Journal of Economics and Finance (2023) 47:15–37
https://doi.org/10.1007/s12197-022-09609-4

Volatility and dependence in energy markets

Jinan Liu[1] · **Apostolos Serletis**[2]

Accepted: 7 November 2022 / Published online: 12 December 2022

Abstract
We use a semiparametric GARCH-in-Mean copula model to examine the price evolution and volatility dynamics of crude oil, natural gas, and hydrocarbon gas liquids markets using data from January 2002 to December 2021. We find that uncertainty has a positive and statistically significant effect on the returns of crude oil and natural gas, but has a negative and statistically significant effect on ethane returns. We also find that the Frank copula is the best copula to describe the (bivariate) dependence structures between the crude oil, natural gas, and hydrocarbon gas liquids markets, except for the relationship between ethane and butane where the Clayton copula is the most fitted copula. It suggests that weak lower and upper tail dependence exists between the energy returns, and there is statistically significant lower tail dependence between ethane and butane. In other words, extremely low crude oil prices are associated with low prices of natural gas and hydrocarbon gas liquids, and vice versa. When ethane returns go down, there is excess comovement in the returns of butane. Moreover, the tail dependence is strongest between crude oil and natural gas.

Keywords Copula · GARCH-in-Mean model · Crude oil price · Natural gas price · Hydrocarbon gas liquids prices

JEL Classification C58; F37; G17

1 Introduction

Energy prices have been variable in the past decades. The monthly crude oil price decreased 51% in February 2020, and increased 28% in April 2020. The wide price fluctuations in crude oil have contributed to the increased natural gas and hydrocarbon gas liquids prices and production. In December 2021, the natural gas production

✉ Apostolos Serletis
Serletis@ucalgary.ca

1 Department of Economics, University of Nebraska at Omaha, Omaha Nebraska, USA

2 Department of Economics, University of Calgary, Calgary, Alberta T2N 1N4, Canada

in the United States had reached the highest daily growth on record, with 118.8 billion cubic feet per day. Moreover, the United States has become the largest global source of hydrocarbon gas liquids supply and a major hydrocarbon gas liquids exporter. Current elevated levels of domestic oil and gas development have pushed hydrocarbon gas liquids production and price to an all-time high as of 2022.

Hydrocarbon gas liquids have not attracted enough attention in academic research. Hydrocarbon gas liquids, which include ethane, propane, butane, isobutane, and natural gasoline, produced in conjunction with natural gas, or as a by-product of crude oil refining, are used as inputs for petrochemical plants, burned for space heat and cooking, and blended into vehicle fuel. In 2020, hydrocarbon gas liquids accounted for about 18% of total petroleum consumption in the United States, and 90% of the total hydrocarbon gas liquids production in the United States is from natural gas processing.

Crude oil, natural gas, and hydrocarbon gas liquids markets can be disrupted by extreme geopolitical events that create uncertainty about future supply or demand, which can lead to higher volatility in prices and excess comovements of prices at extreme values. In this paper, we examine the volatility and dependence across hydrocarbon gas liquids, crude oil, and natural gas. In particular, we examine the following questions: what are the price evolution and volatility dynamics in the crude oil, natural gas, and hydrocarbon gas liquids markets? Does the volatility or uncertainty affect the mean of the energy returns? What is the bivariate dependence structure across the crude oil, natural gas, and hydrocarbon gas liquids markets? Is there any extreme value dependence in these markets? Is the dependence symmetric or asymmetric? By answering these questions, we hope to get a better understanding of the comovements of the crude oil, natural gas, and hydrocarbon gas liquids markets and the risks associated with the dependence structure between these markets.

Extensive volatility and correlation analyses have been performed in the traditional energy markets. For example, Ewing et al. (2002), Efimova and Serletis (2014), and Serletis and Xu (2016) use multivariate GARCH models to estimate the volatility and time-varying dependence structure in energy markets, including the crude oil, natural gas, coal, and electricity markets. However, the multivariate GARCH model is often based on severe restrictions to guarantee a well-defined covariance matrix. First, it is assumed that the stochastic term follows an elliptical (Gaussian or Student's t-) distribution with linear dependence. However, energy data are usually non-elliptically distributed. In this regard, Jahan and Serletis (2019) show that crude oil, natural gas, and hydrocarbon gas liquids returns are skewed, leptokurtic, and fat-tailed. Using an elliptical model to estimate non-elliptical data and inference the potential nonlinear dependence based on linear correlations can be very misleading. Moreover, the actual relationship between crude oil, natural gas, and hydrocarbon gas liquids markets is possibly nonlinear and asymmetric, but the multivariate GARCH models can only measure the linear correlations among variables. The linear correlation coefficient does not carry any information on how the markets are related differently in tranquil periods and in volatile periods, and it fails to model the structure of dependence and the tail dependence. For example, crude oil returns appear to be more related to natural gas returns when the energy markets are highly volatile compared to normal times.

