NORTHWESTERN UNIVERSITY

Pricing Bounds for Bivariate Options with Super- and Submodular Payoff Function.

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Jesse De Lille

EVANSTON, ILLINOIS

December 2008

ABSTRACT

Pricing Bounds for Bivariate Options with Super- and Submodular Payoff Function.

Jesse De Lille

We first introduce the concept of copulas and advocate its use for multivariate option pricing. We focus on four types of bivariate options: basket, rainbow-max, rainbow-min, and spread options. We derive expressions for these options as a function of the copula. We then construct pricing bounds for these bivariate options without imposing any structure at all on the bivariate risk-neutral distribution. Specifically, we introduce the concept of extreme pricing bounds, which are the bounds a bivariate option has to satisfy in order to avoid arbitrage with the univariate options and futures that are trading in the market place. We show that these four types of bivariate options have a super- or submodular payoff function, and that as a result the extreme pricing bounds correspond the Frechet-Hoeffding bounds on the copula. We then proceed to investigate how we can extract partial information about the bivariate risk-neutral distribution from products that are trading in the market place. It is important that when we price a new bivariate option its price is consistent with the products that are currently being traded. We introduce both non-parametric and parametric methods to extract this partial information. Finally, we

show how this partial information about the dependence between two assets translates into tighter pricing bounds.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Pricing Bounds for Bivariate Options

1.1. Motivation

The goal of the paper is to find the price, or to find tight bounds for the price, of options depending on several (at least 2) assets. This paper proposes to use a copula approach to model the interdependency between the underlying assets. The results will be applied to options on energy prices. Why energy prices? First of all, there is a need for a serious empirical investigation of the dependency between 'different' energy prices, where 'different' can apply to product category (oil, gas, power,...), time, or location. Secondly, the significant non-normal and complex behavior of energy prices make it extremely hard to find a multivariate distribution that can simultaneously capture both the behavior of the marginals as well as the dependency between the marginals, and therefore energy prices lend themselves very well to be modelled using copulas. Another big advantage of using copulas to price bivariate options stems from the fact that the margins of the bivariate risk-neutral distribution have to equal the risk-neutral univariate distributions in order to avoid arbitrage. Using copulas to model the risk-neutral bivariate distribution guarantees that this requirement will hold.

The paper is organized as follows. After we introduce the concept of copulas, we focus on four types of bivariate options: basket, rainbow-max, rainbow-min, and spread options. We derive expressions for these options as a function of the copula. We then

construct pricing bounds for these bivariate options without imposing any structure at all on the bivariate risk-neutral distribution. Specifically, we introduce the concept of extreme pricing bounds, which are the bounds a bivariate option has to satisfy in order to avoid arbitrage with the univariate options and futures that are trading in the market place. We show that these four types of bivariate options have a super- or submodular payoff function, and that as a result the extreme pricing bounds correspond the Frechet-Hoeffding bounds on the copula. We then proceed to investigate how we can extract partial information about the bivariate risk-neutral distribution from products that are trading in the market place. It is important that when we price a new bivariate option that its price is consistent with the products that are currently being traded. We introduce both non-parametric and parametric methods to extract this partial information. Finally, we show how this partial information about the dependence between two assets translates into tighter pricing bounds.

1.2. Short intro to copulas

Copulas allow one to disentangle the dependence structure from the marginal distributions. For concreteness, let's take a dataset: X=$\{x_1,x_2,\ldots,x_N\}$, Y=$\{y_1, y_2,\ldots, y_N\}$. Denote by H: the joint distribution of X and Y.

$$H(x,y) = Pr\{X \leq x, Y \leq y\}$$

The multivariate distribution (H) describes the joint behavior of the original dataset (X,Y), whereas a copula describes the joint behavior of the ranks of the original dataset. Denote by r_i : the rank of x_i in X and s_i: the rank of y_i in Y.

Transformed dataset: $R = \{r_1, r_2, \ldots, r_N\}, S = \{s_1, s_2, \ldots, s_N\}$

Now, rescale R and S by dividing every element by (N+1). The empirical joint cdf of R and S after rescaling is called the empirical copula of X and Y (Deheuvels, 1979). $C_N : [0,1]^2 \rightarrow [0,1]$

$$C_N(u,v) = \frac{1}{N} \sum_{i=1}^{N} I(\frac{r_i}{N+1} \leq u, \frac{s_i}{N+1} \leq v)$$

with I(.): indicator function

It can be shown that $C_N(u,v)$ converges to the true copula $C(u,v)$ as N increases. This construction immediately shows that the copula of X and Y does not change after strictly increasing transformations of X and Y (the sets R and S remain the same), as we would expect from a distribution function capturing the dependence between X and Y.

Formally, Sklar's theorem (1959) states that for every joint distribution function H with margins F and G there exists a copula C, mapping $[0,1]^2$ into [0,1] such that for all x,y:

$$H(x,y) = C[F(x), G(y)]$$

When F and G are continuous, the copula C is unique. Knowing the properties of a joint distribution function H it is straightforward to see that:

(1) $C(0,v) = C(u,0) = 0$ and $C(u,1) = u, C(1,v) = v$

(2) C is a 2-increasing (or supermodular) function just as H

Any function satisfying properties (1) and (2) is a (two-dimensional) copula. By a 2-increasing function (or supermodular function) we mean that for every [x1, x2] $\times$ [y1, y2] in the domain of H we have that H(x2,y2) + H(x1,y1) - H(x1,y2) - H(x2,y1) ≥ 0. When the cross derivative of H (the density of H) exists, this condition is identical to saying that the density must be positive on the entire domain.

As a consequence of properties (1) and (2), it can be readily verified that every copula has to satisfy the *Frechet-Hoeffding bounds* :

$$\max(u + v - 1, 0) \leq C(u,v) \leq min(u,v).$$

We will use the following notation throughout this text (with Π called the independence copula):

$$\begin{aligned} W(u,v) &= \max(u + v - 1, 0) \\ M(u,v) &= \min(u,v) \\ \Pi(u,v) &= u.v \end{aligned}$$

Sklar's theorem clearly shows that in order to model the joint behavior of X and Y, one can proceed in two steps. First, select an adequate model for the dependence between X and Y, captured by the copula. Second, and independently from the first step, select an adequate model for the marginal distributions. This gives us a lot more flexibility to model the joint behavior.

It is also clear from Sklar's theorem that a copula captures the entire dependence structure between X and Y. Hence, knowing which copula best describes the dependence structure between X and Y is much more informative than simply knowing some measure of dependence that summarizes the entire dependence structure, such as the linear (Pearson) correlation coefficient between X and Y. In general, many different dependence structures can be summarized by the same linear correlation coefficient. It is only in the class of the spherical or elliptical distributions (such as the multivariate normal or multivariate student-t) that the marginals combined with the linear correlation coefficient determine the joint distribution (Embrechts et al., 2002). To put it differently: only in the class of elliptical distributions does the linear correlation coefficient capture the entire dependence structure.

1.3. Simulation from copulas

The following 4 graphs in Figure 1.1 illustrate this point. All four graphs have uniform(0,1) margins and a linear correlation coefficient of 0.7. As such, all four graphs display positive dependence. The first graph uses the Clayton copula to link the uniform margins, the second graph uses the student-t copula, the third uses the Frank copula, and

the fourth uses the Gumbel copula. Later on, we provide more information about these copulas. Each graph shows 2000 simulated pairs. Even though the marginals and linear correlation coefficient are the same, the joint distribution is different.

The student-t and Frank copulas have a symmetric dependence structure. Symmetric dependence means that the joint probability that margin 1 is in the α^{th}percentile and margin 2 is in the β^{th} percentile is the same as the joint probability that margin 1 exceeds the 1-α^{th} percentile and margin 2 exceeds the 1-β^{th} percentile. Formally, symmetric dependence implies the copula has to satisfy:

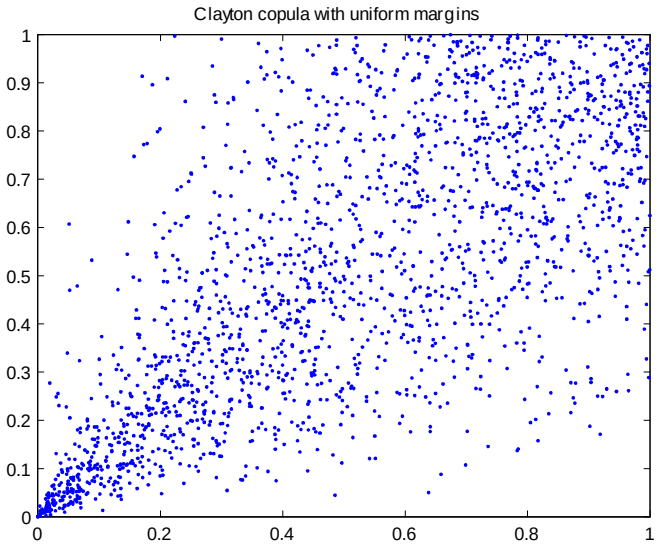
$$\begin{aligned} C(u,v) &= Pr\{U \leq u, V \leq v\} \\ &= Pr\{U \geq 1-u, V \geq 1-v\} \qquad (Symmetry) \\ &= 1 - \{Pr(U < 1-u) + \Pr(V < 1-v) - \Pr(U < 1-u, V < 1-v)\} \\ &= u + v - 1 + C(1-u, 1-v) \\ &= \bar{C}(u,v) \end{aligned}$$

The first and the last copula, Clayton and Gumbel, display asymmetric dependence, but with the direction of asymmetry being different for the two copulas. For the Clayton copula, the joint probability of both margins being in a low percentile exceeds the joint probability of both margins being in a high percentile, which can be observed in the graph, and can be computed as shown in Table 1.1. For the Gumbel copula the asymmetry is in the opposite direction.

Table 1.1. Opposite dependence asymmetry for Clayton and Gumbel copula

Percentile(u)	Clayton		Gumbel	
	C(u,u)	$\bar{C}$(u,u)	C(u,u)	$\bar{C}$(u,u)
1%	0.0074	0.0003	0.0015	0.0059
5%	0.0370	0.0074	0.0145	0.0300
10%	0.0741	0.0269	0.0385	0.0616

Figure 1.1. Clayton simulated pairs



Although the marginals are the same for all four graphs and linear correlation equals 70% for all graphs the difference in dependence structure is clear.

1.4. Copula density contours

In addition to plot simulated pairs from a copula another way to demonstrate the dependence structure of a copula is to plot the copula density contours. The next four

Figure 1.2. Student-t simulated pairs

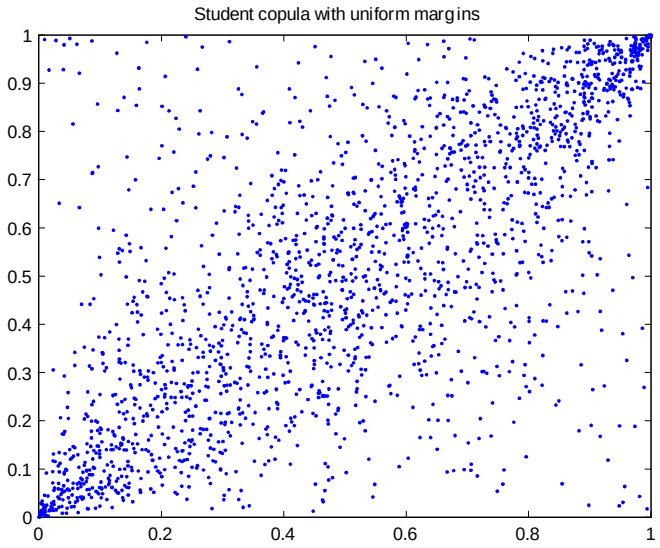
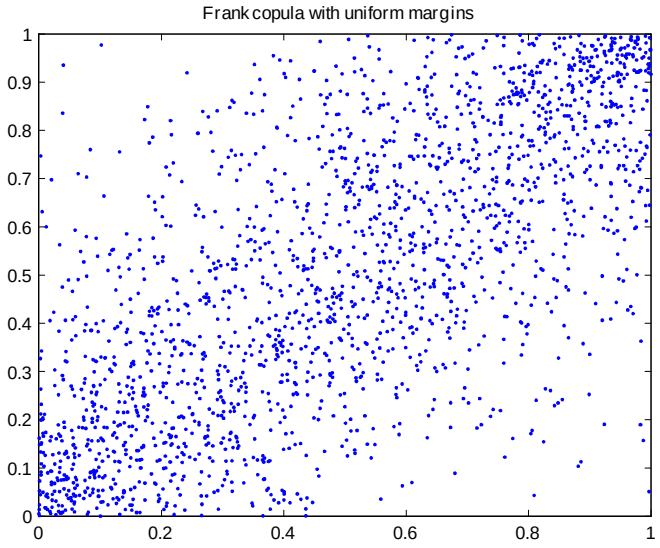
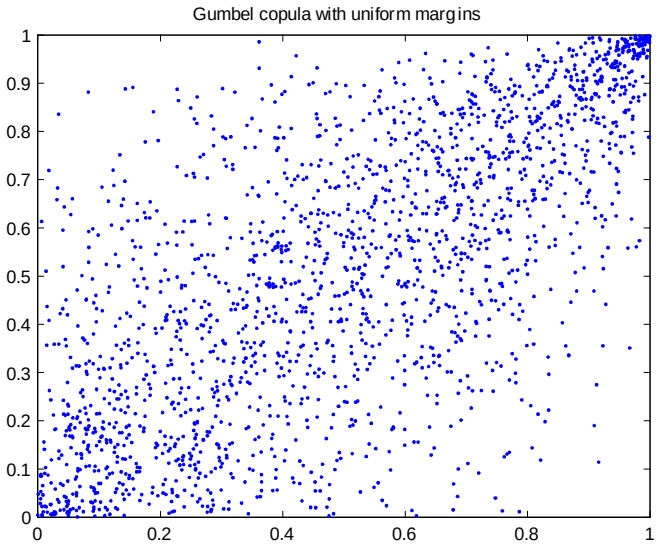


Figure 1.3. Frank simulated pairs



graphs in Figure 1.5 plot the density contours of the same four copulas (Clayton, Student-t, Frank, and Gumbel) using the same parameters as in Figure 1. To obtain the expression

Figure 1.4. Gumbel simulated pairs



for the density of a copula one can use one of several methods depending on the class to which the copula belongs.

1.4.1. Archimedean copulas

The Clayton, Frank, and Gumbel copulas belong to the very important class of Archimedean copulas (Genest and MacKay, 1986a; Genest and Rivest, 1993). Copulas in this class are generated by a continuous function $\phi : [0,1] \rightarrow [0,\infty)$, which is strictly decreasing and convex, and satisfies $\phi(1) = 0$. A generator is said to be a *strict* generator if $\phi(0) = \infty$. Archimedean copulas satisfy the following relation:

$$\begin{aligned} \phi[C(u,v)] &= \phi(u) + \phi(v) \\ C(u,v) &= \phi^{[-1]}[\phi(u) + \phi(v)] \end{aligned} \tag{1.1}$$

with the pseudo-inverse $\phi^{[-1]}(z)$ defined as:

$$\begin{aligned} \phi^{-1}(v) \text{ for } 0 &\leq z \leq \phi(0) \\ 0 \text{ for } \phi(0) &\leq z \leq \infty \end{aligned}$$

In the case of a strict generator the pseudo-inverse and the regular inverse coincide.

1.4.1.1. Clayton copula. First we describe the Clayton copula, which is given by:

$$C(u,v) = max([u^{-\gamma} + v^{-\gamma} - 1]^{-\frac{1}{\gamma}}, 0) \qquad \gamma \in [-1, \infty) \backslash \{0\}$$

A large subclass of the Archimedean copulas have generators that are twice differentiable, such as the generator for the Clayton copula:

$$\phi(t) = \frac{t^{-\gamma} - 1}{\gamma}$$

Taking the total derivative of relation (1.1) one obtains the following expression for the cross derivative (density) of the copula:

$$-\frac{\phi''[C\,(u,v)].\phi'(u)/\phi'(v)}{[\phi'[C(u,v)]^3} \tag{1.2}$$

1.4.1.2. Frank copula. The Frank copula is given by:

$$C(u,v) = -\frac{1}{\gamma}\log(1 + \frac{(e^{-\gamma u} - 1).(e^{-\gamma v} - 1)}{e^{-\gamma} - 1}) \qquad \gamma \in (-\infty, \infty) \backslash \{0\}$$

The generator of the Frank copula is:

$$\phi(t) = -\log(\frac{e^{-\gamma t} - 1}{e^{-\gamma} - 1})$$

Although it is possible to get the density using expression (1.2), it is easier to directly obtain the density as the cross derivative of the copula.

1.4.1.3. Gumbel copula. The Gumbel copula is given by:

$$C(u,v) = exp(-[(-\log u)^{\gamma} + (-\log v)^{\gamma}]^{\frac{1}{\gamma}}) \qquad \gamma \in [1,\infty)$$

Generator of Gumbel copula:

$$\phi(t) = (-\log t)^{\gamma}$$

1.4.2. Elliptical copulas

1.4.2.1. Student-t copula. Its expression is:

$$C^{St}_{r,\mu}(u,v) = \int_{-\infty}^{T_{\mu}^{-1}(u)} \int_{-\infty}^{T_{\mu}^{-1}(v)} \frac{1}{2\pi\sqrt{1-r^2}} \left(1 + \frac{x^2+y^2-2rxy}{\mu(1-r^2)}\right)^{-\frac{\mu+2}{2}} dx.dy$$

with r:correlation, μ:degrees of freedom, T_{μ}(.):univariate student-t cdf

Knowing the expression for the bivariate student-t distribution function ($H^{St}_{r,\mu}$) one can immediately see that:

$$H^{St}_{r,\mu}(x,y) = C^{St}_{r,\mu}[T_{\mu}(x), T_{\mu}(y)]$$

Taking the cross derivative we get:

$$h^{St}_{r,\mu}(x,y) = c^{St}_{r,\mu}[T_{\mu}(x), T_{\mu}(y)].t_{\mu}(x).t_{\mu}(y).$$

Hence,

$$c^{St}_{r,\mu}(u,v) = \frac{h^{St}_{r,\mu}[T^{-1}_{\mu}(u), T^{-1}_{\mu}(v)]}{t_{\mu}[T^{-1}_{\mu}(u)].t_{\mu}[T^{-1}_{\mu}(v)]}$$

Plugging in the univariate and bivariate student-t densities we obtain the student-t copula density:

$$c^{St}_{r,\mu}(u,v) = \frac{\mu}{2}.\frac{1}{\sqrt{1-r^2}}.\frac{\Gamma(\frac{\mu}{2})^2}{\Gamma(\frac{\mu+1}{2})^2}.\frac{[1+\frac{\varepsilon_1^2+\varepsilon_2^2-2r\varepsilon_1\varepsilon_2}{\mu(1-r^2)}]^{-\frac{\mu+2}{2}}}{(1+\frac{\varepsilon_1^2}{\mu})^{-\frac{\mu+1}{2}}.(1+\frac{\varepsilon_2^2}{\mu})^{-\frac{\mu+1}{2}}}$$

with r:correlation coefficient μ:degrees of freedom

$$\varepsilon_1 = T^{-1}_{\mu}(u),\ \varepsilon_2 = T^{-1}_{\mu}(v)$$

1.4.2.2. Normal copula. The normal copula is defined as:

$$C^{Gau}_{r}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-r^2}}.\exp\left(\frac{2rxy-x^2-y^2}{2(1-r^2)}\right) dx.dy$$

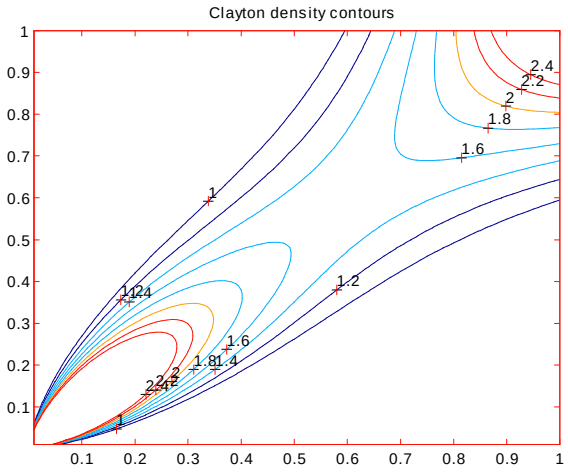
with r:correlation, and $\Phi(.)$:univariate standard normal cdf

The normal copula density has the following form:

$$c^{Gau}_{r}(u,v) = \frac{1}{\sqrt{1-r^2}}.\exp\left(\frac{\varepsilon_1^2+\varepsilon_2^2}{2} + \frac{2r\varepsilon_1\varepsilon_2-\varepsilon_1^2-\varepsilon_2^2}{2(1-r^2)}\right)$$

$$\varepsilon_1 = \Phi^{-1}(u),\ \varepsilon_2 = \Phi^{-1}(v)$$

Figure 1.5. Clayton density contours



The normal copula leads to the bivariate standard normal distribution when we use standard normal marginals.

The four graphs in Figure 1.5 show clearly the symmetric dependence structure of the student-t and Frank copula. The Clayton density contours corroborate that the 2 margins are more often simultaneously low than high, and the Gumbel density contours show that the 2 margins are more often simultaneously high than low.

1.5. Distribution of Sum of Random variables

A third way to demonstrate the difference in dependence structure despite having the same uniform marginals and linear correlation coefficient is to compare the Pr(X+Y $\leq$ t) for t $\in [0, 2]$ for the three different copulas. This probability will also get us closer to one of the objects of interest, $E\{max(aX + bY - K, 0)\}$. Taking into account the

Figure 1.6. Student-t density contours

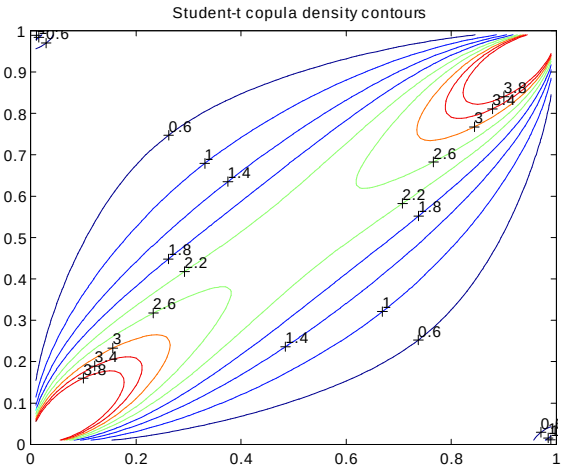
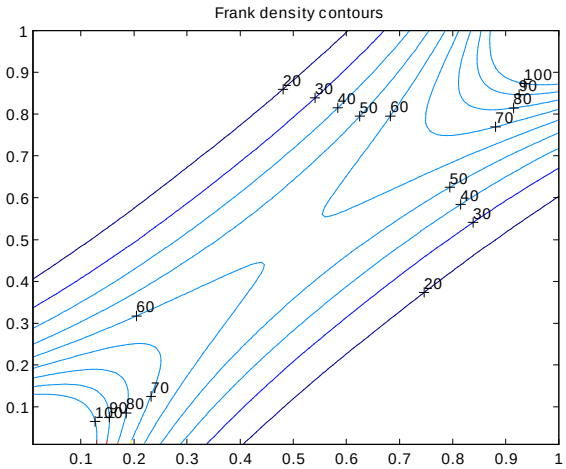


Figure 1.7. Frank density contours



information on the dependence structure between X and Y from Figures 1.1 and 1.5, one would expect that Pr(X+Y $\leq$ t) for small t is highest for the Clayton copula. Table 1.2 shows the results for 50000 simulated pairs from each copula with a linear correlation equal to 0.7 for all four copulas. Given that X and Y are uniform (0,1) marginals, the Pr(X+Y $\leq$ t) for t $\leq$ 0 equals 0, and for t $\geq$ 2 equals 1.

Figure 1.8. Gumbel density contours

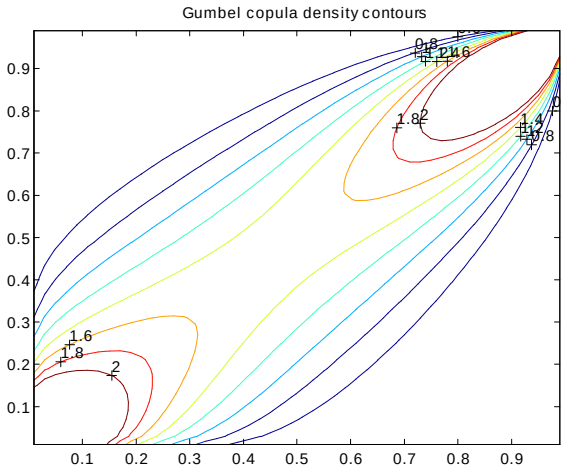


Table 1.2. Table 2: Pr(X+Y = t) using 50000 simulated pairs

t	Clayton	Student-t	Frank	Gumbel
0.10	4.58%	3.55%	2.14%	2.33%
0.25	11.48%	9.38%	8.99%	8.26%
0.50	22.98%	21.04%	22.54%	21.26%
1.00	47.74%	50.01%	49.98%	51.29%
1.50	78.10%	78.70%	77.20%	78.45%
1.75	92.68%	90.16%	91.03%	90.09%
1.90	98.55%	96.19%	97.84%	96.24%

These results corroborate what we have learned from Figures 1.1 and 1.5. The Student-t and Frank copulas are clearly symmetric: $\Pr(X+Y \leq t) \simeq \Pr(X+Y \geq 2\text{-}t)$, which implies that $\Pr(X+Y \leq 1) \simeq 50\%$. The results for the Clayton copula show once again the asymmetry between X and Y, in that $\Pr(X+Y \leq t) > \Pr(X+Y \geq 2\text{-}t)$ for small t, and the opposite sign holds true for the Gumbel copula. Knowing the copula of X and Y and knowing the marginal distribution of X and Y allows us to analytically compute

Pr(X+Y $\leq$ t). In the case that the support of X and Y is the entire real line we get the following expression ($X \sim F_1, Y \sim F_2, (X,Y) \sim F$):

$$\begin{aligned} \Pr(X+Y \leq t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{t-x} f(x,y)dy.dx = \int_{-\infty}^{\infty} \frac{\delta}{\delta X} F_{X,Y}(x, t-x)dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{t-x} c[F_1(x), F_2(y)].f_1(x).f_2(y)dy.\ dx \end{aligned}$$

This expression always holds when the support of X and Y covers the entire real line. It holds for symmetric as well as asymmetric distributions (copulas). Returning to the case of uniform(0,1) marginals one has to be careful to take into account the boundary conditions and the possible asymmetry in the dependence structure. $Pr(X+Y \leq t) =$

$$\begin{cases} 0 & t \leq 0 \\ \int_0^t \int_0^{t-x} c_{X,Y}(x,y)dy.dx = \int_0^t \frac{\delta}{\delta X} C_{X,Y}(x, t-x)dx & 0 \leq t \leq 1 \\ t - 1 + \int_{t-1}^1 \frac{\delta}{\delta X} C_{X,Y}(x, t-x)dx & 1 \leq t \leq 2 \\ 1 & t \geq 2 \end{cases}$$

This result holds for symmetric as well as asymmetric copulas, such as the Clayton and Gumbel copulas. Plugging in the correct copula densities we can numerically evaluate these integrals in a matter of seconds. Doing that confirms the results in Table 1.2. For completeness and as a reference for other parts of this paper we give the results for

Pr(aX+bY $\leq$ t) when both X and Y are uniform(0,1) and have a copula C. There are six separate cases to consider (a, b > 0). $Pr(aX + bY \leq t) =$

$$
\begin{array}{ll}
0 & t \leq 0 \\
\int_0^{\frac{t}{a}} \frac{\delta}{\delta X} C_{X,Y}(x, \frac{t-ax}{b}) dx & 0 \leq t \leq \min(a,b) \\
\int_0^1 \frac{\delta}{\delta X} C_{X,Y}(x, \frac{t-ax}{b}) dx & t \leq b, t > a \\
\frac{t-b}{a} + \int_{\frac{t-b}{a}}^{\frac{t}{a}} \frac{\delta}{\delta X} C_{X,Y}(x, \frac{t-ax}{b}) dx & t > b, t \leq a \\
\frac{t-b}{a} + \int_{\frac{t-b}{a}}^1 \frac{\delta}{\delta X} C_{X,Y}(x, \frac{t-ax}{b}) dx & \max(a,b) \leq t \leq a+b \\
1 & t \geq a+b
\end{array}
$$

It is straightforward to check that for a = b = 1 we recover the results for Pr(X+Y $\leq$ t).

1.6. Option Pricing

The objective is to price bivariate options:

$$
\begin{aligned}
C(k) &= e^{-r(T-t)}.E_t^Q[G(X_T, Y_T)] \\
&= e^{-r(T-t)}.E_t^Q[\max\{\psi(X_T, Y_T) - k, 0\}] = e^{-r(T-t)}.E_t^Q[\psi(X_T, Y_T) - k]^+ .
\end{aligned}
$$

t: time at which the option price is computed

T: final time of the contract

r: risk-free interest rate (T-bill rate at time t for an interval T-t)

k: strike price (fixed parameter)

Q: the *risk-neutral* joint distribution over which the conditional expectation is taken

X, Y: positive random variables

We focus on bivariate options of the European type, which implies that from a pricing perspective we are only interested in the terminal joint distribution of X and Y. For now, ψ can be any function mapping $R_+^2 \rightarrow R$. The crucial issue is to find the terminal risk-neutral distribution or risk-neutral copula of X and Y, which in general is distinct from the joint distribution based on historical prices. It can be shown that under certain conditions the risk-neutral and the historical copula are the same, but more on that later on. There are two ways to compute C(k). One method simply uses the joint density, the other is based on the conditional distribution.

$$\begin{aligned} E_t^Q[G(X_T, Y_T)] &= \int_0^\infty \int_0^\infty G(x,y).h^Q(x,y)dydx \\ &= \int_0^\infty \int_{y \in A(x)} \{\psi(x,y) - k\}.h^Q(x,y)dydx \\ &= \int_0^\infty \int_{y \in A(x)} \{\psi(x,y) - k\}.c^Q[F_1^Q(x), F_2^Q(y)].f_1^Q(x).f_2^Q(y).dydx \end{aligned}$$

$h^Q(x,y)$: risk-neutral joint density of X and Y.

$c^Q(u,v)$: risk-neutral copula density of X and Y.

$F_i^Q(z)$: risk-neutral univariate distribution

$f_i^Q(z)$: risk-neutral univariate density

$A(x) = \{y \in R_+ : \psi(x,y) \geq k\}$

We can also use a conditional distribution approach:

$$\begin{aligned} E_t^Q[G(X_T, Y_T)] &= \int_k^\infty P^Q[\psi(X_T, Y_T) > s]ds \\ &= \int_k^\infty \int_0^\infty P^Q[\psi(X_T, y) > s, Y_T = y]dy.ds \\ &= \int_k^\infty \int_0^\infty P^Q[X_T < \phi(s, y), Y_T = y]dy.ds \\ &= \int_k^\infty \int_{y \in B(s)} \delta_2 H^Q[\phi(s, y), y]dy.ds \\ &= \int_k^\infty \int_{y \in B(s)} \delta_2 C^Q[F_1^Q(\phi(s, y)), F_2^Q(y)].f_2^Q(y).dy.ds \end{aligned}$$

P^Q: probability over the risk-neutral measure

$H^Q(x, y)$: risk-neutral joint distribution of X and Y

$C^Q(u, v)$: risk neutral copula of X and Y

$\delta_i Z(x, y)$: derivative with respect to the ith entry of Z

$B(s)$ =$\{$y$\in$R$_+$: ϕ(s,y)$\geq$0$\}$

It is clear from these expressions that in order to price options with a payoff function that depends on both X and Y we need to obtain the risk-neutral univariate distributions of both X and Y, and we require knowledge about the risk-neutral copula of X and Y. There already exists a vast literature on how to obtain the univariate risk-neutral distributions from market prices on options and futures, which will be summarized in a later section. Methods to obtain the risk-neutral copula from multivariate options do not yet exist. In general it is wrong to work with a joint distribution based on the historical joint behavior of X and Y. In order to ensure that there exist no arbitrage opportunities

between the univariate options on X and Y and the bivariate options on ψ(X,Y), it has to be true that the margins of the risk-neutral joint distribution equal the univariate risk-neutral distributions. In later sections we show some simple examples of how to take advantage of a violation of this requirement. Without going into detail at this point, it is important to note that market prices of bivariate options can never fully identify the implied risk-neutral joint distribution over the entire domain. Bivariate options only convey information about a fraction of the joint distribution. If we do not have a lot of market prices for bivariate options, our non-parametric estimates for the risk-neutral joint distribution will be uninformative. Our only recourse at that point is either to rely on historical behavior or to propose parametric forms for the risk-neutral copula. In the latter case, even a couple of market prices for bivariate options can lead to very tight bounds for the set of permissible parameters. But before we get into the details of estimating the implied univariate and bivariate risk-neutral distributions from option and futures prices, we first work out in detail the expressions for C(k) for different payoff functions G. In particular, we look at (k>0):

$$\begin{aligned} G(X,Y) &= [aX + bY - k]^+ \text{ with a,b>0: Basket option} \\ &= [aX + bY - k]^+ \text{ with a or b<0: Spread option} \\ &= [\max(aX, bY) - k]^+ \text{ with a,b>0: Rainbow-max option} \\ &= [\min(aX, bY) - k]^+ \text{ with a,b>0: Rainbow-min option} \end{aligned}$$

1.7. Basket Option

We now focus our attention on computing bounds for:

$$E\max(aX + bY - k, 0) = E(aX + bY - k)^+ \qquad for \;\; a, b > 0.$$

$$with\ X \sim F_X, Y \sim F_Y,\ and\ (X, Y) \sim F$$

Let us assume for now that the support of X and Y covers the entire real line. Also, define R = aX, S = bY, and R+S = T.

$$\begin{aligned} E(R+S-k)^+ &= E(T-k)^+ = E(T-k|T>k).Pr(T>k) \\ &= \int_k^\infty (s-k).\frac{dF_T(s)}{1-F_T(k)}.(1-F_T(k)) = \int_k^\infty (s-k).dF_T(s) \\ &= \int_k^\infty \Pr(T>u).du \qquad \text{(Integration by parts)} \\ &= \int_k^\infty 1 - \Pr(R+S \le u).du \end{aligned}$$

As stated above, the copula remains the same for strictly increasing transformations of the marginals. For a, b > 0 we have that the copula of X and Y is the same as the copula of R and S. Hence, knowing the copula of X and Y implies a point estimate for $E(R+S-k)^+$.The problem, of course, is that it is a difficult task to exactly determine the copula of X and Y. Uncertainty about the exact copula of X and Y results in bounds for $E(R+S-k)^+$. In what follows we will first compute the bounds without assuming anything about the joint behavior of X and Y. One expects wide bounds in that case. In a second step we show how the bounds shrink when we make specific assumptions about

both the structure and the degree of dependence between X and Y. In both cases we illustrate the results with some examples.

