Noname manuscript No.
(will be inserted by the editor)

Analysis of long-term natural gas contracts with vine copulas in optimization portfolio problems

E. Allevi · L. Boffino · M.E. De Giuli · G. Oggioni

Received: date / Accepted: date

Abstract This paper investigates the dependence risk and the optimal resource allocation of the underlying assets of a long-term natural gas contract through pair-vine copulas and portfolio optimization methods with respect to five risk measures. This analysis takes inspiration from the current situation of the European gas market where both long-term contracts and hub spot prices are applied. The fall of the European natural gas demand combined with the increase of US shale gas exports and the Liquefied Natural Gas availability have led to a reduction of the gas spot prices in Europe. Oil-indexed long-term gas contracts failed to promptly adjust their positions implying significant losses for European gas mid-streamers that asked for a re-negotiation of their existing contracts. With the aim of analyzing this situation and determining whether oil-indexation can still be the solution for the European market we consider both spot gas prices traded at the hub and oil-based commodities as underlying of the LTCs. The portfolio optimization results converge in some commodities.

Keywords Long-term natural gas contracts · Multivariate dependence structure · Pair-Copula Construction · Portfolio optimization · vine copulas

E. Allevi
Department of Economics and Management, University of Brescia, Italy
E-mail: elisabetta.allevi@unibs.it

L. Boffino
Department of Management, Economics and Quantitative Methods, University of Bergamo, Italy
E-mail: l.boffino@studenti.unibg.it

M.E. De Giuli
Department of Economics and Management, University of Pavia, Italy
E-mail: elena.degiuli@unipv.it

G. Oggioni
Department of Economics and Management, University of Brescia, Italy
Tel.: +39-030-2988522
E-mail: giorgia.oggioni@unibs.it

1 Introduction

Natural gas can be sold either through long-term contracts or at spot price at market hubs. Long-term contracts have been historically introduced to allow for risk sharing between gas producers and mid-streamers, which respectively face price and volume risks. These contracts have been traditionally concluded over long periods (typically 20 years or more) and are characterized by quantity and price clauses. The Take or Pay (TOP) quantity clause obligates the buyer to take a certain quantity of natural gas or to pay for it. The Price Indexation (PI) clause relates the price at which gas is bought to some index on the market that has been traditionally represented by the price of crude and oil-products. The price clause provides producers with some price stability and reduction of revenue volatility, which are indispensable for ensuring investments in new infrastructures which are very expensive. Moreover, the quantity and price clauses also allow for hedging the mid-streamers' volume risk (see Abada et al., 2017). The hub pricing approach was firstly introduced in the nineties in the US and in the UK and it is now developing in Europe. In this system, natural gas is traded, every day, on a spot market that determines prices and volume on the short term. International natural gas market is organized in different ways depending on the considered areas. North America is essentially organized on the basis of Henry Hub spot market. The price on this market is currently very low, at about 3$/MBtu,[1] because of the development of unconventional (shale) gas. The Henry Hub is the best-known of all natural gas trading points. It is both a physical distribution hub for pipeline gas and a pricing point, since it is the basis of spot market trading and of futures trading on the New York Mercantile Exchange (NYMEX).

On the other side, Asia is mainly supplied by Liquefied Natural Gas (LNG) that is traded through expensive long-term contracts. The Asian LNG long-term contracts has been set on the basis of the average of Japanese customs-cleared crude oil price that is the Japan Crude Cocktail (JCC). After the Fukushima disaster, Japan has significantly increased its imports of LNG, mainly supplied by Australia, Malaysia and Qatar through long-term contracts (see GIIGNL, 2018). Spot deals exist, but they are bilateral as, up to now, there is no hub in Asia in contrast with America and Europe. Some experts advocate a move from the current system of long-term contracts to hub-pricing system (e.g IEA, 2013). Singapore has proposed itself as a possible gas hub for Asia as well as Tokyo and Shanghai (see Xunpeng, 2016).

Except for the UK, Europe is still dominated by long-term contracts, though spot markets are growing and are expected to develop further. In the UK, gas is largely traded at the National Balancing Point (NBP) spot market. NBP is in operation since the late 1990s and is the longest-established spot-traded natural gas market in Europe. It is characterized by high liquidity[2] and the resulting spot price is widely used as an indicator for European wholesale gas market. In continental Europe, Zeebrugge (ZEE) and the Title Transfer Facility (TTF), respectively located in Belgium and in the Netherlands, are the two dominant spot market places and many others are emerging (see Melling, 2010).

[1] See `http://www.eia.gov/dnav/ng/hist/rngwhhdd.htm`

[2] The liquidity of a gas hub can be defined as the ratio between the total volume of trade on the hub and the volume of gas consumed in the area served by the hub.

The coexistence of long-term contracts and hub-pricing systems on the European market implies that the natural gas is traded at two different prices on the same market. Depending on the conditions, spot price can be higher or lower than long-term contractual gas, implying possibly difficult situations for companies loaded with high TOP gas price against the low spot prices. This is what happened in the last years in Europe, where the combined effects of the increase of the US shale gas exports, the reduction of European gas demand due to the economic crisis, and the increased availability of uncommitted LNG from Qatar led to a new supply/demand balance that was reflected into low gas prices at the European hubs. On the other side, oil-indexed long-term gas contracts failed to promptly adjust their positions implying significant losses for European gas mid-streamers what were committed by the TOP clause to buy quantities of gas higher than those required at higher prices. As a consequence, European mid-streamers have re-negotiationed the long-term gas contracts to make them more flexible and closer to spot gas prices. These re-negotiations have resulted into a decline of oil-indexation and hub-linked pricing has rapidly become the basis for an increasing number of transactions in the UK and in the Northern Western Europe (see Franza, 2014; Kanai, 2011; Stern and Rogers, 2014; Yafimava, 2014). As indicated by Chyong (2015), leading gas suppliers, such as Statoil, GasTerra, Sonatrach and Gazprom, have been forced to modify their LTC price and volume in Europe. However, these gas suppliers have assumed different attitudes: Gazprom and Sonatrach has defended oil indexation and has offered retroactive discounts on existing contracts by introducing either limited degree of spot indexation or reduced minimum TOP provisions. On the other side, GasTerra and Statoil have behaved in a more flexible way, by conceding more spot indexation in their contracts. However, oil is still accounting for the 95% of price formation within the European Community against the 30% of price formation within the UK (see Theisen, 2014). This shows that oil-indexed contract prices still exercises a strong influence over gas prices in Europe. The reliability of prices set on spot markets is one of the main reasons used by the opponents of the spot indexation for LTCs (see Frisch, 2010). Liquidity, transparency, and the ability to attract a significant number of market players are necessary for a hub to become a price maker (see Heather, 2012). For the time being, the NBP and TTF are the only two hubs in Europe with sufficient liquidity (see Heather and Petrovich, 2017).

From these evidence, it turns out that the role of the European spot gas markets and their impacts on LTCs are becoming extremely important issues. In this paper, we focus on LTCs renegotiation as asked by the mid-streamers. We investigate this topic using a combination of vine copulas and portfolio optimization models that are well-known and widely applied approaches in the financial literature. We adopt these relatively standard methodologies on purpose because we want to obtain results that are not biased by the adopted methodology. Our analysis is structured as follows: we estimate, via Pair-Copula Constructions, the dependence risk structure across the underlyings of a long-term natural gas contract. In order to reflect the above mentioned European hybrid pricing system, based on the symbiotic coexistence of oil-indexed contracts and gas-indexed hub prices, we consider the prices of oil based products and natural gas traded at the hub as component of the gas pricing formula. (see Section 3.1 for more details). We then combine vine copula models and classical portfolio optimization methods to construct the optimal underlying portfolio, applying different performance measures. Copula models represent a suitable tool to this scope. In particular, copula is a function that combines marginal distributions to form multivariate distributions. The application of copulas is very popular in several fields, like finance, insurance, financial economics and econometrics (see e.g. Cherubini et al., 2004; Durante and Sempi, 2015; Genest et al. 2009; Malevergne et al., 2006; Krzemienowski et al. 2016, Vaz de Melo Mendes, 2010). Nowadays, the modeling of stochastic dependence via copulas has led to an increasing attention also in the commodity market (see e.g. Accioly and Aiube, 2008; Aloui et al., 2013; Czado et al., 2011; Grégoire et al. 2008; Jäschke, 2014; Lu et al., 2014; Reboredo, 2011; Wen et al., 2012; Wu et al., 2012).

