rolling Correlation



However, in recent months, the volatility in the VIX index has subsided with it reverting back to its historical range (see Figure 16). On the other hand, the correlation between different financial assets has remained elevated. As Figure 17 indicates, the 250-day rolling correlation between crude oil futures and the S&P 500 index has broken its historical range of -20% to 20% in late 2008 and increased steadily to as high 60% in recent months. A similar pattern can also be observed between the S&P 500 and the US dollar index. As the global economy is slowly recovering from the recent recession, asset prices are largely driven by macro factors such as economic growth, unemployment, interest rate changes, inflation expectations, etc. Therefore, during periods of high macro uncertainty, asset prices tend to move together leading to a high level of correlation. Now more than before, this suggests that investors need to pay attention to the co-movements between assets in their portfolios and understand how the dependencies change asset allocation and risk management for their portfolios.

The theory of copulas provides a framework to account for dependencies within a multi-asset portfolio. Introduced by Sklar (1959), 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. In the past decade, many have explored the advantage of applying the copula theory to the estimation of Value-at-Risk. We specifically think of the works of Palaro and Hotta (2006), Fantazzini (2008), Bastianin (2009) and Caillault and Guegan (2009).

Although many studies have demonstrated that copula functions can improve Value-at-Risk estimation, most of these approaches are based on bivariate copulas or very restricted multivariate copula functions. “The curse of dimensionality” still remains a major obstacle for the popularity of copula theory in the field of risk management. This is mainly due to a lack of construction schemes for higher dimension copulas. From 2005 on, a handful of extensions and innovations have appeared, *e.g.*, the Hierarchical Archimedean copula (HAC), see Savu & Trede (2006); the Generalized multiplicative (GMAC), see Morillas (2005); Liebscher copulas, see Liebscher (2008); Fischer and Köck copulas, see Fischer and Köck (2009); Koehler-Symanowski copulas, see Palmitesta and Provasi (2006); and Pair-copulas decomposition (also called vine copulas), see Aas *et al.* (2006). Comparative discussions on these different approaches are limited. The most comprehensive reviews are presented by Berg and Aas (2008) and Fischer *et al.* (2009). They show that pair-copulas provide a better fit to multivariate financial data than do other multivariate copula constructions.

In this chapter, we employ a vine based pair-copula approach to estimate Value-at-Risk and Expected Shortfall for a portfolio of equally weighted crude oil futures, natural gas futures, S&P 500 index and the US Dollar index. The major advantage of vine based copula models is their flexibility in modeling multivariate dependence. They allow for flexible specification of the dependence between different pairs of marginal distributions individually, while specifying an overall dependence between all of the marginal distributions. We utilize both the canonical-vine and the D-vine structures in this chapter

and test the accuracy of the Value-at-Risk and the Expected Shortfall forecasts at different confidence levels. Results are compared with traditional methods, such as RiskMetrics, Historical Simulation, and the conditional Extreme Value Theory, *etc*. This chapter is among a very small number of empirical studies of higher dimension multivariate copula theory. To the best of our knowledge, this study is the first to explore the benefit of using vine copula theory in the estimation of Value-at-Risk and Expected Shortfall for a diversified portfolio including energy futures. Our results show that pair-copula decomposition doesn't provide any added advantage over the competing models in terms of forecasting Value-at-Risk or Expected Shortfall over a long backtest horizon. However, we found that pair-copula can greatly improved the accuracy of Value-at-Risk and Expected Shortfall forecasts in periods of high co-dependence among assets.

This chapter is organized as follows. Section 2 introduces multivariate copula theory. Section 3 discusses pair-copula decomposition of a multivariate distribution. Our empirical results are presented and discussed in Section 4. Section 5 concludes the chapter.

2. Multivariate Copula

In this chapter, we generalize the bivariate copula theory to the multidimensional case. Let us start from the definition of a copula.

Definition of Copula *A d-dimensional copula is a function C:* $[0, 1]^d \rightarrow [0, 1]$ *with the following properties:*

1. *For every* $u_i \in [0, 1]$, $i=1,\ldots,d$

$$C(u_1,\ldots, u_{i-1}, 0, u_{i+1},\ldots, u_d) = 0.$$

2. *For every* $u_i \in [0, 1]$, $i=1,\ldots,d$

$$C(1,\ldots,1,u_i, 1,\ldots,1) = u_i.$$

3. *For all* $(a_1, \ldots,a_d)$, $(b_1, \ldots,b_d) \in [0, 1]^d$ *with* $a_i \le b_i$:

$$\sum_{i_1=1}^{2} \cdots \sum_{i_1=d}^{2} (-1)^{i_1+\ldots+i_d} C(u_{j1},\ldots u_{jd}) \ge 0.$$

where $u_{j1}= a_j$ *and* $u_{j2}= b_j$, *for all* $j=1,\ldots,d$.

Property 1 is also referred to as the grounded property of a copula. It says that the joint probability of all outcomes is zero if the marginal probability of any outcome is zero. Property 3 is the *d*-dimensional analogue of a non-decreasing one-dimensional function. A function with this feature is therefore called *d*-increasing. A *d*-dimensional copula is the distribution function on $[0, 1]^d$ where all marginal distributions are uniform on [0, 1]. The importance of copula theory in the area of multivariate distributions can be summarized in the following theorem by Sklar, which says that not only can joint distribution functions be written in terms of a copula, but the converse is also true: a copula may be used in conjunction with univariate distribution functions to construct multivariate distributions.

Sklar's Theorem *Let F be a d-dimensional joint distribution function with marginal distributions* $F_1, \ldots,F_d$. *Then there exists a copula C such that:*

$$F(x_1,\ldots,x_d) = C(F_1(x_1),\ldots, F_d(x_d)), \ \forall x_i \in [-\infty,\infty]. \qquad (1)$$

If the F_i are continuous for $i=1,\ldots,d$, then C is unique. Conversely, if C is a copula and F_i are distribution functions, then the function F is a multivariate distribution function with marginals $F_1, \ldots,F_d$. The converse of Sklar's theorem is very useful in modeling multivariate distributions in finance. It implies that if we combine *n* different marginal

distributions with any copula, we have defined a valid multivariate distribution. This provides great flexibility when modeling a portfolio as we can use different marginal distributions for each asset class and use a copula to link them together.

Corollary *Let F be a d-dimensional joint distribution function with continuous marginal distributions* $F_1, \ldots, F_d$ *and copula C, satisfying Sklar's Theorem. Then for* $u_1, \ldots, u_d \in [0, 1]$:

$$C(u_1, \ldots, u_d) = F(F^{-1}(u_1), \ldots, F^{-1}(u_d)) \tag{2}$$

Where the F_i^{-1} denote the inverse marginal distribution functions. This corollary provides a motivation for calling a copula a dependence structure as the copula links the quantiles of the marginal distributions rather than the original variables. Based on this corollary, one of the key properties of a copula is that the dependence structure is unaffected by a monotonically increasing transformation of the variables.

Assuming that the joint distribution function F of $x_1, \ldots, x_d$ is n-times differentiable, the density of the multivariate distribution with respect to copula is the n^{th} partial derivative of equation (1):

$$f(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d); \Theta) \prod_{i=1}^{n} f_i(x_i) \tag{3}$$

where c denotes the density of the copula with parameter(s) Θ, which is defined as

$$c(u_1, \ldots, u_d; \Theta) = \frac{\partial^d C(u_1, \ldots, u_d; \Theta)}{\partial u_1 \ldots \partial u_d}. \tag{4}$$

3. Pair-Copula Decomposition of a Multivariate Distribution

The construction of multi-dimensional distribution functions has been recognized as a difficult problem as the dimension increases. Pair-copulas, originally introduced by Joe (1996), provide a flexible way to construct multivariate distributions. See Bedford and Cooke (2001, 2002), Kurowicha and Cooke (2006) and Aas *et al.* (2009) for a detailed discussion of pair-copula construction. The main idea of pair-copula theory is that multivariate copulas can be decomposed into a cascade of bivariate copulas. Let $f(x_1,\ldots,x_d)$ be a d-dimensional density function and $c(u_1,\ldots,u_d)$ be the corresponding copula density function. Therefore, the multivariate density can be factorized as:

$$\begin{aligned} f(x_1,\ldots,x_d) &= f(x_d \mid x_1,\ldots,x_{d-1}) \cdot f(x_1,\ldots,x_{d-1}) \\ &= f(x_d) \cdot f(x_{d-1} \mid x_d) \cdot f(x_{d-2} \mid x_{d-1}, x_d) \cdots f(x_1 \mid x_2,\ldots,x_d). \end{aligned} \tag{5}$$

The above decomposition is unique up to a re-labelling of the variables. To get a pair-copula construction for $f(x_1,\ldots,x_d)$, we only need to replace each of the conditional densities in (5) bit by bit with products of pair-copulas and marginal densities.

We illustrate the construction of pair-copulas for the two-, three-, and four-dimensional cases before we present the general multivariate case. In the bivariate case, (3) can be simplified to:

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2), \tag{6}$$

where c_{12} is the appropriate pair-copula density for the pair of transformed variables $F_1(x_1)$ and $F_2(x_2)$. Therefore, the conditional density can be calculated as:

$$f(x_2 \mid x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2). \tag{7}$$

This formula can be generalized as:

$$f(x_j \mid x_i) = \frac{f(x_i, x_j)}{f_i(x_i)} = c_{ij}(F_i(x_i), F_j(x_j)) \cdot f_j(x_j)\,. \tag{8}$$

Similar to the bivariate case, the joint three-dimensional density can be decomposed using (5):

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f_{2|1}(x_2 \mid x_1) \cdot f_{3|1,2}(x_3 \mid x_1, x_2). \tag{9}$$

The second term on the right-hand side of (9) can be expressed in terms of the product of a copula and a marginal density using (7). The last term of (9) can be factorized as follows:

$$f_{3|1,2}(x_3 \mid x_1, x_2) = c_{13;2}(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)) \cdot f_{3|2}(x_3 \mid x_2)\,. \tag{10}$$

This decomposition involves a pair-copula and the last term can be further decomposed into another pair-copula using (8) and a marginal distribution. Therefore, for the three-dimensional case, the full decomposition is:

$$\begin{aligned} f(x_1, x_2, x_3) &= f_1(x_1) \cdot f_{2|1}(x_2 \mid x_1) \cdot f_{3|1,2}(x_3 \mid x_1, x_2) \\ &= f_1(x_1) \\ &\quad \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2) \\ &\quad \cdot c_{23;1}(F(x_2 \mid x_1), F(x_3 \mid x_1)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_3(x_3). \end{aligned} \tag{11}$$

Note that this decomposition is not unique anymore as we could have used x_1 instead of x_2 for the conditional variable in (10). The choices of different conditional variables lead to three different pair-copula constructions.

A four-dimensional density can be decomposed in a similar way:

$$f(x_1,x_2,x_3,x_4) = f_1(x_1)\cdot f_{2|1}(x_2 \mid x_1)\cdot f_{3|12}(x_3 \mid x_1,x_2)\cdot f_{4|123}(x_4 \mid x_1,x_2,x_3) \qquad (12)$$

Use (8) repeatedly together with the previous results, we rewrite it in terms of six pair-copulas and the four marginal densities:

$$\begin{aligned} f(x_1,x_2,x_3,x_4) = f_1(x_1) & \\ & \cdot c_{12}(F_1(x_1),F_2(x_2))\cdot f_2(x_2) \\ & \cdot c_{23|1}(F(x_2 \mid x_1),F(x_3 \mid x_1))\cdot c_{13}(F_1(x_1),F_3(x_3))\cdot f_3(x_3) \\ & \cdot c_{34|12}(F(x_3 \mid x_1,x_2),F(x_4 \mid x_1,x_2))\cdot c_{24|1}(F(x_2 \mid x_1),F(x_4 \mid x_1)) \\ & \cdot c_{14}(F_1(x_1),F_4(x_4))\cdot f_4(x_4). \end{aligned} \qquad (13)$$

In the general multivariate case, (5) can be expressed in terms of the marginal densities and a product of $n(n\text{-}1)/2$ bivariate pair-copulas, using the following formula iteratively:

$$f(x \mid \nu) = c_{x v_j|\nu_{-j}}(F_{x|\nu_{-j}}(x \mid \nu_{-j}),F_{v_j|\nu_{-j}}(v_j \mid \nu_{-j}))\cdot f(x \mid \nu_{-j}) \qquad (14)$$

where $\nu = (v_1,...,v_d)$ is a vector, v_j with $j=1,\ldots,d$ the j-th element of this vector and $\nu_{-1} = (v_1,...,v_{j-1},v_{j+1},...,v_d)$ the vector ν without the j-th component. The conditional marginal distributions $F(x \mid \nu)$ in (12) can be calculated recursively. Joe (1996) showed that, for every j,

$$F(x \mid \nu) = \frac{\partial C_{x,v_j|\nu_{-j}}(F(x \mid \nu_{-j}),F(v_j \mid \nu_{-j}))}{\partial F(v_j \mid \nu_{-j})}, \qquad (15)$$

where $C_{ij|k}$ is a bivariate copula distribution function. When v is univariate, and x and v are uniformly distributed on the [0, 1] interval, we have

$$F(x \mid v) = \frac{\partial C_{xv}(x,v)}{\partial v}. \qquad (16)$$

In the sections below, we use the function $h(x,v,\Theta)$, to represent this conditional distribution:

$$h(x,v,\Theta) = F(x \mid v) = \frac{\partial C_{xv}(x,v)}{\partial v}, \tag{17}$$

where the second argument in $h(\cdot)$ always corresponds to the conditioning variable and Θ denotes the set of parameters for the copula of the joint distribution function of x and v. For sampling from pair-copula distributions, we need the inverse of the *h*-function with respect to the first variable *u*, which is $h^{-1}(u,v,\Theta)$.