In this paper, we use copula-based GARCH-in-Mean models to address the drawbacks of standard multivariate GARCH analysis. According to Sklar's (1973) theorem, any joint distribution function can be decomposed into its marginal distributions and a copula that describes the dependence between the variables. A copula function can be used to connect the univariate distributions of each energy return to restore the joint distribution of energy returns. First, in contrast to restrictions of multivariate GARCH models on the marginal distributions, copulas do not impose any restrictions on the marginal distributions and even allow marginals to be from different distribution families. Second, compared to linear correlation, copulas are a more informative measure of dependence between two (or more) variables, as they can capture the nonlinear dependence of the marginals. Copulas contain information about the joint behavior of the random variables in the tails of the distribution, which allows us to examine the changes in the dependence structure when extreme values and rare events occur. Third, similar to Chen and Fan (2006), we can use GARCH-in-Mean models and copulas to construct flexible multivariate distributions, exhibiting rich patterns of tail behavior, ranging from tail independence to tail dependence, and different types of asymmetry. Thus, the copula-based GARCH-in-Mean model allows for better flexibility in modeling joint distributions than standrad multivariate GARCH models.

Copulas have been widely used in the finance literature and have been gradually introduced into the empirical analysis of energy markets. Patton (2006), Rodriguez (2007), and Ning (2010) have used copulas to analyze the dependence structure between financial markets and the foreign exchange market. Wu et al. (2012) and Aloui et al. (2013) use a copula-GARCH approach to study the conditional dependence structure between crude oil prices and U.S. dollar exchange rates. Reboredo (2011) uses copulas to examine the dependence structure between benchmark crude oil prices. Tong et al. (2013) investigate the tail dependence and the asymmetry in the propagation of crises (bubbles) between the crude oil market and the refined petroleum markets based on copula models and find evidence of both positive lower and upper tail dependencies between these markets.

In this paper, we investigate the volatility and bivariate dependence between the returns of crude oil, natural gas, and hydrocarbon gas liquids by combining copula functions with GARCH-in-Mean models. We use a GARCH-in-Mean model to study the volatility and price evolution of crude oil, natural gas, and hydrocarbon gas liquids. To estimate the bivariate dependence structure between crude oil, natural gas, and hydrocarbon gas liquids returns, we apply various copulas on the GARCH-in-Mean filtered returns of crude oil, natural gas, and hydrocarbon gas liquids, and select the one with the best goodness of fit based on the Akaike Information Criterion (AIC). We find that volatility has a statistically significant positive effect on crude oil and natural gas returns, but has a statistically significant negative effect on the returns of ethane. We find that the Frank copula is superior to the asymmetric copulas in terms of the description of the bivariate dependence structure between crude oil, natural gas, and hydrocarbon gas liquids returns, suggesting that there is both upper tail and lower tail dependence structure in the energy markets. Clayton copula is the best to capture the dependence structure between ethane and butane. The tail dependence provides a measure of the probability of simultaneous extreme losses. The lower tail dependence and the likelihood of extreme joint losses suggest a higher than normal value-at-risk. The dependence parameter is the highest between crude oil and natural gas.

Our contribution to the literature is three-fold. First, we fill the gap in examining the volatility and price evolution of hydrocarbon gas liquids returns. Using the GARCH-in-Mean model, we find that volatility has a statistically significant negative effect on the returns of ethane, but does not have any statistically significant effects on the returns of butane and propane. Second, we use the copula-based models to analyze the bivariate dependence structure of crude oil, natural gas, and hydrocarbon gas liquids returns. The copula-based models can be used to capture the potential asymmetric and tail dependence between crude oil, natural gas, and hydrocarbon gas liquids returns. We test for both the degree and type of their dependence at extreme levels conditionally on the possibility of extreme events such as market crashes. The tail dependence enables us to examine how crude oil, natural gas, and natural gas liquids returns are related to each other during bearish and bullish markets. Third, the combination of GARCH-in-Mean models and copula functions is of particular interest because they capture a richer volatility and dependence structure than the standard multivariate GARCH framework. A copula is able to describe the dependence structure of marginals from different families of distributions. The volatility and price evolution of crude oil, natural gas, and hydrocarbon gas liquids are very different yet might have common extreme variations. The copula-GARCH model enables us to capture some of the essential empirical features of the data, such as the nonlinear dependence, skewness, and fat tails, while allowing each marginal distribution to vary considerably.

The rest of the paper is organized as follows. Section 2 presents the data. Section 3 presents the empirical results of univariate volatility analysis of crude oil, natural gas, ethane, propane, and butane returns. Section 4 presents the copula approach to the investigation of (bivariate) nonlinear dependence structures as well as tail dependence between the energy returns. The last section concludes.

2 The data

We retrieve the monthly price data of crude oil, natural gas, ethane, propane, and isobutane from Bloomberg and the United States Energy Information Administration. Prices are monthly averages of close-of-day spot prices. The crude oil price is the Brent crude oil price; the natural gas price is the Henry Hub natural gas price; the ethane, propane, and isobutane prices are at Mt. Belvieu non-LST (Lone Star Terminal). Our sample period is from January 2002 to December 2021 with 240 observations.

We compute the return series by taking logarithmic first differences of the monthly prices, that is, $r_t = 100 \times (log P_t - log P_{t-1})$. The log prices and return series are plotted in Figures 1, 2, 3, 4 and 5, with shaded areas indicating NBER recessions. Figure 6 compares the historical evolution of the log prices of crude oil, natural gas, ethane, propane, and butane over the sample period.