While there is a wide range of possible alternative copula functions for the bivariate case, in the multivariate setting the use of families different from Normal and Student-t is rather scarce, due to computational and theoretical limitations (see e.g. Joe, 1997 and Nelsen, 1999). For this reason, in order to represent a multivariate copula with suitable sets of bivariate copulas, Joe (1996) introduced the Pair-Copula Construction (PCC) approach, later discussed in detail by Aas et al., 2009; Bedford and Cooke, 2001, 2002; and Kurowicka and Cooke, 2006. A collection of potentially different bivariate copulas is used to construct the joint distribution of interest via PCCs, allowing to represent different types and strengths of dependence in an easy way. PCCs constitute a flexible and very appealing tool for financial analysis, (see e.g. Allen et al, 2013; Brechmann and Czado, 2013; Dalla Valle et al., 2016; Dißmann et al., 2013). The vine copula models considered for the analysis of the portfolio's dependence risk overcome the restrictive and deterministic features of the bivariate copulas and traditional measures of correlation, due to their suitability in capturing the non-normality, tail dependence and volatility clustering of assets returns. Recently Arreola (2014) and Travkin (2013) show how the vine copula approach can be appropriately used to investigate the dependence structure among the different components of energy portfolio as well as to derive implications for portfolio risk management. In this paper, we investigate via Pair-Copula Constructions, the dependence risk structure across the underlying assets of Long-Term Contracts (LTCs) on natural gas. We define the optimal portfolio composition under different performance measures. Whereas the literature that uses copulas in portfolio optimization is wide and rich, to the best of our knowledge there are no existing studies which combine these two methodologies to determine the optimal composition of the assets commonly used to price the gas LTCs. In doing this, we consider both the traditional oil-based commodities and spot gas prices to address the debate over oil/spot indexation related to the re-negotiation of European LTCs. Our results confirms the effectiveness of the hybrid pricing system currently existing in the continent, but indicates that oil should still play an important role in the definition of the price of the LTCs.

The remainder of this paper is organized as follows: Section 2 briefly presents the PCC and the vine copulae; Section 3 provides the data analysis and introduces the pair vine copula that we use in our study; Section 4 overviews the optimization portfolio problems that we solve to estimate the risk of the optimal portfolio of the underlying assets of LTCs, and illustrates the results of our analysis. Finally, concluding remarks are given in Section 5.

2 PCC approach and vine copulas

PCC is a multivariate copula constructed by using only bivariate copula or pair-copulae as building blocks. All copulae involved in the decomposition may be selected freely among the wide range of bivariate copulae family that are capable of modeling joint distribution with different characteristics. Hence, PCC allows high flexibility in representing complex structures of dependence among multivariate data. It is based on the decomposition of a d-dimensional joint density function $f(x_1,\ldots,x_d)$ of the random vector $X=(X_{1,}\ldots X_d)$, as a product of conditional densities:

$$f(x_1,\ldots,x_d)=f_d(x_d)\times f_{d-1|d}(x_{d-1}\,|x_d)\times\cdots\times f_{1|2\cdots d}(x_1\,|x_2\cdots,x_d)\;. \quad (1)$$

Each term in (1) can be decomposed using Sklar theorem (See Sklar, 1959) to express the conditional density for a generic element x_j conditioned on the d-dimensional vector $\mathbf{v}$ as in (2):

$$f_{x_j|\mathbf{v}}(x_j\,|\mathbf{v})=c_{x_j,v_l|\mathbf{v}_{-l}}(F_{x_j|\mathbf{v}_{-l}}(x_j\,|\mathbf{v}_{-l}),F_{v_l|\mathbf{v}_{-l}}(v_l\,|\mathbf{v}_{-l})\times f_{x_j|v_{-l}}(x_j\,|\mathbf{v}_{-l}), \quad (2)$$

where v_l is an arbitrary component of $\mathbf{v}$, $\mathbf{v}_{-l}$ denotes the $(d-1)$ dimensional vector without v_l, $c_{x_j,v_l}\,|\mathbf{v}_{-l}\,(\cdot,\cdot)$ is the conditional pair copula density and $F_{x_j|\mathbf{v}_{-l}}(\cdot\,|\cdot)$ is the conditional distribution of x_j given $\mathbf{v}_{-l}$. More precisely, for every j, Joe (1996) proves that:

$$F_{x_j|\mathbf{v}}(x_j\,|\mathbf{v})=\frac{\partial C_{x_j,v_l|v_{-l}}(F_{x_j|\mathbf{v}_{-l}}(x_j\,|\mathbf{v}_{-l}),F_{v_l|\mathbf{v}_{-l}}(v_l\,|\mathbf{v}_{-l})\,|\theta)}{\partial F_{v_l|\mathbf{v}_{-l}}(v_l\,|\mathbf{v}_{-l})}, \quad (3)$$

where the bivariate copula function is specified by $C_{x_j,v_l|v_{-l}}$, with parameters θ. In working with copula models a function $h(x,v,\Theta)$ can be defined in order to represent the conditional distribution function when x and v are uniform, i.e. $f(x)=f(v)=1$, $F(x)=x$ and $F(v)=v$. The h-function can be calculated as follows:

$$h(x,v,\Theta)=F(x|v)=\frac{\partial C_{x,v}(x,v,\Theta)}{\partial v}, \quad (4)$$

where the second parameter of $h(x,v,\Theta)$ always corresponds to the conditioning variable and Θ denotes the set of parameters for the copula of the joint distribution function of x and v. For high-dimensional distributions the number of possible pair-copulae constructions is manifold. In order to organize them, the Bedford and Cook(2001, 2002) have introduced a graphical model denoted regular vines that depict multivariate copulas built up using a cascade of bivariate copulas (or pair-copulas). This allows to understand which conditional specifications are used to describe the joint distribution. A regular vine $V(d)$ on d variables is a nested set of trees T_i where $i=1,...,d-1$. In particular the first tree has d nodes and $d-1$ edges that represents the pair-copula densities between the nodes. While, the j trees have

$d+1-j$ nodes deriving from the edges of tree $j-1$ and $d-j$ edges that are the conditional pair-copula densities. Moreover the proximity condition states that if the nodes of tree $j+1$ are connected by an edge, than the corresponding edges in tree j share a common node. According to Kurowicka and Cooke (2006) the joint of a random vector $X=(X_{1,}\dots X_d)$ following an R-vine distribution can be written as:

$$f(x_1,\dots,x_d)=\left[\prod_{k=1}^{d} f_k(x_k)\right]\cdot\left[\prod_{i=1}^{d-1}\prod_{e\in E_i} c_{j(e),k(e)|D(e)}(F(x_{j(e)}|x_{D(e)}),F(x_{k(e)}|x_{D(e)}))\right]. \quad (5)$$

with node set $\mathcal{N}:=\{N_1,\dots,N_{d-1}\}$ and edge set $\mathcal{E}:=\{E_1,\dots,E_{d-1}\}$. Each parameter $e=j(e),k(e)|D(e)$ is an edge, while $c_{j(e),k(e)|D(e)}$ represents a bivariate conditional density copula. $j(e)$ and $k(e)$ are the conditioned nodes and $D(e)$ is the conditioning set. The union {j(e),k(e),D(e)} is called constraint set. The right part of equation 5 which involves $d(d-1)/2$ bivariate copula densities, is called an R-vine copula. Special classes of R-vines are Canonical vines (C-vines) and Drawable vines (D-vines). A C-vine is a regular vine where each tree T_i has a unique node that is connected to $d-i$ edges; while a D-vine is a regular vine where each node is connected to no more than two other nodes. Each tree in a C-vine is a star with one unique node that connects to all other nodes, whereas a D-vine is represented by line trees. In a C-vine, at the first root node at level 1 of the nested set of trees, the key variable presents the highest correlation value in regard to the other variables and governs the dependence structure among the others. Intuitively, we use a C-vine to describe a scenario where one variable dominates the others, whereas in the D-vines we do not assume the existence of a particular node dominating the dependencies.