1.7.1. Bounds without information on dependence structure

A classic problem in statistics (Kolmogorov problem) is to determine bounds for $Pr(R + S \leq u)$ without assuming any information about the dependence structure between R and S (Makarov, 1981; Ruschendorf, 1982; Frank and Schweizer, 1987). These authors show that $Pr(R + S \leq u)$ is bounded by:

$$
\begin{aligned}
(1.3) \qquad F_{\min}(u) &= \sup_x \max\{F_R(x) + F_S(u-x) - 1, 0\} \\
&\leq \Pr(R + S \leq u\} \\
&\leq \inf_x \min\{F_R(x) + F_S(u-x), 1\} = F_{\max}(u)
\end{aligned}
$$

In terms of F_X and F_Y we have (a, b > 0):

$$
\begin{aligned}
F_{\min}(u) &= \sup_x \max\{F_X(\frac{x}{a}) + F_Y(\frac{u-x}{b}) - 1, 0\} \\
&\leq \Pr(aX + bY \leq u\} \\
&\leq \inf_x \min\{F_X(\frac{x}{a}) + F_Y(\frac{u-x}{b}), 1\} = F_{\max}(u)
\end{aligned}
$$

These bounds have analytical solutions in only a limited number of cases (for example: uniform, normal, and exponential). In general these bounds have to be evaluated numerically (Williamson and Downs, 1990).

1.7.1.1. Example 1a: In the first example we take a look at two standard normals.

$X_1 \sim N(0,1), X_2 \sim N(0,1)$

$\implies F_1(x) = F_2(x) = \Phi(x)$

1) Fmin

$F_{\min}(u) = \sup_x \max[\Phi(x) + \Phi(u-x) - 1, 0]$

u=0: $\Phi(x) + \Phi(-x) = 1 \implies \mathrm{F}_{\min}(u) = 0$

u>0: We can always set x=0, which implies that $\Phi(0) + \Phi(u) - 1 > 0$

$\implies F_{\min}(u) = \sup_x \Phi(x) + \Phi(u-x) - 1 \implies F_{\min}(u) = 2\Phi(\frac{u}{2}) - 1$

u<0: $\Phi(x) + \Phi(u-x) < 1$ for all x

$\implies \mathrm{F}_{\min}(u) = 0$

2) Fmax

$F_{\max}(u) = \inf_x \min[\Phi(x) + \Phi(u-x), 1]$

u=0: $\Phi(x) + \Phi(-x) = 1 \implies \mathrm{F}_{\max}(u) = 1$

u>0: $\Phi(x) + \Phi(u-x) > 1$ for all x $\implies \mathrm{F}_{\max}(u) = 1$

u<0: We can always set x=0, which implies that $\Phi(0) + \Phi(u) < 1$

$\implies F_{\max}(u) = \inf_x \Phi(x) + \Phi(u-x)$

$\implies F_{\max}(u) = 2\Phi(\frac{u}{2})$

$$\begin{aligned} F_{\min}(u) &= 0 & u \leq 0 \\ &= 2\Phi(\frac{u}{2}) - 1 & u > 0 \\ F_{\max}(u) &= 2\Phi(\frac{u}{2}) & u < 0 \\ &= 1 & u \geq 0 \end{aligned}$$

1.7.1.2. Example 2a: In the second example we combine one standard normal with one exponentially distributed variable.

$X_1 \sim \exp(1), X_2 \sim N(0,1)$

$$\begin{aligned} \Longrightarrow\ F_1(x) &= 1 - \exp(-x) & x \geq 0 \\ &= 0 & x \leq 0 \\ \Longrightarrow\ F_2(x) &= \Phi(x) \end{aligned}$$

1) Fmin

$F_{\min}(u) = \sup_x \max[F_1(x) + \Phi(u-x) - 1, 0]$

1a) The optimal x such that $F_{\min}(u) > 0$ has to be strictly positive.

$$\begin{aligned} \Longrightarrow\ F_{\min}(u) &= \sup_{x>0} \max[F_1(x) + \Phi(u-x) - 1, 0] \\ &= \sup_{x>0} \max[1 - \exp(-x) + \Phi(u-x) - 1, 0] \\ &= \sup_{x>0} \max[\Phi(u-x) - \exp(-x), 0] \end{aligned}$$

1b) There exists u* such that:

$F_{\min}(u) \geq 0$ for u $\geq$ u*

$F_{\min}(u) = 0$ for u $\leq$ u*

1c) Determine u* and $F_{\min}(u)$ for u $\geq$ u*.

Assume u is such that $F_{\min}(u) > 0$

$$\implies F_{\min}(u) = \sup_{x>0} \Phi(u-x) - \exp(-x)$$
$$= \Phi(u - X^*(u)) - \exp(-X^*(u))$$

with $X^*(u) = 1 + u - \sqrt{1 - \log(2\pi) + 2u}$

$\Phi(u - X^*(u)) - \exp(-X^*(u))$: strictly increasing in u and

$\Phi(u - X^*(u)) - \exp(-X^*(u)) \geq 0$ for $u \geq u^* = 0.6621$

$X^*(u) > 0$ for $u \geq u^*$

2) Fmax

$F_{\max}(u) = \inf_x \min[F_1(x) + \Phi(u-x), 1]$

2a) It is always possible to find x such that $F_1(x) + \Phi(u-x) \leq 1$.

For example, setting x=0 will do the job.

2b) The optimal x cannot be strictly negative as $F_1(x) + \Phi(u-x)$

is always smaller for x=0 than for x < 0.

$$\begin{aligned} \Longrightarrow \; F_{\max}(u) &= \inf_{x\geq 0} F_1(x) + \Phi(u-x) \\ &= \inf_{x\geq 0} 1 - \exp(-x) + \Phi(u-x) \end{aligned}$$

Numerically it can be shown that $x = 0$ is optimal for all u.

$$\begin{aligned} F_{\min}(u) &= \Phi(u - X^*(u)) - \exp(-X^*(u)) && u \geq u^* \\ &= 0 && u \leq u^* \\ with\ X^*(u) &= 1 + u - \sqrt{1 - \log(2\pi) + 2u} \\ F_{\max}(u) &= \Phi(u) \end{aligned}$$

1.7.1.3. Numerical bounds . As mentioned above, in general, the bounds $F_{\min}(u)$ and $F_{\max}(u)$ have to be evaluated numerically. Next, I will test the speed and accuracy of the numerical algorithms developed by Williamson and Downs (1990) in applying them to the previous example 2a for which we know the analytical solution. These algorithms are based on a duality principle by Frank and Schweizer (1979) that transforms the formulation of equation (1.3), which requires optimization over the entire real line, to a bounded optimization problem. Makarov(1981) was the first to establish the result in (1.3), but proved the dual version of it, using a long and cumbersome proof. On the other hand, the proof to obtain the expression (1.4) is fairly straightforward, and the equivalence between the two versions can be shown relying upon a well known duality argument. The advantage of the Makarov version is that it suits itself better to numerical implementation as

it is a bounded optimization problem. Makarov's version, which is the dual of equation (1.3), is as follows:

Denote the left-continuous, quasi-inverse of the cdf F by :

$$F^*(a) = \inf\{x|F(x) \geq a\}$$

$$\inf_{W(s,t)=a}[F_1^*(t) + F_2^*(s)] \leq [cdf(X+Y)]^*(a) \leq \sup_{Z(s,t)=a}[F_1^*(t) + F_2^*(s)] \tag{1.4}$$

$$with\ W(s,t) = max[F_1(s) + F_2(t) - 1, 0]\ and\ Z(s,t) = s + t - W(s,t)$$

The code is implemented in Matlab. In the algorithm the range of the distribution function is split up in N discrete intervals. Williamson and Downs (1990) prove that for N increasing the approximate bounds $\mathrm{F}_{\min}^N$ and $\mathrm{F}_{\max}^N$ converge to $\mathrm{F}_{\min}$ and $\mathrm{F}_{\max}$. As can be judged from Figure 1.9 and 1.10, which displays the bounds for Example 2a, convergence happens very rapidly as for N=100 the true and approximate bounds become indistinguishable.

Staying with example 2a ($X_1 \sim \exp(1), X_2 \sim N(0,1)$) we can now compute bounds for $E(X_1 + X_2 - k)^+$ without assuming anything about the dependence between X_1 and X_2 :

$$\int_k^\infty 1 - F_{\max}(u).du \leq\ E(X_1 + X_2 - k)^+ \leq \int_k^\infty 1 - F_{\min}(u).du$$

Figure 1.9. Bounds for the probability distribution for the sum of 2 variables with no assumption on dependence structure for N=10

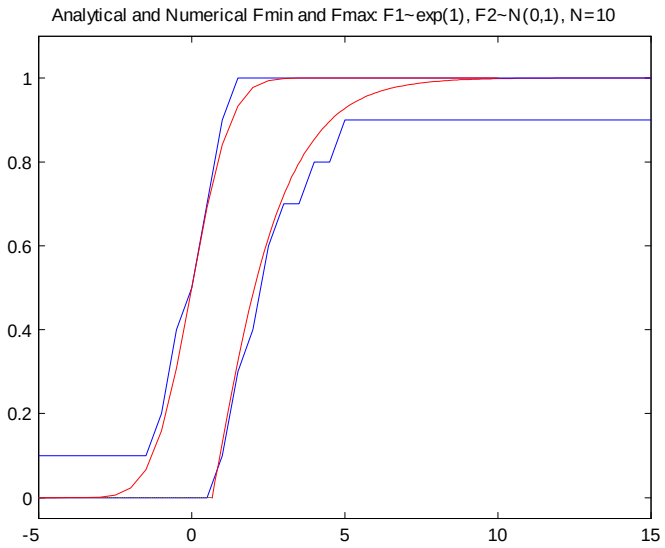


Figure 1.10. Bounds for the probability distribution for the sum of 2 variables with no assumption on dependence structure for N=100

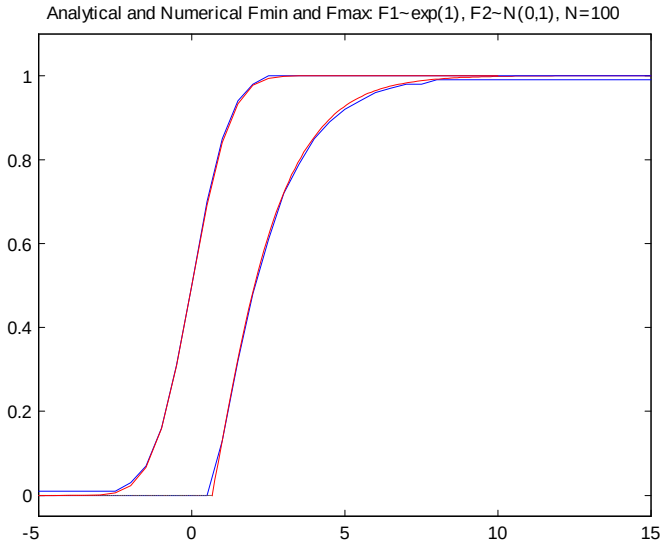
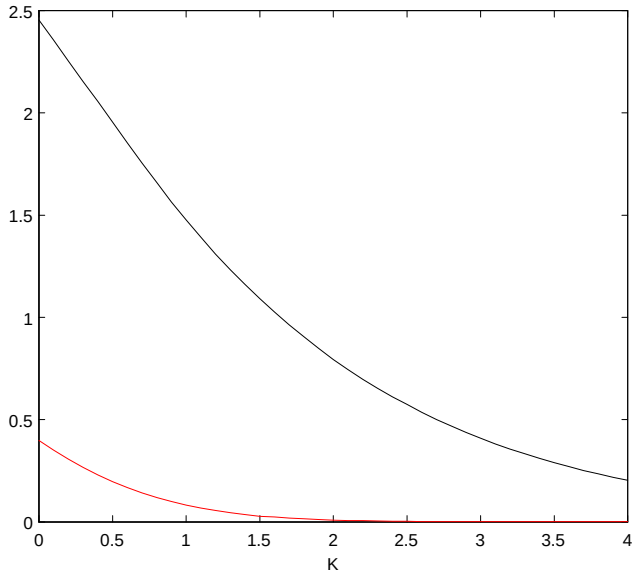


Figure 1.11. Upper and lower bound for E(max(X1+X2-K,0)): no assumption on dependence, F1 N(0,1),F2 exp(1)



For k $\geq$ u* :

$$\int_k^\infty \Phi(-u).du \leq \ E(X_1 + X_2 - k)^+ \leq \int_k^\infty 1 - \Phi(u - X^*(u)) + \exp(-X^*(u)).du$$

For k $<$ u* :

$$\int_k^\infty \Phi(-u).du \leq \ E(X_1 + X_2 - k)^+ \leq u^* - k + \int_{u^*}^\infty 1 - \Phi(u - X^*(u)) + \exp(-X^*(u)).du$$

These bounds have to be numerically integrated, which can be done in a matter of seconds. Figure 1.11 shows the lower and upper bound as a function of k. As expected these bounds are very wide when k is small.

1.7.1.4. Bounds with information on dependence structure . In the next step we investigate how we can shrink the bounds by making assumptions on the structure and degree of dependence between X_1 and X_2. One option is to impose a lower bound on the joint distribution or copula of X_1 and X_2. We can assume that there exists a distribution function L and a corresponding copula C_L such that:

$$Pr(X_1 \leq x, X_2 \leq y) \geq L(x, y) = C_L[F_1(x), F_2(y)].$$

This particular assumption on the dependence between X_1 and X_2 implies the following lower and upper bound on $Pr(X_1 + X_2 \leq u)$:

Claim 1.

$$\begin{aligned} F_{\min}(u) &\leq F_{\min,L}(u) \\ &= \sup_x L(x, u-x) \\ &\leq Pr(X_1 + X_2 \leq u) \\ &\leq \inf_x F_1(x) + F_2(u-x) - L(x, u-x) \\ &= F_{\max,L}(u) \leq F_{\max}(u) \end{aligned} \tag{1.5}$$

Proof.

Lower bound $\forall x : L(x, u-x) \leq \Pr(X_1 \leq x, X_2 \leq u-x) \leq \Pr(X_1 + X_2 \leq u)$

$$\begin{aligned}
\text{Upper bound } \forall x: \quad \Pr(X_1 + X_2 \leq u) &\leq \Pr(X_1 \leq x \text{ or } X_2 \leq u - x) \\
&= \Pr(X_1 \leq x) + \Pr(X_2 \leq u - x) \\
- \Pr(X_1 &\leq x, X_2 \leq u - x) \\
&\leq F_1(x) + F_2(u - x) - L(x, u - x)
\end{aligned}$$

□

Imposing an upper bound on $Pr(X_1 \leq x, X_2 \leq y)$ does not help us to reduce the size of the bounds any further. Recalling the Frechet-Hoeffding bounds for copulas stated above:

$$max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v)$$

This implies similar bounds for any bivariate distribution function H with margins F_1 and F_2:

$$\max[F_1(x) + F_2(y) - 1, 0] \leq H(x, y) \leq \min[F_1(x), F_2(y)]$$

Therefore, making no assumptions about the dependence between X_1 and X_2 is the same as assuming that:

$$Pr(X_1 \leq x, X_2 \leq y) \geq \max[F_1(x) + F_2(y) - 1, 0]$$

It is straightforward to check that when we plug the lower Frechet-Hoeffding bound into (1.5) we recover the result from equation (1.3).

If we were to impose that:

$$Pr(X_1 \leq x, X_2 \leq y) \geq \min[F_1(x), F_2(y)]$$

we would obtain a point estimate for $\Pr(X_1 + X_2 \leq u)$ as there is no longer uncertainty about the copula of X_1 and X_2 in that case, X_1 and X_2 are comonotonic. Given that $F_1(x)$ is increasing in x and $F_2(u-x)$ is decreasing in x, there always exists an interval $[x_-, x^+]$ such that

$$\begin{aligned} F_1(x) &= F_2(u-x) \quad \forall x \in [x_-, x^+] \\ F_1(x) &< F_2(u-x) \quad x < x_- \\ F_1(x) &> F_2(u-x) \quad x > x^+ \end{aligned}$$

When $F_1(.)$ and $F_2(.)$ are strictly increasing, $x_- = x^+$.

The expression for $Pr(X_1 + X_2 \leq u)$ for comonotonic X_1 and X_2 is[1]:

$$\begin{aligned} F_{\min,M}(u) &= \sup_x \min[F_1(x), F_2(u-x)] = F_1(x^*) = F_2(u-x^*) \\ F_{\max,M}(u) &= \inf_x F_1(x) + F_2(u-x) - \min[F_1(x), F_2(u-x)] = F_1(x^*) = F_2(u-x^*) \\ with\ x^* &\in [x_-, x^+] \end{aligned}$$

[1]Another way to show this expression: $\int_{-\infty}^{\infty} \frac{\delta}{\delta X} \min[F_1(x), F_2(u-x)]dx = \int_{-\infty}^{x_-} f_1(x)dx = F_1(x_-) = F_1(x^*)$

1.7.1.5. Positive dependence . In the empirical analysis that will follow, one of the first assumptions we make is positive dependence between power and natural gas prices. To achieve that goal we assume that $Pr(X_1 \leq x, X_2 \leq y) \geq F_1(x).F_2(y)$. This property is called PQD (Positive Quadrant Dependence), and means that small (large) values of X_1 tend to occur more often with small (large) values of X_2 than in the case of independence.

Hoeffding's lemma (joint distribution H, margins $F_1$and F_2) states that:

$$cov(X,Y) = \int\int_{R^2} [H(x,y) - F_1(x).F_2(y)]dx.dy$$

As a result of Hoeffding's lemma, we can conclude that assuming PQD for X_1 and X_2 implies the assumption of positive linear correlation between X_1 and X_2, which is a property that we safely assume to be true between power and natural gas prices based on a supply and demand argument.

The assumption of PQD between X_1 and X_2 leads to the following bounds for $Pr(X_1 + X_2 \leq u)$[2] :

[2]As mentioned above, the independence copula is denoted as Π(u,v)=uv. Therefore, we denote the bounds in the case of PQD as $F_{min,\Pi}$ and $F_{max,\Pi}$.

$$
\begin{aligned}
F_{\min,\Pi}(u) &= \sup_x F_1(x).F_2(u-x) \\
F_{\max,\Pi}(u) &= \inf_x F_1(x) + F_2(u-x) - F_1(x).F_2(u-x) \\
&= \inf_x [1 - \Pr(X_1 > u, X_2 > u-x)] \\
&= \inf_x \{1 - [1 - F_1(x)].[1 - F_2(u-x)]\}
\end{aligned}
$$

Example 1b. In this example 1b we use the same marginals as in example 1a, but now we impose PQD between X_1 and X_2.

$X_1 \sim N(0,1), X_2 \sim N(0,1)$

$Pr(X_1 \leq x, X_2 \leq y) \geq F_1(x).F_2(y)$

$$
\begin{aligned}
F_{\min,\Pi}(u) &= \sup_x \Phi(x).\Phi(u-x) = [\Phi(\frac{u}{2})]^2 \\
F_{\max,\Pi}(u) &= \inf_x \{1 - [1 - \Phi(x)].[1 - \Phi(u-x)]\} \\
&= 1 - \sup_x \Phi(-x).\Phi(x-u) = 1 - [\Phi(-\frac{u}{2})]^2
\end{aligned}
$$

Figure 1.12 displays the bounds on $Pr(X_1 + X_2 \leq u)$ for example 1a (no information on dependence) and 1b (impose PQD on dependence) for the case of 2 standard normal marginal dsitributions.

Figure 1.12. Bounds on Pr(X1+X2=u) as a function of u for 2 standard normals

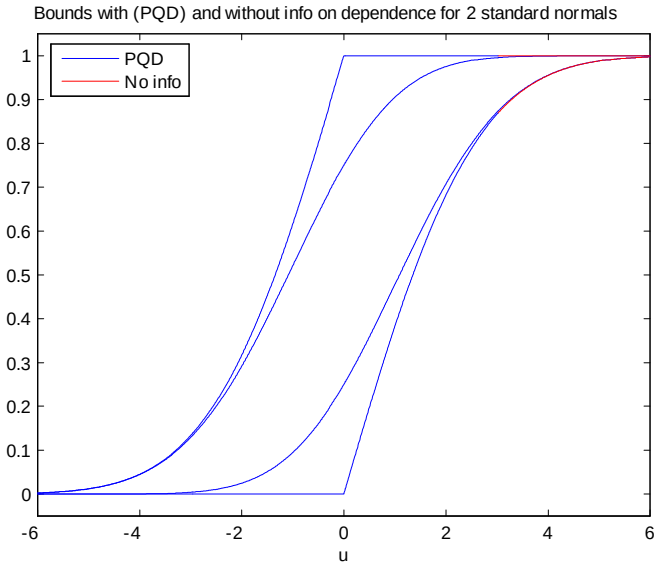


Figure 1.13 shows the bounds for $E(X_1 + X_2 - k)^+$ for 2 standard normals with (PQD) and without information about the dependence structure.

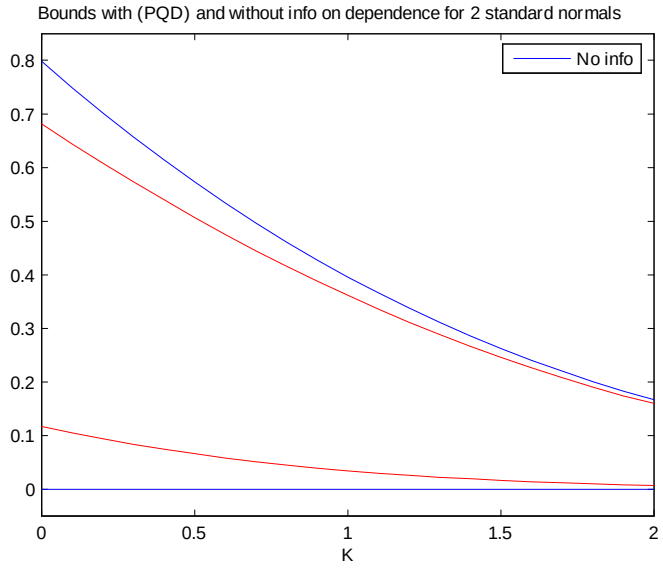
As can be judged from Figure 1.13 the bounds do not shrink by much when we impose PQD between X_1 and X_2. One of the goals of the empirical analysis will be to impose tougher restrictions on the copula that captures the dependence between X_1 and X_2 in order to obtain tight bounds for $E(X_1 + X_2 - k)^+$. This will be achieved by using market prices on bivariate options.

Example 2b. We use the same distributions as in example 1b and impose PQD.

$X_1 \sim \exp(0,1), X_2 \sim N(0,1)$

$Pr(X_1 \leq x, X_2 \leq y) \geq F_1(x).F_2(y)$

Figure 1.13. Bounds on Emax(X1+X2-k,0) as a function of k for 2 standard normals



1) Fmin

$F_{\min,\Pi}(u) = \sup_x F_1(x).\Phi(u-x)$

It is clear that the optimal x must be strictly positive.

$$F_{\min,\Pi}(u) = \sup_{x>0}[1-\exp(-x)].\Phi(u-x)$$

This expression has to be evaluated numerically.

2) Fmax

$F_{\max,\Pi}(u) = \inf_x\{1-[1-F_1(x)].[1-\Phi(u-x)]\} = 1-\sup_x[1-F_1(x)].\Phi(x-u)$

With x $\leq$ 0: $\sup_{x\leq 0}[1-F_1(x)].\Phi(x-u) = \sup_{x\leq 0}\Phi(x-u)$

Therefore, the optimal x cannot be strictly negative.

$$\begin{aligned} F_{\max,\Pi}(u) &= 1 - \sup_{x \geq 0}[1 - (1 - \exp(-x))].\Phi(x-u) \\ &= 1 - \sup_{x \geq 0}[\exp(-x).\Phi(x-u)] \end{aligned}$$

This expression has to be evaluated numerically as well.

Figure 1.14 shows the bounds for $Pr(X_1 + X_2 \leq u)$ as a function of u for the sum of a standard normal and an exp(1) distributed variable. These bounds show the difference between imposing PQD between the two variables and imposing nothing on the dependence structure. It can be seen that the lower bound is more affected by imposing PQD than the upper bound in this particular case.

Figure 1.15 shows the bounds with PQD and without information on the dependence structure for $E(X + Y - k)^+$ for a standard normal and exp(1) variable. Figure 1.15 shows that also in this example imposing PQD does not dramatically tighten the bounds for $E(X + Y - k)^+$.

1.7.1.6. Further tightening of the bounds. From the previous analysis we have learned that in order to further tighten the bounds on $E(X_1 + X_2 - k)^+$ we have to do better than just impose PQD between the variables. We will prove another way to compute $E(X_1 + X_2 - k)^+$, which will allow us to better incorporate extra information on X_1 and X_2. This method will also result in much shorter computation time. First, we

Figure 1.14. Bounds on E(max(X1+X2-k,0)) as a function of k for 2 standard normals

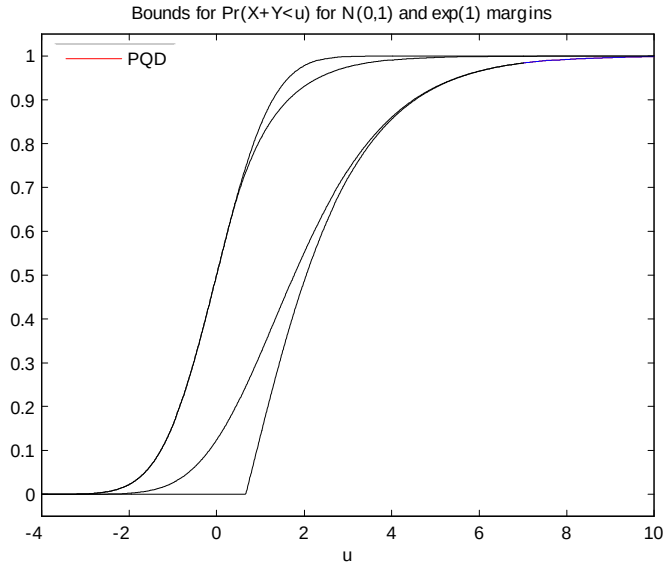
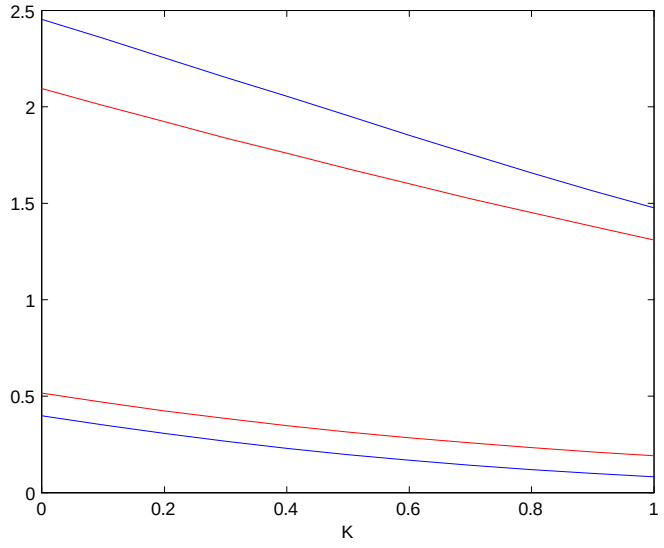


Figure 1.15. Bounds on E(max(X1+X2-k,0)) as a function of k for standard normal and exp(1) variable



show why the current method does not help us much in further reducing the bounds.

$$E(X_1 + X_2 - k)^+ = \int_k^\infty 1 - \Pr(X_1 + X_2 \leq u)du$$

$$\text{Impose } F(x,y) \geq L(x,y) = C_L[F_1(x), F_2(y)]$$

$$\sup_x C_L[F_1(x), F_2(u-x)] \leq \Pr(X_1 + X_2 \leq u) \leq \inf_x F_1(x) + F_2(u-x) - C_L[F_1(x), F_2(u-x)]$$

Imposing an upper bound on F(x,y) does not help us in reducing the bounds. This implies that better information about the true joint distribution function does not easily translate in improved bounds using the current method. In addition, evaluating

$$\begin{aligned} E(X_1 + X_2 - k)^+ &= \int_k^\infty 1 - \Pr(X_1 + X_2 \leq u)du = \int_k^\infty \int_{-\infty}^\infty \int_{u-x}^\infty f(x,y)dy.dx.du \\ &= \int_k^\infty \int_{-\infty}^\infty \int_{u-x}^\infty c[F_1(x), F_2(y)].f_1(x).f_2(y)dy.dx.du \end{aligned}$$

takes a lot of computation time. A better way to compute $E(X_1 + X_2 - k)^+$ is to observe that:

Claim 2.

$$E(X_1 + X_2 - k)^+ = E(X_1) + E(X_2) - k + E(k - X_1 - X_2)^+$$

Proof.

$$T = X_1 + X_2 : E(T) = \int_{-\infty}^{0} t dF_T(t) + \int_{0}^{\infty} t dF_T(t) = -\int_{-\infty}^{0} F_T(t)dt + \int_{0}^{\infty} 1 - F_T(t)dt$$

$$E(k - X_1 - X_2)^+ = E(k - T)^+ = \int_{-\infty}^{k} F_T(t)dt$$

$$E(T) - k + E(k - T)^+ = -\int_{-\infty}^{0} F_T(t)dt + \int_{0}^{\infty} 1 - F_T(t)dt - \int_{0}^{k} dt + \int_{-\infty}^{k} F_T(t)dt$$

$$= \int_{k}^{\infty} 1 - F_T(t)dt = E(T - k)^+ = E(X_1 + X_2 - k)^+$$

□

Claim 3.

$$E(k - X_1 - X_2)^+ = \int_{-\infty}^{\infty} C[F_1(u), F_2(k - u)]du$$

Proof.

$$\begin{aligned} E(k - X_1 - X_2)^+ &= \int_{-\infty}^{\infty} \int_{-\infty}^{k-x} (k - x - y).f(x, y)dy.dx \\ &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{k-x} \int_{u=x}^{k-y} du.f(x, y)dy.dx \\ &= \int_{u=-\infty}^{\infty} \int_{x=-\infty}^{u} \int_{y=-\infty}^{k-u} f(x, y)dy.dx.du = \int_{-\infty}^{\infty} F(u, k - u)du \\ &= \int_{-\infty}^{\infty} C[F_1(u), F_2(k - u)]du \end{aligned}$$

□

The third equality requires some justification. We change the order of integration, and because the integrand stays the same we just have to check that the volume over which

we integrate remains the same. On the left hand side of the third equality x, y, and u cover the entire real line and x and y are such that x+y $\leq$ k. On the right hand side, we have once again that x,y, and u cover the entire real line, and that x+y $\leq$ k. Therefore, we have:

$$E(k - X_1 - X_2)^+ = \int_{-\infty}^{k} \Pr(X_1 + X_2 \leq u)du = \int_{-\infty}^{\infty} F(u, k-u)du$$

Remark: This expression cannot be used in the case of uniform marginals when k>1, because of boundary problems. In the case of uniform marginals we get:

$$E(k-X_1-X_2)^+ = \int_0^{k-1}\int_0^1 (k-x-y).f(x,y).dy.dx + \int_{x=t-1}^{1}\int_{y=0}^{k-x}\int_{u=x}^{k-y} du.f(x,y).dy.dx$$

The assumption of independence between the uniform marginals leads to ($k > 1$):

$$(k - \frac{1}{2}).(k-1) - \frac{2(k-1)^2}{3} + \frac{1}{6}$$

We conclude that we have two different ways to compute $E(X_1 + X_2 - k)^+$:

Method1:

$$\begin{aligned} E(X_1 + X_2 - k)^+ &= \int_k^\infty 1 - \Pr(X_1 + X_2 \leq u) du \\ F_{\min}(u) &\leq \Pr(X_1 + X_2 \leq u) \leq F_{\max}(u) \\ \int_k^\infty 1 - F_{\max}(u) du &\leq E(X_1 + X_2 - k)^+ \leq \int_k^\infty 1 - F_{\min}(u) du \\ F_{\min}(u) &= \sup_x \max\{F_1(x) + F_2(u - x) - 1, 0\} \\ F_{\max}(u) &= \inf_x \min\{F_1(x) + F_2(u - x), 1\} \end{aligned}$$

Method2:

$$E(X_1 + X_2 - k)^+ = E(X_1) + E(X_2) - k + \int_{-\infty}^\infty C[F_1(u), F_2(k - u)] du$$

$$\begin{aligned} & E(X_1) + E(X_2) - k + \int_{-\infty}^\infty \max[F_1(u) + F_2(k - u) - 1, 0] du \\ \leq\ & E(X_1 + X_2 - k)^+ \\ \leq\ & E(X_1) + E(X_2) - k + \int_{-\infty}^\infty \min[F_1(u), F_2(k - u)] du \end{aligned}$$

At first sight one might think that the bounds for both methods are the same, but that is not true. The bounds in method 2 clearly correspond to imposing that the copula equals the lower(W)/upper(M) Frechet-Hoeffding bound. But nothing guarantees in method 1 that $\mathrm{F}_{\min}$ or $\mathrm{F}_{\max}$ correspond to the copulas W or M. Contrary to intuition the $\Pr(X_1 + X_2 \leq u)$ does not necessarily reach its minimum and maximum value for the copulas M and W. Given that the bounds are not the same it is immediate from the construction of

the bounds that the bounds from method 2 must lie inside the bounds from method 1. We show a couple of examples to demonstrate this point.