3.1. Vine Structures

As we have seen in the previous section, there are multiple ways to decompose multivariate density functions into pair-copula constructions. For example, there are 240 different constructions for a five-dimensional density. Bedford and Cooke (2001, 2002) have introduced a graphical model, called a regular vine, to help organize all possible decompositions. An *n*-dimensional vine is represented by *n*-1 trees T_j, $j=1,\ldots,n-1$, which have $n+1-j$ nodes and $n-j$ edges. Each edge of a tree corresponds to a pair-copula density. The edges of tree T_j become the nodes in tree $j+1$. Two nodes in tree T_{j+1} are joined by an edge if the corresponding edges in tree T_j share a node. The whole decomposition is defined by the marginal distributions and the $n(n-1)/2$ bivariate pair-copulas, which do not necessarily need to belong to the same class of copulas. However, regular vine decompositions are very general and include a large number of possible pair-copula decompositions. Here, we concentrate on two special cases of regular vines called *D-vines* and *canonical vines* (Kurowicka and Cooke, (2006)).

Figure 18 A D-Vine with Four Variables, Three Tress and Six Edges. Each Edge May be Associated with a Pair-Copula.

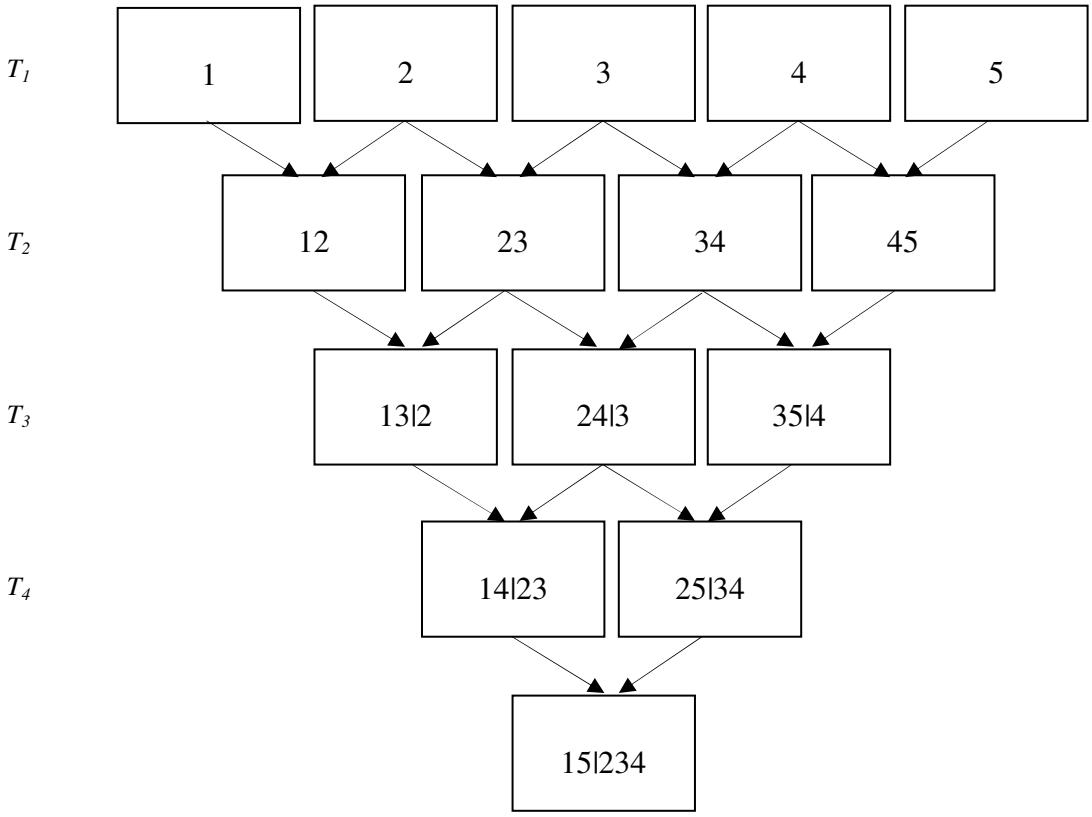
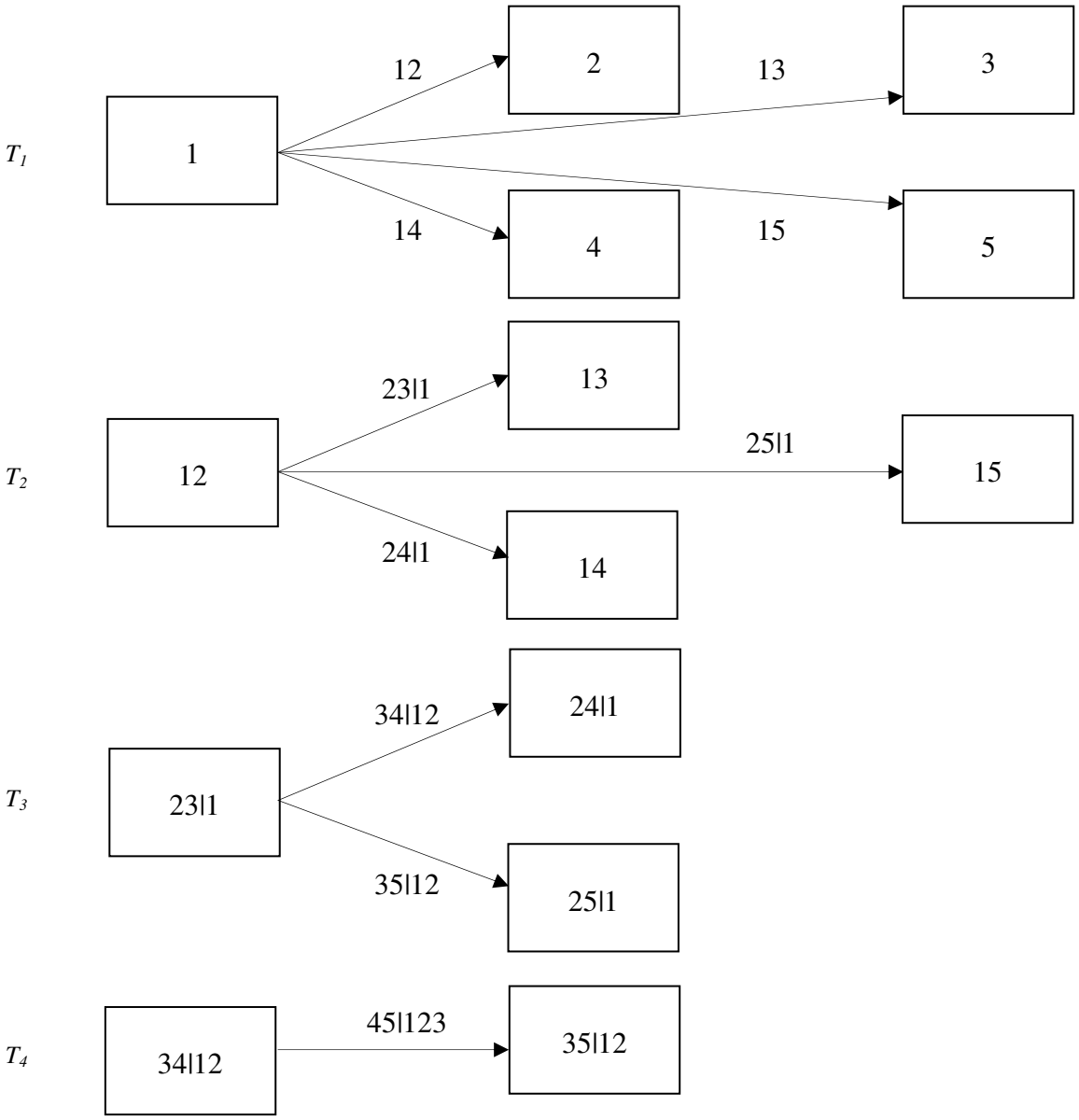


Figure 18 shows the decomposition of a four-dimensional D-vine joint density function into pair-copulas and marginal densities. This D-vine structure is represented by three trees, T_j, j=1,2,3. Tree T_j has 4+1-j nodes and 4-j edges. Each edge corresponds to a pair-copula density and the edge label corresponds to the subscript of the pair-copula density, *e.g.*, edge 14|23 corresponds to the copula density $c_{14|23}(\cdot)$. The whole decomposition is defined by the 4(4-1)/2=6 edges and the marginal densities of each variable. The nodes in tree T_j are only necessary for determining the labels of the edges in tree T_{j+1}.

Figure 19 A Canonical Vine with Four Variables, Three Trees and Six Edges.



For example, in Figure 18, two edges in T_j, which become nodes in T_{j+1}, are joined by an edge in T_{j+1} only if these edge in T_j share a common node. The density $f(x_1,...,x_d)$ of a *d*-dimensional D-vine can be written as:

$$f(x_1,...,x_d) = \prod_{k=1}^{d} f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,...,i+j-1}(F(x_i \mid x_{i+1},...,x_{i+j-1}), F(x_{i+j} \mid x_{i+1},...,x_{i+j-1})), \tag{18}$$

where index *j* identifies the trees, while *i* runs over the edges in each tree.

As we have seen above, in a D-vine, no node in any tree T_j is connected to more than two edges. In a canonical vine, each tree T_j has a unique node that is connected to *n-j* edges. Figure 19 shows a canonical vine with four variables. The *d*-dimensional density corresponding to a canonical vine is given by:

$$f(x_1,...,x_d) = \prod_{k=1}^{d} f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,...,j-1}(F(x_j \mid x_1,...,x_{j-1}), F(x_{j+i} \mid x_1,...,x_{j-1})). \tag{19}$$

As seen in Figure 19, variable 1 is the key variable in the vine and governs the interaction in the data set. Therefore, fitting a canonical vine might be advantageous when a particular variable is known to be a key variable. In such a situation, we can locate the key variable at the root of the canonical vine.

In a three-dimensional vine copula, there are three different decompositions and each of the three decompositions is both a canonical vine and a D-vine. The four-dimensional canonical vine structure can be expressed as:

$$f(x_1, x_2, x_3, x_4) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4)$$

$$\cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot c_{14}(F_1(x_1), F_4(x_4))$$

$$\cdot c_{23|1}(F(x_2 \mid x_1), F(x_3 \mid x_1)) \cdot c_{24|1}(F(x_2 \mid x_1), F(x_4 \mid x_1)$$

$$\cdot c_{34|12}(F(x_3 \mid x_1, x_2), F(x_4 \mid x_1, x_2)), \tag{20}$$

and the D-vine structure can be written as:

$$f(x_1, x_2, x_3, x_4) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4)$$

$$\cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot c_{34}(F_3(x_3), F_4(x_4))$$

$$\cdot c_{13|2}(F(x_1 \mid x_2), F(x_3 \mid x_2)) \cdot c_{24|3}(F(x_2 \mid x_3), F(x_4 \mid x_3)$$

$$\cdot c_{14|23}(F(x_1 \mid x_2, x_3), F(x_4 \mid x_2, x_3)). \tag{21}$$

There are 12 different D-vine decompositions and 12 different canonical vine decompositions. None of the D-vine decompositions are equal to any of the canonical vine decompositions. In total, there are 24 different possible pair-copula decompositions in a four-dimensional case. In the five-dimensional case, there are 60 different D-vines and 60 different canonical vines, and none of the D-vines is equal to any of the canonical vines. In addition, there are also 120 other regular vines for a total of 240 different possible pair-copula decompositions. In general, for the *d*-dimensional case, there are $n!/2$ different canonical vines and $n!/2$ distinct D-vines.

3.2. Estimation of Pair-Copula Decompositions

In this section, we describe how the parameters of the canonical vine density and the D-vine density can be estimated via maximum likelihood. Inference for a regular vine is also possible but the algorithm is not as straightforward. Assume we have *d* variables at *T* time points. The estimation of the parameters can be performed simultaneously for the parameters of the marginal distribution and the pair-copula using the maximum

likelihood method. However, the computation becomes complex very quickly when the dimension gets bigger. Therefore, we use the Inference Function for Margins (IFM) method, where the estimation of the parameters is done in two steps. In the first step, the parameters in the marginal distributions are estimated. In the second step, the copula parameters are estimated conditioned on the previous marginal distributions estimates. For a canonical vine decomposition, the log-likelihood is given by:

$$l(x;\Theta) = \sum_{j=1}^{d-1} \sum_{i=1}^{d-j} \sum_{t=1}^{T} \log(c_{j,j+i|1,\ldots,j-1}(F(x_{j,t} \mid x_{1,t},\ldots,x_{j-1,t}), F(x_{j+i,t} \mid x_{1,t},\ldots,x_{j-1,t}))) . \tag{22}$$

For a D-vine decomposition, the log-likelihood is given by:

$$l(x;\Theta) = \sum_{j=1}^{d-1} \sum_{i=1}^{d-j} \sum_{t=1}^{T} \log(c_{i,i+j|i+1,\ldots,i+j-1}(F(x_{i,t} \mid x_{i+1,t},\ldots,x_{i+j-1,t}), F(x_{i+j,t} \mid x_{i+1,t},\ldots,x_{i+j-1,t}))) . \tag{23}$$

For each copula in the log-likelihood (22) or (23), there is at least one parameter to be estimated. The conditional distributions are determined by using the recursive relation (15) and the appropriate $h(\cdot)$ function (17). For the numerical maximization of the log-likelihood above, we apply the algorithm presented in Aas *et al.* (2009) along with adequate starting values. The starting values for the numerical maximization of the log-likelihood can be determined as follows:

1. Estimate the copula parameters of the first tree using the original data. That is, fit each bivariate copula to the observations.
2. Compute the implied observations for tree two using the copula parameters from the first tree and the appropriate $h(\cdot)$ functions.

3. Estimate the parameters of the copulas in tree two from the observations in step two.
4. Compute the implied observations from tree three using the copula parameters from step three and the appropriate $h(\cdot)$ functions.
5. Continue along the sequence of trees in the pair-copula decompositions.

Note that each estimation step is easy to perform as the data are only two-dimensional. At the end of this procedure, we obtain a set of parameters as starting values for the numerical algorithm.