More precisely, the joint density function $f(x_1,\cdots,x_n)$ of a C-vine of dimension d takes the following form :

$$f(x_1,\dots,x_d)=$$
$$=\prod_{k=1}^{d} f_k(x_k)\cdot\prod_{i=1}^{d-1}\prod_{j=1}^{d-i} c_{i,i+j|1,\dots,i-1}(F(x_i|x_1,\dots,x_{i-1}),F(x_{i+j}|x_1,\dots,x_{i-1})|\theta_{i,i+j|1,\dots,i-1}). \quad (6)$$

In a similar way, the joint density function of a D-vine is given by:

$$f(x_1,\dots,x_d)=$$
$$=\prod_{k=1}^{d} f_k(x_k)\cdot\prod_{i=1}^{d-1}\prod_{j=1}^{d-i} c_{j,j+i|j+1,\dots,j+i-1}(F(x_j|x_{j+1},\dots,x_{j+i-1}),F(x_{j+i}|x_{j+1},\dots,x_{j+i-1})|\theta_{j,j+i|j+1,\dots,j+i-1}). \quad (7)$$

Differently from C-vines and D-vine the joint density function of the R-vine can vary significantly according to the statistical feature of the multivariate distribution being modeled. For this reason, Morales-Napoles (2011) and Dißmann (2010) proposed an efficient method for storing the indices of the pair-copula. It relies on the specification of a lower triangular matrix $M=(m_{i,j}|i,j=1,...,d)\in\{0,...,d\}^{dxd}$, whose diagonal entries $m_{i,i}$ are the nodes $1,...,d$ of the first tree. Each row from the

Table 1 Basic statistics of log return time series referred to the period January 4, 2012 - July 24, 2014

	Mean	**Max**	**Min**	**Std. Dev**	**Skewness**	**Kurtosis**
Brent	-0.01%	6.57%	-5.67%	1.26%	-0.068	2.063
Gasoil	-0.01%	5.08%	-3.11%	1.06%	0.049	1.077
JetF	-0.01%	5.05%	-3.98%	1.03%	0.075	1.599
Naphtha	0.00%	5.27%	-7.72%	1.30%	-0.440	3.654
Lsfo	-0.02%	3.54%	-6.80%	1.13%	-0.758	4.551
Gas NBP	-0.04%	9.10%	-7.69%	1.59%	0.294	3.620
Gas HenryHub	0.04%	13.27%	-11.93%	2.78%	0.290	2.342

bottom up represents a tree. The conditioned sets of a node are determined by a diagonal entry and the corresponding column entry of the row under consideration, while the the column entries below this row provides the conditioning set. Corresponding copula types can also be stored in matrices similar to M (see Section 3.2 for more details).

3 Data and dependence structure

3.1 Data analysis

With the aim of analyzing the re-negotiation of the European LTCs, we consider both the oil-based commodities traditionally used to determine the LTCs price and the main spot gas prices traded at the hub.

In particular the historical daily prices of the following assets have been taken into account: Crude Ice Brent DTD, Gasoil NWECIF, Jet Fuel NWECIF, Naphtha NWECIF, Lsfo 1% NWECIF, Gas NBP 1stMonth, Gas HenryHub 1stMonth. The label "NWE" stands for the reference market "North West Europe", while "CIF" indicates the "Cost, Insurance and Freight" that are the costs included in the prices. For the sake of simplicity, in the rest of the paper, we denote the seven time series as follows: "Brent", "Gasoil", "JetF", "Naphtha", "Lsfo", "Gas NBP" and "Gas HenryHub".

We analyze the period from the 4th of January 2012 to the 24th of July 2014 for a total of 647 observations. The first five time series, referred to oil and its by-products, are spot prices. The last two series, respectively referred to the natural gas traded at the NBP and Henry Hub spot markets, are the first month future prices that can be considered as good proxy of the gas spot prices. Data are provided by Datastream Thomson Reuters and all the numerical computations are run in R 3.3.2.

For each historical daily price time series, we construct the log returns and we report the basic statistics in 1. There is evidence of negative skewness and significant excess kurtosis.

Figure 1 shows the log returns, the 30-days horizon rolling standard deviation on log returns, and the volatility associated with the Gas NBP time series. In particular, figure 2b shows the volatility clustering, namely "*large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small* *changes*", as observed by Mandelbrot, 1963 (for the other series, see Figures 5-10 in Appendix A).

INSERT FIGURE 1

CAPTION: Log returns (a), 30-days horizon rolling standard deviation on log returns (b), and volatility (c) associated to the Gas NBP time series.

The volatility clustering is confirmed also by the sample autocorrelation function (ACF) of the squared mean adjusted log returns (see left hand side of Figures 17-19 in Appendix A). In addition, we observe that all log return series are stationary with respect to the ADF (Dickey and Fuller, 1981) and the PP tests (Phillips et al. 1988). The series are also stationary compared to the KPSS test (Kwiatkowski et al., 1992) with the exception of the Gas NBP log return series. The KPSS is negligible for the latter since only the ACF plot of the differentiated series shows the presence of unit root.

Table 2 Unit root test results

P-Value	Brent	Gasoil	JetF	Naphtha	Lsfo	Gas NBP	Gas HenryHub
ADF	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
PP	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
KPSS	> 0.1	> 0.1	> 0.1	> 0.1	> 0.1	0.013	> 0.1

For modeling the volatility of the series, we test several GARCH models and, for each one, we select the best model according to the significance of parameters, the log-likelihood value and the information criteria. In the case of extreme market events, dummy variables associated to these events have been used to model volatility spikes. Every series is filtered using a TGARCH model with a skewed-t distribution for innovations that allows to capture the asymmetry in volatility (i.e. the leverage effect). ARMA models have been used to compensate for autocorrelation, modeling the conditional mean where needed. Table 3 reports the results of the ARMA-GARCH fitting procedure. The application of the Ljung-Box test on both standardized residuals standardized squared residuals shows the absence of series correlation and the ARCH-LM test (see Fisher and Gallagher, 2012) confirms the adequacy of the ARCH processes. The ACF and the Partial ACF (PACF) of both the log return series and the corresponding residuals are reported in Figures 11-13 and 14-16 in Appendix A. This indicates that there is no evidence of autocorrelation at any lag.

The reduction of the volatility after fitting the ARMA-GARCH models is confirmed by the ACF of both the squared mean adjusted log return series and of the squared mean adjusted residuals reported in Figures 17-19 in Appendix A.

3.2 Pair-Copula Constructions

Using the probability integral transformation on the standardized residuals we obtain copula u-data. In order to avoid the misspecification of the margins that may lead

Table 3 P-value of the ARMA GARCH models. The relative statistics are reported in parentheses. WLB=Weighted Ljung-Box Test, WALM= Weighted Arch LM Test. S.R= Standardized Residuals. S.S.R= Standardized Squared Residuals

	Series	**Brent**	**Gasoil**	**JetF**	**Naphtha**	**Lsfo**	**Gas NBP**	**Gas HenryHub**
	GARCH Model	TGARCH(1,2)	TGARCH(1,1)	TGARCH(2,0)	TGARCH(1,2)	TGARCH(1,2)	TGARCH(1,1)	TGARCH(1,1)
	ARMA Model	ARFIMA(1,0,1)	ARFIMA(0,0,0)	ARFIMA (0,0,0)	ARFIMA(9,0,0)	ARFIMA(14,0,0)	ARFIMA(17,0,0)	ARFIMA(13,0,0)
WLB Test on S.R.	Lag[1]	0,9995 (4.364e-07)	0,9401 (0.005655)	0,7433 (0.1073)	0.9692 (0.001495)	0.8531 (0.03427)	0.8179 (0.05303)	0.7959 (0.0669)
	Lag[2*(p+q)+(p+q)-1]	1 (1.093)	0.9908 (0.008741)	0.9089 (0.1175)	1 (6.642716)	1 (10.24734)	1 (18.76113)	1 (7.8402)
	Lag[4*(p+q)+(p+q)-1]	0.6123 (4.330)	0.83 (1.1283)	0.533 (2.3745)	0.9993 (12.9167)	0,9974 (24.0307)	0.7594 (39.0382)	1 (16.8326)
WLB Test on S.S.R.	Lag[1]	0.2416 (1.371)	0.5683 (0.3256)	0.8934 (0.01796)	0.8808 (0.02248)	0.6537 (0.2013)	0.2902 (1.119)	0.3185 (0.995)
	Lag[2*(p+q)+(p+q)-1]	0.1808 (6.604)	0.9592 (0.4904)	0.9904 (0.22823)	0.6261 (3.32382)	0.286 (5.5291)	0.5591 (2.257)	0.225 (4.245)
	Lag[4*(p+q)+(p+q)-1]	0.2619 (9.213)	0.9871 (0.9792)	0.8128 (2.66476)	0.4987 (7.05544)	0.3085 (8.7125)	0.7089 (3.298)	0.3887 (5.285)
WALM Tests	ARCH Lag[4]	0.1355 (2.228)	0.8308 (0.04564)	0.7293 (0.1198)	0.3062 (1.047)	0.1477 (2.096)	0.1866 (1.745)	0.5247 (0.4047)
	ARCH Lag[6]	0.3469 (2.750)	0.9045 (0.42836)	0.9347 (0.3207)	0.5359 (1.802)	0.1811 (4.078)	0.5012 (1.868)	0.8632 (0.5685)
	ARCH Lag[8]	0.4318 (3.637)	0.9521 (0.73775)	0.856 (1.3236)	0.6851 (2.279)	0.2384 (5.100)	0.7118 (2.026)	0.9105 (1.0185)