Comonotonicity. X and Y are *comonotonic* (Dhaene et al., 2002a, 2002b) if and only if

$$F_{X,Y}(x,y) = C_{X,Y}[F_1(x), F_2(y)] = \min[F_1(x), F_2(y)].$$

The support of this distribution are the pairs (x,y) such that $F_1(x) = F_2(y)$. For $U \sim unif(0,1)$ we get:

$$\begin{aligned} & \Pr\{F_1^{-1}(U) \leq x, F_2^{-1}(U) \leq y\} \\ = \; & \Pr\{U \leq F_1(x), U \leq F_2(y)\} \\ = \; & \min[F_1(x), F_2(y)] \end{aligned}$$

Therefore,

$$(X,Y) \sim [F_1^{-1}(U), F_2^{-1}(U)]$$

Thus, we can also say that X and Y are comonotonic if and only if Y is an increasing function of X, therefore there is perfect positive dependence in the case of comonotonicity.

Countermonotonicity. X and Y are *countermonotonic* if and only if

$$F_{X,Y}(x,y) = C_{X,Y}[F_1(x), F_2(y)] = \max[F_1(x) + F_2(y) - 1, 0].$$

The support of this distribution are the pairs (x,y) such that $F_1(x) + F_2(y) = 1$. For $U \sim unif(0,1)$ we get:

$$\begin{aligned} & \Pr\{F_1^{-1}(U) \leq x, F_2^{-1}(1-U) \leq y\} \\ = \; & \Pr\{U \leq F_1(x), 1-U \leq F_2(y)\} \\ = \; & \Pr\{1-F_2(y) \leq U \leq F_1(x)\} \\ = \; & \max[F_1(x) + F_2(y) - 1, 0] \end{aligned}$$

Therefore,

$$(X,Y) \sim [F_1^{-1}(U), F_2^{-1}(1-U)]$$

Thus, we can also say that X and Y are countermonotonic if and only if Y is a decreasing function of X, therefore there is perfect negative dependence in the case of countermonotonicity.

Example 1c. As in the previous examples 1a and 1b we use two standard normals. Example 1a concluded that without information on the dependence between X and Y we get the following bounds on $Pr(X + Y \leq u)$:

$$\begin{aligned} F_{\min}(u) &= 0 && u \leq 0 \\ &= 2\Phi(\frac{u}{2}) - 1 && u > 0 \\ F_{\max}(u) &= 2\Phi(\frac{u}{2}) && u < 0 \\ &= 1 && u \geq 0 \end{aligned}$$

Comonotonicity

$$\begin{aligned} Pr(X+Y \leq u) &= \Pr\{F_1^{-1}(U) + F_2^{-1}(U) \leq u\} = \Pr\{Z(U) \leq u\} \\ &= \sup\{U \in [0,1] : Z(U) \leq u\} \\ &= \sup\{U \in [0,1] : 2.\Phi^{-1}(U) \leq u\} \\ &= \Phi(\frac{u}{2}) \end{aligned}$$

.

Countermonotonicity. Support $=\{(x,y) : \Phi(y) = 1 - \Phi(x)\} = \{(x,y) : x = -y\}$

$$\begin{aligned} \Pr(X+Y \leq u) &= 0 \quad \text{for u} < 0 \\ &= 1 \quad \text{for u} \geq 0 \end{aligned}$$

We observe that $\mathrm{F_{min}}$ and $\mathrm{F_{max}}$ do not have to correspond to the copulas W and M, and that the $Pr(X+Y \leq u)$ in the case of co- and countermonotonicity lies inside $[\mathrm{F_{min}}(u), \mathrm{F_{max}}(u)]$.

Example 2c. Once again: $X \sim N(0,1)$ and $Y \sim exp(1)$

$$\begin{aligned} F_Y(y) &= 1 - \exp(-y) \ \text{ for y} \geq 0 \\ &= 0 \ \text{ for y} < 0 \\ F_Y^{-1}(t) &= -\log(1-t) \ \text{ for t} \geq 0 \end{aligned}$$

Example 2a concluded that for $\Pr(X+Y \leq u)$:

$$\begin{aligned} F_{\min}(u) &= \Phi(u - X^*(u)) - \exp(-X^*(u)) && u \geq u^* \\ &= 0 && u \leq u^* \\ with\ X^*(u) &= 1 + u - \sqrt{1 - \log(2\pi) + 2u} \\ F_{\max}(u) &= \Phi(u) \end{aligned}$$

Comonotonicity. Support $= \{(x,y) : \Phi(x) = F_Y(y)\} = \{(x,y) : y = F_Y^{-1}[\Phi(x)]\} = \{(x,y) : y = -\log[\Phi(-x)]\}$. We conclude that:

$$\Pr(X+Y \leq u) = \Phi(x^*)$$

$$\text{with } x^* \ : \ u - x^* = -\log[\Phi(-x^*)]$$

As $-\log[\Phi(-x)]$ is strictly increasing in x, x^* always exists and is unique. As $x^* < u$, $\Phi(x^*) < F_{\max}$. Numerically it can be verified that $\Phi(x^*) > F_{\min}$. And again we observe that $F_{\min}$ and $F_{\max}$ do not correspond to the case of comonotonicity.

Countermonotonicity. Support $= \{(x,y) : \Phi(x) = 1 - F_Y(y)\} = \{(x,y) : y = F_Y^{-1}[\Phi(-x)]\} = \{(x,y) : y = -\log[\Phi(x)]\}$.We conclude that:

$$\begin{aligned} \Pr(X+Y \leq u) &= 0 && u \leq u^* \\ &= \Phi(b_u) - \Phi(a_u) && u > u^* \end{aligned}$$

For $u > u^*$: $u - x = -\log[\Phi(x)]$ has two solutions $\{a_u < b_u\}$. As $b_u < u$, $\Phi(b_u) - \Phi(a_u) < F_{\max}$. It can also be verified that $\Phi(b_u) - \Phi(a_u) > F_{\min}$.

Given that $\mathrm{F}_{\min}(u)$ and $\mathrm{F}_{\max}(u)$ do not necessarily correspond to the copulas M (comonotonicity) and W (countermonotonicity), it is now clear why the bounds from Method 1 and Method 2 do not coincide, and as a result why the bounds from Method 1 are wider. It can be shown (Frank et al.) that for each u, there exists a copula C_u such that $\mathrm{F}_{\min}(u) = \Pr(X+Y \leq u)$, where the probability is computed with respect to C_u, the copula of X and Y, and similarly for $\mathrm{F}_{\max}(u)$. Hence, the reason that the bounds for both method 1 and 2 do not coincide is due to the fact that in method 1 we implicitly impose a different copula of X and Y for every u. In method 2, however, we impose the same copula for every u. Therefore, we will from now on use method 2 to price basket options as method 2 has the advantage of less computation time, it is easier to incorporate extra information about the true copula, and it is consistent as the bounds correspond to the same copula no matter the value of u.

Imposing additional information about the structure and degree of dependence. We want to compute (a,b>0):

$$V = E\max(aX + bY - k, 0) = aE(X) + bE(Y) - k + \int_{-\infty}^{\infty} C_{X,Y}[F_X(\frac{s}{a}), F_Y(\frac{k-s}{b})]ds.$$

Obviously, when we know exactly the joint distribution of X and Y, or similarly, when we know the marginal distributions of both X and Y and the copula of X and Y, we have a point estimate for V. In practice, however, we never exactly know this information. Hence, the question: *how does partial information about the dependence structure between X and Y translate into bounds for V?* This question, however, is not very precise as there exist a lot of different ways to impose partial information about the dependence structure.

One of the methods we have been looking at before is to impose bounds on the copula directly:

$$C_L(u,v) \leq C(u,v) \leq C_H(u,v).$$

Looking at the expression for V this type of partial information about the dependence structure between X and Y leads to very straightforward bounds for V. When we set $C_L(u,v) = u.v$ and $C_H(u,v) = M(u,v)$ we impose PQD between X and Y, which implies positive covariance between X and Y. However, we know from the previous analysis that imposing PQD does not lead to tight bounds for V. Therefore, we need to impose additional information about the dependence structure. Later on, we will use market prices on bivariate options to help us determine C_L and C_H.

In the next few sections we derive expressions for the rainbow and spread options. After that we construct bounds that these bivariate options have to satisfy in order to avoid arbitrage.

1.8. Rainbow-max option

For a,b>0 and aX=R, bY=S we have:

$$\begin{aligned} E[\max(aX, bY) - k]^+ &= E\max[(R, S) - k]^+ \\ &= \int_k^\infty 1 - \Pr[\max(R, S) < s]ds \\ &= \int_k^\infty 1 - H_{R,S}(s, s)ds \\ &= \int_k^\infty 1 - C[F_R(s), F_S(s)]ds \\ &= \int_k^\infty 1 - C[F_X(\frac{s}{a}), F_Y(\frac{s}{b})]ds \end{aligned}$$

1.9. Rainbow-min option

For a,b>0 and aX=R, bY=S we have:

$$\begin{aligned} E[\min(aX, bY) - k]^+ &= E\min[(R, S) - k]^+ \\ &= \int_k^\infty \Pr[\min(R, S) > s]ds \\ &= \int_k^\infty \bar{H}_{R,S}(s, s)ds \\ &= \int_k^\infty \bar{C}[\bar{F}_R(s), \bar{F}_S(s)]ds \\ &= \int_k^\infty \bar{C}[\bar{F}_X(\frac{s}{a}), \bar{F}_Y(\frac{s}{b})]ds \end{aligned}$$

$\bar{H}$: survival distribution

$\bar{C}$: survival copula, which is a true copula: $\bar{C}$(u,v)=u+v-1+C(1-u,1-v)

$\bar{F}_X = 1 - F_X$

1.10. Spread option

Without loss of generality we have for b>0 and bY=S:

$$\begin{aligned} E[X-bY-k]^+ &= E[X-S-k]^+ \\ &= E(X)-E(S)-k+E[k-(X-S)]^+ \end{aligned}$$

$$\begin{aligned} E[k-(X-S)]^+ &= \int_{-\infty}^{k} \Pr(X-S<u)du \\ &= \int_{-\infty}^{k}\int_{0}^{\infty} \Pr(X<u+s,S=s)ds.du \\ &= \int_{u=-\infty}^{k}\int_{s=\max(-u,0)}^{\infty} \delta_2 H(u+s,s)ds.du \\ &= \int_{u=-\infty}^{k}\int_{s=\max(-u,0)}^{\infty} \delta_2 C[F_X(u+s),F_S(s)].f_S(s).ds.du \\ &= \int_{s=0}^{\infty}\int_{u=-s}^{k} \delta_2 C[F_X(u+s),F_S(s)].f_S(s).du.ds \\ E[X-bY-k]^+ &= E(X)-bE(Y)-k \\ &\quad + \int_{s=0}^{\infty}\int_{u=-s}^{k} \delta_2 C[F_X(u+s),F_Y(\frac{s}{b})].\frac{1}{b}.f_Y(\frac{s}{b}).du.ds \end{aligned}$$

Another and more convenient expression for a spread option can be found by noticing that

$$(X-bY-k)^+ = X-[\min(X,bY+k)]^+ = [\max(X,bY+k)]^+-(bY+k)$$

Therefore, we also have:

$$\begin{aligned}
E[X-bY-k]^+ &= E(X) - \int_0^\infty \Pr(X>s, bY+k>s)ds \\
&= E(X) - \int_0^\infty \bar{C}[\bar{F}_X(s), \bar{F}_Y(\frac{s-k}{b})]ds \\
&= E(X) - \int_0^k 1 - \Pr(X<s)ds - \int_k^\infty \bar{C}[\bar{F}_X(s), \bar{F}_Y(\frac{s-k}{b})]ds \\
&= E(X) - k + E(k-X)^+ - \int_k^\infty \bar{C}[\bar{F}_X(s), \bar{F}_Y(\frac{s-k}{b})]ds \\
&= E(X-k)^+ - \int_k^\infty \bar{C}[\bar{F}_X(s), \bar{F}_Y(\frac{s-k}{b})]ds
\end{aligned}$$

Or, using the max operator we can write it as:

$$\begin{aligned}
E[X-bY-k]^+ &= \int_0^\infty 1 - C[F_X(s), F_Y(\frac{s-k}{b})]ds - bE(Y) - k \\
&= \int_k^\infty 1 - C[F_X(s), F_Y(\frac{s-k}{b})]ds - bE(Y)
\end{aligned}$$

1.11. Extreme pricing bounds

Definition 4. *Extreme pricing bounds are the bounds a particular bivariate option has to satisfy only by using the information provided by the univariate option prices, or equivalently, only by knowing the univariate risk-neutral distributions.*

Looking closely at the final expressions for basket, rainbow-max, rainbow-min, and spread options we observe that all these final expressions integrate over the copula of X and Y. This implies that the extreme bounds for these options are reached for the minimum (W) and maximum (M) copulas. The upper bound for the basket and rainbow-min options is attained with the maximum (M) copula - the copula enters the final expression with a

positive sign, whereas the upper bound for the spread and rainbow-max option is attained with the minimum (W) copula - the copula has a negative sign. This result holds true more generally for super- and submodular payoff functions.

1.11.1. Concordance order

$C_1 \prec C_2 \Leftrightarrow C_1(u,v) \leq C_2(u,v)$ for all (u,v)$\in[0,1]^2$.A family of copulas $\{C_\gamma\}$ is positively (negatively) ordered if $C_{\gamma_1} \prec C_{\gamma_2}$ when $\gamma_1 \leq (\geq)\ \gamma_2$. One can show for example that the normal and student-t copula are positively ordered. For the large class of Archimedean copulas it is shown (Genest and MacKay, 1986b) that if $\frac{\phi_1(t)}{\phi_2(t)}$ or$\frac{\phi_1'(t)}{\phi_2'(t)}$ is non-decreasing on (0,1) then $C_1 \prec C_2$.

1.11.2. Super- and submodular payoff functions

Definition 5. *For a function f:* $R^2 \rightarrow R$ *we define* $\Delta_i^\delta f(x) = f(x+\delta.e_i) - f(x)$. *The function f is said to be supermodular (or 2-increasing) if and only if* $\Delta_1^\varepsilon \Delta_2^\delta f(x) \geq 0$ *for all* $\delta, \varepsilon > 0$. *The function f is submodular if and only if* $\Delta_1^\varepsilon \Delta_2^\delta f(x) \leq 0$ *for all* $\delta, \varepsilon > 0$.

Claim 6. *For a supermodular (or 2-increasing) payoff function G we have that* $C_1 \prec C_2 \Longrightarrow E_{C_1}[G(X,Y)] \leq E_{C_2}[G(X,Y)]$

Proof. Follows from Theorem 2 in Tchen (1980) and Muller and Scarsini (2000). □

Given that H=-G is submodular for G supermodular, it follows that for a submodular function H we have that $C_1 \prec C_2 \Longrightarrow E_{C_1}[H(X,Y)] \geq E_{C_2}[H(X,Y)]$. It is straightforward to check that the payoff function for the basket and rainbow-min options is supermodular, and that it is submodular for the spread and rainbow-max option. When we denote the

price of the option by C, the claim implies that the option prices have to satisfy the following extreme bounds:

$$e^{-rT}E_W^Q[G(X_T,Y_T)] \leq C \leq e^{-rT}E_M^Q[G(X_T,Y_T)] \text{ for basket and rainbow-min}$$

$$e^{-rT}E_M^Q[G(X_T,Y_T)] \leq C \leq e^{-rT}E_W^Q[G(X_T,Y_T)] \text{ for spread and rainbow-max}$$

In the specific case that X and Y are jointly normal distributed, and given the fact that the bivariate normal distribution is simply the normal copula combined with normal marginals, we recover the known result that the basket option reaches it maximum value at correlation equal to 1, and that the spread option reaches it maximum at correlation equal to -1. We will now look in more detail to the extreme pricing bounds for the different options, and show that given some conditions on the univariate risk-neutral distributions we can write the extreme pricing bounds as a portfolio of univariate options, futures, and cash. In the next section, we then investigate how the inclusion of market prices for bivariate options can shrink the bounds. In the previous sections we derived expressions for the basket, rainbow, and spread options in general terms, i.e. we ignored to specifiy the measure over which we take expectations or compute probabilities. From now on we only work with the risk-neutral distributions.

1.11.2.1. Extreme pricing bounds for the basket option $[C^{Min}(k) \leq C(k) \leq C^{Max}(k)]$**.**

$$
\begin{aligned}
e^{r(T-t)}.C(k) &= E_t^Q(aX_T + bY_T - k)^+ = aE_t^Q(X_T) + bE_t^Q(Y_T) - k \\
&\quad + \int_0^k C^Q[F_X^Q(\frac{s}{a}), F_Y^Q(\frac{k-s}{b})]ds \\
C^{Max}(k) &= e^{-r(T-t)}.\{aF_{t,T}(X) + bF_{t,T}(Y) - k \\
&\quad + \int_0^k \min[F_X^Q(\frac{s}{a}), F_Y^Q(\frac{k-s}{b})]ds\} \\
&= e^{-r(T-t)}.\{aF_{t,T}(X) + bF_{t,T}(Y) - k \\
&\quad + a\int_0^{k^*/a} F_X^Q(u)du + b\int_0^{(k-k^*)/b} F_Y^Q(u)du\} \\
&= e^{-r(T-t)}.\{aF_{t,T}(X) + bF_{t,T}(Y) - k\} \\
&\quad + a.Put_{t,T}(X; \frac{k^*}{a}) + b.Put_{t,T}(Y; \frac{k-k^*}{b})
\end{aligned}
$$

$F_{t,T}(X)$: Futures price of X_T at time t

$Put_{t,T}(X, z)$: Put option at time t on X_T with strike price z

k^*: $F_X^Q(\frac{k^*}{a}) = F_Y^Q(\frac{k-k^*}{b})$

In contrast to C^{Max} there exist a lot of possible scenarios for C^{Min}, which can only be resolved after an empirical investigation of the risk-neutral univariate distribution functions of X and Y. For example, one scenario is that there exists a $k^{'} \in (0, k)$ such that

$F_X^Q(\frac{s}{a}) + F_Y^Q(\frac{k-s}{b}) \geq (\leq)\ 1$ for $s \leq (\geq)\ k'$. Under that scenario we can write out C^{Min} as:

$$\begin{aligned} C^{Min}(k) &= e^{-r(T-t)}.\{aF_{t,T}(X) + bF_{t,T}(Y) - k + \int_0^k \max[F_X^Q(\frac{s}{a}) + F_Y^Q(\frac{k-s}{b}) - 1, 0]ds\} \\ &= e^{-r(T-t)}.\{aF_{t,T}(X) + bF_{t,T}(Y) - k + \int_0^{k'} [F_X^Q(\frac{s}{a}) + F_Y^Q(\frac{k-s}{b}) - 1]ds\} \\ &= e^{-r(T-t)}.\{aF_{t,T}(X) + bF_{t,T}(Y) - (k+k')\} + a.Put_{t,T}(X; \frac{k'}{a}) \\ &\quad + b.Put_{t,T}(Y; \frac{k}{b}) - b.Put_{t,T}(Y; \frac{k-k'}{b}) \end{aligned}$$

1.11.2.2. Extreme pricing bounds for the rainbow-max option.

$$\begin{aligned} e^{r(T-t)}.C(k) &= E_t^Q[\max(aX_T, bY_T) - k]^+ = \int_k^\infty 1 - C^Q[F_X^Q(\frac{s}{a}), F_Y^Q(\frac{s}{b})]ds \\ e^{r(T-t)}.C^{Max}(k) &= \int_k^\infty 1 - \max[F_X^Q(\frac{s}{a}) + F_Y^Q(\frac{s}{b}) - 1, 0]ds \end{aligned}$$

k^*: $F_X^Q(\frac{k^*}{a}) + F_Y^Q(\frac{k^*}{b}) = 1$.

Scenario 1: $k^* > k$.

$$\begin{aligned} C^{Max}(k) &= e^{-r(T-t)}.\{\int_k^\infty 1 - \max[F_X^Q(\frac{s}{a}) + F_Y^Q(\frac{s}{b}) - 1, 0]ds\} \\ &= e^{-r(T-t)}.\{(k^* - k)\} + a.Call_{t,T}(X; \frac{k^*}{a}) + b.Call_{t,T}(Y; \frac{k^*}{b}) \end{aligned}$$

$Call_{t,T}(X; z)$: Call option at time t on X_T with strike price z

Scenario 2: $k^* < k$

$$C^{Max}(k) = a.Call_{t,T}(X; \frac{k}{a}) + b.Call_{t,T}(Y; \frac{k}{b})$$

. For the minimum extreme bound we obtain:

$$e^{r(T-t)}.C^{Min}(k) = \int_k^{\infty} 1 - \min[F_X^Q(\frac{s}{a}), F_Y^Q(\frac{s}{b})]ds$$

Given a single-crossing condition between the risk-neutral distributions on X and Y, which has to be verified empirically, there exists k' such that $F_X^Q(\frac{s}{a}) < (>)\ F_Y^Q(\frac{s}{b})$ for $s < (>)\ k^{'}$. Then we have for $k^{'} > k$:

$$\begin{aligned} C^{Min}(k) &= e^{-r(T-t)}.\{\int_k^{\infty} 1 - \min[F_X^Q(\frac{s}{a}), F_Y^Q(\frac{s}{b})]ds\} \\ &= e^{-r(T-t)}.\{\int_k^{k^{'}} 1 - F_X^Q(\frac{s}{a})ds + \int_{k'}^{\infty} 1 - F_Y^Q(\frac{s}{b})ds\} \\ &= e^{-r(T-t)}.\{(k^{'} - k)\} - a.Put_{t,T}(X; \frac{k^{'}}{a}) + a.Put_{t,T}(X; \frac{k}{a}) + b.Call_{t,T}(Y; \frac{k^{'}}{b}) \end{aligned}$$

1.11.2.3. Extreme pricing bounds for the rainbow-min option.

$$\begin{aligned} e^{r(T-t)}.C(k) &= E_t^Q[\min(aX_T, bY_T) - k]^+ = \int_k^{\infty} \bar{C}[\bar{F}_X(\frac{s}{a}), \bar{F}_Y(\frac{s}{b})]ds \\ e^{r(T-t)}.C^{Max}(k) &= \int_k^{\infty} \min[\bar{F}_X(\frac{s}{a}), \bar{F}_Y(\frac{s}{b})]ds \end{aligned}$$

Given a single crossing condition, there exists k^* such that $F_Y(\frac{s}{b}) < (>)\ F_X(\frac{s}{a})$ for $s < (>)\ k^*$. For $k^* > k$, we have:

$$C^{Max}(k) = e^{-r(T-t)}.\{(k^* - k)\} - b.Put_{t,T}(Y; \frac{k^*}{b}) + b.Put_{t,T}(Y; \frac{k}{b}) + a.Call_{t,T}(X; \frac{k^*}{a})$$

For the minimum extreme bound we get:

$$e^{r(T-t)}.C^{Min}(k) = \int_k^{\infty} \max[\bar{F}_X(\frac{s}{a}) + \bar{F}_Y(\frac{s}{b}) - 1, 0]ds$$

There must exist $k^{'}$ such that $\bar{F}_X(\frac{s}{a}) + \bar{F}_Y(\frac{s}{b}) > 1$ for $s < k^{'}$. In the case that $k^{'} < k$, we have that $C^{Min}(k) = 0$. When $k^{'} > k$, we get

$$\begin{aligned} C^{Min}(k) &= a.Call_{t,T}(X; \frac{k}{a}) - a.Call_{t,T}(X; \frac{k^{'}}{a}) + b.Call_{t,T}(Y; \frac{k}{b}) \\ &\quad -b.Call_{t,T}(Y; \frac{k^{'}}{b}) - e^{-r(T-t)}.(k^{'} - k) \end{aligned}$$

1.11.2.4. Extreme pricing bounds for the spread option.

$$\begin{aligned} e^{r(T-t)}.C(k) &= E_t^Q[X_T - bY_T - k]^+ = \int_k^\infty 1 - C[F_X(s), F_Y(\frac{s-k}{b})]ds - b.Fut_{t,T}(Y) \\ e^{r(T-t)}.C^{Max}(k) &= \int_k^\infty 1 - \max[F_X(s) + F_Y(\frac{s-k}{b}) - 1, 0]ds - b.Fut_{t,T}(Y) \end{aligned}$$

There must exist $k^* > k$ such that $F_X(s) + F_Y(\frac{s-k}{b}) > 1$ for $s > k^*$.

$$C^{Max}(k) = e^{-r(T-t)}.\{(k^* - k) - b.Fut_{t,T}(Y)\} + Call_{t,T}(X; k^*) + b.Call_{t,T}(Y; \frac{k^* - k}{b})$$

Again, given a single crossing condition, there exists $k^{'} > k$ such that $F_X(s) > (<)$ $F_Y(\frac{s-k}{b})$ for $s < (>)$ $k^{'}$. This implies the following minimum extreme bound:

$$\begin{aligned} C^{Min}(k) &= e^{-r(T-t)}.\{\int_k^\infty 1 - \min[F_X(s), F_Y(\frac{s-k}{b})]ds - b.Fut_{t,T}(Y)\} \\ &= Call_{t,T}(X; k^{'}) + b.Call_{t,T}(Y; 0) - b.Call_{t,T}(Y; \frac{k^{'} - k}{b}) \\ &= Call_{t,T}(X; k^{'}) + b.e^{-r(T-t)}.F_{t,T}(Y) - b.Call_{t,T}(Y; \frac{k^{'} - k}{b}) \end{aligned}$$

The analysis of extreme pricing bounds demonstrates how we can use existing market prices on *univariate* options and futures to bound the price of a bivariate option. If a

bivariate option does not satisfy the extreme pricing bounds, arbitrage opportunities exist. For example, if $C(k) > C^{Max}(k)$ for some bivariate option, we go short in the bivariate call option and hedge this position in going long C^{Max}, which can easily be done as it just requires going long and/or short in univariate options and futures. It is important to stress that these extreme pricing bounds do not impose any structure on the dependence between X and Y. In what follows we will show how the inclusion of existing market prices on bivariate options imposes restrictions on the risk-neutral dependence between X and Y. This results in tighter bounds for bivariate options in order to avoid arbitrage. But first we show some results on put-call parity for bivariate options.

1.11.3. Bivariate Put-Call parity

For the basket option, we denote the call by $C^{Basket}(k) = e^{-r(T-t)}.E_t^Q[aX + bY - k]^+$ and the put by $P^{Basket}(k) = e^{-r(T-t)}.E_t^Q[k - aX - bY]^+$. Similar notation applies for the other options. For the basket and spread options the results are immediate.

$$\begin{aligned} C^{Basket}(k) + e^{-r(T-t)}.k &= P^{Basket}(k) + e^{-r(T-t)}.\{a.Fut_{t,T}(X) + b.Fut_{t,T}(Y)\} \\ C^{Spread}(k) + e^{-r(T-t)}.k &= P^{Spread}(k) + e^{-r(T-t)}.\{Fut_{t,T}(X) - b.Fut_{t,T}(Y)\} \end{aligned}$$

For the rainbow-min and rainbow-max options we first get some cross-product restrictions:

$$\begin{aligned}
P^{R-\max}(k) &= a.Put_{t,T}(X;\frac{k}{a}) + b.Put_{t,T}(Y;\frac{k}{b}) - e^{-r(T-t)}.k \\
&\quad + C^{R-\min}(0) - C^{R-\min}(k) \\
P^{R-\min}(k) &= a.Put_{t,T}(X;\frac{k}{a}) + b.Put_{t,T}(Y;\frac{k}{b}) - e^{-r(T-t)}.k \\
&\quad + C^{R-\max}(0) - C^{R-\max}(k) \\
P^{R-\max}(k) - P^{R-\min}(k) &= [C^{R-\max}(k) - C^{R-\max}(0)] - [C^{R-\min}(k) - C^{R-\min}(0)]
\end{aligned}$$

Using the univariate put-call parity:

$$Call_{t,T}(Z;v) + e^{-r(T-t)}.v = Put_{t,T}(Z;v) + e^{-r(T-t)}.Fut_{t,T}(Z)$$

and noting that

$$C^{R-\max}(k) + C^{R-\min}(k) = a.Call_{t,T}(X;\frac{k}{a}) + b.Call_{t,T}(Y;\frac{k}{b})$$

some straightforward algebra leads to

$$\begin{aligned}
C^{R-\max}(k) + e^{-r(T-t)}.k &= P^{R-\max}(k) + C^{R-\max}(0) \\
C^{R-\min}(k) + e^{-r(T-t)}.k &= P^{R-\min}(k) + C^{R-\min}(0)
\end{aligned}$$

1.12. Implied Risk-Neutral Distribution

In order to price bivariate options we need to obtain the bivariate risk-neutral distribution H^Q (density, h^Q). Given Sklar's theorem, we can achieve this by splitting up the bivariate risk-neutral distribution in the risk-neutral copula, C^Q, and the univariate risk-neutral distributions, F_X^Q and F_Y^Q.

$$\begin{aligned} H^Q(x,y) &= C^Q[F_X^Q(x).F_Y^Q(y)] \\ h^Q(x,y) &= c^Q[F_X^Q(x).F_Y^Q(y)].f_X^Q(x).f_Y^Q(y) \end{aligned}$$

1.12.1. Univariate RND

There already exists a vast literature on how to obtain the implied univariate RND from option prices (Breeden and Litzenberger, 1978; Melick and Thomas, 1987; Derman and Kani, 1994; Rubenstein, 1994; Jackwerth and Rubenstein, 1996; Bahra, 1997). Fundamentally, three different approaches are being used. The first method is non-parametric and takes the derivative of the option price with respect to the strike and follows from the observation that:

$$\begin{aligned} Call_{t,T}(X;k) &= e^{-r(T-t)}E_t^Q[X_T-k]^+ = e^{-r(T-t)}\int_k^\infty P^Q(X_T>s)ds \\ F_X^Q(k) &= 1+e^{r(T-t)}.\frac{\delta Call_{t,T}(X;k)}{\delta k} \\ f_X^Q(k) &= e^{r(T-t)}.\frac{\delta^2 Call_{t,T}(X;k)}{\delta k^2}. \end{aligned}$$

In practice we have market option prices for different strikes. So, the first step one has to take is to obtain a smooth function for the call as a function of the strike, for example

by fitting a spline. The problem with this approach in practice is that the smoothed function has to be convex with respect to the price over the entire interval in order to avoid a negative density, which is often hard to realise. Often, practioners first get the implied volatilities from the market option prices, and will then obtain a smooth function for the implied volatilities as a function of the strike. But this is fundamentally the same approach. The second method is parametric and models the terminal risk-neutral density directly. The optimal parameters minimize the distance between the model implied option prices and the market prices. The parametric RND has to be flexible enough so that it allows for both negative and positive skewness, and for leptokurtosis. In addition it is important that the mapping from the RND to the option prices is relatively easy. As a final restriction, the first moment of the RND has to equal the futures price, because $Fut_{t,T}(X) = E_t^Q(X_T)$. Using a mixture of lognormals, as in Alexander & Narayanan (2001), Brigo & Mercurio (2001), and Rebonato & Cardoso (2004), easily allows for skewness and leptokurtosis, and the model implied option price can simply be written as a linear combination of the lognormals. The third method is also parametric and proposes a stochastic process that the underlying variable is assumed to follow. This stochastic process, in turn, implies a terminal distribution. So, the parameters of the stochastic process are then determined in the same way as in the previous method. The advantage of modelling the stochastic process is that we know how the underlying variable evolves over time, and thus we can use it for dynamic hedging. The big problem with this approach, however, is that it is really difficult to come up with a stochastic process that generates a terminal distribution with the preferred properties. In addition, different stochastic processes can map into the same terminal distribution, so in that sense modelling the

terminal distribution is more general. Finally, market option prices of the European type only convey information about the terminal distribution functions, they do not tell us anything about the underlying stochastic process.