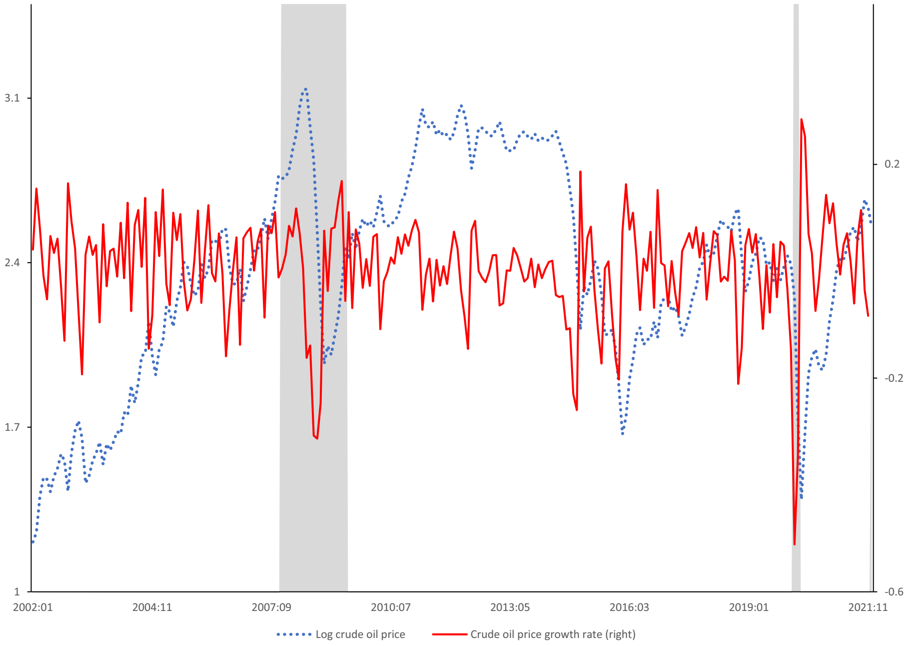

Fig. 1 Log crude oil price and its growth rate

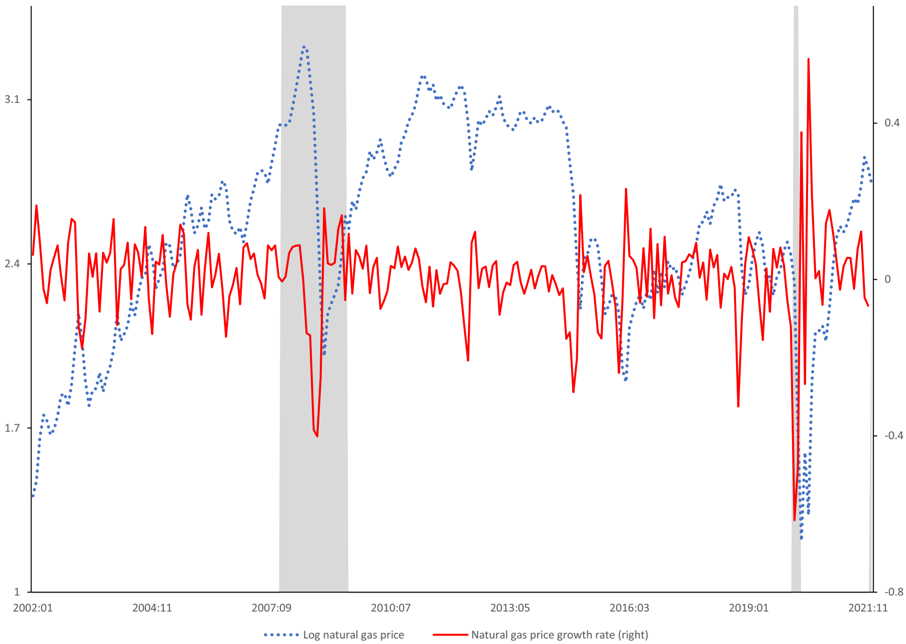

Fig. 2 Log natural gas price and its growth rate

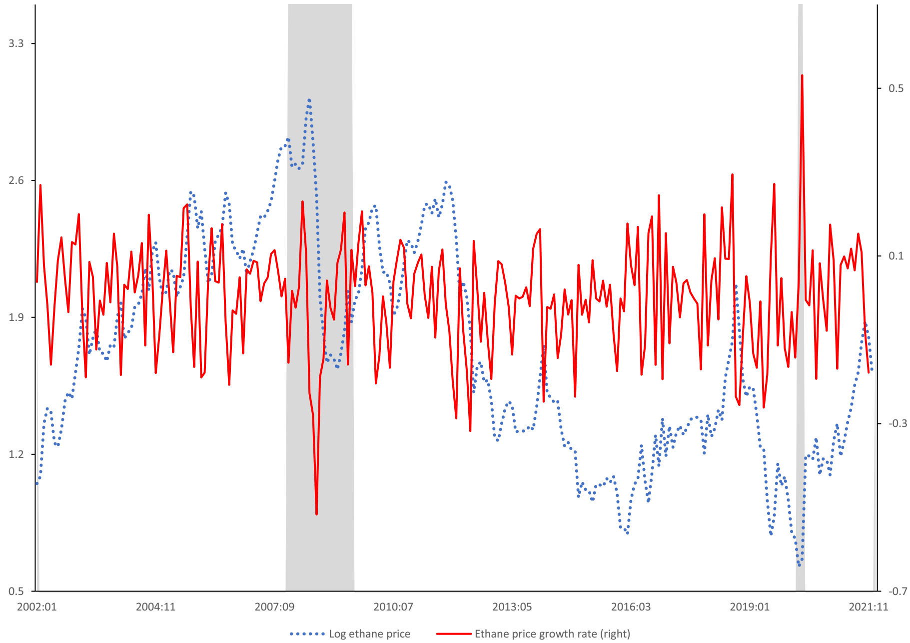

Fig. 3 Log ethane price and its growth rate

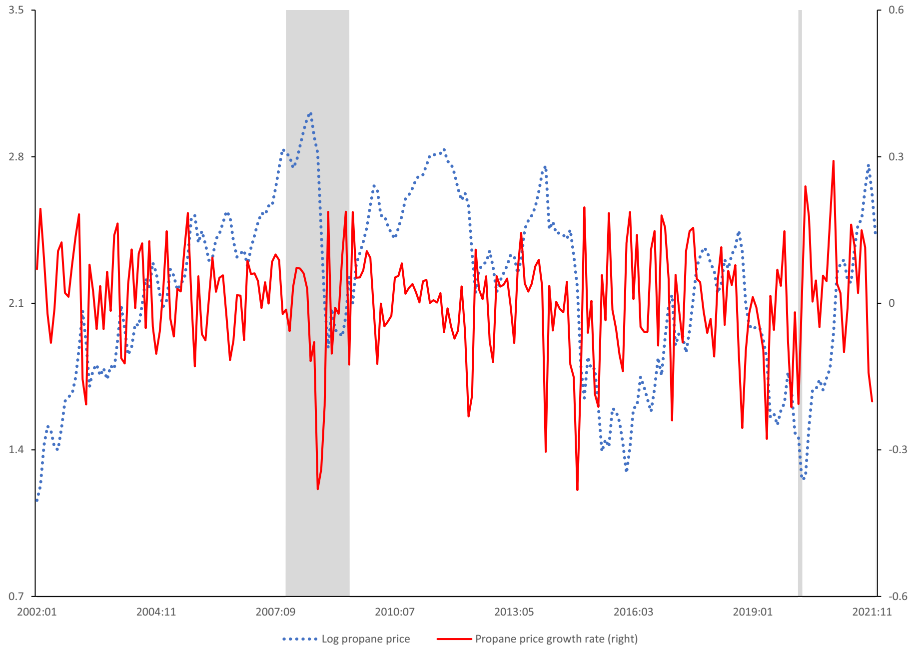

Fig. 4 Log propane price and its growth rate

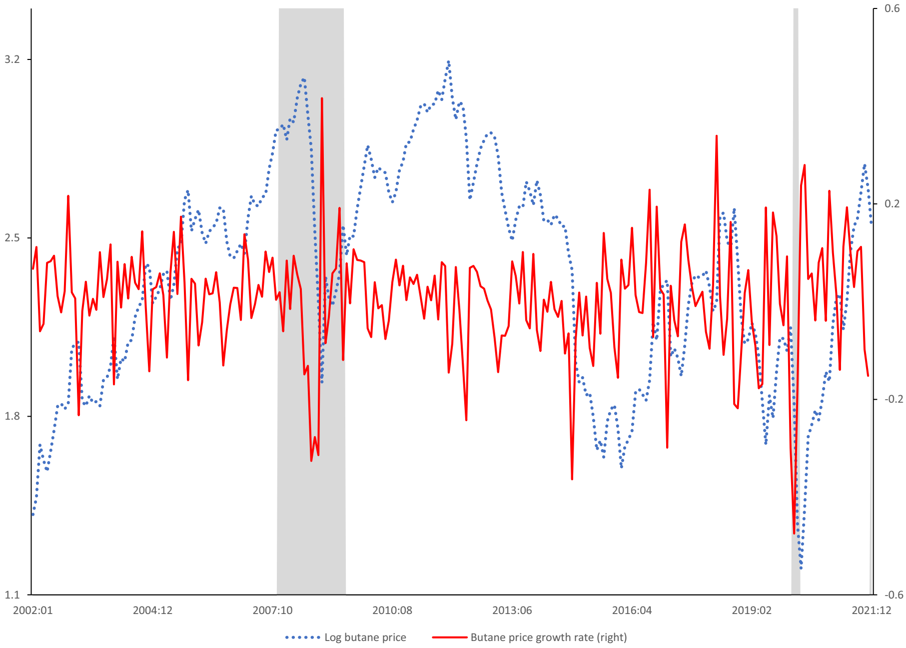

Fig. 5 Log butane price and its growth rate

As shown in Figure 6, until 2009, United States spot prices for natural gas and hydrocarbon gas liquids tracked the price of crude oil closely. According to the United States Energy Information Administration, in 2021, one-third of the United States energy consumption was from natural gas supply, and the United States is the world's largest producer of natural gas. Due to the surging demand in China and the rest of Asia, the global demand for liquefied natural gas has hit record highs each year since 2015. Much of that global appetite has been met by the steadily rising exports of liquefied natural gas from the United States, which have reached new records every year since 2016 and are poised to continue in 2022.

Hydrocarbon gas liquids prices in the United States followed the crude oil prices closely and were bound by international market dynamics until the 2007-2009 financial crisis. This historical relationship, was based on the general assumption that most fuels are interchangeable, and the United States was a net importer of hydrocarbon gas liquids. Since the 2007-2009 financial crisis, the hydrocarbon gas liquids prices began to move away from crude oil prices. Such a divergence reflected the growth production of hydrocarbon gas liquids in the United States and the switch position of the United States from a net importer to a net exporter.