One remaining question regarding the above procedure is which pair-copula to choose. For smaller dimensions, one may want to estimate the parameters for all possible decompositions using the procedure above and compare the resulting log-likelihoods. However, this is practically infeasible when the dimension is high. One should instead determine which pair-copula to use at each tree level. For example, for the decomposition in (11), we need to decide which pair-copula to use for $c_{12}(\cdot,\cdot)$, $c_{13}(\cdot,\cdot)$, and $c_{23|1}(\cdot,\cdot)$. The pair-copulas do not have to belong to the same family. One straightforward approach to select a copula between non-nested parametric copulas estimated by maximum likelihood is to use either the Akaike or (Schwarz) Bayesian information criterion. For example, Akaike's information criterion (AIC) is defined as:

$$AIC(M) = -2\ln(\hat{L}) + 2M$$

where M is the number of parameters being estimated and $\ln(\hat{L})$ is the maximized log likelihood value. Smaller AIC values indicate a better fit. Therefore, for a pair-copula

decomposition with a high dimension, we use the modified sequential estimation procedure mentioned above:

1. Determine which pair-copula to use in tree one by minimizing the AIC values.
2. Estimate the copula parameters of the first tree using the original data.
3. Compute the implied observations from tree two using the copula parameters from the first tree and the appropriate $h(\cdot)$ functions.
4. Determine which copula type to use in tree two in the same way as in tree one.
5. *etc*.

The observations used to select the pair-copula at each tree level depend on the specific pair-copula chosen at the previous level. However, the resulting pair-copula decomposition may not represent a globally optimal fit.

3.3. Value-at-Risk Monte Carlo Simulation using Pair-Copula Decomposition

In order to forecast the Value-at-Risk of the given portfolio for the next day, we need to simulate from the estimated pair-copula decomposition using the algorithm proposed in Aas *et al.* (2009). For simplicity, we assume that the margins of the distribution are uniform. To simulate from a chosen pair-copula, we first sample $w_1,...,w_n$ independent uniform variables. Then set,

$$x_1 = w_1$$

$$x_2 = F^{-1}(w_2 \mid x_1),$$

$$x_3 = F^{-1}(w_3 \mid x_1, x_2),$$

...

$$x_n = F^{-1}(w_n \mid x_1,...,x_{n-1}).$$

Each $F(x_j \mid x_1,...,x_{j-1})$ can be determined using the *h*-function in (17) and the relationship in (15) recursively for the vine structures. The simulated vine copula observations are then transformed using the inverse cumulative distribution function of the chosen marginal distribution. The resulting standardized error terms together with the estimated parameters of the GARCH model are then used to compute the log returns of each asset in the portfolio. Finally, we calculate the value of the portfolio for each of the simulation and use the empirical quantile function to estimate Value-at-Risk and Expected Shortfall at different confidence levels.

4. Empirical Results

4.1. Data Description and Preliminary Analysis

For our empirical study, we consider a portfolio consisting of crude oil futures, natural gas futures, the S&P 500 index as a proxy of a diversified equity portfolio and the US dollar index. End-of-day prices of the front month[24] futures of West Texas Intermediate crude oil (CL), Natural Gas (NG), the S&P 500 Index (SP), and the US dollar index (US) have been obtained from the Energy Information Administration (EIA)[25] and Bloomberg[26] from Jan 13, 1994 to Sept 28, 2010. This time period is chosen to accommodate the longest sample period (Jan 13, 1994 was the first day that Natural Gas

[24] "Front month" refers to the contract month with an expiration date closest to the current date. Front month contracts are generally the most liquid of futures contracts and the closest to the spot price on the underlying commodity.

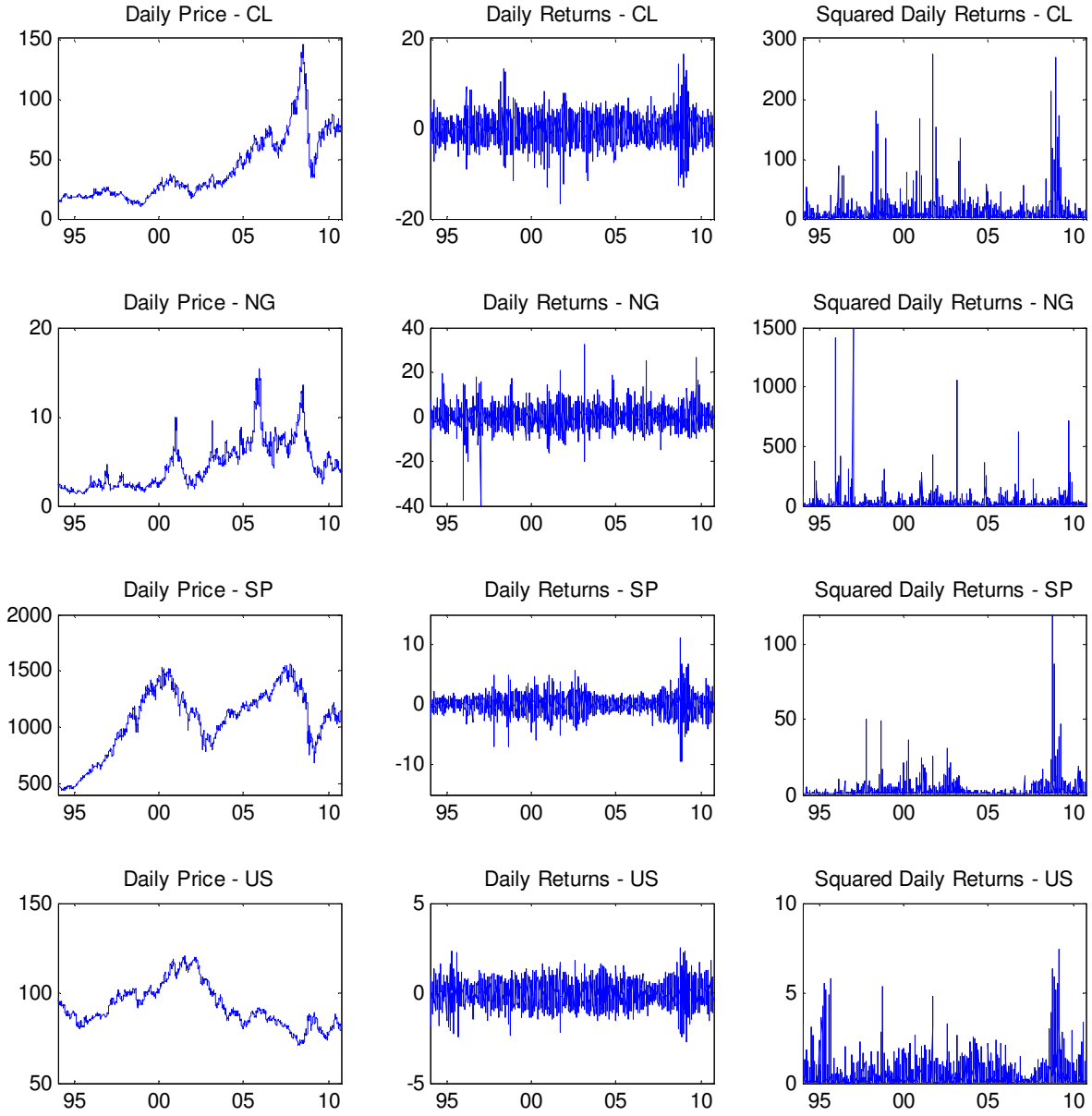
[25] http://www.eia.doe.gov/

[26] http://www.bloomberg.com

Futures price was available *via* EIA's website). Table 25 below gives summary statistics for all returns. As usual, the daily returns are defined as:

$$r_t = \log \frac{P_t}{P_{t-1}} \times 100.$$

Figure 20 Daily Prices, Returns and Squared Returns



** CL: WTI; NG: Natural Gas; SP: S&P 500; US: US Dollar Index*

Table 25 Summary Descriptive Statistics

	CL	NG	SP	US
Sample Size	4,181	4,181	4,181	4,181
Mean	0.040	0.013	0.021	-0.005
Median	0.070	0.000	0.062	-0.004
Max	16.410	32.435	10.957	2.520
Min	-16.545	-40.110	-9.470	-2.726
Std. dev.	2.448	3.820	1.255	0.522
Skewness	-0.110	-0.061	-0.206	-0.099
Kurtosis	7.093	12.361	11.177	4.772
Jarque-Bera Test	2,927*	15,267*	11,677*	554*
p-value	*0.000*	*0.000*	*0.000*	*0.000*
ADF Unit-Root Test	-48.177*	-67.263*	-50.317*	-65.872*
p-value	*0.000*	*0.000*	*0.000*	*0.000*
Auto Corr - *r*				
Lag 1	-0.006	-0.040*	-0.068*	-0.019
Lag 5	-0.035*	-0.021*	-0.036*	-0.001
Lag 10	0.001*	0.023*	0.029*	-0.014
Lag 20	0.012*	-0.006*	0.017*	0.014
Ljung-Box (20)	35.8*	26.0*	101.2*	21.5
p-value	*0.016*	*0.001*	*0.000*	*0.369*
Auto Corr - *r2*				
Lag 1	0.149*	0.055*	0.202*	0.077*
Lag 5	0.151*	0.057*	0.324*	0.054*
Lag 10	0.157*	0.063*	0.255*	0.077*
Lag 20	0.114*	0.037*	0.211*	0.023*
Ljung-Box (20)	1923.8*	291.6*	5522.2*	583.5*
p-value	*0.000*	*0.000*	*0.000*	*0.000*

** Indicates significant at the 5% level*
*** An intercept term (no trend) is included in the integrating regression for the ADF Unit-Root test.*
**** CL: WTI; NG: Natural Gas; SP: S&P 500; US: US Dollar Index*

Figure 20 plots daily prices, returns and squared returns for each series analyzed. Each plot of daily returns confirms the typical empirical time series properties for such data. All series of returns appear to be mean reverting and exhibit periods of low volatility followed by periods of extreme volatility. The squared daily returns exhibit evidence of volatility clustering, *i.e.* large changes tend to be followed by large changes. For risk management purposes, any measurement of risk exposure should be conditional on the current volatility regime. Basic statistics for each series are summarized in Table 25. As is the case with most financial time series, our data exhibit evidence of fat tails. The kurtosis values of all series are far larger than three. The Jarque-Bera normality test also suggests that all returns are not normally distributed.

The test statistics shown in Table 26 suggest that the null hypothesis of no ARCH effect is rejected for all series and the null hypothesis of no GARCH effect is also rejected for all series. This result is not sensitive to the number of lags included in the ARCH-LM and GARCH tests. Given the above test results, the GARCH effects need to be filtered out from the data before the extreme value analysis can be applied to the data.

Table 26 (G)ARCH Effect Test

Testing for ARCH(1)	**CL**	**NG**	**SP**	**US**
F-statistic	95.9*	12.2*	150.4*	25.7*
p-value	*0.0000*	*0.0005*	*0.0000*	*0.0000*
Testing for GARCH(1,1)	**CL**	**NG**	**SP**	**US**
nR^2	51.5*	7.2*	94.4*	13.4*
p-value	*0.0000*	*0.0007*	*0.0000*	*0.0000*

** Indicates significant at the 5% level*
*** CL: WTI; NG: Natural Gas; SP: S&P 500; US: US Dollar Index*

4.2. The Models for the Marginal Distributions

We adopt the Inference Function for Margins (IFM) method, where the estimation of the parameters is done in two steps. The first step is to estimate the marginal models. The model used for the marginal distribution is AR(*1*) – GARCH(*1,1*) – *t* as described below:

$$r_t = \mu + ar_{t-1} + \varepsilon_t$$

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

$$\varepsilon_t \cdot \sqrt{\frac{\upsilon}{\sigma_t^2(\upsilon-2)}} \sim iid\ t_\upsilon$$

where r_t is the return at time *t*. The parameter estimates and standard errors for the marginal distribution models are presented in Table 27. All estimated parameters for the variance equation are highly significant. The autoregressive effect in the volatility specification is strong as it is in the range of 0.90 (natural gas) to 0.97 (US dollar index). In addition, the condition for covariance-stationarity is satisfied since GARCH (1) +

Table 27 Results for the Marginal Distribution

	CL	NG	SP	US
Constant	0.076*	-0.003	0.070*	-0.003
	(0.031)	*(0.046)*	*(0.013)*	*(0.007)*
AR(1)	0.002	-0.020	-0.033*	-0.032*
	(0.016)	*(0.016)*	*(0.017)*	*(0.016)*
GARCH constant	0.045*	0.323*	0.007*	0.001*
	(0.014)	*(0.069)*	*(0.002)*	*(0.000)*
GARCH (1)	0.954*	0.901*	0.926*	0.967*
	(0.005)	*(0.011)*	*(0.007)*	*(0.005)*
ARCH (1)	0.038*	0.077*	0.072*	0.030*
	(0.005)	*(0.010)*	*(0.008)*	*(0.005)*
Degress of freedom	7.643*	5.559*	7.232*	7.442*
	(0.800)	*(0.430)*	*(0.724)*	*(0.905)*
Residual ARCH Test	**CL**	**NG**	**SP**	**US**
F-statistic	4.661	0.259	3.258	0.345
P-value	*(0.170)*	*(0.611)*	*(0.071)*	*(0.557)*

** Indicates significant at the 5% level; standard errors are presented in parentheses*
*** CL: WTI; NG: Natural Gas; SP: S&P 500; US: US Dollar Index*

ARCH (1) <1 for all the series. Table 27 also presents the ARCH effect test statistics on the residuals. As the ARCH-LM test confirms that there are no ARCH effects left in the residuals, the analysis in the rest of this chapter focuses on these standardized residuals.