to the bias of the copula parameter estimates, we perform the Berkowitz test on the u-data. The outcome of this test does not indicate evidence against the uniform (0,1).

Table 4 Kendall's τ correlation between u-data

	Brent.u	**Gasoil.u**	**JetF.u**	**Naphtha.u**	**Lsfo.u**	**Gas NBP.u**	**Gas HenryHub.u**
Brent.u	1.0000	0.6037	0.6002	0.3143	0.2750	0.0851	0.0213
Gasoil.u	0.6037	1.0000	0.7805	0.2887	0.2608	0.0847	0.0341
JetF.u	0.6002	0.7805	1.0000	0.2797	0.2524	0.0863	0.0408
Naphtha.u	0.3143	0.2887	0.2797	1.0000	0.2976	0.1004	-0.0428
Lsfo.u	0.2750	0.2608	0.2524	0.2976	1.0000	0.0990	-0.0025
Gas NBP.u	0.0851	0.0847	0.0863	0.1004	0.0990	1.0000	0.0111
Gas HenryHub.u	0.0213	0.0341	0.0408	-0.0428	-0.0025	0.0111	1.0000

Table 4 reports the pair-wise Kendall's τ correlation between the series. This allow us to construct the vine copula models using the maximum spanning tree algorithm described in Czado et al. (2012). Note that if the strength of dependence is rather small, a good start of a bivariate data analysis is an independent test based on Kendall's τ correlation measure (see Genest and Favre, 2007) In the following, we estimate and analyze three different vine structures, namely C-vine, D-vine, R-vine together with the multivariate Gaussian copula. Note that the latter can be represented as any R-vine with Gaussian pair-copulas, where the parameters are determined by the associated partial correlation (see Czado, 2010). The R package by Schepsmeier et al. (2018) has been used for the analysis conducted in this section.

INSERT Figure 2
CAPTION: First tree of the C-vine with a 5% confidence level. The letters reported between the root nodes indicate the type of the bivariate copulas used to model the dependence, while the numbers refer to the corresponding Kendall's τ correlation.

The C-vine copula model is the first vine used to account for the dependence structure. The application of the Kendall's τ independence test, with a confidence level of 5%, underlines the independence of the pair Gasoil-Gas HenryHub series as shown in the first C-vine tree of Figure 2. Akaike Information Criterion and Bayesian Information Criterion tests are used to select copulas. First, all available copulas are fitted with Maximum Likelihood Estimation (MLE). Then, the criteria are computed for all available copula families and the family with the minimum value is chosen.

The parameters of the PCC can be evaluated using any multivariate copula estimator such as the maximum pseudo likelihood (MPL) estimator or the inference function for margins (IFM). However the computational effort increases exponentially with the dimension. Therefore the sequential method proposed by Aas et al, 2009, is used to estimate the parameters level by level. The selected copulas in the first tree are estimated using the MLE method. To calculate the observations (i.e conditional distributions functions) of the second C-vine tree we used the H-functions (see (4) in Section 2). This procedure is iterated tree by tree. After having fitted all the trees, a joint MLE is provided in order to improve the estimation. This procedure requires the observations to be independent over time, in fact the PCC has been fitted on the standardized residuals obtained by filtering the original series with the ARMA-TGARCH models previously described.

In a straightforward way, we can represent the C-vine copula density factorization using the specification matrix M, while the T matrix is the copula type matrix where each row corresponds to a specific tree and each number denotes the type of pair-copula family.[3]

$$M = \begin{pmatrix} 7 & & & & & & \\ 6 & 6 & & & & & \\ 5 & 5 & 5 & & & & \\ 4 & 4 & 4 & 4 & & & \\ 3 & 3 & 3 & 3 & 3 & & \\ 2 & 2 & 2 & 2 & 2 & 2 & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$T = \begin{pmatrix} & & & & & & \\ 0 & & & & & & \\ 0 & 5 & & & & & \\ 5 & 5 & 2 & & & & \\ 0 & 0 & 1 & 1 & & & \\ 0 & 0 & 13 & 2 & 2 & & \\ 0 & 3 & 1 & 14 & 2 & 2 & \end{pmatrix}$$

We recall that Gaussian and Student-t copulas accounts for a symmetric tail dependence. In particular, Gaussian copula is designed to model the dependence in the center of the joint distributions, while the Student-t copula is both lower and upper-tail dependent. Student-t copula allows for joint extreme events either in both bivariate tails or none of them. Moreover, Frank copula models the dependence in the centre of distribution, i.e. strong dependence in non-extreme scenarios. With respect to the bivariate Gaussian, Frank copula can account for the non linearity in the center of the joint distribution, while the first one only focuses on linear dependence relationship. Finally, Clayton copula is only lower-tail dependent and is characterized by

[3] Copula family type: 0 = Independence copula; 1 = Gaussian copula; 2 = Student-t copula (t-copula); 3 = Clayton copula; 5 = Frank copula; 13 = rotated Clayton copula (180 degrees); 14 = rotated Gumbel copula (180 degrees); 16= Rotated Joe copula (180 degrees);

the asymmetric correlation. Together with the 180° rotated Joe and Gumbel is adequate to model greater dependence in the negative tail. The 180°rotated Clayton is suitable to model dependence in the positive tail.

The D-vine structure is reported in Figure 3. The results of the independence test based on the Kendall's τ correlation measure with a 5% confidence level points out the independence of the pair Gas NPB-Gas HenryHub as shown in the first tree of the D-vine structure in figure 3. In the following the specification matrix M and copula type matrix T are reported for the D-vine structure.

INSERT FIGURE 3

CAPTION: First tree of the D-vine with a 5% confidence level. The letters reported between the root nodes indicate the type of the bivariate copulas used to model the dependence, while the numbers refer to the corresponding Kendall's τ correlation.

$$M = \begin{pmatrix} 7 & & & & & & \\ 1 & 6 & & & & & \\ 2 & 1 & 5 & & & & \\ 3 & 2 & 1 & 4 & & & \\ 4 & 3 & 2 & 1 & 3 & & \\ 5 & 4 & 3 & 2 & 1 & 2 & \\ 6 & 5 & 4 & 3 & 2 & 1 & 1 \end{pmatrix}$$

$$T = \begin{pmatrix} 3 & & & & & \\ 0 & 5 & & & & \\ 0 & 0 & 1 & & & \\ 0 & 0 & 2 & 0 & & \\ 0 & 0 & 5 & 5 & 2 & \\ 0 & 1 & 2 & 1 & 2 & 2 \end{pmatrix}$$

INSERT FIGURE 4

CAPTION: First tree of the R-vine with a 5% confidence level. The letters reported between the root nodes indicate the type of the bivariate copulas used to model the dependence, while the numbers refer to the corresponding Kendall's τ correlation.