1.12.2. Bivariate RN-copula

Some previous work that introduced copulas to price multivariate options includes Rosenberg (1999), Rosenberg & Engle (2002), Cherubini & Luciano (2002), van den Goorbergh, Genest & Werker (2005), and Salmon & Schleicher (2006). As mentioned before, the first requirement the bivariate RND has to satisfy is that the margins of the bivariate RND equal the univariate RNDs. We show with a simple example how one can take advantage of a violation of this requirement. Take the following relation:

$$\begin{aligned} & e^{-r(T-t)}.\{E_t^Q[\max(aX_T, bY_T) - k]^+ + E_t^Q[\min(aX_T, bY_T) - k]^+\} \\ = \;& a.Call_{t,T}(X; \frac{k}{a}) + b.Call_{t,T}(Y; \frac{k}{b}) \end{aligned}$$

By definition, being long in the two bivariate options and short in the two univariate options (called portfolio Π) has to result in a zero payoff at final time T[3]. Denote the univariate RNDs as G_X^Q and G_Y^Q, and denote the margins of the bivariate RND as F_X^Q and F_Y^Q. Writing out the previous relation and cancelling some terms we get:

$$\int_k^\infty \bar{F}_X^Q(\frac{s}{a}) + \bar{F}_Y^Q(\frac{s}{a})ds = \int_k^\infty \bar{G}_X^Q(\frac{s}{a}) + \bar{G}_Y^Q(\frac{s}{a})ds$$

This relation will only hold true for all k if and only if $F_X^Q = G_X^Q$ and $F_Y^Q = G_Y^Q$. Therefore, whenever this is not true, it must be that portfolio Π has a non-zero price at

[3]It is clear that $[\max(aX, bY) - k]^+ + [\min(aX, bY) - k]^+ = [aX - k]^+ + [bY - k]^+$.

time t, which leads to a pure arbitrage opportunity. The copula approach is perfectly suited to model the bivariate RND, because of Sklar's theorem. By linking the univariate RNDs through the copula into the bivariate RND, we are always guaranteed that the margins of the bivariate RND equal the univariate RNDs. In the next sections we investigate how we can learn more about the bivariate RND or the risk-neutral copula from products that are being traded in the market place. Just as in the univariate case, there is both a parametric and a non-parametric approach.

1.12.2.1. Non-parametric approach. The question we try to answer is: *what bounds does the price of a bivariate option have to satisfy in order to be consistent with the products that are being traded in the market place*? The answer to that question depends on the types of products that are being traded in the market place. In the section on extreme pricing bounds we implicitly assume that only univariate options and futures on X and Y are being traded. This means that the market cannot tell us anything about the structure of dependence between X and Y. We then showed for particular types of bivariate options how these bounds can be written as a function of the univariate options and futures. Because the bivariate options we looked at have a super- or submodular payoff function, we showed that these extreme pricing bounds also correspond to assuming that the risk-neutral copula of X and Y equals the minimum (M) or maximum (W) copula. This means that we did not impose any restrictions at all on the bivariate RND. In this section, however, we assume that both univariate and bivariate options on X and Y are being traded in the market place. The fact that we now have market prices of bivariate options imposes some restrictions on the bivariate RND. This implies that no longer all copulas are allowed, and therefore the pricing bounds a bivariate option has to satisfy

to avoid arbitrage are tighter relative to the extreme pricing bounds. More specifically, suppose the following bivariate option is being traded:

$$\begin{aligned} C(k) &= e^{-r(T-t)}.E_t^Q[\psi(X_T, Y_T) - k]^+ \\ &= e^{-r(T-t)}.\int_k^\infty P^Q[\psi(X_T, Y_T) > s]ds \end{aligned}$$

We have market prices for this option for different strikes. As in the univariate case, the first step we take is to obtain a smooth function, $C(k)$. This, in turn, implies that we know the following object for $k \in [k_L, k_U]$[4].

$$P^Q[\psi(X_T, Y_T) < k] = 1 + e^{r(T-t)}.\frac{\delta C(k)}{\delta k}$$

In order to price other bivariate options we want to know the bivariate RND, $H^Q(x, y) = P^Q[X_T < x, Y_T < y]$ for $x, y \in R_+$. Therefore, we ask the following question: what does the knowledge of the risk-neutral distribution of $\psi(X_T, Y_T)$ tell us about $H^Q(x, y)$? In order to answer this question we have to look at specific examples for ψ. It is clear that the RND of $\psi(X_T, Y_T)$ can only convey partial information about $H^Q(x, y)$.

1.12.2.2. Rainbow-max option. When a rainbow-max option is trading in the market place, we know the following information for $k \in [k_L, k_U]$:

$$\begin{aligned} P^Q[\psi(X_T, Y_T) < k] &= P^Q[\max(aX_T, bY_T) < k] = P^Q[X_T < \frac{k}{a}, Y_T < \frac{k}{b}] \\ V_{a,b}^{R-\max}(k) &= P^Q[X_T < \frac{k}{a}, Y_T < \frac{k}{b}] \end{aligned}$$

[4] k_L and k_U depend on the range of strikes that are being traded in the market place.

Given that we know $V_{a,b}^{R-\max}(k)$, how does it restrict the value $H^Q(x,y)$ can take on for $x \in [\frac{k_L}{a}, \frac{k_U}{a}]$ and $y \in [\frac{k_L}{b}, \frac{k_U}{b}]$? The knowledge of $V_{a,b}^{R-\max}(k)$ means that we know the values of $H^Q(x,y)$ along the ray $y = \frac{a}{b}x$. It is straightforward to show that $H^Q(x,y)$ is restricted in the following manner:

$$V_{a,b}^{R-\max}(ax \wedge by) = LB_{a,b}^{R-\max}(x,y) \leq H^Q(x,y) < UB_{a,b}^{R-\max}(x,y) = V_{a,b}^{R-\max}(ax \vee by)$$

$x \wedge y : \min(x,y)$

$x \vee y : \max(x,y)$

1.12.2.3. Rainbow-min option.

$$\begin{aligned} P^Q[\min(aX_T, bY_T) < k] &= P^Q[X_T < \frac{k}{a} \text{ or } Y_T < \frac{k}{b}] \\ &= F_X^Q(\frac{k}{a}) + F_Y^Q(\frac{k}{b}) - P^Q[X_T < \frac{k}{a}, Y_T < \frac{k}{b}] \end{aligned}$$

We know the univariate RNDs for X and Y from market prices on univariate options. Therefore, we can see that the rainbow-min option conveys exactly the same information about $H^Q(x,y)$ as the rainbow-max option.

1.12.2.4. Basket option.

$$P^Q[\psi(X_T, Y_T) < k] = P^Q[aX_T + bY_T < k] = V_{a,b}^{Basket}(k)$$

Knowledge of $V_{a,b}^{Basket}(k)$ restricts $H^Q(x,y)$ for $\{(x,y) \subset R_+^2 : k_L \leq ax + by \leq k_U\}$ in the following manner:

$$V_{a,b}^{Basket}(ax \wedge by) = LB_{a,b}^{Basket}(x,y) \leq H^Q(x,y) \leq UB_{a,b}^{Basket}(x,y) = V_{a,b}^{Basket}(ax + by)$$

1.12.2.5. Spread option.

$$P^Q[\psi(X_T, Y_T) < k] = P^Q[X_T - bY_T < k] = V_b^{Spread}(k)$$

Knowledge of $V_b^{Spread}(k)$ restricts $H^Q(x, y)$ for $\{(x, y) \subset R_+^2 : k_L \leq x - by \leq k_U\}$ in the following manner:

$$H^Q(x, y) \leq UB_b^{Spread}(x, y) = V_b^{Spread}(x) - V_b^{Spread}(-by)$$

Restricted Upper Bound (RUB) and Restricted Lower Bound (RLB). In practice we want to restrict $H^Q(x, y)$ using all the available market prices, i.e. construct bounds that the bivariate RND has to satisfy such that the new product that is being priced is consistent with all the products that are currently trading in the market place. In order to do so we compute

$$\begin{aligned} RLB(x, y) &= \max_{a,b,i}\{LB_{a,b}^i(x, y), [F_X^Q(x) + F_Y^Q(y) - 1]^+\} \\ RUB(x, y) &= \min_{a,b,i}\{UB_{a,b}^i(x, y), F_X^Q(x), F_Y^Q(y)\} \\ RLB(x, y) &\leq H^Q(x, y) \leq RUB(x, y) \end{aligned}$$

$i \in \{R - \max, R - \min, Basket, Spread\}$

Suppose we have market prices on futures, univariate options on X and Y, rainbow options, and spread options. We want to find bounds for the price of a basket option such that it is consistent with the futures, rainbow, spread, and univariate options. These

bounds have the following form:

$$\begin{aligned} & e^{-r(T-t)}.\{a.Fut_{t,T}(X) + b.Fut_{t,T}(Y) - k + \int_0^k RLB(\frac{u}{a}, \frac{k-u}{b})du\} \\ \leq \; & C_{t,T}^{Basket}(k) \\ \leq \; & e^{-r(T-t)}.\{a.Fut_{t,T}(X) + b.Fut_{t,T}(Y) - k + \int_0^k RUB(\frac{u}{a}, \frac{k-u}{b})du\} \end{aligned}$$

1.12.2.6. Parametric approach. The non-parametric approach does not impose any structure at all on H^Q. As a result, the bounds on H^Q, $RLB - RUB$, that stem from the market prices of bivariate options can be wide, especially if only few market prices are available. We propose to add to the information set the historical prices of the underlying variables. We use the historical dataset to impose *structure* on the bivariate RND, but to have the *degree* of dependence determined solely by the forward-looking bivariate option prices[5]. Using Sklar's theorem it is straightforward to impose structure on the dependence between X and Y, while at the same time guaranteeing that the margins of H^Q equal the univariate RNDs, through the use of copulas. Once we fix the structure of dependence (i.e. pick a copula), based on a historical analysis, we need only few market prices to obtain tight bounds for the degree of dependence that is consistent with these market prices.

Quest for the right copula. In helping us select an appropriate copula, it is important that the copula adequately describes the part of the distribution function that is relevant for the application at hand. For example, when we price a basket option, $e^{-r(T-t)}.E(aX +$

[5]This approach is very similar to the traditional approach of modelling the underlying variable as GBM. In the case of incomplete markets, switching from the historical to the risk-neutral measure is done by including a market price of risk factor in the drift component. This does not change the structure of the stochastic process or the implied terminal distribution.

$bY - k)^+$, it is clear that for large k tail dependence becomes crucial. We now introduce the concept of tail dependence and show how this property is completely determined by the copula. We also show some specific results for tail dependence for Archimedean and elliptical copulas, which are the two classes of copulas that we will work with.

Tail dependence. Upper (λ_U) and lower (λ_L) tail dependence coefficients (Nelsen, 2006) are defined as:

$$\begin{aligned} \lambda_U &= \lim_{t\to 1} \Pr[X > F_X^{-1}(t) | Y > F_Y^{-1}(t)] \\ \lambda_L &= \lim_{t\to 0} \Pr[X < F_X^{-1}(t) | Y < F_Y^{-1}(t)] \end{aligned}$$

In other words, the upper tail dependence coefficient gives us the probability that X is large given that Y is large. Consequently, the probability that both X and Y exceed their 99 percentiles (t=0.99) is approximately given by $\lambda_U.(1-t) = \frac{\lambda_U}{100}$. For example, let X and Y represent the daily returns of 2 assets. For $\lambda_U = 0.4$ we expect that the returns simultaneously exceed their 99 percentiles once every year, for $\lambda_U = 0.8$ we expect this to happen once every six months[6]. We can write the expressions for λ_U and λ_L as a function of the copula:

$$\begin{aligned} \lambda_U &= \lim_{u\to 1} \frac{\bar{C}(1-u, 1-u)}{1-u} \\ \lambda_L &= \lim_{u\to 0} \frac{C(u,u)}{u} \end{aligned}$$

This proves that all the properties that apply for the copula also hold true for tail dependence. As a result, it is straightfoward to show that λ_L of a particular copula

[6]We assume 250 trading days per year.

equals λ_H of the corresponding survival copula and vice versa. In addition, $\lambda_U = \lambda_L$ for symmetric distributions (copulas), i.e. copulas for which $C(u,u) = \bar{C}(u,u)$. As mentioned above, examples of symmetric copulas are the normal, student-t, and Frank copulas. When $\lambda_i > 0$ for $i \in \{L, U\}$, the copula has tail dependence, which means that extreme events happen simultaneously more often than in the case of independence. When $\lambda_i = 0$, the variables X and Y are asymptotically independent. The bivariate normal distribution with correlation ρ (normal copula) has $\lambda = 0$ for $\rho \in [-1, 1)$[7]. If we go deep enough into the tail, regardless of the correlation coefficient, extreme events happen independently. In contrast, the student-t copula does have a positive tail dependence coefficient (Joe, 1997):

$$\lambda = 2.\bar{t}_{v+1}\left(\sqrt{v+1}.\sqrt{\frac{1-\rho}{1+\rho}}\right)$$

with t_{v+1}: the univariate student-t distribution with $v+1$ degrees of freedom. As expected, for $v \to \infty$, we revover the results for the normal copula. We know that an Archimedean copula is completely characterized by its generator ϕ_γ. This means that the tail dependence coefficients of Archimedean copulas can be written as a function of the generator:

$$\begin{aligned} \lambda_U &= 2 - \lim_{x \to 0} \frac{1 - \phi_\gamma^{-1}(2x)}{1 - \phi_\gamma^{-1}(x)} \\ \lambda_L &= \lim_{x \to \infty} \frac{\phi_\gamma^{-1}(2x)}{\phi_\gamma^{-1}(x)} \end{aligned}$$

[7]For the normal copula $\lambda = 1$ for $\rho = 1$.

A generator ϕ_γ that is regularly varying[8] at zero with tail index $-\gamma$ ($\gamma \geq 0$) implies that ϕ_γ^{-1} is regularly varying at infinity with tail index $-1/\gamma$. In that case $\lambda_L = 2^{-\frac{1}{\gamma}}$. A generator ϕ_γ that is regularly varying at one with tail index γ ($\gamma \geq 1$) implies that ϕ_γ^{-1} is regularly varying at zero with tail index $1/\gamma$. In that case $\lambda_U = 2 - 2^{\frac{1}{\gamma}}$. Applying these expressions to the Archimedean copulas that we discussed before we get:

$$\begin{aligned}
\lambda_L^{Clayton} &= 2^{-\frac{1}{\gamma}} \text{ and } \lambda_U^{Clayton} = 0 \\
\lambda_L^{Gumbel} &= 0 \text{ and } \lambda_U^{Gumbel} = 2 - 2^{\frac{1}{\gamma}} \\
\lambda^{Frank} &= 0
\end{aligned}$$

These results are in line with the results in Table 1 and Figure 1 where we concluded that the Clayton copula implies that the variables are more often simultaneously low than high, and vice versa for the Gumbel copula. So far, we have only been looking at one parameter Archimedean copulas, which does not offer a lot of flexibility. In that regard, a result from Nelsen (2005) is useful. Let ϕ_γ be a generator, and define $\phi_{\alpha,\gamma}(t) = \phi_\gamma(t^\alpha)$ and $\phi_{\beta,\gamma}(t) = [\phi_\gamma(t)]^\beta$. Then $\phi_{\beta,\gamma}(t)$ is a generator for $\beta \geq 1$, and $\phi_{\alpha,\gamma}(t)$ is a generator for $\alpha \in (0, 1]$. When ϕ_γ is twice differentiable and $t.\phi_\gamma^{'}(t)$ is non-decreasing on (0,1) then $\phi_{\alpha,\gamma}(t)$ is a generator for $\alpha > 0$. The usefulness of this result stems from the fact that these transformations of the generator lead to simple transformations of the tail dependence coefficients. Specifically, the tail dependence coefficients for the copula $C_{\alpha,\gamma}$ are $(\lambda_L)^{1/\alpha}$ and λ_U. For $C_{\beta,\gamma}$ we get $(\lambda_L)^{1/\beta}$ and $2 - (2 - \lambda_U)^{1/\beta}$. We can also combine the two transformations into a single one, $\phi_{\alpha,\beta,\gamma}$. In that case we get the following tail

[8]A positive function L on (0,∞) is regularly varying at ∞ with index α if $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = t^\alpha$ for $t > 0$. Power functions are examples of regularly varying functions.

dependence coefficients for $C_{\alpha,\beta,\gamma}$: $(\lambda_L)^{1/\alpha\beta}$ and $2-(2-\lambda_U)^{1/\beta}$. This result allows us to create Archimedean copulas with arbitrary tail dependence coefficients, by matching α and β, and still have one free parameter, γ.

Parameter restrictions based on historical data and market prices. The first step in the empirical analysis is to determine whether the historical data has upper/lower tail dependence or not, which is an important dependence measure for our application. If no tail dependence is present in the data, we should model the data using the normal or Frank copula, among other possible copulas. If tail dependence is present in the data, one approach is to match α and β to the estimated tail dependence values. Another approach is to use a mixture of copulas to match the tail dependence values. For example, for $C = p.C_{\gamma_1} + (1-p).C_{\gamma_2}$[9]we get that $\lambda_i^C = p.\lambda_i^{C_{\gamma_1}} + (1-p).\lambda_i^{C_{\gamma_2}}$. All these approaches have in common that there is one free parameter remaining after matching the tail dependence coefficients. The second step in the analysis is to choose the copula or mixture of copulas that best describes the overall dependence structure. In the third and final step we restrict the free parameter so that it is consistent with the market prices on options and futures. In that sense the free parameter reflects the forward looking expectations, present in the options and futures market data, about the degree (strength) of dependence. For all $(x,y) \subset R_+^2$ and consistent with k_L and k_U we have:

$$RLB(x,y) \leq H^Q(x,y) = C_\Omega^Q[F_X^Q(x), F_Y^Q(y)] \leq RUB(x,y) \tag{1.6}$$

with Ω: the set of parameters. After fixing some parameters by calibrating them to the tail dependence coefficients, we have to find the permissible set for the remaining free

[9]One can easily prove that a linear combination of copulas results in another copula.

parameters such that they are consistent with the bounds RLB and RUB. In the methods explained above there is just one free parameter γ, for which we have to find the interval $[\gamma_1, \gamma_2]$ such that equation (1.6) holds for all (x,y). It is very useful to restrict our attention to concordance ordered copulas (see definition), a class to which the normal, student-t, Clayton, Gumbel, and Frank copulas all belong. In that case we can apply Claim 6 and bound the price of a bivariate option, $C(k)$, with a supermodular payoff function $G = [\psi - k]^+$, such as a basket option or rainbow-min option, as follows ($\Omega = \{\theta, \gamma_i\}$):

$$e^{-r(T-t)}.E_t^{C_{\theta,\gamma 1}}[\psi(X_T, Y_T) - k]^+ \leq C(k) \leq e^{-r(T-t)}.E_t^{C_{\theta,\gamma 2}}[\psi(X_T, Y_T) - k]^+$$

We impose structure on the bivariate RND, through the choice of a parametric copula, based on the historical joint behavior of X and Y. Therefore, we just require market prices of bivariate options to infer the degree or strength of dependence the market is expecting until the expiration time of the contract. As a result, the bounds in the parametric case are tighter than in the non-parametric case.

CHAPTER 2

Tail Behavior of Natural Gas Futures Returns

2.1. Introduction

The first product we are looking at is natural gas. Given that natural gas is storable (mainly in salt caverns, depleted oil fields, and aquifers) the forward price of a particular month is closely linked to the forward prices of the other months. The January 2009 forward price depends very much on the September 2008 price. The main reason the Jan09 price can exceed a cash-and-carry[1] spread with Sep08 is because of a fear of a shortage of gas come winter 2009. Several factors determine the likelihood of a shortage of gas in the winter: the severity of the winter (measured in Heating Degree Days, HDDs), gas production shut-ins due to hurricanes, and the level of gas storage at the end of the injection season (Apr-Oct). The maximum storage capacity in the US is around 3.7 TCF[2] (Trillion Cubic Feet). The moment that storage drops below 1 TCF, we have a problem. The winter of 2008 was a cold winter and 2.2 TCF was withdrawn during the withdrawal season (Nov-Mar). Early November 2007, the US was at full capacity (3.7 TCF), so at the end of the winter (end of March) there was 1.5 TCF left in storage. Given that at this moment in time, Aug 08, we are at a lower storage level than Aug 07, and add to this the

[1]A reasonable storage cost is a 2% (of the spot gas price) injection fee, and a 2% withdrawal fee. Many operators require that the user empties the facility at the end of the withdrawal season (end of March). This often helps in pushing down April spot gas prices.

[2]This is working gas capacity. At least the same quantity is also being stored, but is defined as base gas, and is needed to maintain the correct pressure levels so that the working gas can be properly injected in and withdrawn from the storage facilities.

potential risk of shut-ins due to hurricanes in Sep-Oct, we could end up with a maximum storage level early November of 3.3-3.4 TCF. If we have a similarly cold winter as in 2007 or even slightly colder, we could withdraw around 2.3-2.4 TCF during the coming winter 2009. This could bring us very close to a 1 TCF storage level at the end of winter 2009, and shows that a scenario in which we experience a shortage of gas at the end of winter 2009 is not totally impossible. At that moment, it becomes very important what is going on in the rest of the world, because we will need to increase our imports of LNG. The price at which we will be able to purchase this LNG will determine the price of gas in the US. If, just as last winter 2008, there is a very cold winter in the south of Europe (big importer of LNG), and there is again a massive nuclear outage in Japan (8000 MW), we will have to pay a lot to convince the LNG tankers to deliver it to the US. Most of the imports of LNG in the US come from Trinidad. If there is a significant chance of a shortage of gas, we will need to import more LNG from the Middle East (with Oman as its biggest exporter), which also puts the US at a transportation cost disadvantage to the south of Europe and South East Asia. All of these factors have to be assessed and will result in fear of a shortage premium that explains why the Jan09 price is usually higher than a Sep08 price plus a simple cash-and-carry argument. Gas forward prices further out in the curve are the result of probability distributions of US gas production, which recently has been growing at an astonishingly high 9% - mainly the result of a dramatic increase in unconventional (horizontal, instead of just vertical) drilling as in for example the enormous Barnett shale in Texas, and the addition of pipeline capacity out of the Rockies to the east. In addition to US gas production, the level of pipeline imports from Canada and LNG imports from Trinidad, the Middle East, and the North Sea, and

the level of pipeline exports to Mexico and LNG exports to the rest of the world are also an important variable. But the main variable, next to US gas production, is US gas consumption. The four main categories of US gas consumption are industrial gas demand, commercial gas demand, residential gas demand, and gas demand by power generators (20-25%). Industrial gas demand is mainly determined by GDP growth. Commercial gas demand also depends on GDP growth, but also has a strong seasonal variation, and residential gas demand is almost purely determined by temperature. These three sources of gas demand are fairly predictable, of course nobody knows what the weather will be like next year, but almost everyone knows how residential gas demand will vary with different weather scenarios. The gas demand for power generation ("gas burn"), which makes up about 20-25% of gas demand, is more complex. The primary drivers for gas burn are temperature and the make-up of the power supply stack. In order to trade gas forwards far out in the curve, it is important to have a view on how the supply stack can change and how this in turn can affect gas burn. A proper evaluation of probability distributions of future US gas production, pipeline imports and exports, LNG imports and exports, US gas demand, and gas burn for power generation will go into the determination of gas forwards far out in the curve.

2.1.0.7. Contango and Backwardation. We focus now on the historical behavior of natural gas futures. We start by plotting the futures curve on the first trading day in January for a couple of years in the period 1992-2007. In the early 90s futures contracts were trading for the next 14-16 months. During the mid 90s contracts were trading for the next 36 months, and in the last couple of years contracts trade out as far as 72 months. The first characteristic that stands out immediately when we plot the futures curve is

the seasonality, prices are higher in the winter months. The reason that seasonality exists in this market is because of limited storage capacity. If there was no constraint on storage capacity you would buy gas in the summer, store it, and then sell it at the higher price in the winter. The degree of seasonality determines to a large extent the level of profit of operating a storage facility. The second characteristic is that the futures curve is for certain years in contango (futures curve is upward sloping), and for other years in backwardation (futures curve is downward sloping). For pure investment assets the futures curves is always in contango, and this can easily be understood with the following no-arbitrage argument. Suppose you want to sell an investment asset at time T (current time is t). In the first scenario you can borrow money to buy the asset now for the spot price P_t for the period (T-t) at the borrowing rate $r_{t,T}$. For certain assets there is also a cost involved to store the asset for the period (T-t), and we denote it by $c_{t,T}$. For financial assets this storage cost is zero (c=0). At final time T we sell the asset for P_T. This first scenario has the following cashflows at final time T: $-c_{t,T} - P_t.(1 + r_{t,T})^{T-t} + P_T$. In the second scenario we enter a futures contract at time t, and agree to buy the asset at final time T for the futures price $F_{t,T}$ and we immediately sell it for P_T. This scenario has cashflows at final time T equal to $-F_{t,T} + P_T$[3]. Given that both scenarios have zero net cash outlays at time t, the cashflows at time T have to be the same for both scenarios, which leads to the following expression:

$$F_{t,T} = c_{t,T} + P_t.(1 + r_{t,T})^{T-t}$$

$c_{t,T}$: cost of storage for the period (T-t)

[3]Equivalently, we could cancel out our futures position at time final T, which leads to the same cashflow: $F_{T,T} - F_{t,T}$, because the futures price converges to the spot price ($F_{T,T} = P_T$).

$r_{t,T}$: borrowing rate for the period (T-t)

P_t : spot price at time t

$F_{t.T}$: futures price at time t for delivery at time T

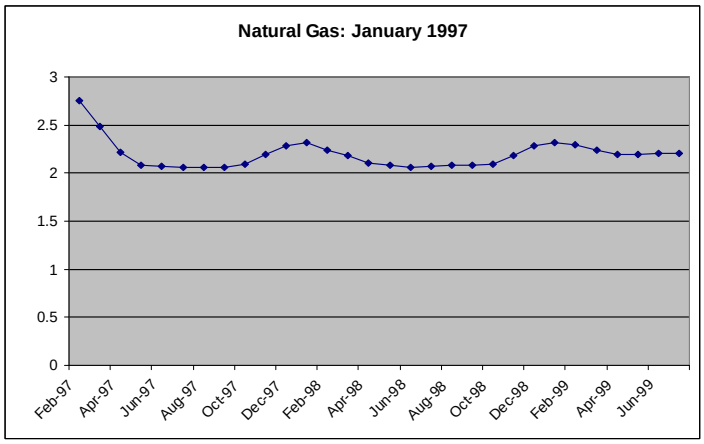
Assets that are being used for consumption or production differ from assets that are held purely for investment purposes. Apart from the cost of storage (c>0) and the cost of borrowing (r>0) there usually is also a benefit to owning the asset now rather than tomorrow. This benefit is called the convenience yield (y). Many intuitive explanations can be given for why there exists a convenience yield (y>0). Introducing this convenience yield leads to the new expression for the futures price:

$$F_{t,T} = c_{t,T} - y_{t,T} + P_t.(1 + r_{t,T})^{T-t}$$

$y_{t,T}$: convenience yield for the period (T-t)

With a high convenience yield, it is thus possible that the futures curve is in backwardation, i.e. downward sloping. The existence of backwardation might seem at odds with the presence of speculators. Take a speculator, who is not interested in using the asset for production purposes. In the presence of backwardation he would short sell the asset today and buy it back later for a lower futures price, locking in a pure profit. Given that we actually observe backwardation in the market means that this action on the margin is not possible. The reason the speculator cannot implement this strategy is that he won't be able to short sell the asset. Backwardation means that there is a high convenience yield, which usually happens when there is fear of a shortage of supply, which is the result

Figure 2.1. Natural Gas Futures curve, Jan1999



of a low level of inventories, and a lot of market participants simply have to hold minimal inventories. In equilibrium, a speculator that wants to sell short the asset cannot convince a current owner of the asset to let him borrow a fraction of the asset now and return it later. Figure 2.1 plots the Henry Hub natural gas futures curve (Nymex) at the first trading day of January 1997, 2002, and 2006. The price quotation is in US dollars and cents per mmBtu, where 1 mmBtu stands for 1 million British thermal units. The three years are chosen to demonstrate that the natural gas market can have flat futures curves, futures curves in contango, and futures curves in backwardation.

Another way to look at futures prices is to plot the price evolution for a given contract. We do not expect to observe any seasonality in the price curve for a particular contract. We just expect an entire (parallel) shift of the price curve for contracts that end in a different season. Figure 2.4 plots the futures prices for the January and July contracts for 2002 and 2006. We plot the futures price for the January and the July contract at the same trading day. This means that Figure 2.4 plots the January contract until expiration and the July contract until six months before expiration. We can observe that the futures

Figure 2.2. Natural Gas Futures curve, Jan2002

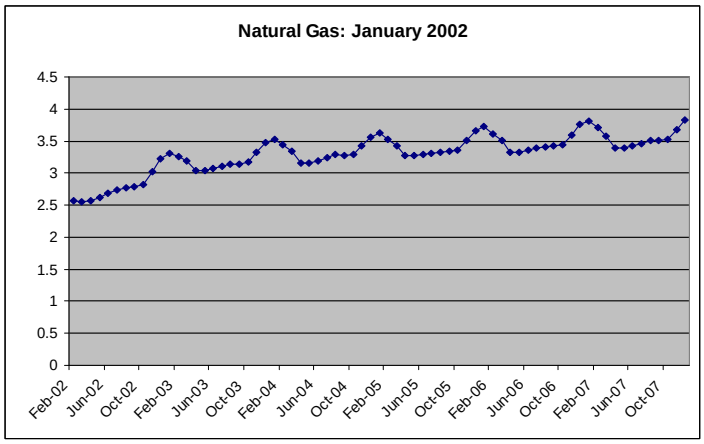
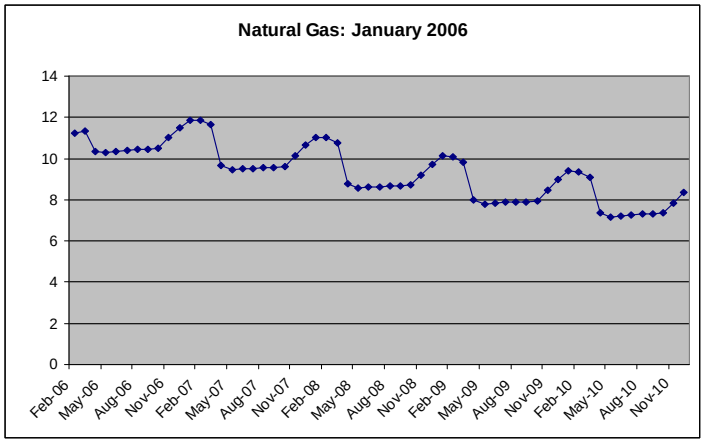


Figure 2.3. Natural Gas Futures curve, Jan2006



prices for the January and July contracts for the same year are closely linked, with the January price just above the July price. We can also see that there is a lot of variation in the futures price for a given contract. Given the fact that the futures price at time t is the (risk-neutral) expectation of the spot price at expiration given the information set at time t, it means that the market reacts strongly to the arrival of new information. The price curves clearly show that the futures price is not always a very good predictor of the spot price at expiration.

Figure 2.4. Natural Gas Jan-Jul 2002 contract price

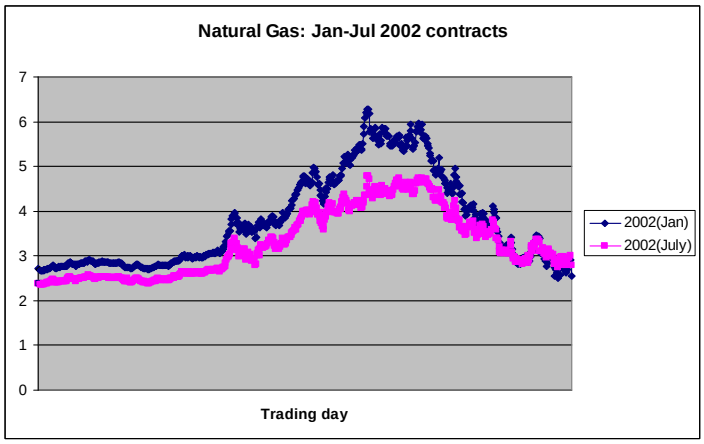
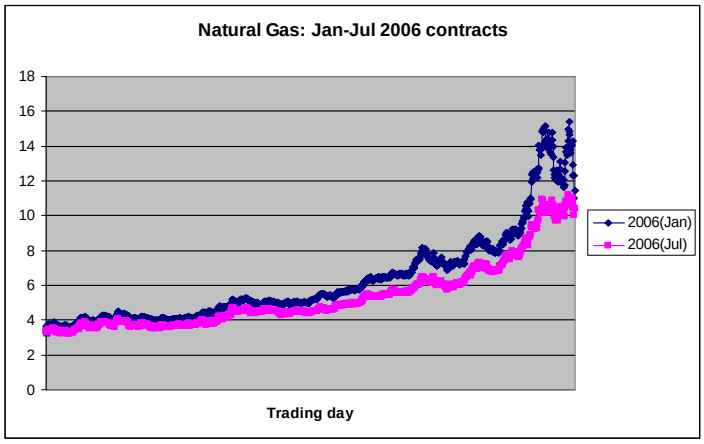


Figure 2.5. Natural Gas Jan-Jul 2006 contract price



We also plot the volatility of the futures price for a particular contract as a function of time to maturity. We can clearly see that volatility rises, sometimes dramatically, as we come close to the end of the contract. Market participants react more strongly to new information as the expiration of the contract is nearby. We give plots for the Jan-2002 and Jul-2006 contracts, but this phenomenon is true in general.