By 2013 and 2014, the continuing increase in the production of propane further depressed the price of propane in the United States. The main consumption of propane in the United States is as a fuel, usually in areas where access to natural gas is limited. Isobutane prices began to fall closer to propane since 2013. The price of ethane delinked from the price of crude oil starting in 2012, and began to follow the

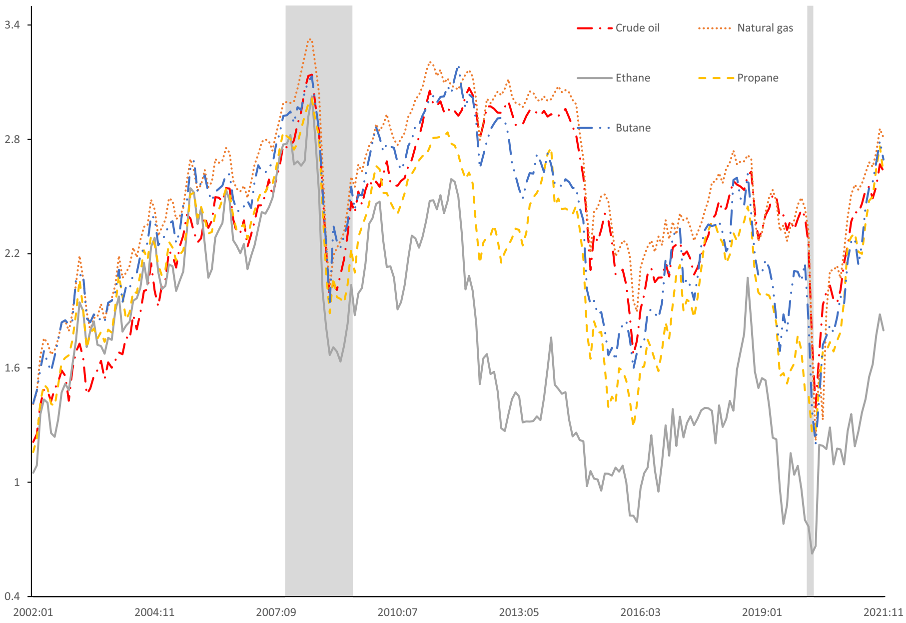

Fig. 6 Log energy prices

price of natural gas closely. Such change is due to the production process of ethane and the lack of alternative markets for ethane, which left natural gas processors with the only option of leaving the ethane as a component of pipeline natural gas and therefore setting ethane prices at the natural gas heating value. The price of ethane started to move away from the link to natural gas prices since late 2017 due to the expansion of ethane export capacity, which allows United States ethane products to reach more distant markets. The ethane consumption in the United States has increased over the past several years due to the lowered cost and increased supply. Ethane is mainly used to produce plastics.

Clearly, there are trend comovements among the energy prices during the recessions and booms. During the 2007-2009 financial crisis and the 2020 Covid-19 recession, all the energy prices were crushed. The high volatility in world oil prices and energy prices in the past decade demonstrates the uncertainty in the global markets.

Descriptive statistics and distributional characteristics of the log price and return series are reported in Table 1. For the price series, the standard deviation of the ethane price is the highest, followed by crude oil, butane, natural gas, and propane. For the return series, the standard deviation of ethane returns is the highest, followed by natural gas, butane, propane, and crude oil. All the return series are negatively skewed at a statistically significant level, except for ethane. The negative skewness indicates that return series are skewed to the left. With respect to the excess kurtosis statistics, the values of all energy returns are positive, with the most pronounced being the ones for natural gas and crude oil, implying that the distribution of returns has larger,

Table 1 Summary statistics

	Mean	Standard deviation	Skewness	Kurtosis	Normality
			A. Log prices		
Crude oil	2.364	0.446	−0.403 (0.011)	−0.506 (0.115)	9.057 (0.011)
Natural gas	2.538	0.423	−0.457 (0.004)	−0.178 (0.579)	8.674 (0.013)
Ethane	1.719	0.538	0.232 (0.145)	−0.930 (0.004)	10.800 (0.005)
Propane	2.169	0.418	−0.282 (0.076)	−0.765 (0.017)	9.028 (0.011)
Butane	2.350	0.427	−0.280 (0.257)	−0.812 (0.011)	7.902 (0.019)
			B. Returns		
Crude oil	0.006	0.098	−1.251 (0.000)	4.142 (0.000)	233.148 (0.000)
Natural gas	0.006	0.116	−0.937 (0.000)	7.200 (0.001)	551.220 (0.000)
Ethane	0.002	0.130	−0.197 (0.215)	1.267 (0.000)	17.529 (0.000)
Propane	0.005	0.109	−0.678 (0.000)	1.185 (0.000)	32.282 (0.000)
Butane	0.005	0.113	−0.519 (0.001)	2.482 (0.000)	72.063 (0.000)

Monthly data: 2002:01-2021:12 (T=240). Numbers in parentheses are p-values

thicker tails than the normal distribution. It indicates that the probability of extreme realizations could be higher than that of a normal distribution. The (Jarque and Bera 1980) test rejects the null hypothesis of normality for all the energy return series.