Finally, the R-vine structure is reported in Figure 4, while the respective specification matrix M and copula type matrix T are as follows.

$$M = \begin{pmatrix} 7 & & & & & & \\ 6 & 1 & & & & & \\ 5 & 6 & 2 & & & & \\ 1 & 5 & 6 & 3 & & & \\ 2 & 4 & 5 & 6 & 5 & & \\ 3 & 3 & 4 & 5 & 6 & 4 & \\ 4 & 2 & 3 & 4 & 4 & 6 & 6 \end{pmatrix}$$

$$T = \begin{pmatrix} 0 & & & & & \\ 0 & 0 & & & & \\ 0 & 0 & 0 & & & \\ 0 & 1 & 0 & 0 & & \\ 1 & 1 & 16 & 2 & 0 & \\ 1 & 2 & 1 & 2 & 2 & 0 \end{pmatrix}$$

The multivariate Gaussian copula is often used for modeling multivariate data, assuming a linear dependence structure and no tail dependencies. The corresponding T and M matrix representation is straightforward and its derivation is left to the reader. For model comparison, a commonly used goodness-of-fit (gof) measure is the value of the log-likelihood of the estimated copula models. This is reported in Table 5. We notice that the values of the three vine models are higher than the one obtained with the multivariate Gaussian copula by up to 178 points.

Table 5 Log-likelihood of the estimated copula models

	C-vine	D-vine	R-vine	Gaussian
Log-likelihood	1336	1321	1328	1158

Others classical comparison measures are AIC and BIC criteria that take into account the model complexity. Looking at the values reported in Table 6, we observe again that vine models outperform the multivariate Gaussian copula by up to 332 and 298 points for AIC and BIC, respectively. C-vine and D-vine are preferred in terms of AIC, while R-vine is preferred in terms of BIC due to its lower number of parameters.

Table 6 AIC and BIC of the estimated copula models

	C-vine	D-vine	R-vine	Gaussian
AIC	-2632	-2632	-2626	-2300
BIC	-2547	-2531	-2563	-2265

Moreover, the comparison of the three T matrices of the considered vine structures provides additional information on tail dependence. More specifically, about 5-30% of the selected pair-copulas have either lower or upper tail dependence, i.e Clayton, Gumbel and Joe, while 20-25% have both upper and lower tail dependence modeled with a Student's t copula. Furthermore, tail dependence is significant since the coefficient calculated from the estimated pair-copula is on average greater than 0.2. Vine-copula structures are strongly preferred in terms of log-likelihood, AIC, BIC with respect to the multivariate Gaussian copula. Moreover, C-vine, D-vine and R-vine account for the significance estimated tail dependence. For these reasons, we decide to exclude the multivariate Gaussian copula from further analysis.
As an additional step, we run the ECP and ECP2 gof tests on the three selected vine structures. These are non-parametric tests based on the Cramer-von Mises (CvM) and Kolmogorov-Smirnov (KS) test statistics (see Schepsmeier, 2014 for more details on the tests). The resulting p-values are reported in Table 7. None of the structures can be rejected at a 5% significance level, i.e all of them fit the data quite well. As suggested by Gaupp et al. (2017), future developments of improved goodness-of-fit tests are needed to help distinguishing between alternative structuring approaches.

We also perform the likelihood ratio tests by Vuong (1989) and Clarke (2007) to select the structure that better accounts for the dependence among the assets. The

Table 7 Goodness-of-fit test on C-vine, D-vine and R-vine with bootstrap repetition rate $x = 200$.

	ECP (CvM)	ECP (KS)	ECP2 (CvM)	ECP2 (KS)
D-vine	p=0.49 ts=1.43	p=0.93 ts=0.61	p=1 ts=0.01	p=0.94 ts=0.61
C-vine	p=0.42 ts=1.47	p=0.78 ts=0.68	p=1 ts=0.01	p=0.84 ts=0.68
R-vine	p=0.33 ts=1.53	p=0.38 ts=4.91	p=1 ts=0.01	p=0.97 ts=0.50

results for each possible couple of vine structures are reported in Table 8 and Table 9. These show that, in almost all cases, no decision among the models is possible, i.e the null hypothesis that both models are statistically equivalent cannot be rejected. The only exception is represented by the result of the Clark test where the R-vine is preferred to the C-vine.

Table 8 Vuong test results at level $\alpha = 5\%$

	Statistic	Statistic Akaike	Statistic Schwarz	P-Value	P-Value Akaike	P-Value Schwarz
D-vine VS C-vine	1.121	1.637	2.791	0.262	0.201	0.005
R-vine VS C-vine	-1.094	-0.413	1.107	0.273	0.679	0.267
R-vine VS D-vine	-1.641	-1.416	-0.911	0.100	0.156	0.361

Table 9 Clark test results

	Statistic	Statistic Akaike	Statistic Schwarz	P-Value	P-Value Akaike	P-Value Schwarz
D-vine VS C-vine	306	311	320	0.181	0.345	0.813
R-vine VS C-vine	293	302	329	0.018	0.098	0.694
R-vine VS D-vine	312	317	338	0.387	0.637	0.270

4 The optimal composition of long-term natural gas contract

4.1 Optimization portfolio problems

In the following, we summarize the optimization portfolio problems that we solve to compute the optimal weights of each underlying asset of a long-term natural gas contract under the minimum portfolio risk. We consider five well-known risk measures that are represented by Variance, Mean Absolute Deviation (MAD), MiniMax, Conditional Value-at-Risk (CVaR), and Conditional Drawdown at Risk (CDaR). We simulate the portfolio returns based on the dependence structures specified in the C-vine and D-vine models described above and estimate the risk of the seven-dimensional long-term natural gas contract. We assume to have M assets ($m = 1, ..., M$) and T time periods ($t = 1, ..., T$). Recall that, in our analysis, the assets are represented by the seven time series indicated in Section 3.1. More precisely, we denote with w_m the weights associated to each asset m of the portfolio; $r_{t,m}$ the return of each asset m

in time period t; μ_m the average return of asset m that is $\mu_m = \frac{1}{T}\sum_{t=1}^{T} r_{t,m}$ and μ_p the portfolio target return. The mean variance (EV) nonlinear optimization problem (see Markowitz, 1952) is the following:

$$\min_w \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{m=1}^{M} w_m(r_{t,m}-\mu_m)\right)^2 \tag{8a}$$

$$s.t. \tag{8b}$$

$$\sum_{m=1}^{M} w_m\mu_m = \mu_p \tag{8c}$$

$$\sum_{m=1}^{M} w_m = 1 \tag{8d}$$

$$w_m \geq 0 \quad \forall j = 1, \cdots, M. \tag{8e}$$

This optimization problem aims at minimizing portfolio variance (8a) under the portfolio target return (8c), assuming that the sum of the asset weights has to be equal to one (8d) and the non-negativity of weights w_m (8e).

We then consider the portfolio optimization model that is based on the MAD risk measure (see Konno et Al., 1993):

$$\min_w \frac{1}{T}\sum_{t=1}^{T}\left|\sum_{m=1}^{M}(r_{t,m}-\mu_m)w_m\right| \tag{9a}$$

$$s.t. \tag{9b}$$

$$\sum_{m=1}^{M} w_m\mu_m = \mu_p \tag{9c}$$

$$\sum_{m=1}^{M} w_m = 1 \tag{9d}$$

$$w_m \geq 0 \quad \forall m = 1, \cdots, M. \tag{9e}$$

Problem (9a)-(9e) can be transformed in the following linear optimization problem:

$$min_{w,y} \frac{1}{T} \sum_{t=1}^{T} y_t \tag{10a}$$

$$s.t. \tag{10b}$$

$$\left| \sum_{m=1}^{M} (r_{t,m} - \mu_m) w_m \right| \leq y_t \tag{10c}$$

$$\sum_{m=1}^{M} w_m \mu_m = \mu_p \tag{10d}$$

$$\sum_{m=1}^{M} w_m = 1 \tag{10e}$$

$$w_m \geq 0 \quad \forall m = 1, \cdots, M. \tag{10f}$$

The MiniMax model proposed by Young (1998) aims at maximizing the minimum return L_p, namely minimizing the maximum loss, defined as:

$$L_p = \min_t \left(\sum_{m=1}^{M} w_m r_{t,m} \right) \quad \forall t = 1, .., T.$$