Figure 2.6. Natural Gas Jan 2002 volatility curve

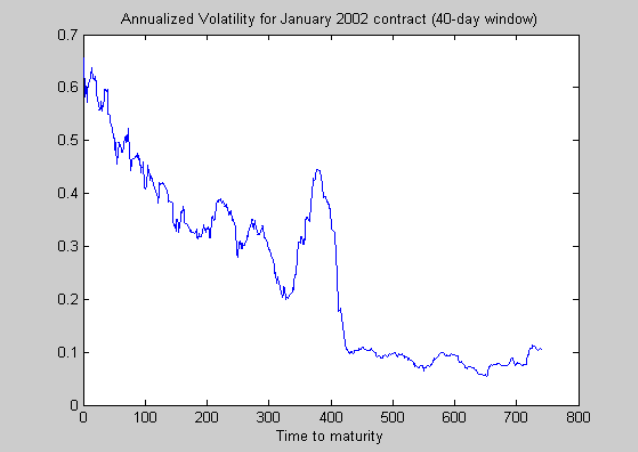
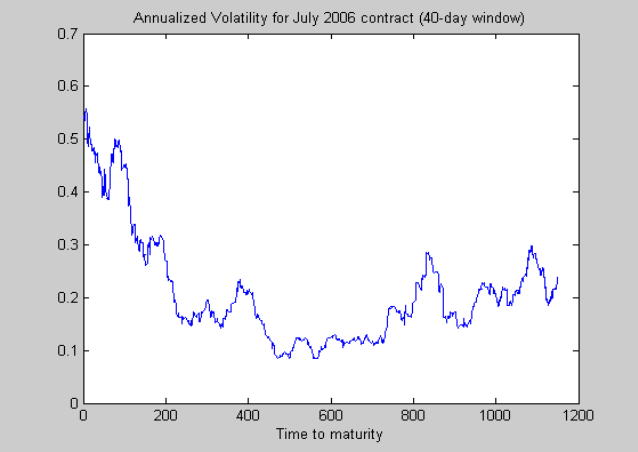


Figure 2.7. Natural Gas Jul2006 volatility curve



2.2. Historical Distribution of Natural Gas Futures Returns

2.2.1. Empirical density function and Summary statistics

First off we want to check whether there is a significant difference between the returns of futures contracts that expire in the summer or winter months. Two standard tests can be used to test whether two sample datasets come from the same distribution: the Kolmogorov-Smirnov (KS) test and the Cramer-von Mises (CvM) test. The KS test uses

the following test statistic, which has asymptotically a χ^2distribution:

$$TS_{K-S} = \max_x |F_1(x) - F_2(x)|$$

We can see that the KS test just looks at the maximum distance between the two cdfs, which means that this test is sensitive to outliers. As an alternative, we have the Cramer-von Mises test, which is basically an average Kolmogorov test:

$$TS_{C-vM} = \int |F_1(x) - F_2(x)| dF(x)$$

Both tests do not reject the null hypothesis that the summer and winter daily returns originate from the same distribution. Both tests take the difference between two cdfs. By plotting two cdfs one can immediately observe the problem with this approach. Any discrepancies between the two datasets in the tails of the distribution are almost impossible to be picked up by the KS and CvM tests. Figure 2.8 shows the QQ-plot of the summer (horizontal axis) vs. winter (vertical axis) returns. This QQ-plot clearly shows that in the bulk of the distribution the summer and winter returns seem to originate from the same distribution. In both the upper and lower part of the distribution, however, the winter returns have fatter tails than the summer returns, but this difference is not picked up by the KS and CvM test. In the subsequent analysis we choose not to discriminate between summer and winter returns.

In Figure 2.9 we plot the empirical density function for the daily returns of the natural gas futures from June 1993 until December 2010 and compute the first four moments.

Figure 2.8. QQ-plot: summer vs. winter daily returns

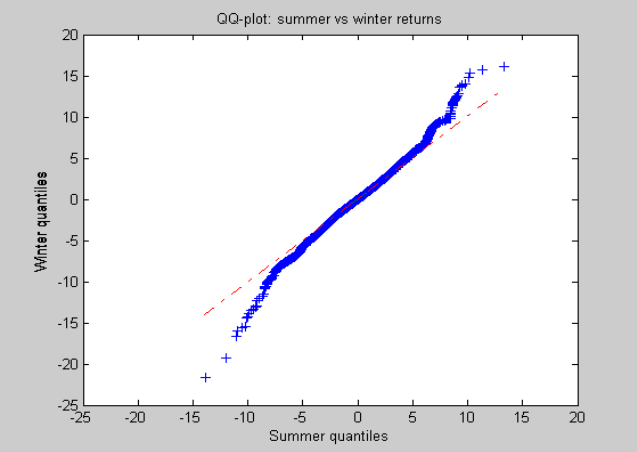
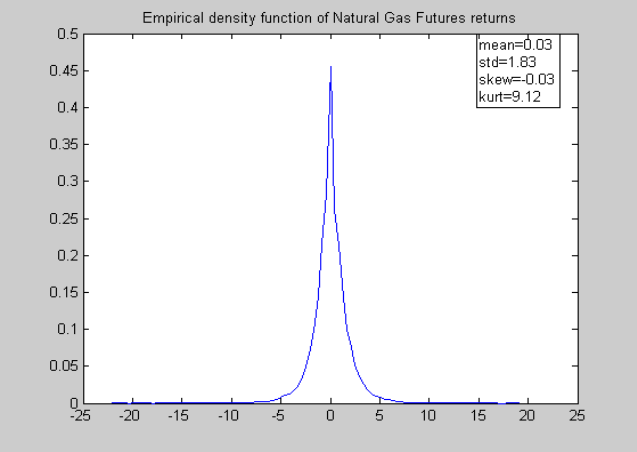


Figure 2.9. Empirical density function for daily returns of natural gas futures



Given a kurtosis of 9.12, it is clear that natural gas futures returns are not normally distributed, and as a result gas futures prices are not lognormally distributed.

2.2.2. Exponential and Subexponential distributions

In the subsequent analysis we look separately at the positive returns and the negative returns. We have about 38,000 observations for positive returns and a similar number for the negative returns. In the next couple of sections we want to find the most adequate

distribution for the positive and negative returns of natural gas futures. In particular, we want to find the distribution that best captures the tail behavior of these returns. Having a good model for the tail behavior of the returns is very important both from a risk management point of view (VaR and Expected Shortfall[4]) and an option pricing point of view. Value-at-Risk calculations at high percentiles and deep out-of the-money option prices can vary dramatically for different models of tail behavior. Given a kurtosis of 9.17 it is clear that the return distribution has heavier tails than the normal distribution. A convenient way to model phenomena with high variability is through the use of subexponential distributions. Following Goldie and Kluppelberg (1997), a distribution function F belongs to the class of subexponential distributions if it has support (0,∞) and for X_1, X_2,...X_n iid with cdf F we have

$$\lim_{x\to\infty} \frac{\Pr(X_1 + X_2 + ... + X_n > x)}{\Pr\{\max(X_1, X_2, ..., X_n) > x\}} = 1, \quad n \geq 2.$$

This intuitive condition shows that deep in the tail the maximum determines the sum. This means that there is a significant probability that X takes on a very large value. It is also shown that for subexponential distributions F we have that

$$\lim_{x\to\infty} \frac{\Pr(X > x)}{e^{-\delta x}} = \frac{\bar{F}(x)}{\bar{F}_{\exp(\delta)}(x)} = \infty, \quad \text{for all } \delta > 0.$$

This condition explains the name subexponential distribution. The tail of a subexponential distribution decays more slowly than any exponential distribution. This implies for a subexponentially distributed variable X that $E[(e^X)^\beta] = \infty$ for all $\beta > 0$. As

[4]For a position X with distribution function F_X we have that $VaR_\alpha = \inf\{x : F_X(x) \geq \alpha\}$ and $ES_\alpha = E[X - VaR_\alpha | X \geq VaR_\alpha]$.

we will see in the next section $E(X^{\beta})$ can be both finite and infinite within the class of subexponentials. Examples of subexponential distributions are the Pareto, Weibull, lognormal, and loggamma distributions among others. The normal, exponential, and gamma distributions do not belong to the class of subexponential distributions, as the tails of these distributions decay too quickly. Within the class of subexponential distributions, i.e. distributions for which the tail decays more slowly than the exponential distribution, there are two important subclasses that we will focus on. The regularly varying distributions (Bingham et al., 1987), to which the Pareto distribution belongs, and the stretched-exponentials (SE), to which the Weibull distribution belongs. In our analysis we will investigate how slowly the tail of the natural gas futures returns decays: more slowly than a power tail, similar to a power tail, in between a power tail and an exponential tail, similar to an exponential tail, or in between an exponential and a normal tail[5]. One extra property of subexponential distributions F that we will use later on is:

$$\lim_{x\to\infty} \frac{\bar{F}(x-y)}{\bar{F}(x)} = 1, \text{ for all } y > 0 \tag{2.1}$$

2.2.2.1. Regularly varying distribution. The tail of a distribution function F is said to be regularly varying with index $-\alpha$ ($\alpha > 0$, called the tail index), denoted $\bar{F} \in R_{-\alpha}$, if 1-F(x) is a regularly varying function, i.e. if

$$\lim_{x\to\infty} \frac{1-F(tx)}{1-F(x)} = t^{-\alpha}, \quad t > 0.$$

An important fact about regularly varying distributions discusses the finiteness of the moments of X. For X, a positive random variable with distribution tail $\bar{F} \in R_{-\alpha}$, with

[5]It is well known that a normal tail decays a rate of $exp(-x^2)$.

$\alpha > 0$, we have that

$$\begin{aligned} E(X^\beta) &< \infty \text{ for } \beta < \alpha \\ E(X^\beta) &= \infty \text{ for } \beta > \alpha. \end{aligned}$$

As a result, for $\alpha < 2$, X has a finite mean, but infinite variance. For $\alpha < 4$, X has infinite kurtosis, but finite mean, variance, and skewness. As such, the lower is α, the heavier is the tail. It can be shown that as the tail index, α, goes to infinity, the tail becomes an exponential tail[6]. As we will see in later sections, regularly varying distributions play an important role in extreme value theory as $\bar{F} \in R_{-\alpha}$ implies that the distribution function F belongs to the domain of attraction of the Frechet distribution. An example of a regularly varying distribution that we will use in this paper is the *Pareto* distribution: $F(x) = 1 - (\frac{u}{x})^\alpha$, for $x \geq u$, and $\alpha > 0$. Other names for regularly varying distributions that one can find in the literature are fractal, scaling, and power law distributions. The names "fractal", "self-similar", and "scaling" all have a very similar meaning: we use these terms to denote geometrical objects, processes, or distributions that have a similar structure independent of the size or scale. For example, looking at the Pareto distribution, we can easily see that the $\Pr(X > \rho x) = \rho^{-\alpha}.\Pr(X > x)$, no matter the value of x. Therefore, the $\Pr(X > \rho x | X > x)$ is independent of x. This property does not hold true for the exponential or normal distribution. This fractal or scaling property

[6]The exponential distribution belongs to the class of rapidly varying distributions, i.e. $\bar{F}_{\exp} \in R_{-\infty}$. For this class we have that

$$\begin{aligned} \lim_{x\to\infty} \frac{1-F(tx)}{1-F(x)} &= 0, t > 1 \\ &= \infty, 0 < t < 1. \end{aligned}$$

This class plays an important role in extreme value theory, as distributions that are rapidly varying belong to the domain of attraction of the Gumbel distribution.

seems to ubiquitous in daily life. Many studies have demonstrated the presence of the scaling property: internet traffic flows (Tadaki, 2007), natural disasters (Meier, 2006), word frequencies (Li, 1992), income distributions (Di Guilmi et al., 2003), city populations (Gabaix, 1999), and many other applications. Therefore, the power law distribution is probably more normal than the normal distribution, a point that Mandelbrot (1963, 1997) has been advocating for the last forty-five years. Several studies in finance have also demonstrated the presence of power law tails in financial returns (Gabaix et al., 2003; Fama, 1965; Farmer and Lillo, 2004; Lux and Sornette, 2002; Mantegna and Stanley, 1995; Mittnik et al., 1998; Muzy et al., 2001).

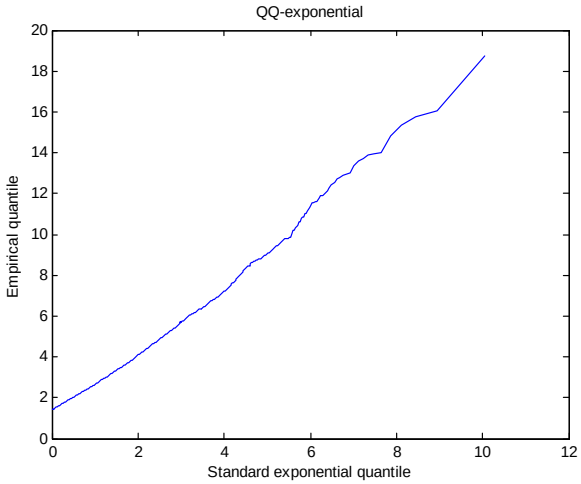
2.2.2.2. Stretched-Exponential distributions (Weibull). The class of stretched-exponential (SE) distributions is also an important subset of the subexponentials. Given that they are a subset of the subexponentials, we know that SE distributions decay at a slower rate than the exponential distribution, and that their exponential moments are infinite. However, in contrast to the regularly varying distributions, all their moments are finite. As such, the SE distributions have thinner tails than the power law distributions and fatter tails than the exponential distribution. A well-known example is the Weibull distribution for parameters within a specified range. The Weibull distribution: $F(x) = 1 - \exp(-\lambda x^{\gamma})$, for $x \geq 0$, and $\lambda, \gamma > 0$. The Weibull distribution belongs to the class of SE distributions when $\gamma < 1$, i.e. when the distribution decays more slowly than the exponential distribution.

2.3. Preliminary graphical analysis of tail behavior of gas futures returns

Before going into the details of modelling the tail distribution, it is instructive to use some graphical tools to get a rough idea of how heavy-tailed the distribution of gas futures returns is. As a first step we want to see whether the tail behavior of the gas futures returns resembles a power tail, a SE tail, or an exponential tail. The techniques we will use are the quantile-quantile (QQ) plots, the mean excess (or expected shortfall) plots, the large claim index, and one other technique that relies on results about how the ratio of the max over the sum behaves for different levels of fat-tails. The mean excess plots, the large claim index, and the max/sum ratio compare the data with asymptotic results. Given that we only have a finite sample, there is room for variability in these plots. Therefore, it is important to look at several different techniques with the hope to see a consistent pattern of how heavy-tailed the distribution is. We treat the positive and negative returns separately.

Quantile-Quantile plots. In this section we compare the empirical quantile function Q_n with the theoretical quantile function Q that corresponds to the exponential, pareto, and weibull distribution. For a distribution function F we define the quantile function as: $F^{-1}(p) = Q(p) = \inf\{x : F(x) \geq p\}$. Given the exponential distribution: $F(x) = 1 - \exp[-\lambda(x-u)]$, for $x \geq u$, and $\lambda > 0$, we have the corresponding quantile function: $Q(p) = u - \frac{1}{\lambda}\log(1-p)$. If the data resembles the exponential distribution, then the plot $[-\log(1-p), Q_n(p)]$ should be linear with the slope equal to $\frac{1}{\lambda}$. Recall that we are interested in what happens in the tail. Figures 2.10 and 2.11 show the QQ-plots for the top 30% futures returns (positive).

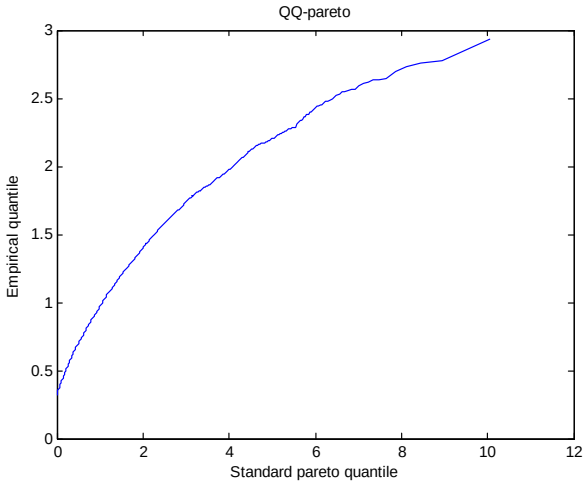
Figure 2.10. QQ-plot: data vs. exponential distribution



From Figure 2.10 we learn that the tail of the gas futures returns is close to the exponential tail. The curve is slightly convex, indicating that the empirical tail decays slightly more slowly than its exponential counterpart. Next, we compare the tail behavior of the data with a power tail, through the pareto distribution. We use the following version for the pareto distribution: $F(x) = 1 - (\frac{x}{u})^{-\alpha}$, for $x \geq u$, and $\alpha > 0$. This results in the quantile function: $\log Q(p) = \log u - \frac{1}{\alpha}\log(1-p)$. Hence, the plot $[-\log(1-p), \log Q_n(p)]$ should be linear if the data resembles a power tail.

It is clear from Figure 2.11 that the tail of gas futures returns decays faster than a power tail. Finally, we take a look at the weibull distribution: $F(x) = 1-\exp[-\lambda(x^\gamma - u^\gamma)]$, for $x \geq u$, and $\lambda, \gamma > 0$. This leads to the quantile function: $\log Q(p) = \frac{1}{\gamma}.\log[-\frac{1}{\lambda}\log(1-p) + u^\gamma]$. The problem with this expression is that we cannot obtain a linear relation between a function of p and the parameters γ and λ as long as $u > 0$. As a result, we cannot use a QQ-plot for the weibull distribution when $u > 0$.

Figure 2.11. QQ-plot: data vs. pareto distribution



Max/Sum ratio. O'Brien (1980) proved the following result:

$$R_n(p) = \frac{M_n(p)}{S_n(p)} \rightarrow^{a.s.} 0 \Leftrightarrow EX^p < \infty$$

$M_n(p) = \max\{X_1^p, X_2^p, ..., X_n^p\}$

$S_n(p) = |X_1|^p + |X_2|^p + ... + |X_n|^p$

Plotting $R_n(p)$ as a function of n for different levels of p, we learn whether the p-th moment of X is finite or not. As mentioned before, random variables that belong to the exponential and the SE class, have all their moments finite. For regularly varying distributions with tail index α, we have that $EX^\beta < \infty$ for $\beta < \alpha$. In Figure 2.12 we plot $R_n(p)$ with p=1, 3, 5, and 7 for the natural gas futures returns.

We observe that, according to this graphical tool, not all moments of gas futures returns are finite. The finiteness of the fifth moment is questionable, but $R_n(7)$ is clearly

Figure 2.12. R for different levels of p: gas futures returns

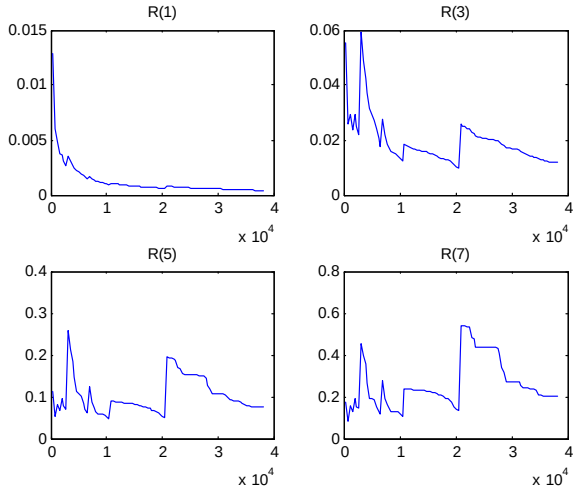
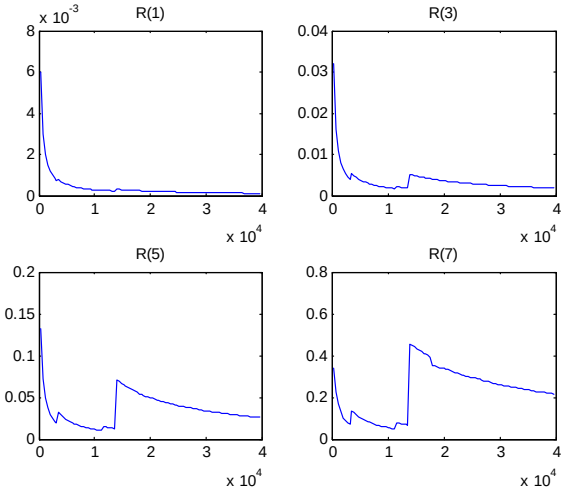
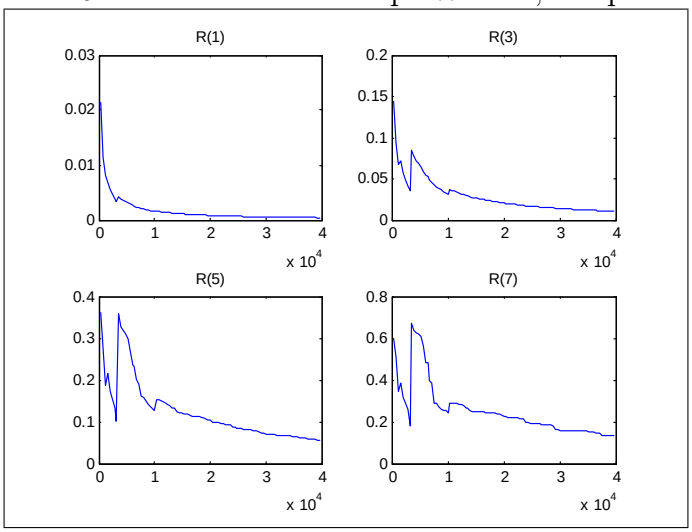


Figure 2.13. R for different levels of p: Pareto, sample size 40,000



not close to zero. However, we know that the O'Brien result holds true asymptotically ($n \to \infty$). In order to put Figure 2.12 in better perspective for a finite sample we will plot two similar figures with 40,000 simulated points from a pareto distribution with tail exponent $\alpha = 8$, and from a weibull distribution ($\gamma = 0.8$).

Figure 2.14. R for different levels of p: Weibull, sample size 40,000



We know that the first seven moments of the pareto distribution with tail index $\alpha = 8$ are finite, which implies that $\lim_{n\to\infty} R_n(p) = 0$ for $p < 8$. We also know that all moments of the weibull distribution with $\gamma = 0.8$ are finite. However, we can see in Figures 2.13 and 2.14 that for a finite sample (n=40,000), $R_n(5)$ and definitely $R_n(7)$ are not equal to zero, although theoretically the fifth and seventh moments are finite. Actually, both Figure 2.13 and 2.14 resemble very much Figure 2.12, which plots the data. Therefore, we cannot conclude from Figure 2.12 with certainty that the fifth and seventh moment of the gas futures returns are infinite, nothwithstanding the fact that $R_n(5)$ and $R_n(7)$ are different from zero.

Mean excess function. Another method to investigate to which distribution class the data belongs is based on the mean excess function[7]:

$$e(u) = E(X - u|X \geq u).$$

[7]In a risk management context and switching to the lower tail, this is referred to as Expected Shortfall.

The mean excess function has a different shape depending on the distribution class to which X belongs. It can be rewritten as:

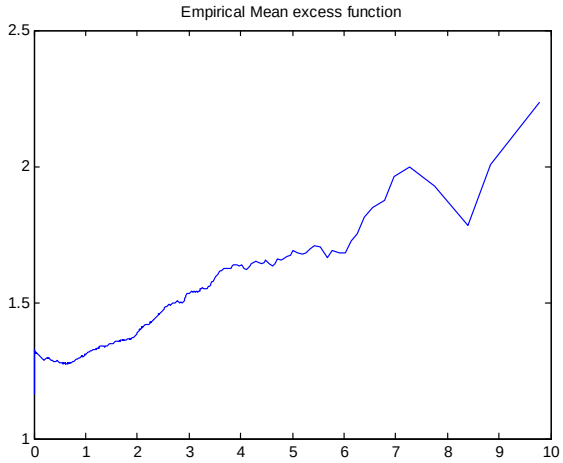
$$e(u) = \frac{1}{\bar{F}_X(u)} \int_0^{x^F} (x-u).dx \tag{2.2}$$
$$= \frac{1}{\bar{F}_X(u)} . \int_u^{x^F} \bar{F}_X(x).dx$$

x^F: right end-point of the distribution F_X, which can be ∞.

Applying Karamata's theorem[8], we show the following useful fact about mean excess functions. If for all $y > 0$, $\lim_{x\to\infty} \frac{\bar{F}(x-y)}{\bar{F}(x)} = e^{\mu y}$ for $\mu \in [0, \infty]$, then $\lim_{u\to\infty} e(u) = \mu^{-1}$. As we can see from equation (2.1), $\mu = 0$ for the class of subexponential distributions, i.e. the mean excess function tends to infinity for the subexponentials, to which the Pareto and the Weibull ($\gamma < 1$) distributions belong. For the class of superexponentials ($\bar{F}(x) \sim \exp(-x^c)$, with $c > 1$), the tails of which decay faster than an exponential, to which also the normal distribution and the Weibull distribution ($\gamma > 1$) belong, we have that $\mu = \infty$, and as result the mean excess function tends to zero for the superexponentials. It is easy to check that for the exponential distribution itself with parameter λ we have that $\lim_{u\to\infty} e(u) = \lambda^{-1}$. Actually, using equation(2.2), we have that $e(u) = \lambda^{-1}$ for all u for the exponential distribution (λ). The mean excess function for the Pareto distribution is: $e(u) = \frac{1+u}{\alpha-1}$ for $\alpha > 1$. Thus, it is linearly increasing in u. For the weibull distribution we get: $e(u) \simeq \frac{u^{1-\gamma}}{\lambda\gamma}$. As a result, for $\gamma < 1$, i.e. when the weibull belongs to the SE class, it is increasing and concave in u, and it tends to ∞. When $\gamma > 1$, i.e. when the weibull

[8]See Embrechts et al., 1997 (Theorem A3.6).

Figure 2.15. Empirical mean excess function for gas futures returns



belongs to the superexponential class, it is decreasing and convex, and it tends to zero. In Figure 2.15 we plot the empirical mean excess function for the gas futures returns.

The empirical mean excess function for gas futures returns is clearly increasing, which implies that the tails are heavier than an exponential distribution. However, it is not clear from Figure 2.15, whether the tail behavior of the data better resembles a power tail or an SE tail. Even though the empirical mean excess function might be more concave than linear, the next Figures 2.16 and 2.17 show how difficult it is to draw definitive conclusions from the mean excess function based on a finite sample. Figures 2.16 and 2.17 plot 20 mean excess functions for 40,000 simulated points from the weibull distribution ($\gamma = 0.5$) and pareto distribution ($\alpha = 8$).

As we can see from Figure 2.16 some of the mean excess functions are close to linear for high u, and in Figure 2.17 certain mean excess functions are concave and might therefore be mistaken as coming from the weibull distribution. As a result, the only statement that

Figure 2.16. 20 mean excess functions of sample size 40000 from the weibull distribution

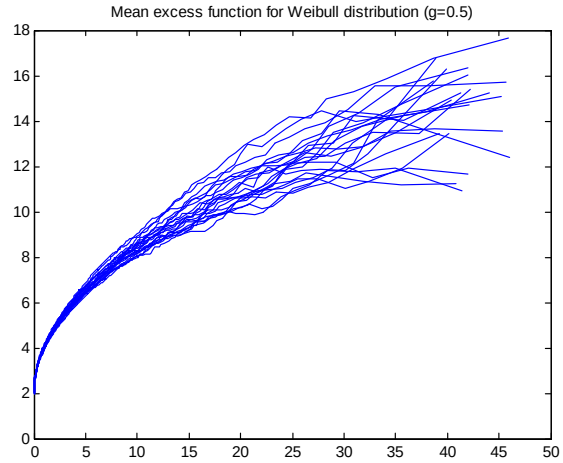
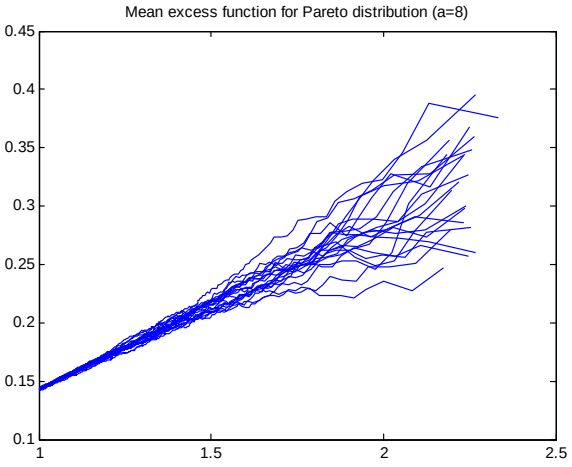


Figure 2.17. 20 mean excess functions of sample size 40000 from the pareto distribution



we can safely make from the empirical mean excess function is that the tail of gas futures returns is heavier than an exponential tail.

Large Claim index. This term originates in the insurance industry where it is commonly used to quantify the fraction x of claims that is responsible for the fraction y of total pay-outs. Take a sample of returns: $r_1, r_2, ..., r_N$ iid with distribution function F

and mean $\mu_F < \infty$[9]. We denote the order statistics by $r_{N,N} \leq r_{N-1,N} \leq r_{2,N} \leq r_{1,N}$. We define the empirical large claim index as:

$$LC_n(p) = \frac{r_{1,N} + r_{2,N} + ... + r_{[pN],N}}{\sum_{i=1}^{N} r_i}$$

The large claim index is the fraction of the $p.N$ largest returns to the sum of all returns. It is the opposite of the Lorentz curve, which gives the fraction of the pN smallest returns to the sum of all returns. Next, we define the following object:

$$D_F(p) = \frac{1}{\mu_F} \int_{1-p}^{1} F^{-1}(s)ds.$$

The next theorem (see proof in Embrechts et al.) shows the connection between both objects. For $n \to \infty$,

$$\sup_{p \in [0,1]} |LC_n(p) - D_F(p)| \to^{a.s.} 0$$

From this theorem it is clear how we can use the large claim index to help us determine the tail behavior of the gas futures returns. We limit ourselves to for example the top 30 or 40% of the returns, and we plot the empirical large claim index. We then find the distribution function F such that $D_F(p)$ is close to $LC_n(p)$. For the exponential distribution $F(x) = 1 - exp(-\lambda x)$, we have that $\mu_F = \lambda^{-1}$, and $F^{-1}(s) = -\frac{1}{\lambda}\ln(1-s)$. This results in $D_F^{\exp}(p) = p - p.\ln(p)$. For the Pareto distribution $F(x) = 1 - (1+x)^{-\alpha}$, we have that $F^{-1}(s) = (1-s)^{-\frac{1}{\alpha}} - 1$, and $\mu_F = \int_0^1 F^{-1}(y)dy = (\alpha - 1)^{-1}$. This leads to $D_F^{Par}(p) = \alpha.p^{-\frac{1}{\alpha}+1} - (\alpha - 1).p$. Finally, for the weibull distribution $F(x) = 1 - \exp(-x^{\gamma})$,

[9]This implies that in the case of the Pareto distribution the tail index α must be bigger than 1.

Figure 2.18. Empirical Large Claim Index vs. exponential distribution

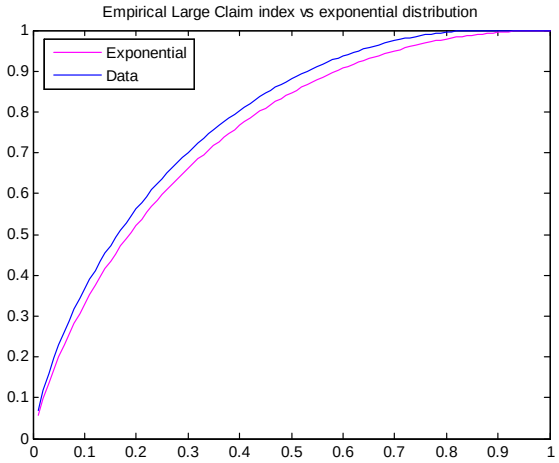
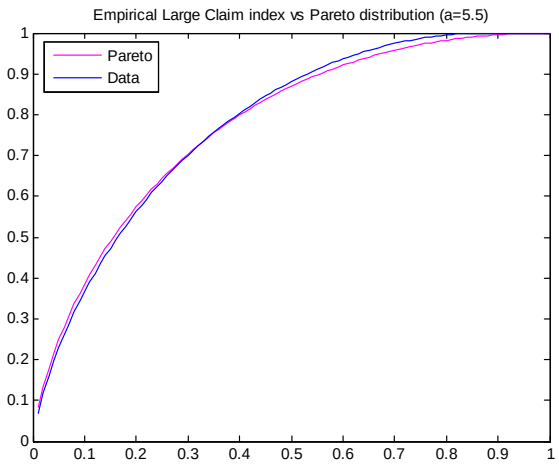


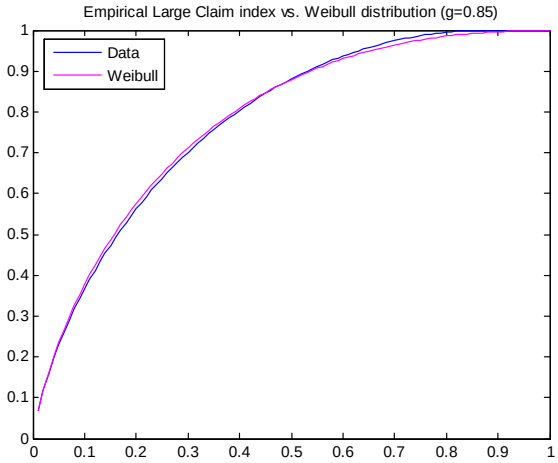
Figure 2.19. Empirical Large Claim Index vs. pareto distribution



we get that $F^{-1}(s) = [-\ln(1-s)]^{\frac{1}{\gamma}}$. For the weibull distribution there is no closed form for $D_F(p)$, so we have to numerically integrate over $F^{-1}(s)$.