4.3. Pair-copula Decomposition Estimation

The standardized residuals obtained from the previous section have been transformed to standard uniform series to estimate the pair-copulas. For a four-dimensional pair-copula decomposition, there are 12 canonical vine structures. For each canonical vine structure,

there are six pair-copulas. We consider three different copulas for the selection of pair-copula, these being the *t*-copula, Clayton copula and SJC copula. Therefore, the number of different ways of estimating a four-dimensional canonical vine copula is 3^6 x 12= 8,748, which is also the same number of possible ways of estimating a D-vine copula. Therefore, for our empirical analysis, we use the modified sequential estimation procedure mentioned above in section three.

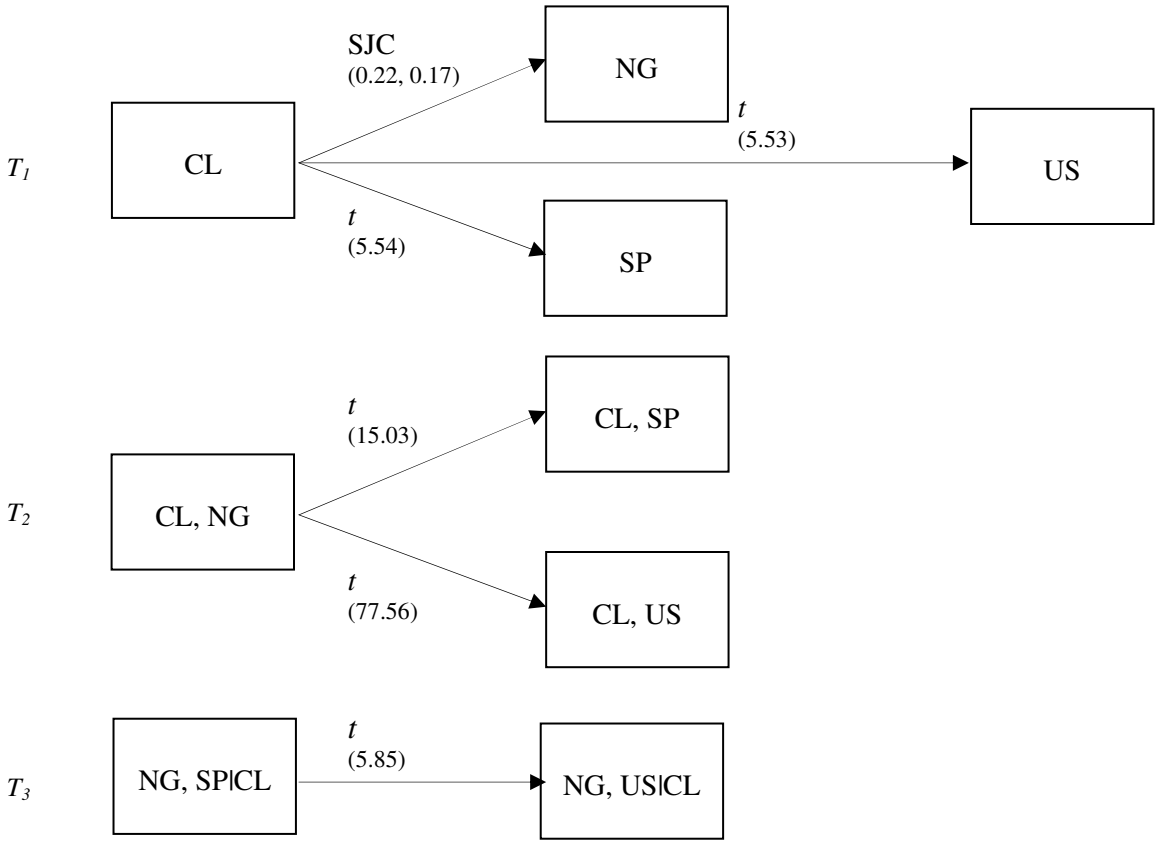
In addition, before conducting the modified sequential estimation procedure for pair-copula selection, we need to determine which canonical vine structure to use. The advantage of the canonical vine structure is that we can locate the key variable that governs interactions in the dataset at the root of the vine. We identify this variable by selecting the series that has the highest dependence relationship with other series. As shown in Table 28, crude oil futures exhibit the highest dependence among the series. We have decided to locate crude oil futures at the root of the canonical vine. The chosen pair-copula decomposition, along with best copula fits and parameter estimates for the canonical vine is shown in Figure 21. In the case of the *t*-copula, the parameter is the degrees of freedom. For the SJC copula, the estimates are the upper and lower tail dependence coefficients. Note that except for one pair-copula in tree one (SJC for crude oil futures and natural gas futures), all other pair-copulas are all *t*-copula. The chosen D-vine pair-copula decomposition, along with best copula fits and parameter estimates is shown in Figure 22.

Table 28 Correlation Measure Matrix of the Transformed Margins

	Linear Correlation			Spearman's Rho			Kendall's Tau		
	CL	**NG**	**SP**	**CL**	**NG**	**SP**	**CL**	**NG**	**SP**
NG	0.28*			0.28*			0.19*		
	(0.000)			*(0.000)*			*(0.000)*		
SP	0.04*	0.01		0.04*	0.01		0.03*	0.01	
	(0.012)	*(0.493)*		*(0.013)*	*(0.531)*		*(0.012)*	*(0.527)*	
US	-0.14*	-0.05*	0.08*	-0.14*	-0.05*	0.08*	-0.09*	-0.03*	0.05*
	(0.000)	*(0.001)*	*(0.000)*	*(0.000)*	*(0.002)*	*(0.000)*	*(0.000)*	*(0.002)*	*(0.000)*

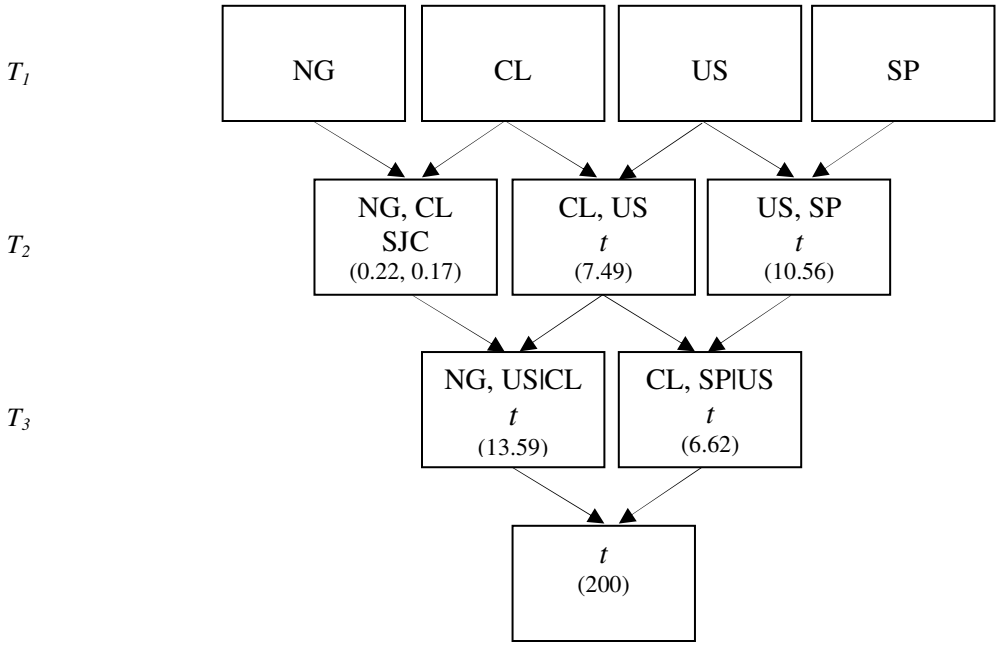
** Indicates significant at the 5% level; standard errors are presented in parentheses*
*** CL: WTI; NG: Natural Gas; SP: S&P 500; US: US Dollar Index*

Figure 21 Canonical Vine Decomposition. Best Copula Fits Along with Their Parameters' Estimates.



** CL: WTI; NG: Natural Gas; SP: S&P 500; US: US Dollar Index*

Figure 22 D-Vine Decomposition. Best Copula Fits Along with Their Parameters' Estimates



** CL: WTI; NG: Natural Gas; SP: S&P 500; US: US Dollar Index*

4.4. Value-at-Risk Backtest

The performance of different vine copula models is evaluated in terms of their out-of-sample Value-at-Risk forecasts. In addition to the multivariate copula and pair-copula models, we also use the traditional methods on portfolio returns to estimate VaR. The competing traditional models included in this chapter are:

1. The traditional benchmark models on portfolio returns such as the Normal model, the Student-*t* model, the model based on the historical simulation, and the unconditional Extreme Value Theory model;
2. The GARCH processes based models such as the RiskMetrics model, the conditional Normal model, the conditional Student-*t* model and the conditional Extreme Value model;

3. The multivariate copula models considered are: multivariate the Gaussian copula and the multivariate *t*-copula;
4. The pair-copula models such as the canonical vine copula and D-vine copula models based on three different pair-copula families: *t*-copula, Clayton copula and SJC copula where the Clayton copula exhibits lower tail dependence and the SJC copula exhibits both upper and lower tail dependence.

For all models, we use a rolling window of 1,000 observations to calibrate our models and forecast VaR and ES for the confidence levels of 5%, 1%, and 0.5%. The dynamic nature of this process allows us to capture the time-varying characteristics of the data. For each vine copula model, we re-estimate the parameters for each 1,000 observations and simulate 50,000 scenarios to calculate one-step-ahead VaR forecast. Another important consideration during the backtest procedure is to choose an appropriate combination of pair-copula families in the vine copula decomposition. In such a long backtest period, it is not feasible to carefully examine the best pair-copula fit for each case. Therefore, we use the same pair-copula family for the vine copula simulation. Table 29 summarizes the Value-at-Risk violation ratios and the competing models are compared and ranked by the absolute deviation from the expected violation ratios. The unconditional and conditional coverage tests results are shown in Tables 30 and 31 respectively. The null hypothesis of Kupiec's unconditional coverage test assumes that the probability of occurrence of the violations equals the expected significance level. A *p*-value above 5% for the unconditional coverage test is taken to indicate that the number of violations is statistically equal to the expected one.

Table 29 VaR Violation Ratio and Model Ranking

	5.00%		1.00%		0.50%	
Right Tails						
	Panel A: Traditional Benchmark Models					
Normal	5.38%	3	1.73%	15	1.26%	16
Student-*t*	5.69%	7	1.26%	6	0.72%	12
HS	5.63%	6	1.16%	5	0.66%	6
EVT	5.72%	8	1.29%	8	0.63%	5
	Panel B: GARCH Processes					
Risk Metrics	5.88%	10	1.73%	15	1.04%	15
Conditional Normal	5.03%	1	1.26%	6	0.82%	14
Conditional *t*	5.75%	9	0.88%	3	0.47%	1
Conditional EVT	5.22%	2	0.88%	3	0.38%	4
	Panel C: Multivariate Copula Models					
MV Gaussian Copula	3.58%	13	0.60%	9	0.28%	8
MV t Copula	3.68%	11	0.57%	10	0.28%	8
	Panel D: Pair-copula Models					
C-Vine t Copula	3.58%	13	0.57%	10	0.28%	8
C-Vine Clayton Copula	4.62%	4	0.97%	1	0.53%	2
C-Vine SJC Copula	3.14%	16	0.44%	14	0.25%	13
D-Vine t Copula	3.62%	12	0.53%	12	0.31%	7
D-Vine Clayton Copula	4.62%	4	0.91%	2	0.53%	2
D-Vine SJC Copula	3.18%	15	0.53%	12	0.28%	8
Left Tails						
	Panel A: Traditional Benchmark Models					
Normal	4.94%	2	1.79%	16	1.35%	16
Student-*t*	5.44%	5	1.38%	4	0.57%	3
HS	5.53%	7	1.38%	4	0.72%	5
EVT	5.50%	6	1.45%	7	0.82%	7
	Panel B: GARCH Processes					
Risk Metrics	4.97%	1	1.10%	3	0.63%	4
Conditional Normal	4.46%	8	0.91%	1	0.53%	1
Conditional *t*	4.90%	3	0.57%	6	0.25%	6
Conditional EVT	5.28%	4	0.91%	1	0.44%	2
	Panel C: Multivariate Copula Models					
MV Gaussian Copula	3.46%	14	0.47%	9	0.16%	8
MV t Copula	3.58%	11	0.47%	9	0.16%	8
	Panel D: Pair-copula Models					
C-Vine t Copula	3.55%	13	0.44%	12	0.16%	8
C-Vine Clayton Copula	3.93%	9	0.50%	8	0.16%	8
C-Vine SJC Copula	3.27%	15	0.38%	14	0.13%	14
D-Vine t Copula	3.58%	11	0.44%	12	0.16%	8
D-Vine Clayton Copula	3.77%	10	0.47%	9	0.16%	8
D-Vine SJC Copula	3.05%	16	0.28%	15	0.13%	14

** CL: WTI; NG: Natural Gas; SP: S&P 500; US: US Dollar Index*

*** All models are ranked based on the deviation from the expected significance levels. The best models are highlighted by rectangles.*

Table 30 Unconditional Coverage Test

	5.00%	1.00%	0.50%
Right Tails			
Panel A: Traditional Benchmark Models			
Normal	0.337**	0.000	0.000
Student-*t*	0.080	0.161**	0.095
HS	0.111**	0.367**	0.222**
EVT	0.068	0.117**	0.322**
Panel B: GARCH Processes			
Risk Metrics	0.027	0.000	0.000
Conditional Normal	0.938**	0.161**	0.020
Conditional *t*	0.057	0.488**	0.818**
Conditional EVT	0.574**	0.488**	0.305**
Panel C: Multivariate Copula Models			
MV Gaussian Copula	0.000	0.014	0.059
MV t Copula	0.000	0.007	0.059
Panel D: Pair-copula Models			
C-Vine t Copula	0.000	0.007	0.059
C-Vine Clayton Copula	0.321**	0.885**	0.785**
C-Vine SJC Copula	0.000	0.000	0.028
D-Vine t Copula	0.000	0.004	0.111**
D-Vine Clayton Copula	0.321**	0.611**	0.785**
D-Vine SJC Copula	0.000	0.004	0.059
Left Tails			
Panel A: Traditional Benchmark Models			
Normal	0.867**	0.000	0.000
Student-*t*	0.263**	0.040	0.606**
HS	0.175**	0.040	0.095
EVT	0.201**	0.018	0.020
Panel B: GARCH Processes			
Risk Metrics	0.932**	0.576**	0.322**
Conditional Normal	0.158**	0.611**	0.785**
Conditional *t*	0.803**	0.007	0.028
Conditional EVT	0.470**	0.611**	0.625**
Panel C: Multivariate Copula Models			
MV Gaussian Copula	0.000	0.001	0.001
MV t Copula	0.000	0.001	0.001
Panel D: Pair-copula Models			
C-Vine t Copula	0.000	0.000	0.001
C-Vine Clayton Copula	0.004	0.002	0.001
C-Vine SJC Copula	0.000	0.000	0.000
D-Vine t Copula	0.000	0.000	0.001
D-Vine Clayton Copula	0.001	0.001	0.001
D-Vine SJC Copula	0.000	0.000	0.000