On the basis of this assumption, the model is formulated as follows:

$$\max_{L_p, w} L_p \tag{11a}$$

$$s.t. \tag{11b}$$

$$\sum_{m=1}^{M} w_m r_{t,m} - L_p \geq 0 \quad \forall t = 1, .., T \tag{11c}$$

$$\sum_{m=1}^{M} w_m \mu_m = \mu_p \tag{11d}$$

$$\sum_{m=1}^{M} w_m = 1 \tag{11e}$$

$$w_m \geq 0 \quad \forall m = 1, \cdots, M. \tag{11f}$$

Following Rockafellar and Uryasev (2000), the portfolio optimization problem with respect to the CVaR measure can defined as follows:

$$\min_{w,d,v} \frac{1}{(1-\alpha)T} \sum_{t=1}^{T} d_t + v \tag{12a}$$

$$s.t. \tag{12b}$$

$$\sum_{m=1}^{M} w_m r_{t,m} + v \geq -d_t \quad \forall t = 1, .., T \tag{12c}$$

$$\sum_{m=1}^{M} w_m \mu_m = \mu_p \tag{12d}$$

$$\sum_{m=1}^{M} w_m = 1 \tag{12e}$$

$$w_m \geq 0 \quad \forall m = 1, \cdots, M \tag{12f}$$

$$d_t \geq 0 \quad \forall t = 1, \cdots, T. \tag{12g}$$

where v represents the VaR, $(1-\alpha)$ is the coverage rate and d_t is the deviation value below the VaR.

Following Chekhlov et al. (2005), the CDaR optimization problem is as follows:

$$\min_{w,u,v,z} \frac{1}{(1-\alpha)T} \sum_{t=1}^{T} z_t + v \tag{13a}$$

$$s.t. \tag{13b}$$

$$\sum_{m=1}^{M} w_m r_{t,m} + u_t - u_{t-1} \geq 0, \quad u_0 = 0 \quad \forall t = 1, .., T \tag{13c}$$

$$z_t - u_t + v \geq 0 \quad \forall t = 1, .., T \tag{13d}$$

$$\sum_{m=1}^{M} w_m \mu_m = \mu_p \tag{13e}$$

$$\sum_{m=1}^{M} w_m = 1 \cdots, M \tag{13f}$$

$$w_m \geq 0 \quad \forall m = 1, \cdots, M \tag{13g}$$

$$z_t \geq 0 \quad \forall t = 1, \cdots, T \tag{13h}$$

$$u_t \geq 0 \quad \forall t = 1, \cdots, T \tag{13i}$$

$$\tag{13j}$$

where z is an auxiliary vector of variables of the conditional drawdowns, u is the auxiliary vector of variables used to model the cumulative returns and v represents the Drawdown Risk at the quantile $(1-\alpha)$.

4.2 Results

In our analysis, we combine pair-vine copula models and portfolio optimization methods to define the optimal allocation of the underlying assets of a long-term natural gas contract. The integration of the PCC into the portfolio optimization allows to capture the complete multivariate dependence risk structure across the considered assets. As mentioned in Section 3.2, the PCC exploits the relationship between the pair-copula family and the corresponding Kendall's τ to compute the correlations coefficients among the assets. This methodology does not constraint the returns to be normal, but it captures the asymmetry and nonlinear dependence among the commodities. For each of five risk measures (EV, MAD, MiniMax, CVaR, and CDaR) we minimize the portfolio risk by fixing the same target return μ_p for the three structures. Tables 10, 11 and 12 report the optimal assets allocation for C-vine, D-vine and R-vine structures, respectively. These weights can be interpreted as the proportions to attribute to the different underlyings of the long-term natural gas contract, according to the minimum risk optimal portfolio.

Table 10 Optimal weights for long-term natural gas portfolio C-vine

	EV	**MAD**	**MiniMax**	**CVaR**	**CDaR**
Brent	0.13	0.14	0.26	0.13	0.38
Gasoil	0.06		0.06		
JetF	0.11	0.17	0.07	0.17	0.62
Naphtha	0.17	0.16	0.12	0.18	
Lsfo	0.31	0.31	0.27	0.30	
Gas NBP	0.12	0.13	0.18	0.13	
Gas HenryHub	0.09	0.09	0.04	0.09	
Min Risk	0.01%	0.03%	0.12%	0.07%	5.32%

Table 11 Optimal weights for long-term natural gas portfolio D-vine

	EV	**MAD**	**MiniMax**	**CVaR**	**CDaR**
Brent	0.16	0.17	0.26	0.17	0.14
Gasoil	0.06				
JetF	0.11	0.17		0.18	
Naphtha	0.16	0.16		0.16	0.19
Lsfo	0.29	0.28	0.58	0.28	0.41
Gas NBP	0.13	0.14	0.16	0.13	0.19
Gas HenryHub	0.09	0.08		0.09	0.07
Min Risk	0.01%	0.03%	0.14%	0.07%	3.67%

The analysis of the optimal asset allocation shows that, in general, there is a convergence in the weight of the same asset within the same risk measure among the three structures. This is in line with the findings of Section 3.2, where it is shown that all the three structures are appropriate for modeling the considered series.

Table 12 Optimal weights for long-term natural gas portfolio R-vine

	EV	**MAD**	**MiniMax**	**CVaR**	**CDaR**
Brent	0.16	0.16	0.03	0.17	0.27
Gasoil	0.06				0.59
JetF	0.12	0.19		0.31	0.17
Naphtha	0.14	0.14	0.24	0.14	
Lsfo	0.31	0.29	0.26	0.29	
Gas NBP	0.13	0.13	0.04	0.14	0.01
Gas HenryHub	0.09	0.09	0.11	0.09	0.14
Min Risk	0.01%	0.03%	0.13%	0.07%	4.59%

The combination of vine copula models with optimization methods leads to optimal portfolios with total risk close to zero.

By focusing on the single risk measure, we analyze the optimal asset allocation in the three structures. We first observe that the optimal asset allocation in the three vine copula models is very similar when applying the EV, MAD, and CVaR risk measures (compare Tables 10-12). In addition, these three risk measures lead to optimal portfolios with equal risk when analyzing the same risk measure. There is also a similar portfolio composition when comparing the aforementioned three risk measures within the same vine structure. Starting from the EV, the resulting portfolio is the one with the minimum risk. All assets are considered with Lsfo constituting approximately 30% of the portfolio, followed by Naphtha, Brent and Gas NBP. Gasoil plays instead a marginal role. This is registered in all the three vine copula models considered. A similar composition results by applying MAD and CVaR risk measures to the three structures. Lsfo is still the asset with the highest weight followed by Jetf, Naphtha and Brent; Gasoil is not included among the optimal underlyings.

The application of MiniMax and CDaR risk measures generates divergences in the components of the optimal portfolio compared to what obtained with the EV, MAD, and CVaR risk measures. A similar behavior for these risk measures is also observed in Bekiros et al. (2015). We recall that MiniMax considers the maximum loss in the portfolio, while CDaR takes into account a number of draw down events in the historical return distribution. Both measures are sensible to large losses occurring with low probability, which may differ in the simulated distributions of the three structures. The MiniMax portfolio includes Lsfo, Brent and Gas NBP with different weights in the three structures. In the CDaR portfolio, Brent is the only common underlying among the three structures that show also a significance difference in the total risk.

Table 13 re-elaborates the information provided by Tables 10-12 in order to better quantify the impact of oil-based commodities and spot gases within the optimal portfolios. In particular, the values denoted as “Oil-based commodities" are determined by summing up the optimal weights assigned to Brent, Gasoil, JetF, Naphtha, and Lsfo, while the terms “Spot Gases" results from the sum of the optimal weights of Gas NBP and Gas HenryHub.

It is worth noting that, when the EV, MAD and CVaR measures are applied, there is a perfect convergence in the composition of the optimal portfolios resulting from

Table 13 Oil and gas composition of the optimal portfolios

Copula	Asset	EV	MAD	MiniMax	CVaR	CDaR
C-vine	Oil-based commodities	0.78	0.78	0.78	0.78	1.00
	Spot Gases	0.22	0.22	0.22	0.22	0.00
D-vine	Oil-based commodities	0.78	0.78	0.84	0.78	0.74
	Spot Gases	0.22	0.22	0.16	0.22	0.26
R-vine	Oil-based commodities	0.78	0.78	0.85	0.78	0.85
	Spot Gases	0.22	0.22	0.15	0.22	0.15

the three structures: 78% is constituted by oil-based commodities and the remaining 22% is represented by gas traded on spot markets. On the contrary, the MiniMax and CDaR risk measures lead to optimal portfolios that differ in the three vine copula models, even though the oil-based underlyings still cover the larger share. This confirms the results discussed above.