From Figures 2.18, 2.19, and 2.20 we can see that according to the large claim index the gas futures returns have a heavier tail than the exponential distribution, but that it is hard to tell the difference between a power tail and an SE tail. In summary, all

Figure 2.20. Empirical Large Claim Index vs. weibull distribution



graphical tools are consistent in the conclusion that gas futures have a fatter tail than the exponential tail. However, the results are mixed about gas futures returns having a power tail or an SE tail. On the one hand QQ-plots indicate that the data has a thinner tail than the pareto distribution, but on the other hand the max/sum ratio indicates that not all moments are finite, which implies that the tail is heavier than an SE tail. The results from the large claim index and mean excess function corroborate that the data is heavier-tailed than the exponential, but both techniques do not help us in discriminating between a power tail and an SE tail.

2.4. Non-parametric results from Extreme Value Theory

A classic, non-parametric method to determine the tail behavior of a random variable is based on results from Extreme Value Theory (EVT). In order to make this text self-contained we introduce the concept and give the main results. We then demonstrate how difficult it is to make any reliable conclusions about the tail behavior for finite samples

using this non-parametric method. We will see, that by construction, EVT cannot help us to discriminate between an SE tail and an exponential tail. In addition, we demonstrate that in finite samples it is is very difficult to even reliably tell the difference between a power tail and an exponential tail. The idea behind the EVT is similar to that of the Central Limit Theorem (CLT). The CLT states that the limiting distribution of the (properly normalised) sum off iid random variables (rvs) coincides with the class of α-stable distributions with $\alpha \in [0, 2]$, to which the normal distribution belongs ($\alpha = 2$). In case the iid rvs are not too heavy-tailed, the limiting distribution for the (properly normalised) sum is the normal distribution. For more heavy-tailed iid rvs (for example a Pareto distribution with tail index < 2), the limiting distribution is the α-stable distribution with $\alpha < 2$. EVT answers the question: what are the limiting distributions of the (properly normalised) maxima of iid rvs? The answer to this question is given by the Fisher-Tippett theorem, which states that the limiting distribution of the (properly normalised) maxima of iid rvs has to be one of the three extreme value distributions, i.e. the Frechet, Weibull, or Gumbel distribution.

Theorem 7. *$M_n = \max\{X_1, X_2, ..., X_n\}$ with $\{X_i\}$ iid rvs. If there exist normalising constants $a_n > 0$, $b_n \in R$, and a non-degenerate distribution function H such that $\frac{M_n - b_n}{a_n} \rightarrow H$, then H must be the generalized extreme value distribution (GEV), defined by:*

$$H_\gamma(x) = \exp[-(1 + \gamma.x)^{-\frac{1}{\gamma}}],$$

with $1 + \gamma.x > 0$. The parameter $\gamma \in R$ is called the extreme value index (EVI). The sign of γ leads to the three possible forms of the GEV distribution.

(1) $\gamma > 0$: the underlying distribution belongs to the maximum domain of attraction (MDA) of the extreme value Frechet distribution, and is regularly varying, i.e. a power tail.

(2) $\gamma < 0$: the underlying distribution belongs to the MDA of the extreme value Weibull distribution which means that with probability one the underlying distribution has a finite right endpoint.

(3) $\gamma = 0$: the underlying distribution belongs to the MDA of the Gumbel distribution, and is rapidly varying, i.e. an exponential tail.

When $\gamma = 0$, $H_\gamma(x) = \exp[-\exp(-x)]$. The use of the term Weibull is confusing here, because the actual Weibull distribution we defined before as $F(x) = 1 - \exp(-x^\gamma)$ for $x \geq 0$ and $\gamma > 0$. If there is any confusion possible we will refer to the weibull distribution in the theorem above as the extreme value weibull. Actually, the weibull distribution belongs to the maximum domain of attraction of the Gumbel distribution. Some examples of distributions that belong to the MDA of the extreme value Frechet distribution are the Frechet, Pareto, Cauchy, Fisher, inverse gamma, and loggamma distributions. The uniform and beta distributions belong to the MDA of the extreme value Weibull distribution. The weibull, exponential, gamma, normal, and lognormal distributions belong to the MDA of the Gumbel distribution. Let's demonstrate with three simple examples.

(1) Take the standard Frechet distribution: $F(x) = \exp(-\frac{1}{x})$. Set $a_n = n$ and $b_n = 0$. As $n \to \infty$:

$$\begin{aligned} \Pr(\frac{M_n - b_n}{a_n} &\leq x) = F^n(n.x) \\ &= [\exp(-\frac{1}{n.x})]^n \\ &= \exp(-\frac{1}{x}) \end{aligned}$$

(2) Take the uniform $U(0,1)$ distribution. Set $a_n = 1/n$ and $b_n = 1$. As $n \to \infty$:

$$\begin{aligned} \Pr(\frac{M_n - b_n}{a_n} &\leq x) = F^n(n^{-1}.x + 1) \\ &= (1 + \frac{x}{n})^n \\ &\to e^x \end{aligned}$$

(3) Take the standard exponential $Exp(1)$ distribution: $F(x) = 1 - \exp(-x)$. Set $a_n = 1$ and $b_n = \log n$. As $n \to \infty$:

$$\begin{aligned} \Pr(\frac{M_n - b_n}{a_n} &\leq x) = F^n(x + \log n) \\ &= [1 - \exp(-x - \log n)]^n \\ &= [1 - \frac{e^{-x}}{n}]^n \\ &\to \exp[-e^{-x}] \end{aligned}$$

We see that in the limit the properly normalised maxima of the standard Frechet distribution converge to the extreme value Frechet distribution, with $\gamma = 1$. The normalised

maxima of the uniform distribution converge to the extreme value Weibull distribution with $\gamma = -1$, and the normalised maxima of the exponential distribution converge to the Gumbel distribution with $\gamma = 0$. There exists a second method to estimate the EVI, the Peaks-over-Threshold method (Beirlant et al., 2004). Let's define the function

$$\Pr\{X - u > x | X > u\} = \overline{F}_u(x)$$

the excess distribution function.

Theorem 8. *The excess distribution function $\overline{F_u}(x)$ with right endpoint x_F belongs to the MDA of H_γ, the GEV, if and only if there exists a positive scale function s(u), such that*

$$\lim_{u \to x_F} \sup_{0 \le x \le x_F - u} \left| \overline{F_u}(x) - \overline{G_u}(x|\gamma, s(u)) \right| = 0$$

with

$$G(x|\gamma, s) = 1 + \log H_\gamma(\frac{x}{s}) = 1 - (1 + \gamma.\frac{x}{s})^{-1/\gamma}.$$

$G(x|\gamma, s)$ is called the generalized Pareto distribution (GPD).

Using the same three examples as above we demonstrate the relationship between the GPD and the GEV approach.

(1) Take the standard Frechet distribution: $F(x) = \exp(-\frac{1}{x})$. As $u \to \infty$

$$\begin{aligned} \Pr\{X - u > x | X > u\} &= \frac{1 - F(x+u)}{1 - F(u)} \\ &= \frac{1 - \exp(-\frac{1}{u+x})}{1 - \exp(\frac{1}{u})} \\ &\to (1 + \frac{x}{u})^{-1} \end{aligned}$$

(2) Take the uniform $U(0,1)$ distribution.

$$\begin{aligned} & \frac{1-F(x+u)}{1-F(u)} \\ = & \frac{1-(u+x)}{1-u} \\ = & 1-\frac{x}{1-u} \end{aligned}$$

(3) Take the standard exponential $Exp(1)$ distribution: $F(x) = 1 - \exp(-x)$.

$$\begin{aligned} & \frac{1-F(x+u)}{1-F(u)} \\ = & \frac{\exp[-(x+u)]}{\exp[-u]} \\ = & \exp[-x] \end{aligned}$$

We see that the standard Frechet distribution corresponds to the GPD with $\gamma = 1$ and $s(u) = u$. The uniform distribution corresponds to the GPD with $\gamma = -1$ and $s(u) = 1 - u$, and in the case of the exponential distribution we get $\gamma = 0$ and $s(u) = 1$. We obtain the same extreme value index as in the GEV case.

From the graphical tools we have learned that the tail behavior for gas futures returns is in between a power and an exponential tail. It is clear that EVT cannot help us discriminate between an SE tail and an exponential tail (weibull and exponential distribution both belong to the MDA of the Gumbel distribution). But in theory the EVT should be able to tell us whether the distribution of gas futures returns belongs to the MDA of the extreme value Frechet distribution (and therefore has a power-like tail) or whether it belongs to the MDA of the Gumbel distribution (and therefore has an exponential tail).

Before we estimate the EVI of gas futures returns, we first estimate the EVI based on 40,000 simulated values from both a pareto distribution and an exponential distribution. In what follows we choose to estimate the EVI using the Peaks-over-Threhold method. This can be done using Maximum Likelihood, based on GPD, or using Pickands' (1975) estimator. We will use Pickands' estimator, which is given by

$$\gamma_{k,N} = \frac{1}{\log 2}.\log\left(\frac{X_{N-k/4+1,N} - X_{N-k/2+1,N}}{X_{N-k/2+1,N} - X_{N-k+1,N}}\right)$$

where $X_{1,N} \leq X_{2,N} \leq ... \leq X_{N,N.}$ We now estimate the EVI for simulated Weibull values $F(x) = 1-\exp(-x^c)$ with c = 0.2 / 0.4 / 0.8 / 1. Given that we use the Peaks-over-Threshold method we do this for X > 1, X > 3, and X > 5. We make sure that we always have 40,000 simulated values in excess of the threshold. We set $k = N/2$, but changing k does not alter the results by much. We know from the EVT that the Weibull and the exponential distributions belong to the MDA of the Gumbel distribution, and that as a result the theoretical value of the EVI equals 0. In Figure 2.21 we can see that for low c values the convergence towards zero for the EVI is slow. If the underlying distribution of an actual dataset (with a realistic number of data points) is the Weibull with a low tail index, then using this method we will erroneously conclude that it originates from a power-like distribution. However, for simulated values from the exponential distribution (c = 1), we obtain the theoretical value of the EVI immediately.

We will now do the same for simulated values from the Pareto distribution $F(x) = 1-(\frac{2}{x})^b$. We know that the Pareto distribution belongs to the MDA of the extreme value Frechet distribution with $\gamma > 0$. We can see in Figure 2.22 that the EVI is always bigger than 0.

Figure 2.21. EVI estimates using Pickands' estimator for simulated values from the Weibull distribution

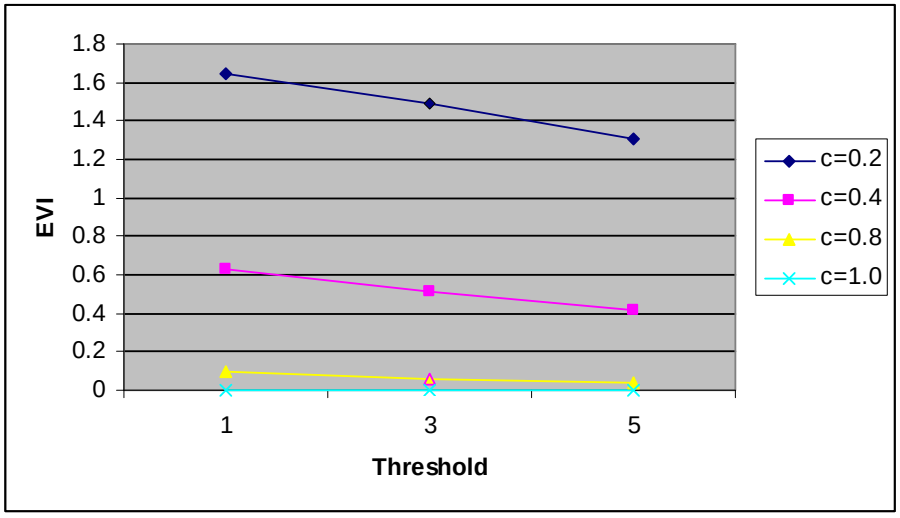
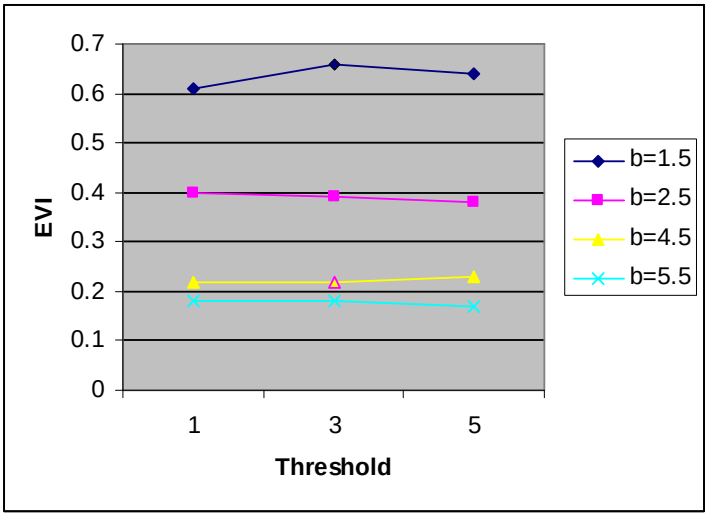
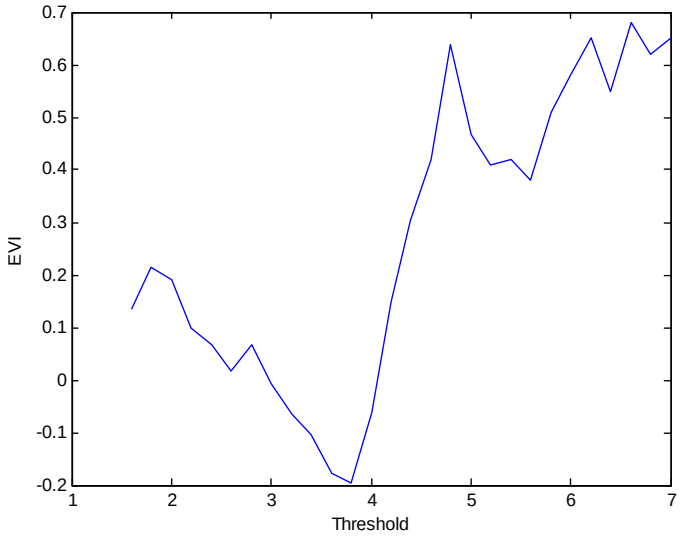


Figure 2.22. EVI estimates using Pickands' estimator for simulated values from the Pareto distribution



From these simple tests we conclude that if we obtain an estimated EVI close to zero for our natural gas futures returns, we can be reasonably confident that the underlying distribution of the data has an exponential-like tail. However, if the estimated EVI is

Figure 2.23. EVI estimates using Pickands' estimator for natural gas futures returns data



clearly bigger than zero, we can reasonably reject the notion that the underlying distribution is the exponential distribution or a weibull distribution with a high tail index, but we cannot say much about whether the underlying distribution belongs to the MDA of the extreme value Frechet or the Gumbel distribution. From Figure 2.23 we can see that up to a threshold of 4, the estimate of the EVI is approximately zero, which implies that the underlying gas futures data for the bulk of the distribution belongs to the MDA of the Gumbel distribution, i.e. the gas futures data has a rapidly varying or exponential-like behavior. For daily gas returns exceeding 5% the EVI estimate jumps up substantially, which can lead us to believe that deep in the tail the gas data belongs to the MDA of the extreme value Frechet distribution. However, the results from Figure 2.21 tell us that we have to be careful making definitive conclusions when the estimated EVI is bigger than zero.

2.5. Parametric fitting

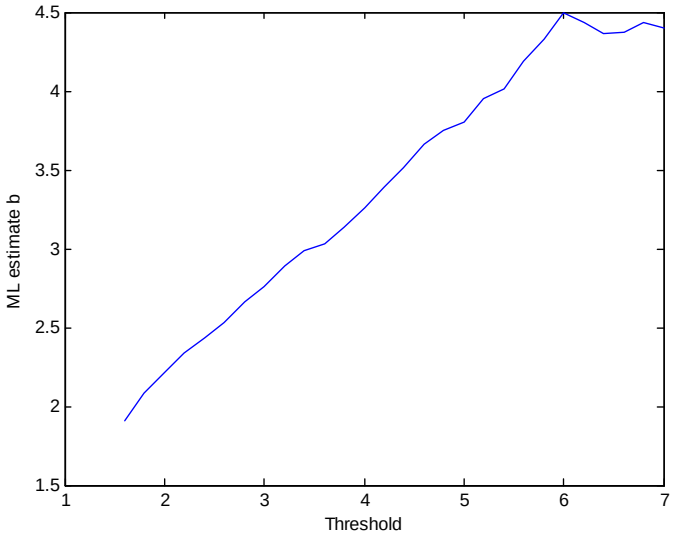
Both the graphical tools and the results from EVT point to similar conclusions: 1) the underlying distribution of the natural gas futures returns has a tail that decays more slowly than the exponential distrubution, and 2) it is difficult to conclude with sufficient certainty whether the data behaves more like a power distribution or whether the tail decays faster like a weibull distribution. The results from EVT also show us that the underlying behavior might be different for the bulk of the data and for the data deep in the upper tail. Again, the reason we care is because it has important implications for VaR and Expected Shortfall analysis at high probability levels and for deep out-of-the-money option pricing. If we want to obtain more information about the tail behavior of gas futures returns we should also investigate a parametric approach, given the limited dataset. Of course, in addition to sampling error we now add misspecification risk. In the next section we will introduce the log-weibull distribution which for certain parameters either becomes the weibull or the pareto distribution, and this will hopefully allow us to have more of an inclination towards one or the other. First, we will estimate using ML the parameters of the pareto, weibull, and log-weibull distributions for different threshold levels u. Afterwards we will do a pairwise comparison of the models using Wilks' likelihood ratio test.

Pareto distribution. $F_u(x) = 1 - \left(\frac{u}{x}\right)^b$ with $x \geq u$.

It is easy to show that when there are N_u observations, the ML estimator yields

$$\widehat{b} = \left[\frac{1}{N_u}\sum_i \log(\frac{x_i}{u})\right]^{-1}.$$

Figure 2.24. ML estimate of b for daily gas futures returns as a function of the threshold u



Weibull distribution. $F_u(x) = 1 - \exp[-\left(\frac{x}{d}\right)^c + \left(\frac{u}{d}\right)^c]$ with $x \geq u$.

One can show that the ML estimator for c is the solution to the following equation

$$\frac{1}{c} = \frac{\frac{1}{N_u}\sum_i \left(\frac{x_i}{u}\right)^c \log\frac{x_i}{u}}{\frac{1}{N_u}\sum_i \left(\frac{x_i}{u}\right)^c - 1} - \frac{1}{N_u}\sum_i \log\frac{x_i}{u}$$

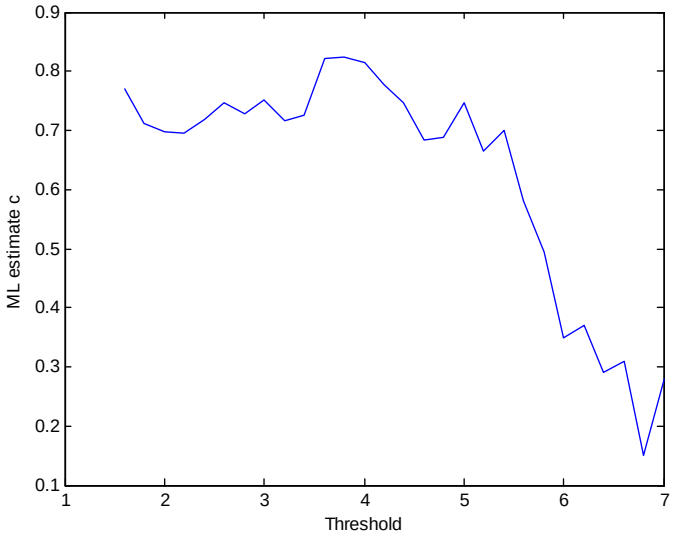
In addition, one can show after some calculations that in order to have a solution with $c > 0$, the following condition is both necessary and sufficient

$$2.\left(\frac{1}{N_u}\sum_i \log\frac{x_i}{u}\right)^2 > \frac{1}{N_u}\sum_i [\log\frac{x_i}{u}]^2.$$

A finite sample may not always satisfy this condition.

Looking at both Figure 2.24 and 2.25 we can observe that something is going on when the threshold is around 6. If the data was truly generated by a Pareto distribution the tail coefficient b would be constant for all u. However, the coefficient is increasing until $u = 6$,

Figure 2.25. ML estimate of c for daily gas futures returns as a function of the threshold u



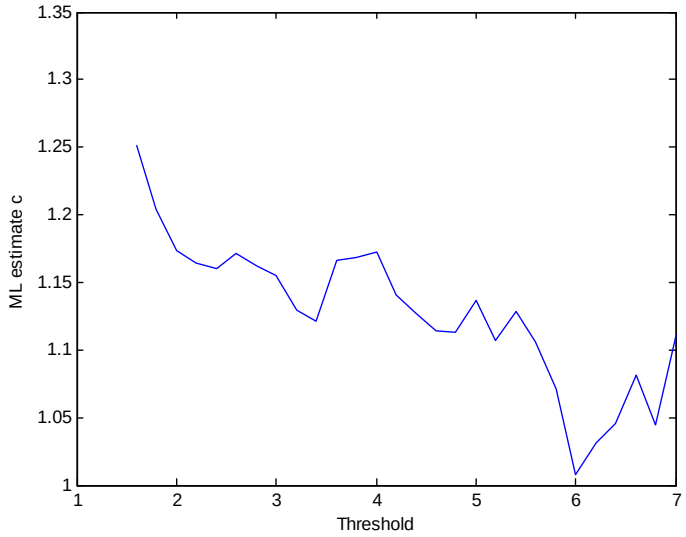
beyond which it seems to settle down. In Figure 2.25 we see the opposite phenomenon. The coefficient c is constant until $u = 5.5$, beyond which it falls. The combination of both Figure 2.24 and 2.20 seems to tell us that the weibull distribution describes well the bulk of the data (98% of the data), but that beyond a high threshold level, i.e. daily returns exceeding 6%, the data behaves more like a regularly varying or power-like tail.

Log-weibull distribution. $F_u(x) = 1 - \exp[-b.\{\log(\frac{x}{u})\}^c]$ with $x \geq u$.

For $c = 1$, it recovers the Pareto distribution, with b the tail coefficient. For $c = 2$, it becomes the lognormal distribution. When $c > 1$ the log-weibull has a thinner tail than the Pareto, but a heavier tail than the SE family, to which the weibull belongs. The ML estimator for c is the solution to the following equation

$$\frac{1}{c} = \frac{\frac{1}{N_u}\sum_i \left(\log \frac{x_i}{u}\right)^c \log[\log \frac{x_i}{u}]}{\frac{1}{N_u}\sum_i \left(\log \frac{x_i}{u}\right)^c} - \frac{1}{N_u}\sum_i \log[\log \frac{x_i}{u}]$$

Figure 2.26. ML estimate of c (log-weibull) for daily gas futures returns as a function of the threshold u



and b is given by

$$b = \frac{1}{N_u} \sum_i \left(\log \frac{x_i}{u} \right)^c$$

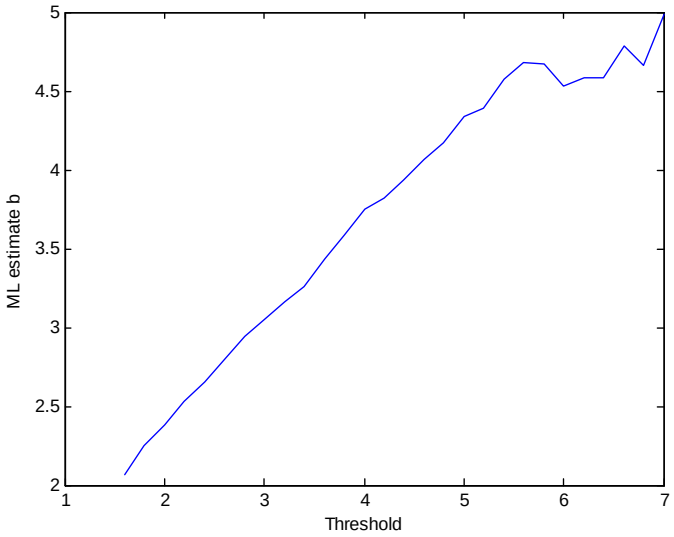
Figure 2.26 corroborates what we have learned from previous tests. Up to the threshold $u = 5.5$ the tail is definitely thinner than a power tail, but beyond this threshold the tail behavior of Henry Hub gas futures returns is well approximated by the pareto distribution.

Direct Comparison of Parametric Distributions. As a final test we will directly compare the SE distribution with the exponential and the pareto distribution. Wilks' theorem states that the log-likelihood ratio LR:

$$LR = 2. \log \frac{\max_\Delta L(H_0, X, \Delta)}{\max_\delta L(H_1, X, \delta)}$$

asymptotically has a χ^2-distribution. $L(.)$ is the log-likelihood function, Δ & δ are the parameter spaces corresponding to the hypotheses H_0 & H_1. The theorem works if

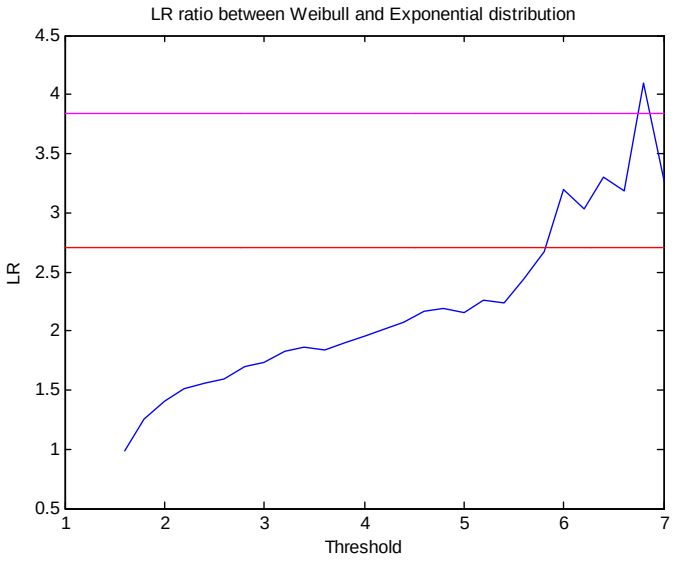
Figure 2.27. ML estimate of b (log-weibull) for daily gas futures returns as a function of the threshold u



H_1 is true for particular parameters in the Δ-parameter space. The number of degrees of freedom is given by the difference in dimension between the Δ-space and the δ-space. We know that LR is always positive, because $\delta \subset \Delta$. If LR does not exceed some high threshold level, either the 90th- or the 95th percentile of the χ^2-distribution, we favor the parsimonious version H_1.

Weibull vs. Exponential. We can use Wilks' theorem directly to test the weibull distribution against the exponential distribution, because setting the tail index parameter c, from the weibull distribution, equal to 1, one obtains the exponential distribution. Figure 2.28 plots the LR ratio between the weibull and the exponential distribution as a function of the threshold level. It also plots the 90th- and the 95th percentile of the $\chi^2(1)$-distribution. We can see that only at high threshold levels, i.e. daily gas returns beyond 5.5%, we can start to reject the exponential in favor of the weibull distribution.

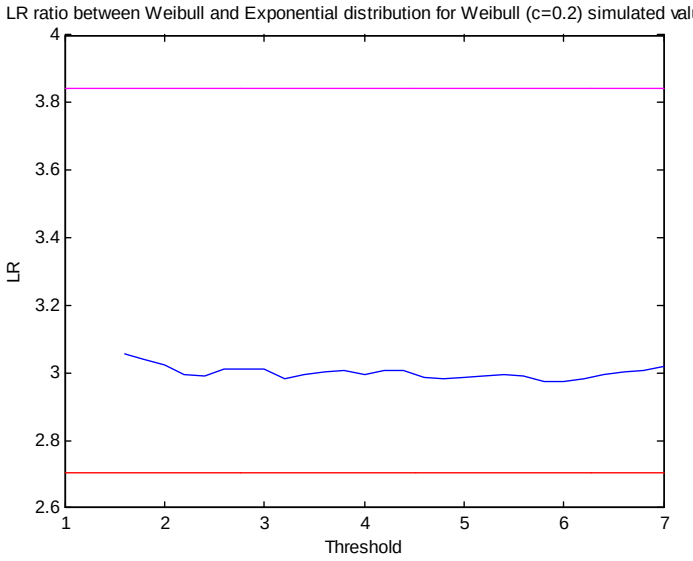
Figure 2.28. Log-likelihood ratio between the weibull and the exponential distribution for daily natural gas returns



To put this in perspective we simulate around 10,000 values from a weibull distribution with tail index parameter c equal to 0.2. We then compute again the LR ratio between the weibull and the exponential distribution based on these weibull generated values. As we can see in Figure 2.29, even in the case of 10,000 weibull generated values with c = 0.2, i.e. clearly a heavier tail than the exponential distribution, we cannot reject the exponential model in favor the weibull model at the 95th percentile using Wilks' theorem.

Weibull vs. Pareto. We cannot simply use Wilks' theorem to test the weibull distribution (2-parameters) against the pareto distribution (1-parameter), because they are not exactly nested. There exists no subset of parameters for the weibull distribution that recovers the pareto distribution. There is, however, a very close link between the two distributions, which will allow us to still use Wilks' theorem. Recall that the weibull distribution is given

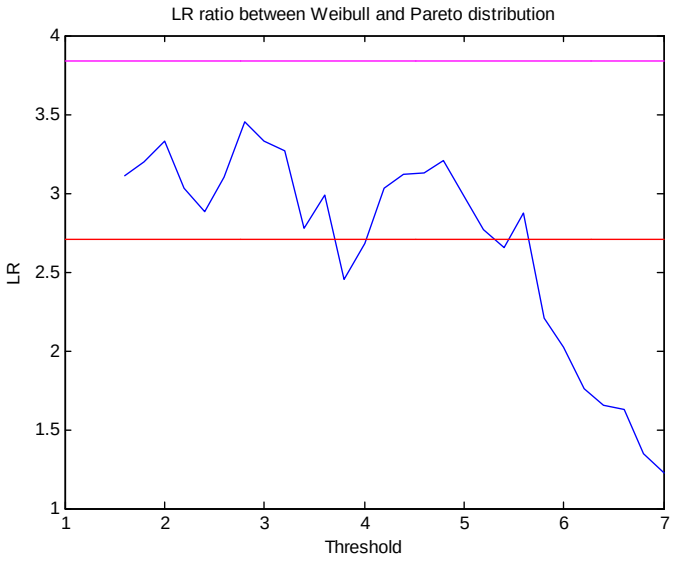
Figure 2.29. Log-likelihood ratio between the weibull and the exponential distribution for weibull (c = 0.2) simulated values



by $F_u(x) = 1-\exp[-\left(\frac{x}{d}\right)^c+\left(\frac{u}{d}\right)^c]$, and the pareto distribution is given by $F_u(x) = 1-\left(\frac{u}{x}\right)^b$. As a result, the pareto density is described by $b.\frac{u^b}{x^{b+1}}$, and the weibull density is described by $\exp[-\left(\frac{x}{d}\right)^c + \left(\frac{u}{d}\right)^c].c.(\frac{x}{d})^{c-1}.d^{-1}$. One can show that as $c \to 0$, the weibull density can be approximated by the pareto density provided that $c.(\frac{u}{d})^c \to k > 0$ as $c \to 0$. Therefore, technically, the weibull and the pareto distribution are not nested. One does not recover the pareto distribution simply by setting $c = 0$ from the weibull distribution. However, Malevergne and Sornette (2006) have shown that in this particular case Wilks' theorem can still be applied.

We can see from Figure 2.30 that for the bulk of the distribution the weibull is preferred over the pareto, but that deep in the tail we reject the weibull in favor of the pareto distribution. As a final test we compare the pareto with the log-weibull distribution. The

Figure 2.30. Log-likelihood ratio between the weibull and the pareto distribution for daily natural gas returns



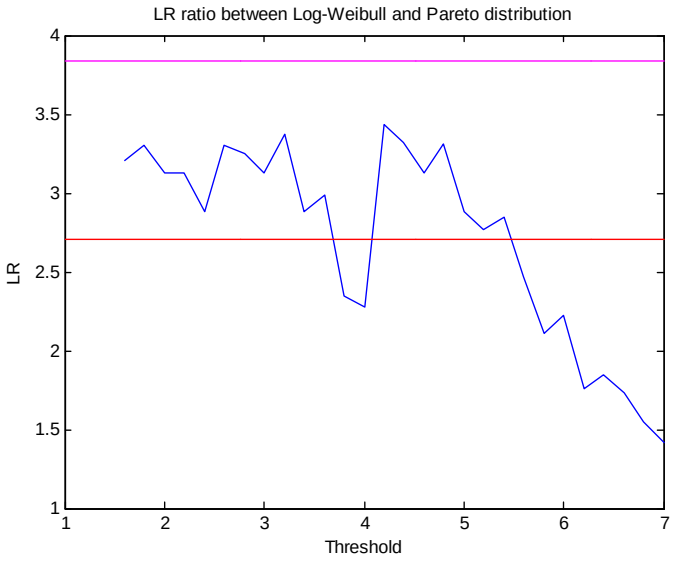
log-weibull is described by $F_u(x) = 1 - \exp[-b.\{\log(\frac{x}{u})\}^c]$, and recall that for $c = 1$, we recover the pareto distribution with tail coefficient b. Therefore, Wilks' theorem can be applied directly as the pareto and log-weibull are nested.

Figure 2.31 not only parallels Figure 2.30, but also Figure 2.26, where we estimated the c coefficient from the log-weibull distribution. In Figure 2.26 we can see that the estimated tail coefficient c gets close to 1 at high threshold levels, where $c = 1$ corresponds to the pareto distribution. This is also reflected in Figure 2.31, where at high threshold levels we can no longer reject the pareto in favor of the log-weibull distribution.