*** A p-value above 5% for the unconditional coverage test is taken to indicate that the number of violations is statistically equal to the expected one.*

Table 31 Conditional Coverage Test

	5.00%	1.00%	0.50%
Right Tails			
Panel A: Traditional Benchmark Models			
Normal	0.308**	0.000	0.000
Student-*t*	0.066	0.083	0.076
HS	0.103**	0.191**	0.182**
EVT	0.062	0.059	0.265**
Panel B: GARCH Processes			
Risk Metrics	0.002	0.000	0.000
Conditional Normal	0.210**	0.083	0.016
Conditional *t*	0.012	0.318**	0.651**
Conditional EVT	0.119**	0.318**	0.283**
Panel C: Multivariate Copula Models			
MV Gaussian Copula	0.000	0.012	0.057
MV t Copula	0.000	0.007	0.057
Panel D: Pair-copula Models			
C-Vine t Copula	0.000	0.007	0.057
C-Vine Clayton Copula	0.276**	0.420**	0.605**
C-Vine SJC Copula	0.000	0.000	0.027
D-Vine t Copula	0.000	0.003	0.106**
D-Vine Clayton Copula	0.199**	0.368**	0.605**
D-Vine SJC Copula	0.000	0.003	0.057
Left Tails			
Panel A: Traditional Benchmark Models			
Normal	0.559**	0.000	0.000
Student-*t*	0.236**	0.002	0.487**
HS	0.144**	0.012	0.076
EVT	0.184**	0.001	0.016
Panel B: GARCH Processes			
Risk Metrics	0.643**	0.291**	0.265**
Conditional Normal	0.147**	0.368**	0.605**
Conditional *t*	0.444**	0.007	0.027
Conditional EVT	0.393**	0.368**	0.542**
Panel C: Multivariate Copula Models			
MV Gaussian Copula	0.000	0.001	0.001
MV t Copula	0.000	0.001	0.001
Panel D: Pair-copula Models			
C-Vine t Copula	0.000	0.000	0.001
C-Vine Clayton Copula	0.004	0.002	0.001
C-Vine SJC Copula	0.000	0.000	0.000
D-Vine t Copula	0.000	0.000	0.001
D-Vine Clayton Copula	0.001	0.001	0.001
D-Vine SJC Copula	0.000	0.000	0.000

*** A p-value above 5% for the conditional coverage test is taken to indicate that the forecasting ability of the VaR model is adequate.*

Table 32 VaR Backtest Summary Table

	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC
	5.00%			**1.00%**			**0.50%**		
Right Tails									
				Panel A: Traditional Benchmark Models					
Normal	3	**	**	15			16		
Student-*t*	7	**	**	6	**	**	12	**	**
HS	6	**	**	5	**	**	6	**	**
EVT	8	**	**	8	**	**	5	**	**
				Panel B: GARCH Processes					
Risk Metrics	10			15			15		
Conditional Normal	1	**	**	6	**	**	14		
Conditional *t*	9	**		3	**	**	1	**	**
Conditional EVT	2	**	**	3	**	**	4	**	**
				Panel C: Multivariate Copula Models					
MV Gaussian Copula	13			9			8	**	**
MV t Copula	11			10			8	**	**
				Panel D: Pair-copula Models					
C-Vine t Copula	13			10			8	**	**
C-Vine Clayton Copula	4	**	**	1	**	**	2	**	**
C-Vine SJC Copula	16			14			13		
D-Vine t Copula	12			12			7	**	**
D-Vine Clayton Copula	4	**	**	2	**	**	2	**	**
D-Vine SJC Copula	15			12			8	**	**
Left Tails									
				Panel A: Traditional Benchmark Models					
Normal	2	**	**	16			16		
Student-*t*	5	**	**	4			3	**	**
HS	7	**	**	4			5	**	**
EVT	6	**	**	7			7		
				Panel B: GARCH Processes					
Risk Metrics	1	**	**	3	**	**	4	**	**
Conditional Normal	8	**	**	1	**	**	1	**	**
Conditional *t*	3	**	**	6			6		
Conditional EVT	4	**	**	1	**	**	2	**	**
				Panel C: Multivariate Copula Models					
MV Gaussian Copula	14			9			8		
MV t Copula	11			9			8		
				Panel D: Pair-copula Models					
C-Vine t Copula	13			12			8		
C-Vine Clayton Copula	9			8			8		
C-Vine SJC Copula	15			14			14		
D-Vine t Copula	11			12			8		
D-Vine Clayton Copula	10			9			8		
D-Vine SJC Copula	16			15			14		

** Kupiec denotes the unconditional coverage test*

*** CC denotes the conditional coverage test*

**** All models are ranked based on the deviation from the expected significance levels. The best models are highlighted by rectangles.*

The null hypothesis of the conditional coverage test says that the probability of occurrence of the violations equals the expected significance level and the violations are independently distributed through time. A *p*-value below 5% for the conditional coverage test is taken to indicate that the model generates too few or too many clustered violations.

Table 32 summarizes the ranking of the violation ratios along with the two statistical tests.

1. For the right tail, among all of the models that correctly forecast the VaR (*i.e.*, passes both the unconditional and conditional coverage tests), the canonical vine Clayton copula, the D-vine Clayton copula, and the conditional EVT models are consistently ranked among top four.
2. For the left tail, all copula models fail to correctly forecast the VaR. The conditional EVT and the RiskMetrics models perform better than the rest of the models.
3. Therefore, for the full sample of 3,181 observations used for backtesting, none of the models overwhelmingly outperforms the rest. The conditional EVT model stands out as it correctly forecasts the out-of-sample VaR and is ranked among the top four models for all confidence levels for both tails. On the other hand, the multivariate copula and pair-copula models fail to correctly forecast VaR, especially for the left tail.
4. In addition, the copula models tend to overestimate the risk at all confidence levels, while the traditional benchmark models are more likely to underestimate the VaR.

4.5. VaR Backtest during Recent Financial Crisis Period

Since the Federal Reserve Banks took control of Fannie Mae and Freddie Mac and Lehman Brothers went bankrupt in September 2008, the correlations between returns of different asset classes rose to an unprecedented level due to elevated macroeconomic uncertainty. Now more than ever, the co-movement of the assets in a portfolio needs to be modeled properly to forecast portfolio Value-at-Risk. Therefore, we take a closer look at the sample period from September 2008 to the end of the data period (September 2010). Table 33 below summarizes the Value-at-Risk violation ratios at the significance levels of 5%, 1%, and 0.5% for the recent financial crisis period. Table 34 summarizes the ranking of the violation ratios along with the unconditional and conditional coverage test statistics.

1. It is clear that the traditional unconditional univariate models such as Normal, Student-*t*, historical simulation and Extreme Value Theory models underestimate the risk for this period. The violation ratios for these models are 200 to 350 basis points higher than the violation ratio for the expected VaRs for the 5% and 1% confidence levels.
2. The models that take autoregressive conditional heteroskedasticity into consideration tend to perform better than those that do not allow for it.
3. The multivariate copula models also correctly forecast out-of-sample VaRs.

However, due to the small sample size, it is hard to distinguish the forecasting performance of the models that correctly forecast out-of-sample VaRs. A superior risk measurement, Expected Shortfall is then used here to judge the forecasting performance of the competing models.

Table 33 VaR Violation Ratio and Model Ranking – Recent Financial Crisis

α	5.00%		1.00%		0.50%	
Right Tails						
	Panel A: Traditional Benchmark Models					
Normal	8.57%	13	4.19%	16	3.43%	16
Student-*t*	8.76%	15	3.43%	15	1.90%	15
HS	8.76%	15	2.29%	13	1.52%	13
EVT	8.57%	13	3.05%	14	1.52%	13
	Panel B: GARCH Processes					
Risk Metrics	6.48%	11	1.90%	12	0.95%	9
Conditional Normal	5.90%	8	1.14%	8	0.95%	9
Conditional *t*	6.48%	11	0.95%	1	0.38%	1
Conditional EVT	5.90%	8	0.95%	1	0.38%	1
	Panel C: Multivariate Copula Models					
MV Gaussian Copula	5.14%	3	0.95%	1	0.38%	1
MV t Copula	5.14%	3	0.95%	1	0.38%	1
	Panel D: Pair-copula Models					
C-Vine t Copula	5.14%	3	0.95%	1	0.38%	1
C-Vine Clayton Copula	5.90%	8	1.52%	10	0.95%	9
C-Vine SJC Copula	4.95%	1	0.76%	9	0.38%	1
D-Vine t Copula	5.14%	3	0.95%	1	0.38%	1
D-Vine Clayton Copula	5.71%	7	1.52%	10	0.95%	9
D-Vine SJC Copula	4.95%	1	0.95%	1	0.38%	1
Left Tails						
	Panel A: Traditional Benchmark Models					
Normal	7.43%	13	4.38%	16	3.43%	16
Student-*t*	7.81%	16	3.62%	14	1.52%	13
HS	7.43%	13	3.24%	13	1.71%	14
EVT	7.62%	15	3.62%	14	2.10%	15
	Panel B: GARCH Processes					
Risk Metrics	4.19%	10	0.57%	9	0.38%	1
Conditional Normal	4.76%	1	0.76%	1	0.38%	1
Conditional *t*	5.52%	4	0.38%	11	0.19%	10
Conditional EVT	5.90%	11	0.76%	1	0.38%	1
	Panel C: Multivariate Copula Models					
MV Gaussian Copula	4.38%	5	0.76%	1	0.38%	1
MV t Copula	4.38%	5	0.76%	1	0.38%	1
	Panel D: Pair-copula Models					
C-Vine t Copula	4.38%	5	0.76%	1	0.38%	1
C-Vine Clayton Copula	4.76%	1	0.76%	1	0.38%	1
C-Vine SJC Copula	4.38%	5	0.76%	1	0.19%	10
D-Vine t Copula	4.38%	5	0.57%	9	0.38%	1
D-Vine Clayton Copula	4.57%	3	0.76%	1	0.38%	1
D-Vine SJC Copula	4.00%	12	0.38%	11	0.19%	10

** All models are ranked based on the deviation from the expected significance levels. The best models are highlighted by rectangles.*

Table 34 VaR Backtest Summary Table – Recent Financial Crisis

	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC	Violation Ratio	Kupiec	CC
Right Tails	**5.00%**			**1.00%**			**0.50%**		
				Panel A: Traditional Benchmark Models					
Normal	13			16			16		
Student-*t*	15			15			15		
HS	15			13			13		
EVT	13			14			13		
				Panel B: GARCH Processes					
Risk Metrics	11	**	**	12	**		9	**	**
Conditional Normal	8	**	**	8	**	**	9	**	**
Conditional *t*	11	**	**	1	**	**	1	**	**
Conditional EVT	8	**	**	1	**	**	1	**	**
				Panel C: Multivariate Copula Models					
MV Gaussian Copula	3	**	**	1	**	**	1	**	**
MV t Copula	3	**	**	1	**	**	1	**	**
				Panel D: Pair-copula Models					
C-Vine t Copula	3	**	**	1	**	**	1	**	**
C-Vine Clayton Copula	8	**	**	10	**	**	9	**	**
C-Vine SJC Copula	1	**	**	9	**	**	1	**	**
D-Vine t Copula	3	**	**	1	**	**	1	**	**
D-Vine Clayton Copula	7	**	**	10	**	**	9	**	**
D-Vine SJC Copula	1	**	**	1	**	**	1	**	**
Left Tails									
				Panel A: Traditional Benchmark Models					
Normal	13			16			16		
Student-*t*	16			14			13		
HS	13			13			14		
EVT	15			14			15		
				Panel B: GARCH Processes					
Risk Metrics	10	**	**	9	**	**	1	**	**
Conditional Normal	1	**	**	1	**	**	1	**	**
Conditional *t*	4	**	**	11	**	**	10	**	**
Conditional EVT	11	**		1	**	**	1	**	**
				Panel C: Multivariate Copula Models					
MV Gaussian Copula	5	**	**	1	**	**	1	**	**
MV t Copula	5	**	**	1	**	**	1	**	**
				Panel D: Pair-copula Models					
C-Vine t Copula	5	**	**	1	**	**	1	**	**
C-Vine Clayton Copula	1	**	**	1	**	**	1	**	**
C-Vine SJC Copula	5	**	**	1	**	**	10	**	**
D-Vine t Copula	5	**	**	9	**	**	1	**	**
D-Vine Clayton Copula	3	**	**	1	**	**	1	**	**
D-Vine SJC Copula	12	**	**	11	**	**	10	**	**

** Kupiec denotes the unconditional coverage test*

*** CC denotes the conditional coverage test*

**** All models are ranked based on the deviation from the expected significance levels. The best models are highlighted by rectangles.*

Table 35 Backtest Expected Shortfall using Loss Function Approach – Recent Financial Crisis