We also conduct a set of unit root and stationary tests for each of the logarithmic energy prices. Panel A of Table 2 shows that the null hypothesis of the presence of a unit root for all the (log) energy price series cannot be rejected by the Augmented

Table 2 Unit root and stationary tests

Series	ADF	PP	KPSS	Decision
A. Log prices				
Crude oil	−2.968	−2.675	0.755	$I(1)$
Natural gas	−3.283	−3.108	0.476	$I(1)$
Ethane	−3.337	−3.211	0.476	$I(1)$
Propane	−3.336	−3.069	0.521	$I(1)$
Butane	−3.042	−2.895	0.675	$I(1)$
B. Returns				
Crude oil	−10.996	−10.781	0.054	$I(0)$
Natural gas	−12.223	−12.162	0.053	$I(0)$
Ethane	−13.204	−13.218	0.085	$I(0)$
Propane	−11.358	−11.198	0.119	$I(0)$
Butane	−12.591	−12.532	0.073	$I(0)$

Monthly data: 2002:01-2022:12 (T=240). The 1% and 5% critical values are −4.000 and −3.430 for the ADF test and the PP test, and 0.216 and 0.146 for the KPSS test

Dickey Fuller (ADF) test (see Dickey and Fuller 1981) nor the Phillips-Perron (PP) test (see Phillips and Perron 1988), suggesting nonstationarity in the price series. The optimal lag length in the ADF test is selected based on the Bayesian information criterion (BIC) with a maximum lag length of 4. Moreover, given that unit root tests have low power against trend stationary alternatives, we also use the KPSS test (see Kwiatkowski et al. 1992) to test the null hypothesis of stationarity around a trend. As shown in panel A of Table 2, the null hypothesis of trend stationarity is rejected at the 1 percent statistical significance level. We thus conclude that none of the log energy price series is stationary.

We repeat the unit root and stationary tests on the first differences of the logarithms of the energy price series. As shown in panel B of Table 2, the ADF and PP tests reject the null hypotheses of a unit root for all the return series. Moreover, the KPSS test cannot reject the null hypothesis of stationarity for all the return series, suggesting that all the energy return series are stationary. Therefore, the energy price series are integrated of order one, I(1), and the energy return series are integrated of order zero, I(0).

3 Univariate volatility analysis

We adopt the GARCH-in-Mean model, developed by Engle (1982) and Bollerslev (1987), to examine the evolution path and volatility of the crude oil, natural gas, ethane, propane, and butane return series. The GARCH-in-Mean model provides a natural and convenient way to model the dynamic trade-off between expected return and risk by including the conditional standard deviation term into the conditional mean equation. The (univariate) GARCH-in-Mean model is commonly used in financial time series analysis, and it allows volatility to directly affect the conditional mean. It is given by

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \psi \sqrt{h_t} + \epsilon_t \tag{1}$$

$$h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{2}$$

$$\epsilon_t = \sqrt{h_t} v_t$$

$$v_t \sim N(0, 1) \tag{3}$$

where y_t is the return of the crude oil, natural gas, ethane, propane, and butane prices, respectively. ϵ_t is the shock, and h_t is the conditional variance of returns at time t. The standardized innovations, v_t, is distributed with $E(v_t) = 0$ and $E(v_t^2) = 1$. Equation 1 gives the relationship between the expected return and the risks. Equation 2 provides the dynamics of the conditional variance assuming a GARCH(1,1) process.

We examine univariate ARMA(1, 1)-GARCH(1,1)-in-Mean models for each of the energy return series. As can be seen in panel A of Table 3, the GARCH-in-Mean term is positive and statistically significant in the case of crude oil (0.280 with a p-value 0.000) and natural gas (0.196 with a p-value 0.000), suggesting that the conditional volatility has a positive and statistically significant effect on crude oil and

Table 3 Univariate GARCH-in-Mean models

Coefficient	Crude oil	Natural gas	Ethane	Propane	Butane
A. Conditional mean equation					
constant	−0.010 (0.044)	0.000 (0.000)	0.192 (0.000)	0.011 (0.138)	−0.016 (0.000)
y_{t-1}	0.290 (0.000)	−0.063 (0.000)	0.031 (0.663)	0.057 (0.462)	0.025 (0.940)
ϵ_{t-1}	−0.042 (0.594)	0.180 (0.000)	0.074 (0.301)	0.234 (0.002)	0.138 (0.679)
$\sqrt{h_t}$	0.280 (0.000)	0.196 (0.000)	−1.500 (0.000)	−0.003 (0.965)	0.278 (0.630)
B. Conditional variance equation					
constant	0.004 (0.000)	0.000 (0.000)	0.007 (0.000)	0.004 (0.000)	0.008 (0.041)
ϵ^2_{t-1}	0.573 (0.000)	0.959 (0.000)	0.129 (0.000)	0.303 (0.000)	0.368 (0.124)
h_{t-1}	0.010 (0.870)	−0.019 (0.000)	0.474 (0.000)	0.372 (0.000)	0.023 (0.958)
C. Standardized residual diagnostics					
$\hat{\epsilon}$ mean	−0.107	−0.150	−0.000	−0.045	−0.055
$\hat{\epsilon}$ standard error	0.996	1.006	1.002	1.002	1.001
$Jarque-Bera$	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$Q(8)$	(0.269)	(0.022)	(0.775)	(0.072)	(0.781)
$Q^2(10)$	(0.706)	(0.840)	(0.419)	(0.712)	(0.803)
AIC	−2.103	−2.008	−1.245	−1.679	−1.616

Monthly data: 2002:01-2021:12 (T=240). Numbers in parentheses are p-values

natural gas returns. However, the GARCH-in-Mean term is negative and statistically significant in the case of ethane (−1.500 with p-value 0.000). The GARCH-in-Mean term is not statistically significant for propane or butane returns.