2.6. Conclusion

In this chapter we have investigated the distribution of daily Henry Hub natural gas futures returns, and more specifically its tail behavior. Having the correct model for tail behavior has important implications for Value-at-Risk measures at high probability levels

Figure 2.31. Log-likelihood ratio between the log-weibull and the pareto distribution for daily natural gas returns



and for deep out-of-the-money option pricing. Real datasets have a limited number of datapoints, whereas most existing theorems are only true asymptotically. It is, therefore, important to have a sense of how well a theorem performs in the presence of a finite sample. Moreover, because we work with a real dataset, and most theorems hold true as the number of datapoints approaches infinity, it is important to look at the problem from many different vantage points with the hope of seeing a recurring pattern emerge. This is exactly what we have done in this chapter. First we looked at some preliminary graphical tools (some of them based on theorems that hold true asymptotically) such as QQ-plots, max/sum ratios, mean excess functions, and the large claim index. Then we investigated a non-parametric approach based on Extreme Value Theory, and finally we delved deeper into a few promising parametric distributions (exponential, weibull, log-weibull, and pareto distributions), and we compared them pair-wise based on Wilks'

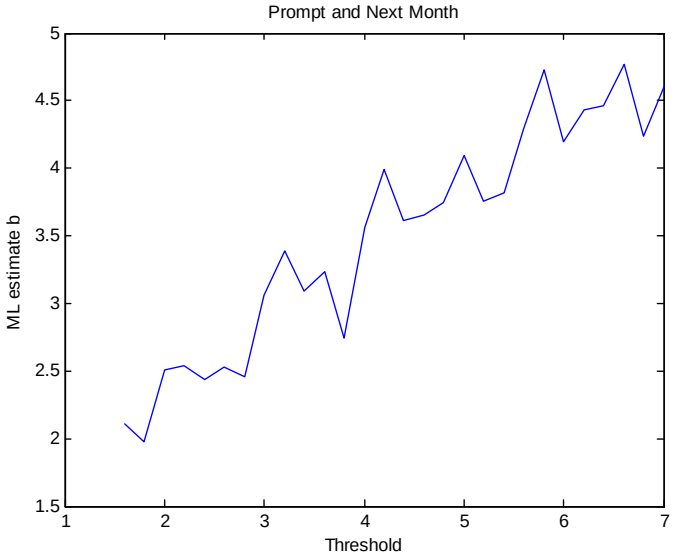
log-likelihood ratio test. As a result of this analysis we can make the following statements. All graphical tools are consistent in the conclusion that daily gas futures returns have a fatter tail than the exponential tail. However, the results are mixed about gas futures returns having a power tail or an SE tail. On the one hand QQ-plots indicate that the data has a thinner tail than the pareto distribution, but on the other hand the max/sum ratio indicates that not all moments are finite, which implies that the tail is heavier than an SE tail. The results from the large claim index and mean excess function corroborate that the data is heavier-tailed than the exponential, but both techniques do not help us in discriminating between a power tail and an SE tail. In the section on Extreme Value Theory, we plot the results as function of the threshold level in order to check whether all data is generated by the same underlying distribution or whether two or more different distributions can better explain different parts of the gas data distribution. We can see that up to a threshold of 4, the estimate of the EVI (based on Pickands' estimator) is approximately zero, which implies that the underlying gas futures data for the bulk of the distribution belongs to the MDA of the Gumbel distribution, i.e. the gas futures data has a rapidly varying or exponential-like behavior. For daily gas returns exceeding 5% the EVI estimate jumps up substantially, which can lead us to believe that deep in the tail the gas data belongs to the MDA of the extreme value Frechet distribution, i.e. the gas data has a heavier or power-like tail. However, the results from Figure 2.21 tell us that we have to be careful making definitive conclusions when the estimated EVI is bigger than zero. In the third section on parametric distributions we find confirmation of the result that the gas data is best explained by two different distributions. When we plot the tail coefficient of the pareto distribution as a function of the threshold level, we see

that the tail coefficient settles down at a high threshold level. This is in exact parallel with what we observe when we plot the tail coefficient of the weibull as a function of the threshold. In the latter case the coefficient is relatively constant for the bulk of the distribution, but falls dramatically at high threshold levels. And a similar result can be seen when we plot the results of the log-weibull. These results are corroborated when we plot the Wilks' log-likelihood ratio based on direct pair-wise comparison between these distributions. When we compare the weibull with the exponential we see that at high threshold levels the exponential is rejected, and when we compare the weibull with the pareto and the log-weibull with the pareto, the pareto is rejected at low and intermediate threshold levels, but is favored at high threshold levels. Therefore, almost all of these tests point to the same conclusion: for the bulk of the distribution, up to daily gas futures returns of 5-6%, which make up around 98% of the data, the weibull distribution with tail coefficient $c = 0.7 - 0.8$ or the log-weibull with $c = 1.10 - 1.15$ explain the data well. At high threshold levels, i.e. daily gas futures returns exceeding 5.5-6%, the data is better described with an underlying distribution that has a power-like tail, such as the pareto distribution with tail coefficient $b = 4.5$.

2.7. Sensitivity Analysis

Up until now we have investigated the tail behavior of daily gas returns for futures contracts of all maturities combined. Most of the price movements, however, are experienced for futures contracts with short maturities (prompt month and next month). It is, therefore, instructive to split up the data sample in two groups: the first group contains the daily returns for futures contracts with maturity up to two months and the second

Figure 2.32. ML estimate of b for prompt and next month daily gas futures returns as a function of the threshold u



group contains all the longer dated contracts. It is important to split up the data because the underlying model used for the pricing of a deep out-of-the-money option on a prompt or next month versus a further dated month will be different. We do not want to split up the sample further because of lack of data deep in the tails. We will focus on the parametric fitting for both groups and determine how different the tail coefficients are.

Pareto. We start by comparing the tail coefficient, b, of the pareto distribution for the prompt and next month (Figure 2.32) returns versus the later contracts (Figure 2.33).

We notice that Figure 2.32 resembles the original Figure 2.24 very well, which means that the prompt and next month returns dominate what happens deep in the tails. Figure 2.33 is not totally different, but the tail coefficient is higher overall, and it is not that clear that the tail coefficient stabilizes at high thresholds.

Figure 2.33. ML estimate of b for future months daily gas futures returns as a function of the threshold u

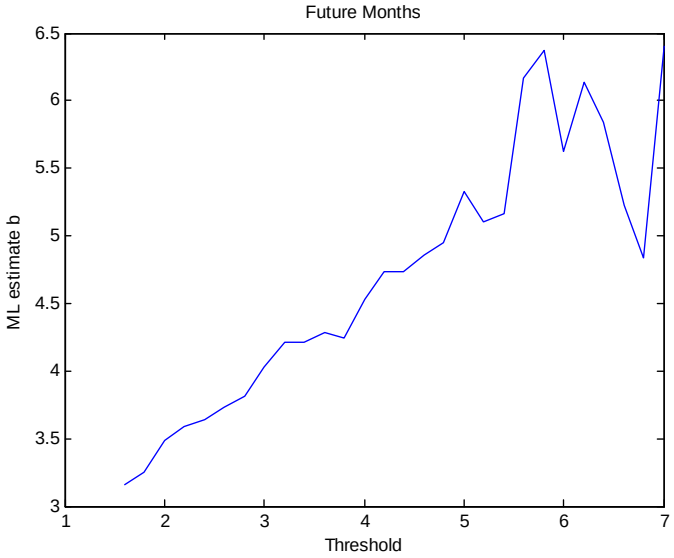
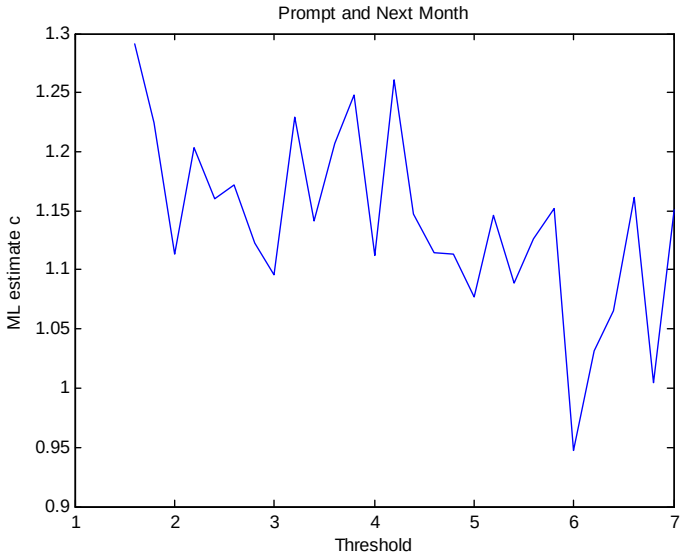
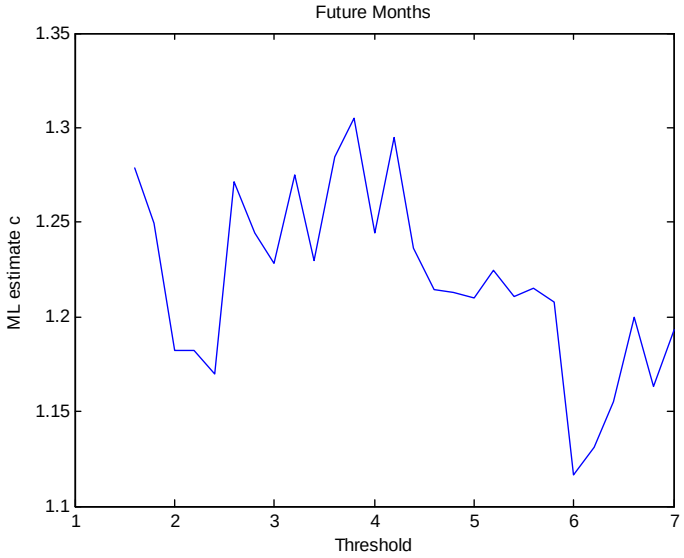


Figure 2.34. ML estimate of c (log-weibull) for prompt and next month daily gas futures returns as a function of the threshold u



Log-weibull. As the weibull and the log-weibull coefficients follow a similar pattern we will just compare the 'c' coefficent of the log-weibull.

Figure 2.35. ML estimate of c (log-weibull) for future months daily gas futures returns as a function of the threshold u



We observe that Figure 2.34 has a similar pattern as Figure 2.26, just as we observed for the pareto coefficient. And we observe here as well that Figure 2.35 is shifted upwards compared to the original Figure 2.26, and that the shape has also changed to some extent. Given that the prompt and next month contracts determine the tail coefficients for the entire dataset, it is not surprising that the pairwise comparison tests are very similar to the ones described for the entire dataset. The results for the future month contracts are slightly different. The exponential tail is mostly rejected, but we cannot reject the weibull/log-weibull in favor of the more parsimonious pareto distribution. These results from the pairwise comparison for the future months are a reflection of what we observe in Figures 2.33 and 2.35.

CHAPTER 3

Tail Behavior of Power Forwards Returns

3.1. Introduction

3.1.1. Independent System Operators (ISOs)

The power market in the US is organized through regional Independent System Operators (ISOs), who are responsible for the dispatch and reliability of power in their region. Examples of ISOs are PJM (biggest ISO, which comprises pieces of 13 states - going from New Jersey to Virginia to Kentucky and Illinois), NEPOOL (New England), NYISO (New York), CAISO (California), MISO (Midwest), and ERCOT (Texas). For the remainder of this section I will go into more detail of the inner workings of the power market, and more specifically of PJM. I choose PJM because it is the biggest and most established ISO, and also because most of the other ISOs seem to converge towards the PJM model. No two ISOs have exactly the same rules, but if you understand one ISO well, it will help you understand the inner workings of the other ISOs. Currently, there exist two models to determine the spot price of power at each location within the ISO. Some ISOs use a zonal model (ERCOT & CAISO). The other ISOs that I mentioned use a nodal model. PJM was one of the first ISOs to use the nodal model and the other ISOs seem to follow suit. In fact, both ERCOT and CAISO are near the transition towards a nodal model. A nodal model computes a power spot price for every node (=busbar) within its region. PJM, for example, has about 8000 different nodes, and for every single node a specific spot price

is calculated. The spot price of power in PJM is called the Locational Marginal Price (LMP). With each node is associated a specific LMP. For the purpose of trading, nodes are grouped together into Hubs. The Hub LMP is then simply the arithmetic average of the LMPs of the nodes that make up the Hub. WestHub is an example of the most liquidly traded Hub in PJM, or even in the entire US, and it consists of 110 individual nodes located mainly in the western part of PJM (western Pennsylvania), but also in the Washington DC area.

3.1.2. LMP calculation

Now we get to the question: how is an LMP determined? Before we answer this question we have to know that there exist two types of LMP at every node, i.e. a Day-Ahead LMP (DA-LMP) and a Real-Time LMP (RT-LMP). For each node there exists a DA-LMP for every hour of the day. It is impossible to explain in exact detail how this hourly DA-LMP is computed, but I will try to explain the basic idea. The calculation of an LMP, whether it is DA or RT, is the result of an optimization algorithm performed by the ISO. The ISO tries to minimize the total cost of serving the power demand (load) subject to a whole range of constraints. Even though it is an Alternating Current (AC) system, all the constraints are linearized, so the ISO is constantly solving a massive LP problem. For the Day-Ahead market it roughly works as follows. All generators and load servers that want to participate in the DA market have to submit their generation offers and load bids to the ISO by 12pm. Generation offers are supply schedules (we are willing to generate 50 MW @ \$60/MWh, 60 MW @ \$67/MWh, etc., up to ten increments are allowed with an offer cap of \$1000/MWh). In addition, a start-up cost offer (cold, warm, and hot) and a

no-load cost offer[1] are part of the offer schedule. Finally, a unit needs to provide the ISO with its ramp rate (both up and down), which is expressed in MW/min, minimum run time, and its maximum and minimum capacity levels. Load bids also consist of a demand schedule. The ISO receives these bid and offer schedules for specific locations (nodes) by 12pm and solves the LP problem, the result of which are 24 DA-LMPs (one for each hour of the following day) for every node in the PJM system. Generators that are picked up to run the next day for a certain numbers of hours and load servers that are picked up will receive/pay the DA-LMP during those hours for the quantity that they are picked up for. The DA market allows participants to lock in a price for the next day. When the next day rolls around, participants are expected to execute the DA results and this is enforced with monetary penalties. Generators and load servers also submit to the ISO Real-time schedules. If the ISO notices that not enough generation is online to satisfy the RT load, more units will have to be scheduled (or the same units but more MW) and those will be paid the RT LMP, and the same holds true for load servers. For example, a load server Comed submits in the DA market a bid of 5000 MW for a specific location for 7am. Comed bid in the load @ ∞, meaning that it is willing to serve the load at any price, and therefore it is picked up for all 5000 MW for 7am, and the DA-LMP for 7am turns out to be \$100/MWh. The RT load at 7am that Comed is supposed to serve turns out to be 5200 MW and the RT-LMP for 7am for that location is \$160/MWh. Comed will have to pay to PJM for that hour 5000x100 + 200x160 = \$532,000. In reality a penalty will be added for the 200 MW deviation of the RT load from the DA load, which is called a RT

[1]A start-up cost is expressed in \$ and it is the amount the generator needs to be paid in order to start up the unit. Cold, warm, and hot start-up costs refer to how long the unit has been idle before it is being asked to start running. A no-load cost is expressed in \$/MWh and is the amount that a unit needs to be paid when it is asked to keep running off-line.

balancing charge. So, every 5 minutes the ISO solves the LP problem and computes 5-min RT-LMPs for every node based on the generation schedules and PJM's RT load forecasts. These 5-min LMPs are averaged into hourly RT-LMPs, and these hourly RT-LMPs are used for settlements and are published. In solving the LP problem an important set of constraints are reliability and transmission constraints. Reliability requirements are set by NERC (North American Electric Reliability Council), and are set in such a way that the system can withstand contingencies (a unit that trips, sudden transmission outage, etc.) One of the rules is that every ISO needs to maintain 10-minute reserves, i.e. generation that is ready to provide power to the system within 10 minutes, of an amount equal to the first contingency, and 30-minute reserves equal to 50% of the second contingency. The first contingency is usually the biggest unit that is currently running, probably a nuclear unit of 1500 MW. We will not go into details but PJM also has a market based solution for providing the necessary level of reserves and other reliability requirements. These markets are called Ancillary Services markets. This adds to the complexity of the LP problem that is being solved for, because the Energy market and the Ancillary Services markets are co-optimized and solved for simultaneously. The other set of constraints are the transmission constraints. Transmission constraints are very important and can have a very substantial impact on LMPs (both DA and RT). The LMP is the sum of three components:

$$LMP_h = SEP_h + MCC_h + MLC_h$$

with SEP: System Energy Price

MCC: Marginal Congestion Component

MLC: Marginal Loss Component

The SEP is the same for every node. Technically, it is the LMP of the reference bus or slack bus. Every node is priced relative to the reference bus, which is just a computational trick. If we forget about the MLC for a moment, if in solving for the optimal solution none of the transmission constraints are binding, the MCC would be zero for every node, and the LMP would be exactly the same throughout the system. The MLC is the answer the following question: if we inject 1 extra MW at node i, and we withdraw it at the reference bus, how much does this cost in terms of losses. The longer power travels, the more power we dissipate. If power generated at node i has to travel long distances to reach the load, node i will penalized with a very negative MLC, thereby lowering the LMP at node i and discouraging more generation at node i. Most of the generation in PJM is located in the western part (coal units in the Appalachians), whereas most of the load is located in the east. Therefore, the MLC of the nodes in the west are negative and positive in the east. The same holds true in general for the MCC: most of the congestion in PJM is the result of too much power flowing west to east causing transmission constraints to bind. If a transmission constraint binds, the optimal solution is no longer feasible. This situation has to be resolved through re-dispatch. Some units will have to back down and other units will have to ramp up in order to reduce the power flow along the transmission line that was binding. Re-dispatch has to increase the cost of serving the load, because this solution is suboptimal. Nodes that contribute to congestion problems will be penalized by negative MCCs, nodes that help relieve the congestion problems will be encouraged to generate more by making the MCC positive. The hope is also to send a signal to the market to invest in new generation at nodes with consistently higher prices. PJM publishes not only

a schedule of transmission line outages, but also of other elements, such as transformers, circuit breakers, capacitors, etc. Outages or derates of these elements can significantly reduce the amount of power that can flow on certain lines, thereby causing re-dispatch, which can significantly increase the LMP at certain nodes and decrease the LMP at other nodes. This also impacts Hub prices, recall that WestHub is made up of only 110 nodes. If 20 of these nodes that are located close to each other blow up, WestHub will blow up.

3.1.3. Virtual bidding (INCs & DECs)

PJM allows for Virtual bidding through Decrement bids (DECs) and Increment Offers (INCs). The goal is to make sure that the average of the DA-LMPs is close to the average of the RT-LMPs most of the time. This is realised by allowing market participants to speculate on the spread between the DA-LMP and the RT-LMP at all individual nodes. If a trader thinks that the DA-LMP at node i for hour j will be higher than the RT-LMP, he will enter an INC in the DA market for node i and hour j. An INC is seen by PJM as a virtual generation offer. The trader will receive the DA-LMP (just as any other generation unit that is picked up). But in RT he will not deliver the power, which means that he will be charged the RT-LMP. We also know that he will have to pay a RT balancing charge for the full MW amount, which in PJM averages around $2/MWh. So, in effect, the pay-off to the trader equals (DA-LMP minus RT-LMP minus 2) x MW amount. If he thinks the RT-LMP will be higher than the DA-LMP, he will enter a DEC, which is seen as a virtual load bid. The hope is that on average the RT-LMP stays withing $2 of the DA-LMP, and this is realised for most of the nodes most of the time. Of course, for most hours, the

RT-LMP deviates substantially from the DA-LMP, but it is not systematically higher or lower.

3.1.4. Financial Transmission Rights (FTRs)

PJM also organizes an FTR market (once a month), which allows market players to participate in an FTR auction. The products offered in the auction are monthly, quarterly, and yearly FTR obligations and FTR options. FTR obligations pay the spread between the MCC of the sink node of the FTR and the source node of the FTR, which can be both positive and negative. FTR options are the same thing, but the pay-off can never be negative. As a result an FTR option is always at least as expensive as an FTR obligation on the same path. We now introduce the concept of peak and off-peak power. Power trading in general and also the FTR markets in particular are done for separate time buckets: 5x16, 2x16H, 7x8, wrap (= 2x16H + 7x8), and 7x24. For example, when one buys a Sep08 5x16 WestHub RT financial forward contract for $100, it means that for the weekdays (Mon-Fri) in Sep08 (except holidays) during the 16 peak hours from 7am to 10pm one pays $100 per MW and receives the RT-LMP of WestHub for those same hours. 2x16H stands for the 16 peak hours of the weekend days plus NERC holidays. 7x8 stands for the 8 off-peak hours (midnight till 6am +11pm) for all days. FTRs options and obligations are offered for a month, for a quarter, and for a whole year as 5x16, wrap, and 7x24. The purpose of FTRs is mainly twofold: 1) it allows market participants to lock in a congestion premium between two nodes: a load server that is constantly buying power at node A and delivering this power to its customers at node B does not want to be at the mercy of sudden transmission outages that blow up the spread between node A and

B. 2) it allows traders to take speculative positions on spreads between individual nodes. Individual nodes are not traded at all, so that the only way to gain exposure to them is through FTRs. Participants submit bids for FTRs. PJM then runs an LP problem that maximizes the revenue it receives from the auction subject to a feasibility constraint. The first letter in FTR is important, it is a purely financial contract. Owning an FTR from node A to B does not grant any right for physical power from A to B. Suppose 50 MW of physical power flows from A (generation node) to B (load node). Congestion takes place and $MCC_B - MCC_A = 20$. PJM will receive congestion revenues equal to 50 x 20 = \$1000. If PJM awarded 100 MW of FTRs from A to B, it would have to pay out 100 x 20 = \$2000. This simple example illustrates the feasibility constraint. PJM wants to maximize the revenue it can receive from the FTR auction, but it has to compute whether it will be able to pay out the winning FTR bids at all times. PJM forecasts how much power will flow on specific lines and it will use this estimate to cap the FTR quantity that it awards in the auction.

3.1.5. Capacity Market (Reliability Pricing Model - RPM)

A capacity market has existed for a while in PJM, but recently it has been transformed into RPM. The goal of RPM is to induce new generation investments. A capacity market is needed because load can peak dramatically during very hot summer days. Peaking generators are needed to cover this load, but they are needed only a few hours per year. Because offer prices are capped at \$1000/MWh, and as result LMPs are capped, these peaking units cannot recover their fixed costs over time by simply running a few hours per year in the summer. Still, they are needed for reliability reasons. There are two possible

solutions. On the one hand, one can remove the offer cap, and LMPs would reach high enough values such that peaking units can recover their fixed costs over time. On the other hand, one can provide generation units extra money simply for being available to run during the summer period. The latter solution is exactly what a capacity market is doing. It provides generation units extra money simply for being available during the summer. Inefficient generation units, i.e. high heat rate units, are out-of-the-money in the energy market the vast majority of the time. For these units the capacity market revenues make up most of their revenues. Power plants with low short-run marginal costs, for example nuclear units, are in-the-money in the energy market practically 100% of the time, and capacity market revenues make up only a small fraction of the entire revenue stream. In the long run the sum of energy market and capacity market revenues should be fairly constant. An increase in capacity prices induces new generation investments or delays mothballing/retiring older plants and should therefore decrease energy market revenues down the road. The RPM market in PJM is organized as a single round sealed-bid uniform price auction with no price floor. The Forward Capacity Market (FCM) in Nepool is organized as a multiple round descending clock uniform price auction with a predetermined price floor. Both capacity markets in PJM and Nepool are organized three years in advance of the Delivery Year. For example, if an existing or currently non-existing unit clears in the upcoming capacity auction in April 2009, then that unit has to be available to generate power starting in June 2012. Non-performance penalties are very stiff. The three year period was chosen to reflect the time needed to build a peaking Combustion Turbine (CT) unit. Capacity markets should also be understood as a wealth transfer from consumers to producers. The consumers pay a capacity price

to the generation owners, which can be seen as an insurance premium, in order to avoid reliability problems and extreme energy prices during peak demand hours. It is important for the consumer to not focus solely on either capacity prices or energy prices, but to look at the sum of the two in the medium to long run. Still, it is important that questions are asked and research is conducted on whether the current capacity market structure and rules lead to the desired results in the most cost effective way[2].

3.1.6. Fundamental Drivers for Peak & Off-Peak Power

Contrary to natural gas, power is not storable, at least not efficiently and not in high volume. Pump storage facilities and batteries do exist, and although they play an important role locally, in the grand scheme of things they are not that important. This implies that, again in contrast to natural gas and many other storable commodities, January 2009 power does not have to be closely linked with September 2008 power through a cash-and-carry argument. In reality it is closely linked, due to the fact that power, at least peak power, trades off of natural gas, which does have this cash-and-carry spread between Jan09 and Sep08. In Figure 3.1 we plot the forward curve of WestHub 5x16 power on 07/15/2008.

When there is no congestion in the electrical system the LMPs at every node are identical. The LMP at a particular node answers the question: how much does it cost to serve 1 extra MW of load at that node. As long as there is no congestion, the cost of 1 extra MW of load is the same for every node and equals the marginal cost of the

[2]The capacity price in PJM for the 2010-2011 Delivery Year cleared at $175/MW-day and approximately 134,000 MW cleared in the auction. This means a wealth transfer of roughly $8.5 Billion from the PJM consumers to producers for the period June 2010-May 2011. PJM has approximately 50 Million consumers, which implies a capacity price of $171 per consumer for the year 2010-2011.

Figure 3.1. WestHub peak power forward curve

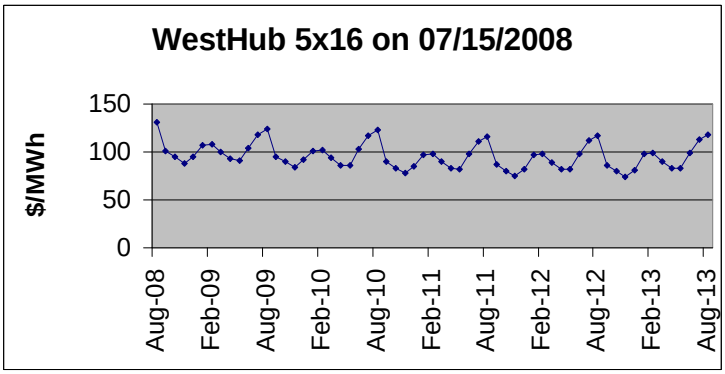
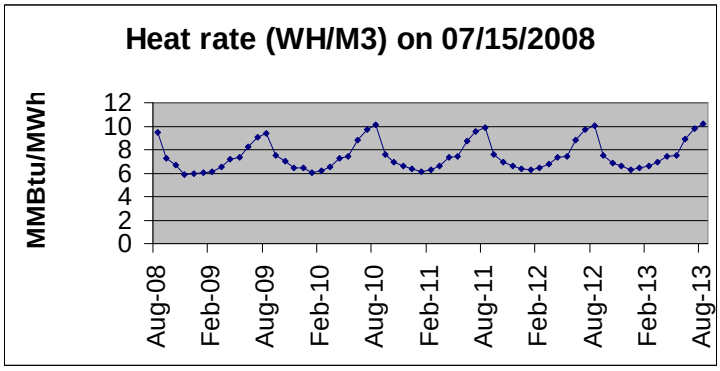


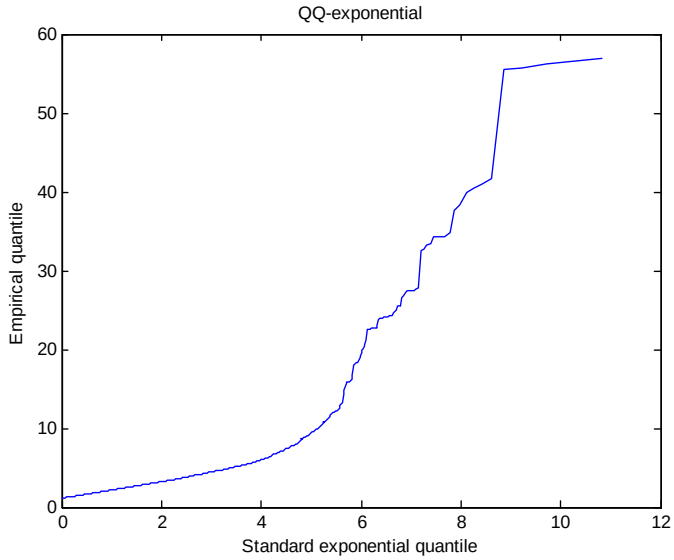
Figure 3.2. Heat rate (WestHub power / TETCO-M3 gas) forward curve



unique marginal unit at that time. During peak times, the marginal unit in PJM is almost always a gas-fired unit, in extreme times it can be an oil-fired unit. As a result peak power trades off of gas. The market implied heat rate, i.e. the forward power price divided by the forward gas price, should reflect the heat rate, i.e. the efficiency, of the average marginal gas-fired unit. During off-peak times the coal units are mostly the marginal units. However, during off-peak times something more complex is happening. Very often, one sees off-peak average prices in PJM around $20, even though we know that the marginal unit is a coal unit with a marginal cost of $40-$50. The reason we observe this phenomenon is because PJM tries to minimize the cost of serving the load.

PJM knows that base load coal units are willing to continue to run overnight at a loss, because if they shut down, they incur a start-up cost the following day. Off-peak power definitely trades off of coal, but the relationship is not that simple. Looking at Figure 3.2 we see that the term structure of the implied market heat rate is not flat, it is higher during the summer than during the winter. Clearly, load is higher during the summer than during the winter. As a result, during the summer we will have to go deeper into the stack and use less efficient gas-fired plants with higher heat rates. At times, we might even have to use oil-fired units. In addition, during hot summer days and in combination with transmission constraints, market power can become an issue, leading to LMPs deviating from marginal cost. All of these reasons combined lead to higher market implied heat rates during the summer. This paragraph, so far, ignored congestion. In the presence of congestion, things become more complex. LMPs are no longer identical throughout the system and there is no longer a unique marginal unit. The moment that there is a single transmission constraint you have at least two marginal units. One marginal unit has to back down and another one has to ramp up. The relationship between gas price and power price is no longer a simple linear relationship. Figure 3.2 plots the market implied heat rate forward curve for PJM over TETCO-M3 gas. When looking at the market implied heat rates, it is important to use the gas price that gas-fired units use that make up the PJM power price. Gas price in the Pennsylvania area is called TETCO-M3 (Texas Eastern Company gas pipeline - Zone M3) and trades as a premium over Henry Hub gas, which is the main hub located in the Houston area, and is used for the Nymex gas contract. In the winter, when gas demand in the NorthEast can be high, and given the fact that pipeline capacity is limited, the basis spread between Henry Hub and Tetco-M3

Figure 3.3. QQ-plot: power data vs. exponential distribution



can blow up. This is reflected in the Tetco-M3 forward gas curve, and hence in the PJM WestHub power forward curve.

3.2. Preliminary Graphical Tools

Just as we did for the Henry Hub natural gas futures returns, we will now focus on the tail behavior of daily PJM WestHub 5x16 power forward returns. The dataset starts in 2005, and we have around 84,000 daily returns. We first look at some of the grahical tools to get a rough sense of how heavy tailed the daily power returns are. We look at the QQ-plots, mean excess function, max/sum ratio, and the large claim index.

Quantile-Quantile plots. We compare the empirical quantile function with the theoretical quantile function of the exponential and the pareto distribution.

From the QQ-plots we learn that the tail behavior of daily power forwards is in between a power tail and an exponential tail.

Figure 3.4. QQ-plot: power data vs. pareto distribution

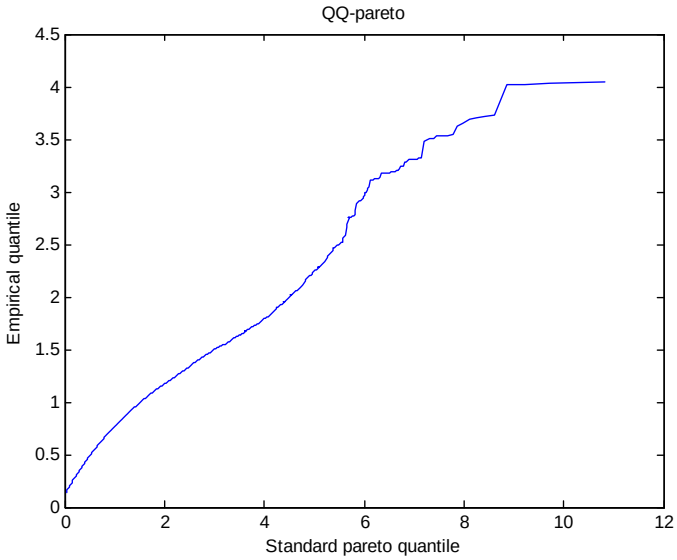
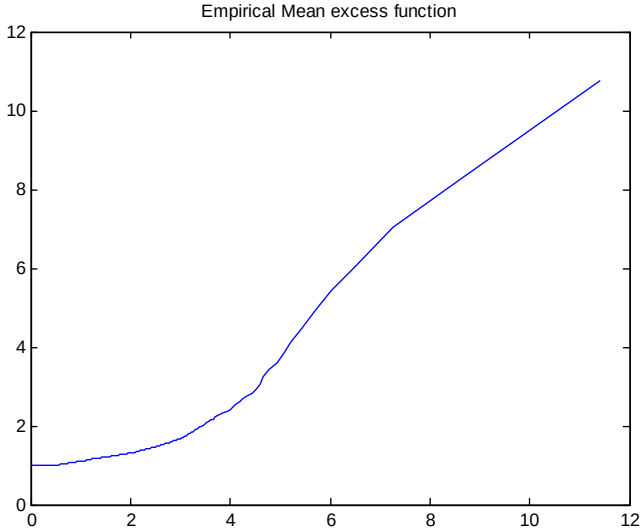


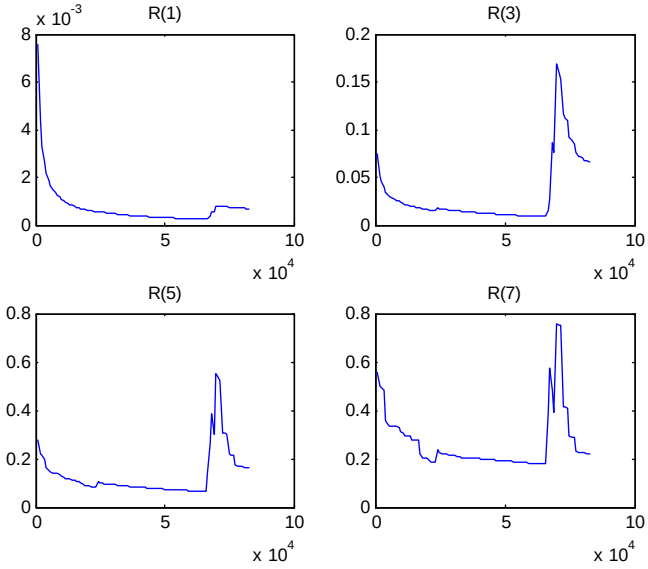
Figure 3.5. Empirical mean excess function of power forward returns



Mean excess function. Next, we plot the empirical mean excess function for power returns.