Right Tail	**5.00%**		**1.00%**		**0.50%**	
	Panel A: Traditional Benchmark Models					
Normal	n.a.		n.a.		n.a.	
Student-*t*	n.a.		n.a.		n.a.	
HS	n.a.		n.a.		n.a.	
EVT	n.a.		n.a.		n.a.	
	Panel B: GARCH Processes					
Risk Metrics	2,866	9	n.a.		467	12
Conditional Normal	2,809	7	665	10	430	11
Conditional *t*	3,258	12	462	8	205	7
Conditional EVT	2,855	8	451	7	227	8
	Panel C: Multivariate Copula Models					
MV Gaussian Copula	2,240	4	391	5	159	4
MV t Copula	2,269	5	357	3	154	3
	Panel D: Pair-copula Models					
C-Vine t Copula	2,279	6	403	6	176	6
C-Vine Clayton Copula	2,934	11	696	11	353	10
C-Vine SJC Copula	2,195	3	315	1	125	1
D-Vine t Copula	2,184	2	358	4	154	2
D-Vine Clayton Copula	2,892	10	665	9	347	9
D-Vine SJC Copula	2,068	1	349	2	163	5
Left Tail	**5.00%**		**1.00%**		**0.50%**	
	Panel A: Traditional Benchmark Models					
Normal	n.a.		n.a.		n.a.	
Student-*t*	n.a.		n.a.		n.a.	
HS	n.a.		n.a.		n.a.	
EVT	n.a.		n.a.		n.a.	
	Panel B: GARCH Processes					
Risk Metrics	1,505	6	293	10	200	10
Conditional Normal	1,801	10	363	12	231	12
Conditional *t*	2,123	11	236	9	132	9
Conditional EVT	n.a.		334	11	215	11
	Panel C: Multivariate Copula Models					
MV Gaussian Copula	1,479	5	193	6	58	7
MV t Copula	1,515	7	157	4	49	5
	Panel D: Pair-copula Models					
C-Vine t Copula	1,440	4	163	5	34	4
C-Vine Clayton Copula	1,793	9	205	7	50	6
C-Vine SJC Copula	1,403	2	119	2	17	1
D-Vine t Copula	1,403	3	135	3	30	3
D-Vine Clayton Copula	1,703	8	207	8	60	8
D-Vine SJC Copula	1,231	1	108	1	19	2

** All models are ranked based on the deviation from the expected significance levels. The best models are highlighted by rectangles.*

*** The models that do not pass both the unconditional and conditional coverage tests are marked as "n.a."*

We use a two-stage approach in this chapter to backtest the ES estimates. The first stage selects models that pass both the unconditional and conditional coverage tests. In the second stage, the values of the loss function (as defined in chapter two) for the selected models are calculated. These are presented in Table 35. All competing models are ranked by minimizing the sum of the loss function across all 1,000 days used in each backtest period. It is clear that the canonical vine SJC and the D-vine SJC copula models outperform the other models in terms of minimizing the Expected Shortfall based loss function by allowing for different upper and lower tail dependence between each pair in the portfolio.

5. Conclusions

Most of the traditional methods for the Value-at-Risk estimation are based on the assumption that the risk factors are multivariate normally distributed. However, previous empirical evidence has shown the return distributions are asymmetric and often show tail dependence. In this chapter, we have introduced a new method to forecast Value-at-Risk and Expected Shortfall of a multivariate portfolio based on the theory of pair-copula. Pair-copula models allow for flexible modeling of the marginal distributions of each risk factor in the portfolio and at the same time capture the dependence between each asset within the portfolio. We estimate the 95%, 99% and 99.5% VaR and ES for an equally weighted portfolio composed of the crude oil futures, natural gas futures, S&P 500 and the US dollar index using 50,000 Monte Carlo simulations for each one step ahead forecast. We then compare the accuracy of the different VaR estimates using the Kupiec's and Christoffersen's tests in the first stage and Expected Shortfalls in the second stage to compare the forecasting performances of the competing models.

Our empirical analysis shows that, for the whole sample period under study, pair-copula models do not offer any added advantage in forecasting VaR even though the canonical vine and D-vine Clayton copula performs as well as the conditional EVT model for the right tail of the returns distribution. Overall, for both tails at all confidence levels, the conditional EVT model tend to provide a consistently accurate and better VaR forecast than other models. In addition, the pair-copula models tend to overestimate the risk while the traditional benchmark models are more likely to underestimate the risk. The multivariate Gaussian and *t* copula models offer a very poor prediction for portfolio risk.

The time period of the recent financial crisis brought about elevated dependence between financial assets. The traditional benchmark models grossly under-estimate the portfolio Value-at-Risk during the financial crisis period, with the violation ratios that are greater by as much as 350 basis points. Even the GARCH process based models tend to underestimate the risk as well. On the other hand, all copula based models accurately forecast out-of-sample VaR. In particular, the canonical vine and D-vine SJC copula models offer a significant improvement over the other models for the Expected Shortfall forecasting. It appears that multivariate copulas models can capture the risk much better than the traditional models in the post crisis period. However, as the economy enters into next stage of the business cycle, the correlation between assets is expected to return to historical levels. A possible future research direction would be to use a regime-switching multivariate copula model to capture the different phases of the business cycle.

CHAPTER FIVE: SUMMARY

1. Conclusion

This dissertation examines the estimation of risk measures, Value-at-Risk (VaR) and Expected Shortfall (ES) of energy futures investments using conditional Extreme Value Theory (EVT) and copulas. We test the performance of eight competing models for calculating VaR and ES for single asset investments. Our results indicate that the GARCH-EVT approach outperforms the competing models in forecasting VaR and ES by a wide margin. This approach provides a significant improvement over the widely used Normal distribution based VaR and ES models, which tends to underestimate the true risk and fail to provide statistically accurate VaR estimates. Our results are similar to Marimoutou, Raggad, and Trabelsi (2009), where the authors illustrate that the conditional Extreme Value Theory and Filtered Historical Simulation approaches outperform the traditional methods. However, our results show that the GARCH-EVT approach is overwhelmingly better than the competing models, especially at lower significance levels (1%, 0.5%, and 0.1%).

However, when energy futures are added into investors' portfolios for the benefit of diversification, we need to properly account for the dependence structure between assets as the dependence structure provides insights into portfolio risk management, portfolio diversification, pairs trading and exotic derivatives pricing especially when returns are non-Normal, and simply linear correlation fails to capture the degree of association between assets. First, we employ a time-varying conditional copula method which only

includes lagged endogenous explanatory variables. The results imply that energy futures are more likely to move together during bull markets than in bear markets. In contrast, the previous research on copulas found financial assets are more likely to move down together than move up together. Similar to other studies utilizing Patton's (2006a) dynamic evolution structure for the time-varying copula, the dependence parameters can be very volatile over time and deviate from their constant levels frequently. In addition, we also investigate whether the estimates are affected by supply and demand balance of oil and gas by utilizing a time-varying copula models which also include the fundamental supply and demand balance as the explanatory variables, the most theoretically associated pairs, WTI/Brent and Heating Oil/Natural Gas appear to be more likely to move together during down markets than up markets. As far as we know, this is the first study that introduces supply and demand balance in modeling the dependence structure of energy futures using copulas. Our results confirm our hypothesis that it is important to consider fundamental supply-demand relationship for modelling energy futures.

Finally, in order to capture portfolio risks while taking into consideration portfolio dependence structure, we introduce a pair-copula approach to a multivariate portfolio consisting of energy futures, the S&P 500 index and the US Dollar index. Although the pair-copula approach does not provide any added advantage in terms of portfolio VaR and ES estimation for the whole backtest period, we found that the pair-copula approach outperforms other models during the period of elevated co-movements among financial assets. This chapter contributes to the literature of application of multivariate copula theory by exploring the benefits of using vine copula theory in the estimation of the

Value-at-Risk and the Expected Shortfall for a diversified portfolio of energy futures, equity index and currency index.

2. Future Research

This dissertation suggests several directions for future research. As discussed in chapter two, selecting a threshold value in the Peaks-over-Threshold approach graphically using the mean excess plot is not an exact science. More research is needed to develop a parametric method to estimate the threshold value. In addition, a multi-step-ahead VaR estimation framework is needed. The Bank for International Settlement (BIS) requires banking institutions provide the BIS with 99% 10-day VaR estimates (which are used by the BIS in calculating regulatory capital requirements). Therefore, future research is needed to extend current one-step-ahead VaR forecasts and backtesting to a multi-step-ahead VaR framework. In addition, it is interesting to consider market microstructure in this research to see if it can improve the out-of-sample risk forecast.

The results in chapter three suggest that the time-varying dependence parameters are very volatile over time for all copula models and all pairs under study. Therefore, it is important to explore different ways of capturing the time dependence other than the Patton's (2006a) dynamic evolution structure used in this dissertation. Our experience during the estimation process also tells us that the choice of initial values turns out to be very important and the computation quickly becomes very complex as we introduce more variables to the model. Therefore, much more remains to be done to develop more efficient estimation procedures. In addition, with increasing numbers of copula models

available, it is becoming important to advance goodness-of-fit tests and be able to choose the most appropriate copula model for the assets under study.

Our empirical results in Chapter four show that the multivariate copula models can capture risk much better than the traditional risk models in the post crisis period. However, as the economy enters into the next phase of the business cycle, we expected the correlation between different assets to return to historical levels. Therefore, it is possible to extend current models to regime-switching copula models which may capture different stages of the business cycle.

We employ a pair-copula approach to capture portfolio risk exposure. A natural extension is to optimally choose the weights of assets in the portfolio to find the expected return-risk frontier (with respect to the Value-at-Risk measure or the Expected Shortfall measure while endogenously taking into account the induced capital reserve requirement). However, in order to achieve the next stage of portfolio optimization, we need to greatly improve the speed of out-of-sample Monte Carlo simulation. It took us over 20 days (on a laptop with 1.6 GHz Intel Core 2 Duo CPU, 2 GB of RAM using Matlab) for VaR simulation using the D-vine *t* copula for the full sample. Computational efficiency will be even more important if one wants to implement the pair-copula method in a portfolio with a much higher dimension.

BIBLIOGRAPHY

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44, 182-198.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.

Alexander, C. (2005). Correlation in crude oil and natural gas markets. In *Managing Energy Price Risk: The New Challenges and Solutions*, 3rd edition, 573-604. London: Risk Books.

Ang, A. and Chen, J. (2002). Asymmetric correlation of equity portfolios. *Journal of Financial Economics*, 63, 443-94.

Angelidis, T. and Degiannakis, S. (2006). Backtesting VaR models: An expected shortfall approach. Athens University of Economics and Business, Department of Statistics, Technical Report, 223.

Aragones, J.R., Dowd, K. and Blanco, C. (2000). Extreme value VaR. *Derivatives Week,* 7-8.

Artzner, P., Delbaen, F., Eber, J. and Heath, D. (1997). Thinking coherently. *Risk*, 10, 68–71.

Artzner, P., Delbaen, F., Eber, J. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance,* 9, 203–228.

Bali, T. G. (2003). An extreme value approach to estimating volatility and value at risk. *Journal of Business*, 76, 83–107.

Bali, T. G., and Neftci, S. N. (2003). Disturbing extremal behaviour of spot price dynamics. *Journal of Empirical Finance*, 10, 455– 477.

Balkema, A.A. and de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2, 792-804.

Bastianin, A. (2009). Modelling asymmetric dependence using copula functions: An application to Value-at-Risk in the energy sector. Foundazione Eni Enrico Mattei Working Papers.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32, 245-268.

Bedford, T. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 30, 1031-1068.

Berg, D. and Aas, K. (2008). Models for construction of multivariate dependence. Working Paper of the Norwegian Computing Center.

Bhattacharyya, M. and Ritolia, G. (2008). Conditional VaR using EVT – Towards a Planned Margin Scheme, *International Review of Financial Analysis*, 17, 382-395.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31, 307–327.

Byström, H. (2005). Extreme value theory and extremely large electricity price changes. *International Review of Economics and Finance*, 14, 41– 55.

Caillault, C. and Guegan, D. (2009). Forecasting VaR and expected shortfall using dynamical systems: a risk management strategy. *Frontiers in Financial and Economics*, 6, 26-50.

Chan, F. and Gray, P. K. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting*, 22, 283-300.

Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39, 841-862.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65, 141-151

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.

Degiannakis, S. and Xekalaki, E. (2007). Assessing the performance of a prediction error criterion model selection algorithm in the context of ARCH models. *Applied Financial Economics*, 17, 149-171.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427–431.

Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Embrechts, P., McNeil, A. and Straumann, D. (1999). Correlation: Pitfalls and alternatives. *Risk*, 5, 69-71.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In M. Dempster (ed.), *Risk Management: Value at Risk and Beyond*. Cambridge: Cambridge University Press, 176-223.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. *Econometrica*, 50, 987-1008.

Fernandez, V. (2005). Risk management under extreme events. *International Review of Financial Analysis*, 14, 113– 148.

Fisher, R. A. (1932). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of largest or smallest members of a sample. *Proceedings of Cambridge Philosophic Society*, 24, 180-190.

Fischer, M. and Köck, C. (2007). Multivariate copula models at work: Dependence structure of energy prices. EcoMod 2007, Moscow, September.

Fischer, M., Köck, C., Schlüter, S. and Weigert, F. (2009). Multivariate copula models at work: outperforming the "Desert Island Copula"?. Universität Erlangen.

Frank, M. J. (1979). On the simultaneous associativity of F(x,y) and x + y - F(x, y). *Aequationes Mathamatics*. 19, 194-226.

Krehbiel, T., and Adkins, L. C., (2005). Price risk in the NYMEX energy complex: an extreme value approach. *Journal of Futures Markets*. 25, 309-337.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3, 73-84.

Gençay, R., and Selçuk, F. (2004). Extreme value theory and value-at-risk: Relative performance in emerging markets. *International Journal of Forecasting*, 20, 287– 303.

Genest, C. and Favre, A.C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12, 347-368.

Genest C., and Rémillard B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 44, 1096-1127.

Genest, C., Rémillard, B. and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44, 199-213.