Panel B of Table 3 shows the parameter estimates in equation (2). The parameter α_1 in front of the ARCH terms in the variance specification is statistically significant for all return series except for Butane (0.573 with a p-value 0.000 for crude oil, 0.959 with a p-value 0.000 for natural gas, 0.129 with a p-value 0.000 for ethane, 0.303 with a p-value 0.000 for propane). The parameter β_1 in front of the GARCH term h_{t-1} is statistically significant in the case of natural gas (−0.019 with a p-value of 0.000), ethane (0.474 with a p-value of 0.000), and propane (0.372 with a p-value of 0.000), reflecting the persistence of volatility in natural gas, ethane, and propane returns.

Panel C of Table 3 reports diagnostic test statistics based on the standardized residuals, $\hat{\epsilon}_t = \epsilon_t/\sqrt{h_t}$. The Ljung-Box Q test cannot reject the null hypothesis that the residuals are independently distributed (with p-values of 0.269, 0.022, 0.775, 0.072, and 0.781 for crude oil, natural gas, ethane, propane, and butane, respectively). Also, the McLeod-Li Q^2 test cannot reject the null hypothesis that the squared residuals are independently distributed (with p-values of 0.706, 0.840, 0.419, 0.712, and 0.803 for crude oil, natural gas, ethane, propane, and butane, respectively). Both diagnostic tests suggest that the standardized residuals are serially uncorrelated and are approximately i.i.d. However, the Jarque-Bera test rejects the null hypothesis that the residuals are normally distributed. Overall, the diagnostic tests show that the GARCH-in-Mean model can capture the nonlinearity in the conditional variance and is correctly specified for each of the five return series.

4 Correlation and dependence analysis

Copulas are a powerful tool for modelling nonlinear dependence between random variables, and in particular dependence at extremely values and in the tails of the distributions. Two measures of tail dependence related to copulas, known as the upper and the lower tail dependence coefficients, are particularly helpful for measuring the tendency of random variables to move together—Trivedi and Zimmer (2007). Upper tail dependence, λ_U, and lower tail dependence, λ_L, are defined as

$$\begin{aligned} \lambda_U &= \lim_{a\to 1} Pr[\epsilon_{2t} > F_2^{-1}(a)|\epsilon_{1t} > F_1^{-1}(a)] \\ \lambda_L &= \lim_{a\to 0} Pr[\epsilon_{2t} \le F_2^{-1}(a)|\epsilon_{1t} \le F_1^{-1}(a)]. \end{aligned}$$

where $F^{-1}(q) = inf\{x \in R : F(x) \ge a\}$, that is, the inverse of the cumulative probability distribution function for a. Tail dependence measures the probability that one event is extreme conditional on another extreme event. When $\lambda_L = \lambda_U$, there is symmetric tail dependence. We can interpret $\lambda_U^{c_1} > \lambda_U^{c_2}$ as copula c_1 is more concordant than copula c_2.

According to Sklar's Theorem (1973), for continuous multivariate distributions, the modeling of the univariate marginals and the dependence structure can be separated, and the multivariate structure can be represented by a copula. Copulas can be used to express a multivariate distribution in terms of its marginal distributions—see Joe (2014, p. 7). In this paper, we consider a number of copulas which are widely used in the literature. They are the Gaussian copula, the Clayton (1978) copula, the Gumbel (1960) copula, and the Frank (1979) copula.

4.1 The Gaussian copula

The copula of $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t})$ is assumed to be the normal copula with unknown correlation matrix $\boldsymbol{\Sigma}$. Let Ψ denote the univariate standard normal distribution and $\Psi_{\boldsymbol{\Sigma},2}$ the bivariate normal distribution with correlation matrix $\boldsymbol{\Sigma}$. Then the bivariate normal copula with correlation matrix $\boldsymbol{\Sigma}$ is

$$C(\boldsymbol{u}; \boldsymbol{\Sigma}) = \Psi_{\boldsymbol{\Sigma},2}(\Psi^{-1}(u_1), \Psi^{-1}(u_2)).$$

More explicitly,

$$C(u_1, u_2; \alpha) = \int_{-\infty}^{\Psi^{-1}(u_1)} \int_{-\infty}^{\Psi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\alpha^2}} \exp\left(-\frac{\epsilon_{1t}^2 - 2\alpha\epsilon_{1t}\epsilon_{2t} + \epsilon_{2t}^2}{2(1-\alpha^2)}\right) d\epsilon_{1t} d\epsilon_{2t}$$

where $u = (u_1, u_2)$, and $u_1 = \Psi(\epsilon_{1t})$ and $u_2 = \Psi(\epsilon_{2t})$ are the cumulative distribution functions of ϵ_{1t} and ϵ_{2t}, respectively. The copula dependence parameter, $\alpha \in (-1, 1)$, is the collection of all the unknown correlation coefficients in $\boldsymbol{\Sigma}$. If

$\alpha \neq 0$, then the normal copula generates joint symmetric dependence, but no tail dependence.

4.2 The Clayton copula

For $0 < \alpha < \infty$, the Clayton copula (1978) can capture the lower tail dependence. The Clayton copula takes the form:

$$C(u_1, u_2, \alpha) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-1/\alpha}.$$

The density of the Clayton copula is

$$c(u_1, u_2; \alpha) = \frac{(1+\alpha)(u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}-2}}{(u_1 u_2)^{\alpha+1}}.$$

The lower tail dependence can be calculated as $\lambda_L = 2^{-1/\alpha}$.