A synthesis of the portfolio optimization indicates that oil-based commodities, such as Lsfo, Brent, Jetf, and Naphtha, appear to be fundamental picks in our asset allocation, together with the Gas NBP. This is an evidence of the important role still played by the oil-indexation in the long-term gas contracts. The influence of oil-based commodities in gas contracts is measured and included in our analysis through the pair-copulas. However, the choice of Gas NBP and Gas HenryHub gases, that together, on average, account for more that 20% of the optimal portfolio, reflects the fact that the re-negotiation policy advocated by mid-streamers in Europe is possible (see Franza, 2014).

5 Conclusions

This analysis takes inspiration from the current situation of the European natural gas market where both long-term contracts and hub spot price systems are applied. The fall of the European gas demand combined with the increase of the US shale gas exports and the rise of LNG availability on international markets have led to a reduction of the gas-hub prices in Europe. On the other side, oil-indexed long-term gas contracts failed to promptly adjust their positions implying significant losses for European gas mid-streamers that asked for a re-negotiation of their existing contracts and obtained new contracts linked also to hub spot prices. The debate over the necessity of the oil-indexed pricing is still ongoing and the main issue is that in the early days of the European gas industry this was the only option to mitigate the risk of launching such a highly capital expensive industry. The supporters of the gas-indexation state that nowadays the gas industry is mature enough and for this reason hub-based pricing reflects the true supply and demand dynamics in natural gas market. This paper investigates the dependence risk and the optimal resource allocation of the underlying assets of a long-term natural gas contract through pair-vine copulas and portfolio optimization methods with respect to five risk measures (EV, MAD, MiniMax, CVaR, and CDaR). In order to address the above mentioned debate both spot gas and oil commodities are included as underlyings. The usage of the PCC allows modeling the conditional dependence structure, overcoming the drawbacks of the mean-variance Markowitz optimization, including normally distributed returns and linear correlation among the assets of the same portfolio. The results of our simulations suggest that the weight allocation across portfolios obtained by implementing different vine structures, in almost all cases, converge within the same risk measure. A general finding is that oil commodities still cover the largest share of the optimal portfolios, but spot gas are also included. This suggests that European LTCs will most likely remain indexed to oil-based commodities. In other words, both spot gas and oil-based commodities can be included among the underlyings of long-term gas contract, but the latter will still exercise a major impact. This is a crucial point because increasing the share of spot gas in LTCs would artificially make long-term prices closer to hub price levels. From an economic perspective, this can leave room to a spiral mechanism of downward price adjustment to hub prices that are, in turns, influenced by long-term contract prices.

Acknowledgements

The work of the third author was supported by MIUR, Italy, PRIN MISURA 2010RHAHPL.

References

1. Aas, K., C. Czado, A. Frigessi and H. Bakken (2009) "Pair-copula constructions of multiple dependence". *Insurance: Mathematics and Economics*, 44:182-198.
2. Abada, I., A. Ehrenmann and Y. Smeers (2017) 'Modeling Gas Markets with Endogenous Long-Term Contracts". *Operations Research*, 65(4): 856-877.
3. Accioly, R.M.S. and F.A.L. Aiube (2008) "Analysis of crude oil and gasoline prices through copulas". *Cadernos do IME- Série Estatìstica*, 24:5-28.
4. Allen, D.E., M.A. Ashraf, M. McAleer, R.J. Powell and A.K. Singh (2013) "Financial dependence analysis: applications of vine copulas". *Statistica Neerlandica*, 67:403–435.
5. Aloui, R., M.S. Ben Aïssa and D.K. Nguyen (2013) "Conditional dependence structure between oil prices and exchange rates: a copula-garch approach". *Journal of International Money and Finance*, 32:719-738.
6. Arreola Hernandez, J. (2014) "Are oil and gas stocks from the Australian market riskier than coal and uranium stocks? Dependence risk analysis and portfolio optimization". *Energy Economics*, 45:528-536.
7. Bedford T. and R.M. Cooke (2001) "Probability density decomposition for conditionally dependent random variables modeled by vines". *Annals of Mathematics and Artificial Intelligence*, 32:245-268.
8. Bedford T. and R.M. Cooke (2002) "vines- a new graphical model for dependent random variables". *Annals of Statistics*, 30:1031-1068.
9. Bekiros, S., Hernandez, J. A., Hammoudeh, S., Nguyen, D. K. (2015) "Multivariate dependence risk and portfolio optimization: An application to mining stock portfolios.". *Resources Policy*, 46, 1-11.
10. Brechmann E.C. and C. Czado (2013) "Risk management with high-dimensional vine copulas: an analysis of the Euro Stoxx 50". *Stat. Risk Model*, 30:307-342.
11. Chekhlov A., S. Uryasev and M. Zabarankin (2005) "Drawdown measure in portfolio optimization". *International Journal of Theoretical and Applied Finance*, 8:13-58.
12. Cherubini U., E. Luciano and W. Vecchiato (2004) *Copula Methods in Finance.*. John Wiley: London.
13. Clarke, K. A. (2007). "A Simple Distribution-Free Test for Nonnested Model Selection". *Political Analysis* 15, 347-363.
14. Chyong C-K. (2015). "Markets and long-term contracts: The case of Russian gas supplies to Europe" *EPRG Working Paper*, 1524.

15. Czado, Claudia (2010) "Pair-copula constructions of multivariate copulas.". *Copula theory and its applications*, Springer, Berlin, Heidelberg, 2010. 93-109.
16. Czado C., F. Gärtner and A. Min (2011) "Analysis of Australian Electricity Loads using Joint Bayesian Inference of D-vines with Autoregressive Margins".
In Kurowicka, D., Joe. H.(Ed) Dependence Modeling: vine Copula Handbook, 265-280, World Scientific, London.
17. Czado C., U. Schepsmeier and A. Min (2012) "Maximum likelihood estimation of mixed C-vines with application to exchange rates". *Statistical Modelling*, 12(3):229-255.
18. Dalla Valle L. , M.E. De Giuli, C. Tarantola and C. Manelli (2016) "Default probability estimation via pair-copula constructions". *European Journal of Operational Research*, 249:298-311.
19. Dickey, D. and W. Fuller (1981) "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root". *Econometrica*, 49:1057-1072.
20. Dißmann, J., E. Brechmann, C. Czado and D. Kurowicka (2013) "Selecting and estimating regular vine copulae and application to financial returns". *Computational Statistics & Data Analysis*, 59:52-69.
21. Durante, F. and C. Sempi (2015) *Principles of copula theory*. Chapman and Hall/CRC: London
22. Fisher, T.J. and O. Gallagher C.M. (2012) "New weighted portmanteau statistics for time series goodness of fit testing". *Journal of the American Statistical Association*, 107(498):777–787.
23. Franza, L. (2014) "Long-term gas import contracts in Europe". *CIEP paper*, 8. Available at http://www.clingendaelenergy.com/inc/upload/files/Ciep_paper_2014-08_web_1.pdf.
24. Frisch, M. (2010) "Current European gas pricing problems: solutions based on price review and price reopener provisions.". *International energy law and policy research paper series, No 3*
25. Gaupp, F., Pflug, G., Hochrainer-Stigler, S., Hall, J., Dadson, S. (2017) "Dependency of crop production between global breadbaskets: a copula approach for the assessment of global and regional risk pools.". *Risk Analysis*, 37(11), 2212-2228.
26. Genest, C. and A.C. Favre (2007) "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask". *Journal of Hydrologic Engineering*, 12:347–368.
27. Genest, C. M. Gendron and M. Bourdeau-Brien (2009) "The Advent of Copulas in Finance". *The European Journal of Finance*, 15:609-618.
28. GIIGNL (2018) "The LNG industry". Available at https://giignl.org/publications#webform-client-form-1161
29. Grégoire, V., C. Genest and M. Gendron (2008) "Using copulas to model price dependencies in energy markets." *Energy Risk*, 5:58-64.
30. Heather, P. (2012) "Continental European Gas Hubs: Are they fit for purpose?." *NG: Oxford Institute for Energy Studies*. 63.
31. Heather P, B. Petrovich (2017). "European traded gas hubs: an updated analysis on liquidity, maturity and barriers to market integration". *The Oxford Institute for Energy Studies paper-Energy Insight*, 13.
32. IEA (International Energy Agency), (2013) Developing a Natural Gas Trading Hub in Asia-Obstacles and Opportunities. Available https://www.iea.org/publications/freepublications/publication/AsianGasHub_FINAL_WEB.pdf.
33. Jäschke S. (2014). "Estimation of risk measures in energy portfolios using modern copula techniques." *Computational Statistics and Data Analysis*, 76:359-376.
34. Joe, H. (1996) "Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters." *IMS lecture notes*, 76: 359-376.
35. Joe, H. (1997). *"Multivariate model and dependence concepts"*. Monographs on Statistics an Applied Probability, 73, Chapman, Hall, London.
36. Kanai, M. (2011) "Decoupling the Oil and the Gas Prices". *IFRI papers*. Available at http://www.ifri.org/sites/default/files/atoms/files/noteenergiemiharukanai.pdf.
37. Konno, H., Shirakawa, H., and Yamazaki, H. (1993). "A mean-absolute deviation-skewness portfolio optimization models." *Annals of Operations Research*, 45(1), 205-220.
38. Konoplyanik, A.A. (2010). "Evolution of Gas Pricing in Continental Europe: Modernization of Indexation Formulas Versus Gas to Gas Competition." , *University of Dundee, International Energy Law and Policy Research Paper Series*, 01. Available at http://www.konoplyanik.ru/ru/publications/articles/465_Evolution_of_Gas_Pricing_in_Continental_Europe.pdf
39. Krzemienowski A and Szymczyk S. (2016). "Portfolio optimization with a copula-based extension of conditional value-at-risk." *Annals of Operations Research*, 237.1-2 (2016): 219-236.

40. Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, (1992). "Testing the Null Hypothesis of Stationarity Against the Alternative of a UnitRoot" *Journal of Econometrics*, 54:159-178.
41. Kurowicka, D. and R. Cooke (2006). *"Uncertainty analysis with high dimensional dependence modelling"*. Wiley: Chichester.
42. Lu, X.F., K.K. Lai and L. Liang (2014). "Portfolio value-at-risk estimation in energy futures markets with time-varying copula-GARCH model" *Annals of Operations Research*, 219: 333-357.
43. Malevergne, Y. and D. Sornette (2006). *"Extreme financial risks: From dependence to risk management"*. Berlin: Springer-Verlag.
44. Mandelbrot, B.B. (1963). "The Variation of Certain Speculative Prices." *The Journal of Business*, 36:394-419.
45. Melling, A.J (2010). *"Natural gas pricing and its future-Europe ad the battleground."* Carnegie Endowment for international peace. Available at http://carnegieendowment.org/files/gas_pricing_europe.pdf
46. Morales-Napoles, O. (2010). "Counting vines" *Dependence modeling: vine copula handbook*, 189-218.
47. Nelsen, R.B. (1999). *"An introduction to copulas. Lecture Notes in Statistics, 139"*. Springer-Verlag: New York.
48. Phillips, P.C.B. and P. Perron (1988). "Testing for Unit Roots in Time Series Regression'.' *Biometrika*, 75:335-346.
49. Reboredo, J.C. (2011). "How do crude oil prices co-move?: A copula approach." *Energy Economics*, 33:95-113.
50. Rockafellar, R.T. and S. Uryasev (2000). " Optimization of conditional value-at-risk" *Journal of risk*, 2:21-42.
51. Sklar, M. (1959). "Fonctions de répartition á *n*dimensions et leurs marges". *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229-231.
52. Stern, J. and H.V. Rogers (2014). "The Dynamics of a Liberalised European Gas Market: Key determinants of hub prices, and roles and risks of major players." *The Oxford Institute for Energy Studies*, 94. Available at http://www.oxfordenergy.org/wpcms/wp-content/uploads/2014/12/NG-94.pdf.
53. Schepsmeier, Ulf (2014). "Efficient goodness-of-fit tests in multi-dimensional vine copula models". *ArXiv:1309.5808.*
54. Schepsmeier, Ulf, and Claudia Czado (2016). "Dependence modelling with regular vine copula models: a case-study for car crash simulation data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(3): 415-429.
55. Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., Erhardt, T. (2018). "vineCopula: statistical inference of vine copulas, 2018." Available at https://CRAN.R-project.org/package=vineCopula.
56. Theisen (2014). "Natural Gas Pricing in the EU: From Oil-indexation to a Hybrid Pricing System." *Regional Centre for Energy Policy Research*. Available at http://rekk.hu/downloads/projects/2014_rekk_natural\%20gas\%20pricing.pdf.
57. Vaz de Melo Mendes, B., M. Mendes Semeraro and R.P. Camara Leal (2010). "Pair-copulas modeling in finance". *Financial Market and Portfolio Management*, 24:193-213.
58. Vuong, Q.H. (1989). "Ratio tests for model selection and non-nested hypotheses". *Econometrica*, 57:307-333.
59. Wen, X., Y. Wei and D. Huang (2012). "Measuring contagion between energy market and stock market during financial crisis: a copula approach". *Energy Economics*, 34:1435-1446.
60. Wu, C.C., H. Chung and Y.H. Chang (2012). "The economic value of co-movement between oil price and exchange rate using copula-based GARCH models". *Energy Economics*, 34:270-282.
61. Xunpeng, S. (2016). "Gas and LNG pricing and trading hub in East Asia: An introduction". *Natural Gas Industry B*, 3:352-356.
62. Yafimava, K. (2014). "Outlook for the Long Term Contracts in a Globalizing Market (focus on Europe)" Presentation available at http://www.unece.org/fileadmin/DAM/energy/se/pp/geg/gif5_19Jan2015/s1_1_Yafimava.pdf
63. Young, M.R. (1998). "A minimax portfolio selection rule with linear programming solution". *Management science*, 44:673-683.

A Appendix: additional results

INSERT FIGURE 5
CAPTION: Log returns (a), 30-days horizon rolling standard deviation on log returns (b), and volatility (c) associated to the Brent series.

INSERT FIGURE 6
CAPTION: Log returns (a), 30-days horizon rolling standard deviation on log returns (b), and volatility (c) associated to the Gasoil series.

INSERT FIGURE 7
CAPTION: Log returns (a), 30-days horizon rolling standard deviation on log returns (b), and volatility (c) associated to the JetF series.

INSERT FIGURE 8
CAPTION: Log returns (a), 30-days horizon rolling standard deviation on log returns (b), and volatility (c) associated to the Naphtha series.

INSERT FIGURE 9
CAPTION: Log returns (a), 30-days horizon rolling standard deviation on log returns (b), and volatility (c) associated to the Lsfo series.

INSERT FIGURE 10
CAPTION: Log returns (a), 30-days horizon rolling standard deviation on log returns (b), and volatility (c) associated to the Gas HenryHub series.

INSERT FIGURE 11
CAPTION: ACF and PACF of Gas Nbp log return series.

INSERT FIGURE 12
CAPTION: ACF and PACF of Brent (a), Gasoil (b), and JetF (c) log return series.

INSERT FIGURE 13
CAPTION: ACF and PACF of Naphtha (a), Lsfo (b), and Gas HenryHub (c) log return series.

INSERT FIGURE 14
CAPTION: ACF and PACF of Gas Nbp residuals.

INSERT FIGURE 15
CAPTION: ACF and PACF of Brent (a), Gasoil (b), and JetF (c) residuals.

INSERT FIGURE 16
CAPTION: ACF and PACF of Naphtha (a), Lsfo (b), and Gas HenryHub (c) residuals.

INSERT FIGURE 17
CAPTION: ACF of the squared mean adjusted log return series and ACF of the squared mean adjusted residuals of Gas Nbp log return series.

INSERT FIGURE 18
CAPTION: ACF of the squared mean adjusted log return series and ACF of the squared mean adjusted residuals of Brent (a), Gasoil (b), and JetF (c) log return series.

INSERT FIGURE 19
CAPTION: ACF of the squared mean adjusted log return series and ACF of the squared mean adjusted residuals of Naphtha (a), Lsfo (b), and Gas HenryHub (c) log return series.