The empirical mean excess function clearly shows that the data has a heavier tail than the exponential tail, and given the near linearity of the empirical mean excess function it

Figure 3.6. R as a function of n for different levels of p: power forward returns

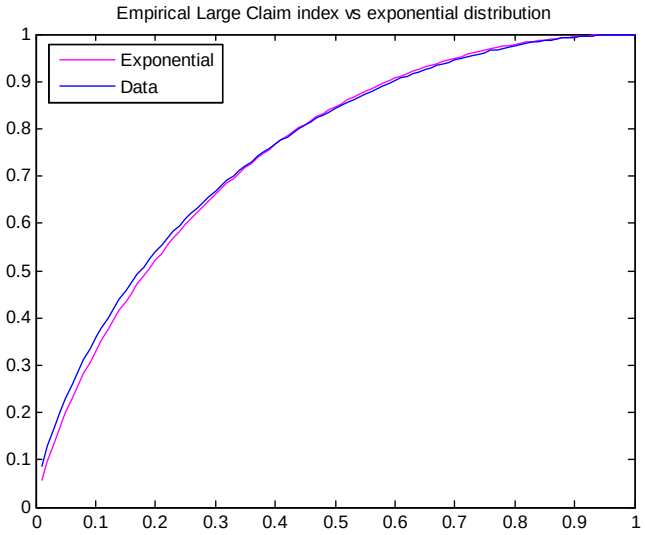


points in the direction of a power-like tail. Recall that the mean excess function is flat for an exponential tail, increasing and concave for the weibull distribution in the SE class, and linearly increasing for a power-like tail.

Max/Sum ratio. Plotting $R_n(p)$ as a function of n for different levels of p, we learn whether the p-th moment of X is finite or not. As mentioned before, random variables that belong to the exponential and the SE class, have all their moments finite. For regularly varying distributions with tail index α, we have that $EX^\beta < \infty$ for $\beta < \alpha$.

Given that $R_n(5)$ and $R_n(7)$ are clearly not zero, the max/sum ratio test tells us that power forward returns have a heavier tail than an exponential and an SE tail. However, we know that we have to be careful making definitive conclusions from these max/sum ratio tests. Figures 2.14 and 2.13also have $R_n(7)$ different from zero even though theoretically all moments should be finite.

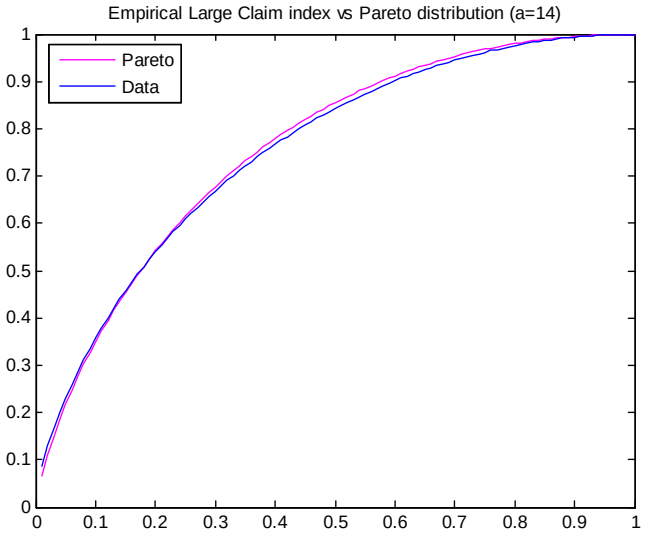
Figure 3.7. Empirical Large Claim index vs. exponential distribution



Large Claim index. The final graphical tool that we use is the large claim index. We limit ourselves to the top 30% of the returns, and we plot the empirical large claim index. We then find the distribution function F such that $D_F(p) = \frac{1}{\mu_F}\int_{1-p}^{1} F^{-1}(s)ds$ is close to the empirical claim index.

From the large claim index we learn that the power forward returns resemble an exponential tail very well. Note that the tail coefficient, a=14, is very high in order to get a good fit with the data. We know that as the tail coefficient approaches ∞, the pareto distribution tends to the exponential. In conclusion, QQ-plots lead us to believe that the power returns data behaves in between an exponential and a power tail. The mean excess function and the max/sum ratio test also conclude that the data decays faster than an exponential tail. The large claim index, however, seems to point to an exponential tail.

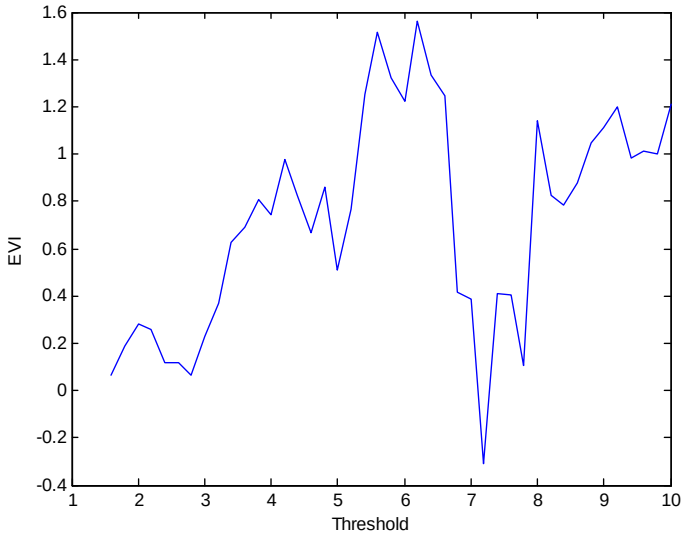
Figure 3.8. Empirical Large Claim index vs. pareto distribution



3.3. Non-Parametric Extreme Value Theory

Just as we have done for the Henry Hub natural gas futures data, we now compute the extreme value index (EVI) for different thresholds for PJM Westhub 5x16 power forward returns. Recall that EVI > 0 implies that the underlying distribution belongs to the MDA of the extreme value Frechet distribution, which means that it has a regularly varying or power-like tail. EVI $= 0$ implies that the distribution belongs to the MDA of the Gumbel distribution, which means that it has a rapidly varying or exponential-like tail. By construction, EVT cannot discriminate between an exponential and an SE tail, because both the exponential and weibull distribution belong to the MDA of the Gumbel distribution. But the estimate of the EVI should shed more light on whether the data tends more towards an exponential-like tail or towards a power-like tail. In the next Figure 3.9 we plot the EVI for the power returns for different threshold levels.

Figure 3.9. EVI estimates using Pickands' estimator for power forward returns data



We can see in Figure 3.9 that overall the EVI is in excess of zero. There is no clear split in the data with two different regimes as we had with the natural gas futures returns. The EVI estimates clearly indicate that the power forward returns have a power-like tail.

3.4. Parametric Fitting

The overall conclusion from the QQ-plots, mean excess function, max/sum ratio, and the EVI estimates is that the power forward returns data definitely decays more slowly than an exponential tail, and is even pointing towards a power-like tail. In that regard, we will now zero in on the pareto, weibull, and log-weibull distributions to figure out whether a power-like tail or an SE tail is more likely and to quantify the tail coefficient. Pareto. $F_u(x) = 1 - \left(\frac{u}{x}\right)^b$ with $x \geq u$. In Figure 3.10 we plot the tail coefficient, b, using ML, for different levels of u. We observe that the estimate of the tail coefficient is relatively stable for the entire threshold range, and fluctuates around 2.2.

Figure 3.10. ML estimate of tail coefficient, b, for power forward returns as a function of u

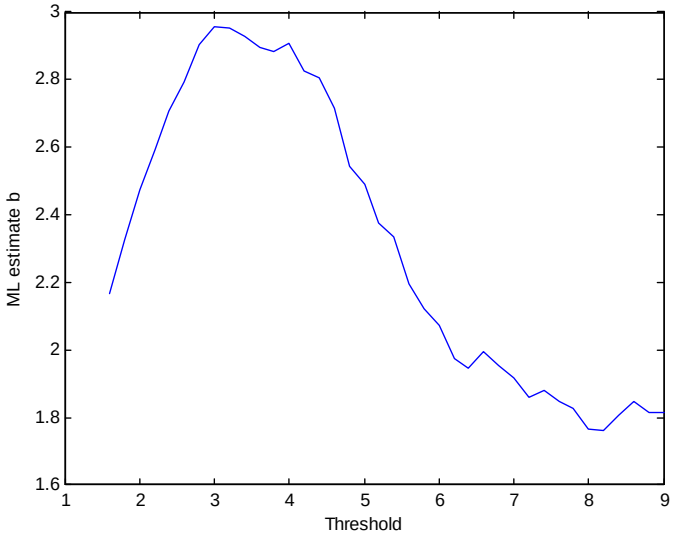
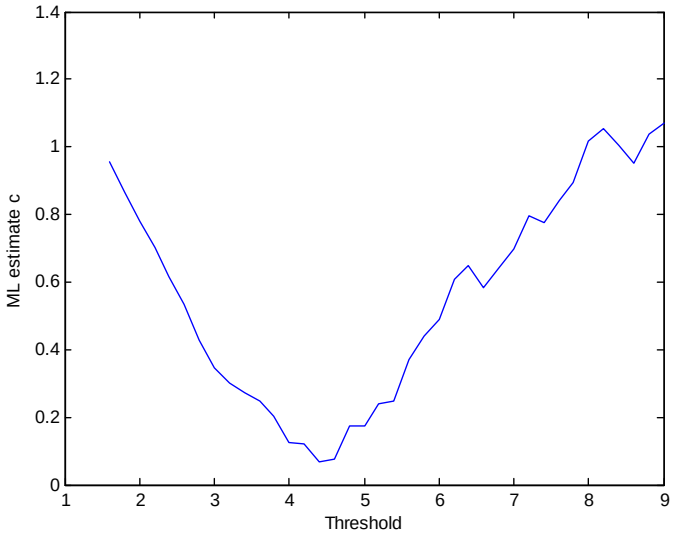
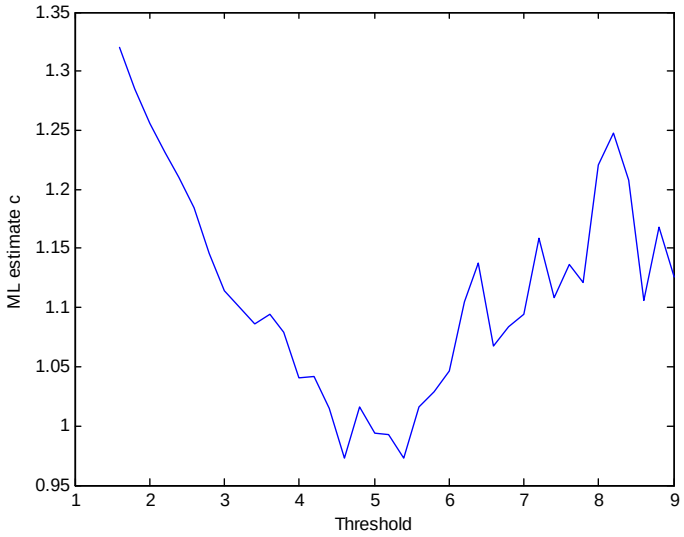


Figure 3.11. ML estimate of tail coefficient, c, for power forward returns as a function of u



Weibull . $F_u(x) = 1 - \exp[-\left(\frac{x}{d}\right)^c + \left(\frac{u}{d}\right)^c]$ with $x \geq u$.

Figure 3.12. ML estimate of tail coefficient c (log-weibull) for power forward returns as a function of u

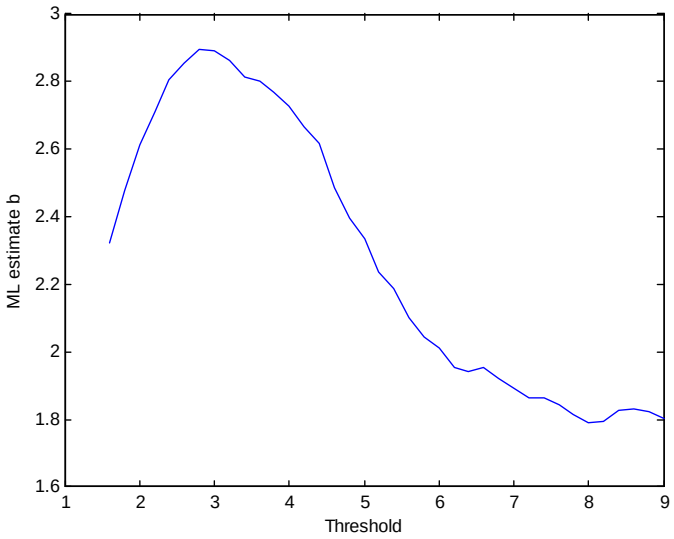


Log-weibull. $F_u(x) = 1 - \exp[-b.\{\log(\frac{x}{u})\}^c]$ with $x \geq u$. Recall that for $c = 1$, it recovers the Pareto distribution, with b the tail coefficient. For $c = 2$, it becomes the lognormal distribution. When $c > 1$ the log-weibull has a thinner tail than the Pareto, but a heavier tail than the SE family, to which the weibull belongs. In Figures 3.12 & 3.13 we plot the coefficients c and b, respectively.

Given that in Figure 3.13 the coefficient c is larger than 1 for almost the entire threshold range, the log-weibull tells us that the underlying distribution decays faster than a power-like tail. In order to gain some more definitive conclusions, we will compare the different models pairwise using Wilks' LR test.

Weibull vs. Exponential. As we have mentioned before, we can use Wilks' theorem directly to test the weibull distribution against the exponential distribution, because they

Figure 3.13. ML estimate of tail coefficient b (log-weibull) for power forward returns as a function of u



are nested. Figure 3.14 plots the LR ratio between the weibull and the exponential distribution as a function of the threshold level. It also plots the 90th- and the 95th percentile of the $\chi^2(1)$-distribution. We can see that, apart from the very low thresholds, we can reject the exponential for the entire threshold range. This corroborates the conclusion from the previous tests, i.e. daily power forward returns decay more slowly than an exponential tail.

Weibull vs. Pareto. We know that the weibull and the pareto distribution are not nested, and that therefore Wilks' theorem cannot be used directly. However, Malevergne and Sornette (2006) have shown that in this case Wilks' theorem can still be applied.

Figure 3.15 tells us that we favor the weibull over the pareto for most of the range, but that in the middle of the threshold range we have a hard time rejecting the more

Figure 3.14. Log-likelihood ratio between the weibull and the exponential distribution for the power forward returns

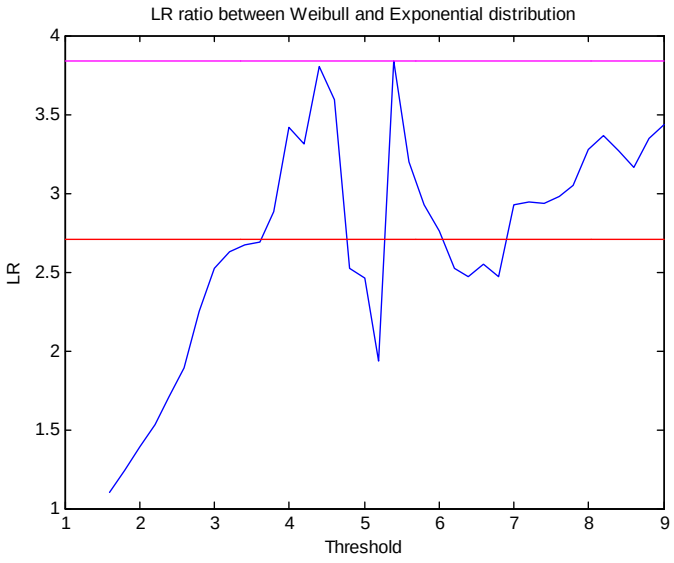
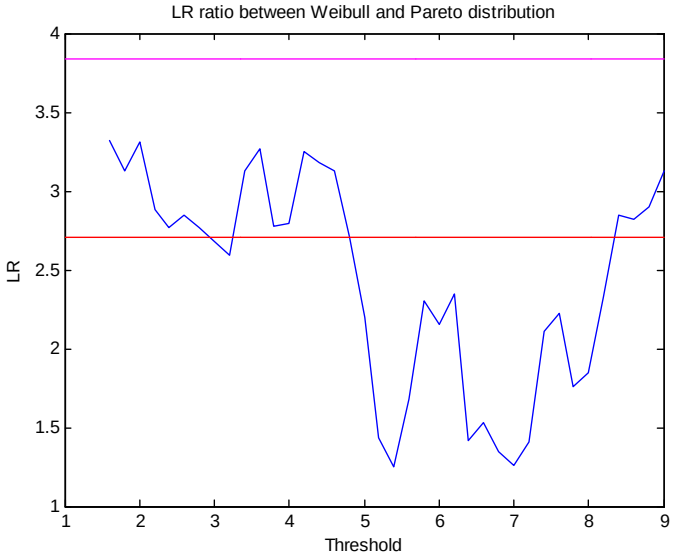


Figure 3.15. Log-likelihood ratio between the weibull and the pareto distribution for the power forward returns



parsimonious pareto distribution. In that regard Figure 3.15 reflects to some extent Figure 3.11. In Figure 3.11, we plot c, the tail coefficient of the weibull distribution, and we see a similar pattern. In the middle of the range the tail coefficient is close to zero, reflecting a power-like tail. We do not plot the LR ratio between the log-weibull and the pareto distribution, because the graph is very similar to Figure 3.15. This also reflects Figure 3.12 to some extent. In Figure 3.12 we plot the coefficient c of the log-weibull, and we observe that in the middle of the range it is close to 1, reflecting a power-like tail.

3.5. Conclusion

Almost all the tests (QQ-plots, max/sum ratios, mean excess function, EVI, & parametric tests) conclude that daily power forward returns have heavier tails than an exponential tail. However, there is not really a consistent pattern to have a definitive view on whether a power-like or an SE tail describes the data best. Using the Wilks' LR test we cannot reject the weibull in favor the pareto distribution. As a result, we conclude that a good model for the tail behavior of daily WestHub 5x16 power forward returns is the weibull distribution with a tail coefficient around 0.6.

3.6. Sensitivity Analysis

We now split up the entire sample in two groups, just as we have done for the natural gas data.

Pareto. In Figures 3.16 and 3.17 we plot the tail coefficient, b, for the pareto distribution for the prompt/next month forwards and for the future month forwards respectively.

Figure 3.16. ML estimate of tail coefficient, b, for power forward returns for the prompt and next month contracts as a function of u

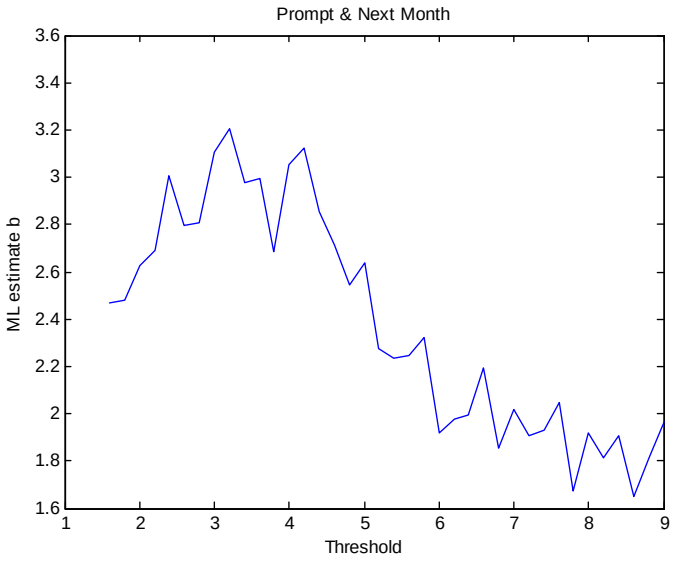
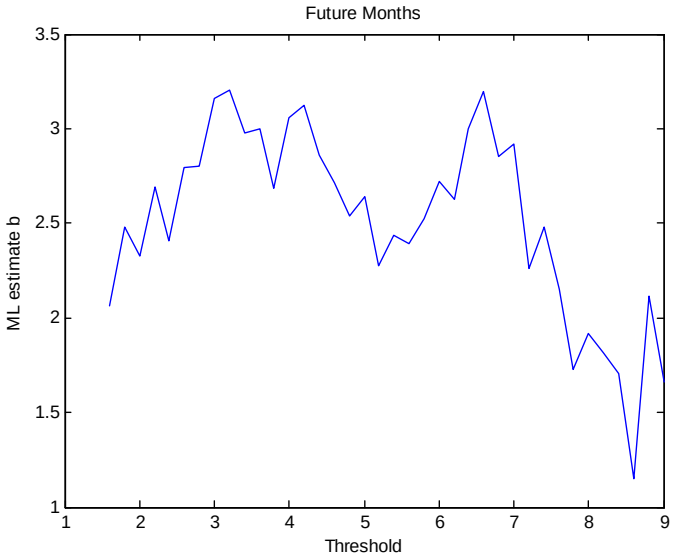
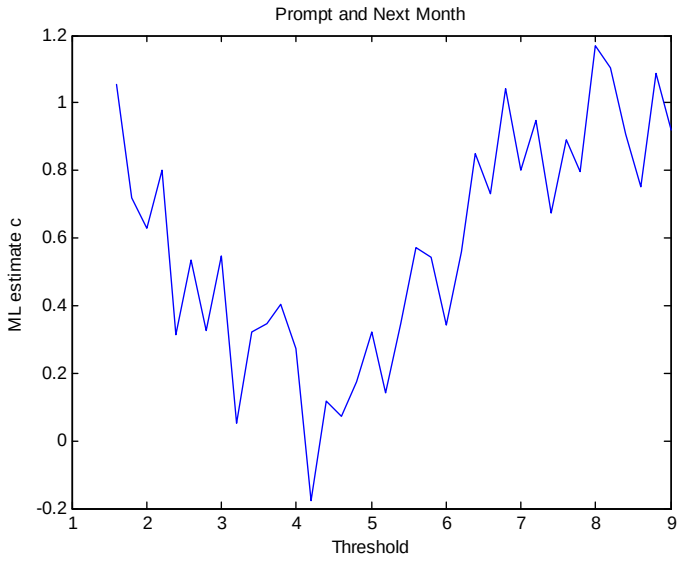


Figure 3.17. ML estimate of tail coefficient, b, for power forward returns for the future month contracts as a function of u



We can see that the tail coefficients for the prompt and next month contracts are very similar to the ones in Figure 3.10, i.e. for the entire dataset. Figure 3.17, which plots the tail coefficient for the future month contracts has a different pattern.

Figure 3.18. ML estimate of tail coefficient, c, for power forward returns for the prompt and next month contracts as a function of u

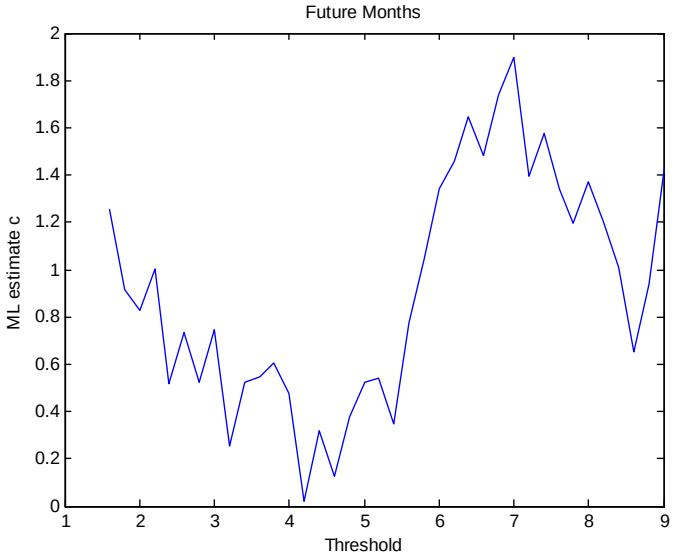


Weibull. We plot the tail coefficient c of the weibull for both groups in Figure 3.18 and 3.19.

We notice again that the tail coefficients for the prompt and next month contracts are very similar to the ones for the entire dataset (Figure 3.11). The average tail coefficient equals 0.52. It is therefore not surprising that the pairwise comparison tests of the prompt and next month contracts closely match the results using the power returns from the entire dataset. The average tail coefficient for the future month contracts equals 0.78.

We know that peak power trades mainly off of gas. In particular in the longer run gas is the most important driver for power forwards. Far out market implied heat rates are relatively stable. Market implied heat rates several months/years out can change because

Figure 3.19. ML estimate of tail coefficient, c, for power forward returns for the future month contracts as a function of u



of large new generation projects expected to come online or large new transmission lines coming into play. A change in market rules can also affect heat rates in the future. Probably the biggest reason forward heat rates change is the result of financial flows. When a big generation company decides to hedge a substantial fraction of its fleet going forward, by selling heat rates far out, these heat rates can fall dramatically. Still, heat rates far out are relatively stable. As a result of all of this we would expect the peak power forward returns and the gas futures returns for the months further out in the curve to have a similar behavior, which in this case turns out to be a weibull distribution/SE tail with tail coefficient around 0.8. In the short term, i.e. prompt and next month, the relationship between power and gas becomes weaker as weather comes into play and transmission/generation outages, most of which are announced one to two months in advance, become a factor. We conclude that short term gas futures returns have an SE

tail up to a 5.5% threshold and a power-like tail in excess of this threshold. Short-term power forward returns do not differ much from the power forward returns using the entire dataset, and we conclude that a good model for them is a weibull with a tail coefficient around 0.5.

3.7. References

- Alexander and Narayanan (2001), Option pricing with normal mixture returns: modelling excess kurtosis and uncertainty in volatility,

  ICMA Centre Discussion Papers in Finance 10.

- Bahra (1997), Implied risk-neutral density functions from option prices, *Bank of England.*

- Beirlant et al. (2004), Statistics of Extremes: Theory and Applications, Wiley Series in Probability and Statistics.

- Bingham, Goldie, and Teugel, Regular Variation, Cambridge University Press.

- Breeden and Litzenberger (1978), Prices of state contingent claims implicit in option prices, *Journal of Business 51.*

- Brigo and Mercurio (2001), Lognormal mixture dynamics and calibration to market volatility smiles, *International Journal of*

  Theoretical & Applied Finance.

- Cherubini and Luciano (2002), Multivariate option pricing with copulas, *International Centre for Economic Research Working*

  Paper.

- Deheuvels (1979), La fonction de dependence empirique et ses proprietes. Un test non-parametrique d'independence, *Acad Roy Belg Bul Cl Sci 5.*
- Derman and Kani (1994), The volatility smile and its implied tree, *Goldman Sachs Quantitative Strategies Research Notes.*
- Dhaene et al. (2002a), The concept of comonotonicity in actuarial science and finance: theory, *Insurance: Mathematics and Economics 31.*
- Dhaene et al. (2002b), The concept of comonotonicity in actuarial science and finance: applications, *Insurance: Mathematics and Economics 31.*
- Di Guilmi et al. (2003), Power law scaling in world income distributions, *Economics Bulletin 15.*
- Embrechts, Kluppelberg, and Mikosch (1997): Modelling Extremal Events for Insurance and Finance, Springer-Verlag
- Embrechts, McNeil, and Straumann (2002), Correlation and dependence in risk management: properties and pitfalls, In: *Risk Management: Value at Risk and Beyond, ed. M.A.H. Dempster, Cambridge University Press, Cambridge.*
- Fama (1965), The behavior of stock market prices, *Journal of Business 38*
- Farmer and Lillo (2004), On the origin of power-law tails in price fluctuations, *Quantitative Finance 4.*

- Frank and Schweizer (1987), Best possible bounds for the distribution of a sum - a problem of Kolmogorov, *Probability Theory 74.*
- Gabaix (1999), Zipf's law for cities: an explanation, *Quarterly Journal of Economics 114.*
- Gabaix et al. (2003), A theory of power-law distributions in financial market fluctuations, *Nature 423.*
- Genest and Rivest (1993), Statistical inference procedures for bivariate Archimedean copulas, *Journal of American Statistical*

  Association 88.
- Genest and MacKay (1986a), The joy of copulas: bivariate distributions with uniform marginals, *The American Statistician 40.*
- Genest and MacKay (1986b), Copules archimediennes et familles de lois bidimensionnelles dont les marges sont donees, *Canad J*

  Statist 14.
- Goldie and Kluppelberg (1997), Subexponential Distributions, In: *A Practical Guide to Heavy Tails: Statistical Tails for Analysing*

  Heavy-Tailed Distributions (Adler, Feldman, and Taqqu).
- Jackwerth and Rubenstein (1996), Recovering probability distributions from option prices, *Journal of Finance 51.*
- Joe (1997), Multivariate Models and Dependence Concepts, Chapman & Hall.
- Li (1992), Random texts exhibit Zipf's law-like word frequency distribution, *IEEE Transactions on Information Theory 38.*

- Lux and Sornette (2002), On rational bubbles and fat tails, *Journal of Money, Credit & Banking 34.*
- Makarov (1981), Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed,

  Theory Probability Applied 26.
- Malevergne and Sornette (2006), Extreme Financial Risks, Springer
- Mandelbrot (1963), The variation of certain speculative prices, *Journal of Business 36, 392-417.*
- Mandelbrot (1997), A multifractal model of asset returns, *Coles Foundation Discussion Paper 1146.*
- Mantegna et al. (1995), Scaling behavior of an economic index, *Nature 376.*
- Meier (2006), Natural disasters, casualties, and power laws, *Annual Meeting of the American Political Science Association.*
- Melick and Thomas (1987), Recovering an asset's implied PDF from option prices: an application to crude oil during the Gulf Crisis,

  Journal of Financial and Quantitative Analysis 32.
- Mittnik et al. (1998), Stable paretian modeling in finance, In: *A Practical Guide to Heavy Tails: Statistical Tails for Analysing*

  Heavy-Tailed Distributions (Adler, Feldman, and Taqqu).
- Muller and Scarsini (2000), Some remarks on the supermodular order, *Journal of Multivariate Analysis 73.*
- Muzy et al. (2001), Multifractal returns and hierarchical portfolio theory, *Quantitative Finance 1.*

- Nelsen (2005), Dependence modeling with Archimedean copulas, *Proceedings of the Second Brazilian Conference on Statistical*

  Modeling in Insurance and Finance: Institute of Mathematics and Statistics, University of São Paulo (2005).
- Nelsen (2006), An Introduction to Copulas, Springer.
- O'Brien (1980), A limit theorem for sample maxima and heavy branches in Galton-Watson trees, *Journal of Applied Probability 17.*
- Pickands (1975), Statistical inference using extreme order statistics, *Annals of Statistics 3.*
- Rebonato and Cardoso, Unconstrained fitting of implied volatility surfaces using a mixture of normals, *RBS Quantitative Research*

  Centre and Oxford University.
- Rosenberg (1999), Semiparametric pricing of multivariate contingent claims, *Stern School of Business Working Paper Series 28.*
- Rosenberg & Engle (2002), Empirical pricing kernels, *Journal of Financial Economics.*
- Rubenstein (1994), Implied binomial trees, *Journal of Finance 49.*
- Ruschendorf (1982), Random variables with maximum sums, *Advanced Applied Probability 14.*
- Salmon and Schleicher (2006), Pricing multivariate currency options with copulas, *Bank of England and Warwick Business School.*
- Sklar (1959), Fonctions de repartitions a n dimensions et leurs marges, *Publ Inst Statist Univ Paris 8.*

- Tadaki (2007), Long term power law fluctuation in internet traffic, *Journal of Physical Society 76.*
- Tchen (1980), Inequalities for distributions with given marginals, *Annals of Probability 8.*
- van den Goorbergh, Genest, and Werker (2005), Bivariate option pricing using dynamic copula models, *Insurance: Mathematics and Economics 37.*
- Williamson and Downs (1990), Numerical methods for calculating convolutions and dependency bounds, *International Journal of Approx. Reasoning 4.*
- Wilks (1938), The large sample distribution of the likelihood ratio for testing composite hypotheses, *Annals of Mathematical Statistics 9.*