Gouriéroux, C. (1997). *ARCH-models and financial applications*. New York: Springer.

Grégoire, V., Genest, C. and Gendron, M. (2008). Using copulas to model price dependence in energy markets. *Energy Risk*, 5, 58-64.

Gumbel, E.J. (1960a) Bivariate exponential distributions. *Journal of American Statistical Association*. 55, 698-707

Guégan, D and Maugis, P.A. (2010). An econometric study of vine copulas. *International Journal of Economics and Finance,* 2, 2-14.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6, 255–259.

Jenkinon, A.F. (1955). The frequency distribution of the annual maximum (mimimum) values of meteorological events. *Quarterly Journal of the Royal Meteorological Society*, 81, 158-172.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman and Hall.

Joe, H. and Xu, J.J. (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.

Jondeau, E. and Rockinger, M. (2006). The copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25, 827-853.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30, 81-89.

Köhler, K. J. and Symanowski, J. T. (1995). Constructing multivariate distributions with specific marginal distributions. *Journal of Multivariate Analysis*, 55, 261-282.

Kurowicha, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. New York: Wiley.

Liebscher, E. (2006). Modelling and estimation of multivariate copulas. Working Paper, University of Applied Sciences, Merseburg.

Ljung , G. M., and Box, G. E. P. (1978). On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, 65, 297–303.

Longin, F. M. (1996). The asymptotic distribution of extreme stock market returns. *Journal of Business*, 69, 383– 408.

Longin, F., and Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance*, 56, 659-676.

Lopez, J.A. (1999). Methods for Evaluating Value-at-Risk Estimates. *Economic Review*, Federal Reserve Bank of San Francesco, 2, 3-17.

Nelsen, R. (2006). *An Introduction to Copulas*, 2nd edition. New York: Springer.

Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance*, 7, 77–91.

Markowitz, H. M. (1990). Foundations of portfolio theory. Nobel Prize in Economics documents 1990-1, Nobel Prize Committee.

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

McNeil, A. and Frey, R. (2000). Estimation of tail-related risk measures for heterocedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7, 271-300.

Michelis, L. and Ning, C. (2010). The dependence structure between the Canadian stock market and the US/Canada exchange rate: A copula approach. *Canadian Journal of Economics*, 43, 1016-1039.

Morgan, J. P. (1996). Riskmetrics (4^{th} ed.). Technical Document, New York: J. P. Morgan & Co. Incorporated.

Morillas, P. M. (2005). A method to obtain new copulas from a given one. *Metrika*, 61, 169-184.

Müller, U. A., Dacorogna, M. M. and Pictet, O. V. (1998). Heavy tails in high-frequency financial data. In R. J. Adler, R. E. Feldman, and M. S. Taqqu (eds.), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, 55-77. Boston: Birkhäuser.

Palaro, H.P. and Hotta, L.K. (2006). Using conditional copula to estimate value at risk. *Journal of Data Science*, 4, 93-115.

Palmitesta, P. and Provasi, C. (2005). Aggregation of dependent risks using the Koehler-Symanowski copula function. *Computational Economics*, 25, 189-205.

Patton, A. J. (2002). Applications of Copula Theory in Financial Econometrics. Unpublished Ph.D. dissertation, University of California, San Diego.

Patton, A.J. (2006a). Modelling asymmetric exchange rate dependence, *International Economics Review*, 47, 527-556.

Patton, A.J. (2006b). Estimation of copula models for time series of possibly different lengths. *Journal of Applied Econometrics*, 21, 147-173.

Pickands, J.I. (1975). Statistical inference using extreme value order statistics. *Annals of Statististics*, 3, 119-131.

Pictet, O. V., Dacorogna, M. M. and Müller, U. A. (1998). Hill, bootstrap and jackknife estimators for heavy tails. In Taqqu, M. (ed.), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, 283–310. Boston: Birkhäuser.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23, 470–72.

Sarma, M., Thomas, S. and Shah, A. (2003). Selection of VaR models. *Journal of Forecasting*, 22, 337-358.

Savu, C. and Trede, M. (2006). Hierarchical Archimedean copulas, Working Paper, University of Münster.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

Shih, J.H. and Louis, T.A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51, 1384-1399.

Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229-231.

Smith, M. (2003). Modelling sample selection using Archimedean copulas. *Econometrics Journal*, 6, 99-123.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101.

So, M.K.P. and Yu, P.L.H. (2006). Empirical analysis of GARCH models in Value at Risk estimation. *Journal of International Markets, Institutions and Money*, 16, 180-197.

Wang, W. and Wells, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data (with discussion). *Journal of American Statistics Association*, 95, 62-76.

von Mises, R. (1954). La distribution de la plus grande de n valeurs. In *Selected Papers*, Volume II, pages 271-294. Providence, RI: American Mathematical Society.

APPENDIX A

Figure 23 Contour plots of the t-Copula and Normal Copula with various correlation coefficient parameters

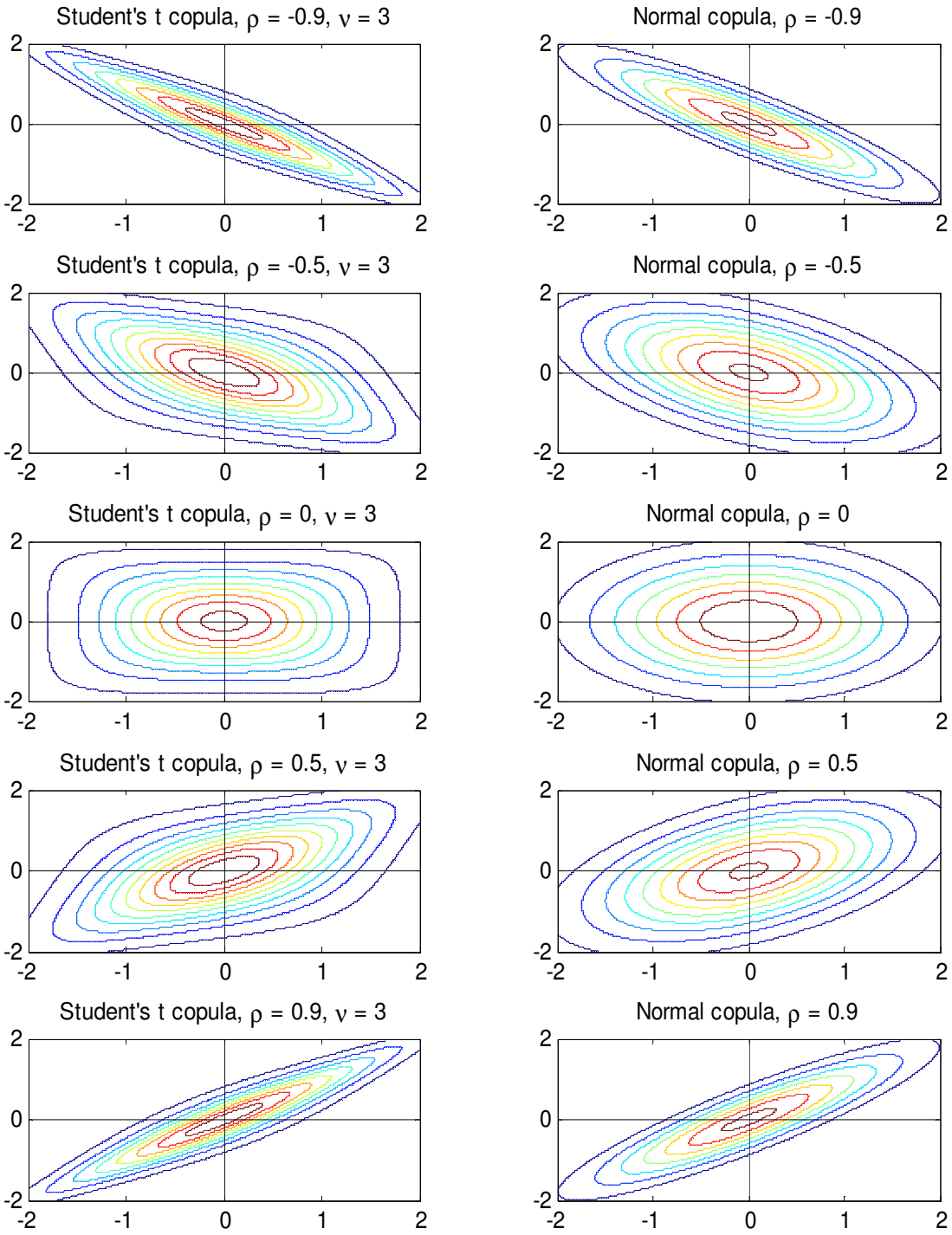


Figure 24 Contour plots of the Symmetrized Joe-Clayton Copula with various tail dependence parameters

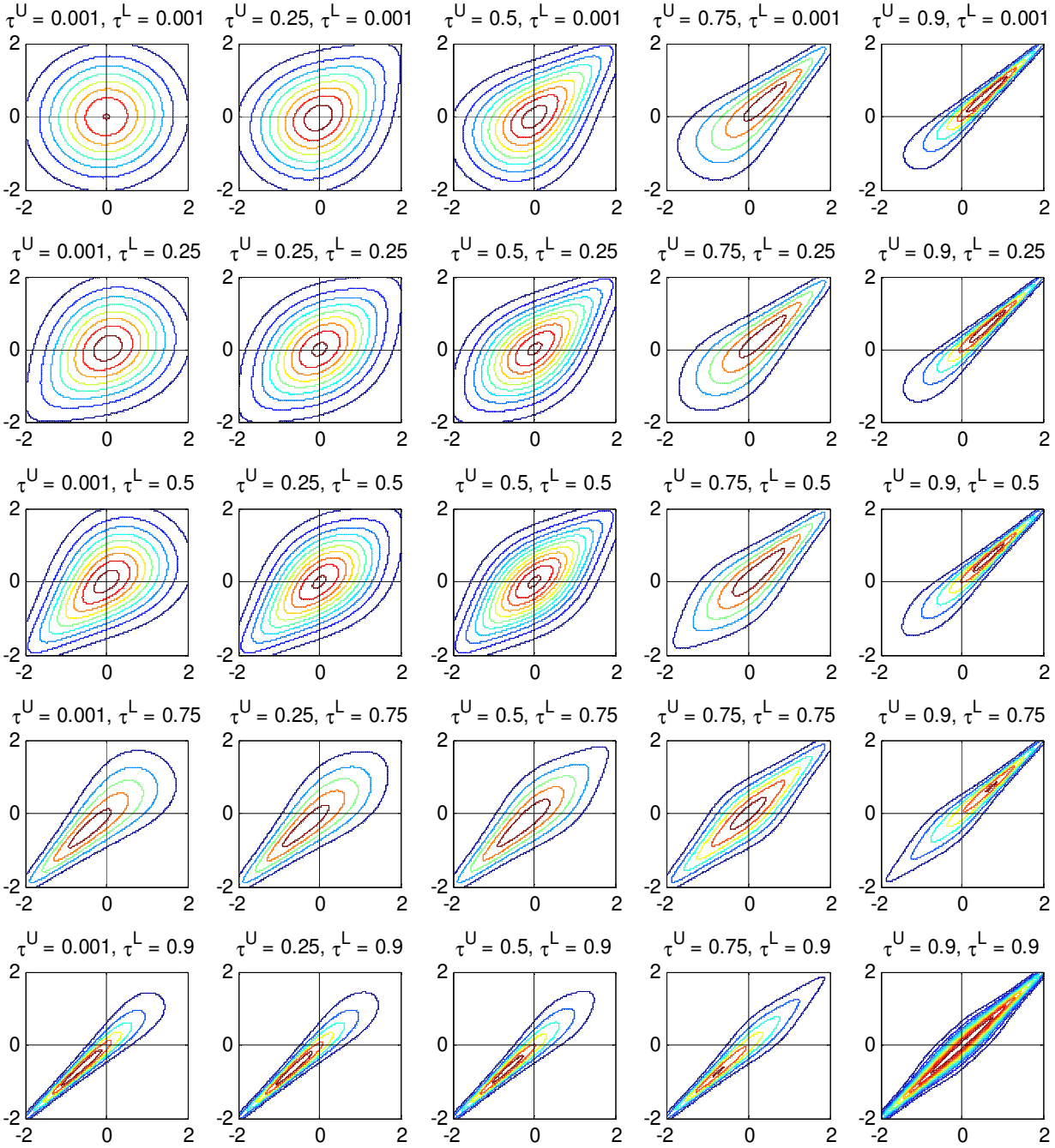
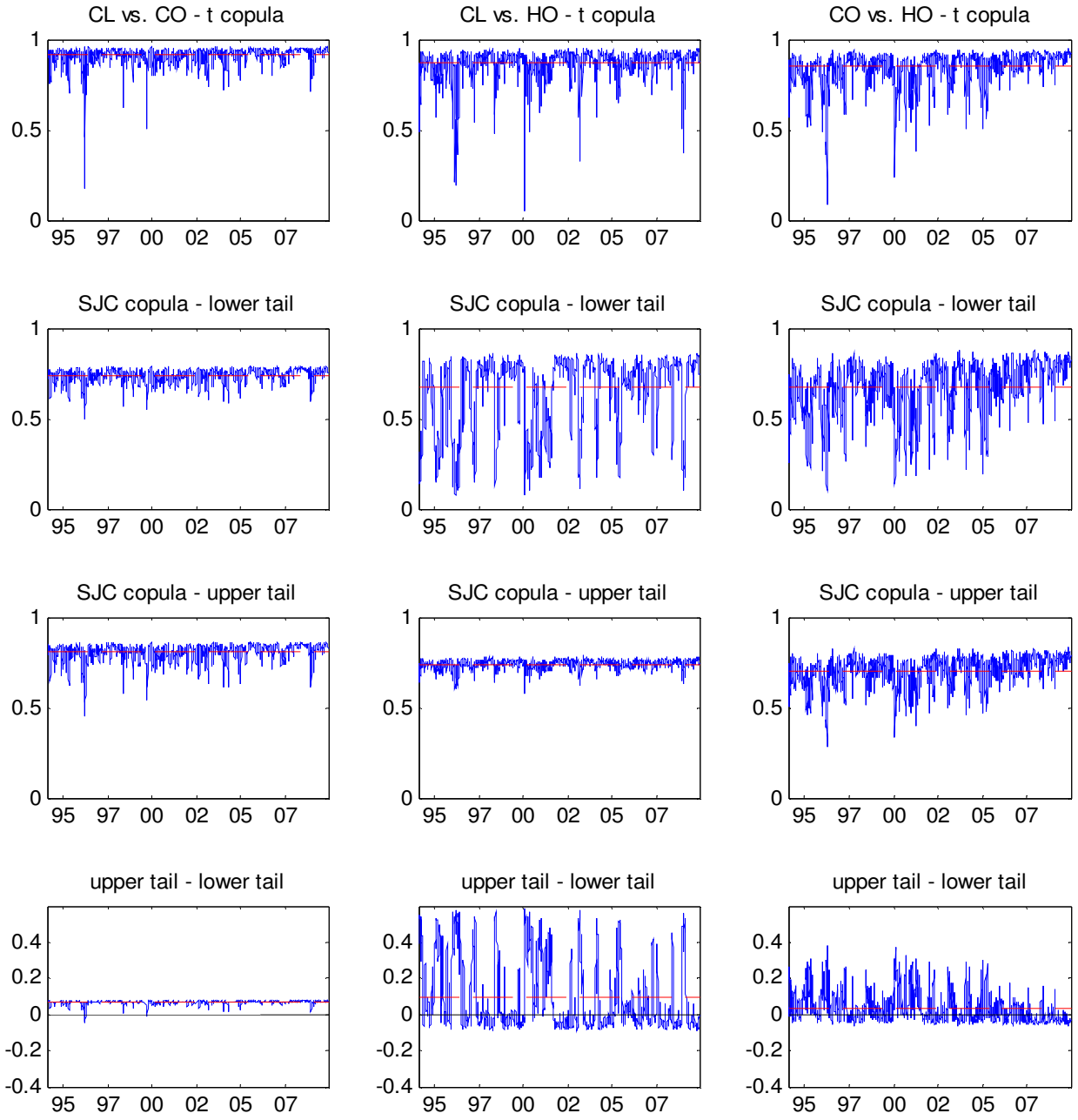
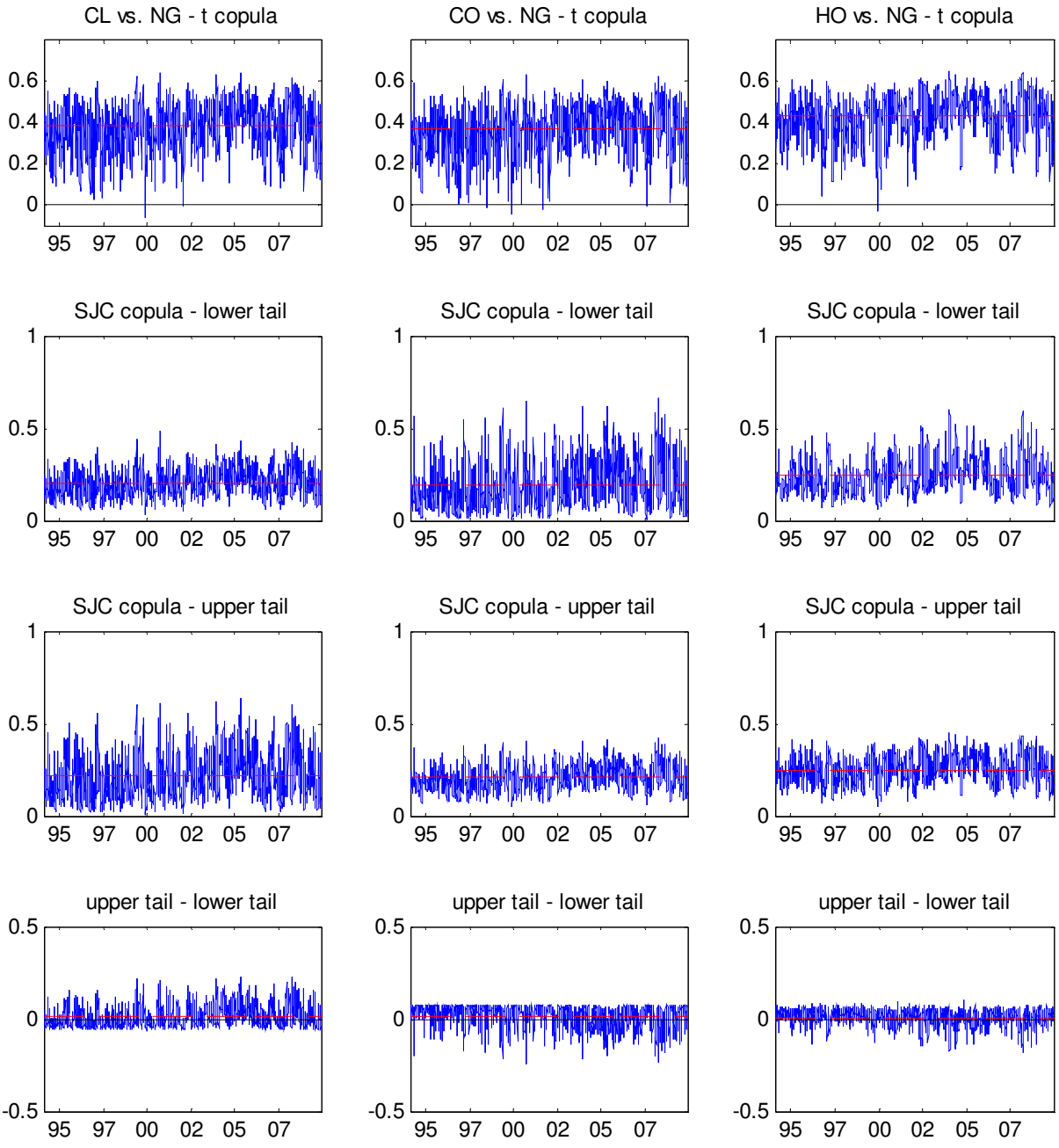


Figure 25 Time path of the time-varying correlation coefficient and tail dependence




CL vs. NG - t copula
CO vs. NG - t copula
HO vs. NG - t copula
SJC copula - lower tail
SJC copula - lower tail
SJC copula - lower tail
SJC copula - upper tail
SJC copula - upper tail
SJC copula - upper tail
upper tail - lower tail
upper tail - lower tail
upper tail - lower tail
95 97 00 02 05 07

APPENDIX B

Bivariate Gaussian Copula

The bivariate Gaussian copula can be written as:

$$C_{Ga}(u,v;\rho) = \Phi_\rho(\Phi^{-1}(u),\Phi^{-1}(v)),$$

where Φ is the standard Normal CDF, Φ^{-1} is the inverse of the standard Normal CDF, and ρ is the usual linear correlation coefficient of the corresponding bivariate Normal distribution Φ_ρ. The density of the bivariate Gaussian copula is given by:

$$c(u,v;\rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\ -\frac{\rho^2(u^2+v^2)-2\rho uv}{2(1-\rho^2)} \quad .$$

For this copula, the *h*-function is given by:

$$h(u,v;\rho) = \Phi\ \frac{\Phi^{-1}(u)-\rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}} \quad .$$

Bivariate Student-t Copula

The bivariate student-*t* copula (or briefly *t* copula) with υ degrees of freedom and correlation $\rho = Corr[t_\upsilon^{-1}(u), t_\upsilon^{-1}(v)]$:

$$C_t(u,v;\upsilon,\rho) = \int_{-\infty}^{t_\upsilon^{-1}(u)} \int_{-\infty}^{t_\upsilon^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}}\ 1+\frac{r^2-2\rho rs+s^2}{\upsilon(1-\rho^2)}^{-(\upsilon+2)/2} drds$$

where t_υ^{-1} denotes the inverse of the cdf of the standard univariate *t*-distribution with υ degrees of freedom. The parameter υ controls the heaviness of the tails and as $\upsilon \to \infty$, $C_t(u,v;\upsilon,\rho) \to \Phi_G(u,v;\rho)$, where Φ_G is the standard bivariate Normal distribution.

The density of the bivariate *t* copula is given by:

$$c(u,v;\rho,\upsilon)=\frac{\Gamma(\frac{\upsilon+2}{2})/\Gamma(\frac{\upsilon}{2})}{\upsilon\pi dt(x_1,\upsilon)dt(x_2,\upsilon)\sqrt{1-\rho^2}}\left(1+\frac{x_1^2+x_2^2-2\rho x_1x_2}{\upsilon(1-\rho^2)}\right)^{-\frac{\upsilon+1}{2}}.$$

where υ and ρ are the parameters of the copula, $x_1=t_\upsilon^{-1}(u)$, $x_2=t_\upsilon^{-1}(v)$, and $dt(\cdot,\upsilon)$ and $t_\upsilon^{-1}(\cdot)$ are the probability density and the quantile function, respectively, for the standard univariate *t*-distribution with υ degrees of freedom. For this copula, the *h*-function is given by:

$$h(u,v;\rho,\upsilon)=t_{\upsilon+1}\left(\frac{t_\upsilon^{-1}(u)-\rho t_\upsilon^{-1}(v)}{\sqrt{\frac{(\upsilon+(t_\upsilon^{-1}(v))^2)(1-\rho^2)}{\upsilon+1}}}\right).$$

Bivariate Clayton Copula

The Clayton copula is defined as:

$$C_C(u,v;\theta)=(u^{-\theta}+v^{-\theta}-1)^{-1/\theta},$$

with the dependence parameter θ restricted on the region (0, ∞). The density of the bivariate Clayton copula is given by:

$$c(u,v;\theta)=(1+\theta)(uv)^{-\theta-1}(u^{-\theta}+v^{-\theta}-1)^{-2-1/\theta}.$$

For this copula, the *h*-function is given by:

$$h(u,v;\theta)=((uv^{\theta+1})^{-\theta/(1+\theta)}+1-v^{-\theta})^{-1/\theta}.$$

Bivariate Symmetrized Joe-Clayton Copula (SJC) Copula

The Symmetrized Joe-Clayton (SJC) copula proposed by Patton (2006a) is a slight modification of original Joe-Clayton (JC) copula. Joe-Clayton copula proposed by Joe (1997) is a Laplace transformation of Clayton's copula. It is defined as:

$$C_{JC}(u,v;\tau^U,\tau^L) = 1-(1-\{[1-(1-u)^{\kappa}]^{-\gamma}+[1-(1-v)^{\kappa}]^{-\gamma}-1]\}^{-1/\gamma})^{1/\kappa}$$

where $\kappa = 1/\log_2(2-\tau^U)$, $\gamma = -1/\log_2(\tau^L)$ and tail dependence parameter $\tau^U \in (0,1], \tau^L \in (0,1]$. By construction, the Joe-Clayton copula always exhibits asymmetry even when the two tail dependence measures are equal. In order to overcome this drawback, Patton (2006a) proposed the "Symmetrized Joe-Clayton" copula, which has the tail dependence measures completely determining the presence or absence of asymmetry. The SJC copula is given by:

$$C_{SJC}(u,v;\tau^U,\tau^L) = 0.5(C_{JC}(u,v;\tau^U,\tau^L)+C_{JC}(1-u,1-v;\tau^U,\tau^L)+u+v-1),$$

where C_{JC} represents the Joe-Clayton copula. The advantage of the Symmetrized Joe-Clayton copula is that it is symmetric when $\tau^U = \tau^L$. The density of the bivariate SJC copula is given by:

$$c(u,v;\tau^U,\tau^L) = \frac{\partial^2 C_{SJC}}{\partial u \partial v} = \frac{1}{2}\left(\frac{\partial^2 C_{JC}(u,v\mid\tau^U,\tau^L)}{\partial u\partial v} - \frac{\partial^2 C_{JC}(1-u,1-v\mid\tau^L,\tau^U)}{\partial u\partial v}\right),$$

where:

$$\frac{\partial^2 C_{JC}(u,v\mid\tau^U,\tau^L)}{\partial u\partial v} = A - B,$$

$$A = \frac{\gamma\cdot\kappa(1-\frac{1}{Z^{1/\gamma}})^{-2+1/\kappa}\cdot(\frac{1}{\gamma}-1)\cdot(1-u)^{\kappa-1}\cdot(1-v)^{\kappa-1}}{(1-(1-u)^{\kappa})^{1+\gamma}\cdot(1-(1-v)^{\kappa})^{\gamma+1}\cdot Z^{2+1/\gamma}},$$

$$B = \frac{\kappa(1-\frac{1}{Z^{1/\gamma}})^{-2+1/\kappa}\cdot(\frac{1}{\kappa}-1)\cdot(1-u)^{\kappa-1}\cdot(1-v)^{\kappa-1}}{(1-(1-u)^{\kappa})^{1+\gamma}\cdot(1-(1-v)^{\kappa})^{\gamma+1}\cdot Z^{2+2/\gamma}}.$$

For this copula, the *h*-function is given by:

$$h(u,v;\tau^U,\tau^L) = \frac{\partial C_{SJC}}{\partial v} = \frac{1}{2}(\frac{\partial C_{JC}(u,v\mid\tau^U,\tau^L)}{\partial v} - \frac{\partial C_{JC}(1-u,1-v\mid\tau^L,\tau^U)}{\partial v}+1),$$

where:

$$\frac{\partial C_{JC}(u,v\mid\tau^U,\tau^L)}{\partial v} = \frac{(1-\frac{1}{Z^{1/\gamma}})^{-\frac{1}{\kappa}+1}}{(1-(1-v)^{\kappa})^{\gamma+1}\cdot Z^{\frac{1}{\gamma}+1}},$$

$$Z = \frac{1}{(1-(1-u)^{\kappa})^{\gamma}} + \frac{1}{(1-(1-v)^{\kappa})^{\gamma}} - 1$$