4.3 The Gumbel copula

For $1 \leq \alpha < \infty$, the Gumbel (1960) copula takes the form:

$$C(u_1, u_2; \alpha) = \exp\left\{-[(-\ln u_1)^{\alpha} + (-\ln u_2)^{\alpha}]^{1/\alpha}\right\}.$$

The density of the Gumbel copula is

$$c(u_1, u_2; \alpha) = \frac{C(u_1, u_2; \alpha)(\ln u_1 \ln u_2)^{\alpha-1}\{[(-\ln u_1)^{\alpha} + (-\ln u_2)^{\alpha}]^{1/\alpha} + \alpha - 1\}}{u_1 u_2[(-\ln u_1)^{\alpha} + (-\ln u_2)^{\alpha}]^{2-1/\alpha}}.$$

The Gumbel copula can capture positive upper tail dependence, but it cannot capture neither negative dependence nor lower tail dependence. The upper tail dependence can be calculated as $\lambda_U = 2 - 2^{1/\alpha}$.

4.4 The Frank copula

Frank copula captures the symmetric dependence. For $-\infty < \alpha < \infty$, the CDF of the Frank (1979) copula takes the form:

$$C(u_1, u_2; \alpha) = -\alpha^{-1} \ln\left(\frac{1 - e^{-\alpha} - (1 - e^{-\alpha u_1})(1 - e^{-\alpha u_2})}{1 - e^{-\alpha}}\right).$$

The density is

$$c(u_1, u_2; \alpha) = \frac{\alpha(1 - e^{-\alpha})e^{-\alpha(u_1+u_2)}}{[1 - e^{-\alpha} - (1 - e^{-\alpha u_1})(1 - e^{-\alpha u_2})]^2}$$

where the dependence parameter α captures the symmetric dependence.

5 Copula estimation

In this paper, we estimate the copula model using the two-stage semiparametric procedure similar to that described in Chen and Fan (2006). In particular, having estimated, for each energy returns series, the univariate GARCH-in-Mean model using quasi-maximum likelihood estimation, we estimate F_j, $j = 1, 2$, using the empirical cumulative distribution functions of the residuals, $\epsilon_{jt}(\boldsymbol{\theta})_{t=1}^{n}$

$$F_j(x) = \frac{1}{n}\sum_{t=1}^{n}\mathbf{1}\left(\epsilon_{jt}(\boldsymbol{\theta}) \leq x\right), j = 1, 2 \tag{4}$$

where n is the number of observations. As pointed out in Chen and Fan (2006), since the estimation of the marginals is nonparametric, our copula estimation is robust and free of specification errors.

Note that the dependence captured by a copula is invariant with respect to increasing and continuous transformations of the marginal distributions. The bivariate copula dependence parameter α is estimated by

$$\hat{\alpha} = \arg\max \frac{1}{n}\sum_{t=1}^{n}\ln c\left(F_1(\epsilon_{1t}), F_2(\epsilon_{2t}); \alpha\right)$$

and $\hat{\alpha}$ is obtained by solving the score equations, $\partial L_c/\partial\alpha = 0$. The equation is nonlinear in general, and standard quasi-Newton iterative algorithms are employed. One advantage of the two-step estimation approach compared to the fully parametric approach is that the dependence statistics of the two-step estimated parameters are not affected by the models of the conditional mean and variance.

In sum, our semiparametric GARCH-in-Mean copula model specifies the conditional mean and conditional variance parametrically, but specifies the distribution of the marginal (standardized) innovations semiparametrically. A parametric copula is evaluated at the nonparametric univariate marginals. The copula function captures the concurrent dependence between the components of the multivariate innovation. Thus, a semiparametric GARCH-in-Mean copula model is very flexible in capturing a wide range of nonlinear, asymmetric dependence structures and the marginal behavior of multivariate time series.

6 Dependence analysis using Copulas

Conditional on the marginal specifications, we estimate copula dependence structures for each pair of energy returns. We estimate the empirical distribution of the marginals based on Eq. 4. Table 4 summarizes the results of the linear and rank correlation coefficients for energy returns. As Poon et al. (2004) notes, the conventional dependence measure, which is the linear correlation calculated as the average of deviations from the mean, assumes a linear relationship of the variables which follow a joint Gaussian distribution. The risk from joint extreme events could be underestimated. Moreover, it cannot distinguish between positive and negative returns, neither the large nor small values.

Table 4 Bivariate dependence between crude oil, natural gas, ethane and natural gas liquids

	Overall			Non-recession			Recession		
	Correlation	Kendall's τ	Spearman's ρ	Correlation	Kendall's τ	Spearman's ρ	Correlation	Kendall's τ	Spearman's ρ
Crude oil-Natural gas	0.826	0.664	0.827	0.769	0.642	0.809	0.976	0.829	0.921
Crude oil-Ethane	0.509	0.347	0.497	0.435	0.321	0.465	0.735	0.543	0.694
Crude oil-Propane	0.669	0.459	0.632	0.629	0.435	0.604	0.822	0.676	0.810
Crude oil-Butane	0.669	0.459	0.632	0.629	0.435	0.604	0.822	0.676	0.810
Natural gas-Ethane	0.521	0.397	0.550	0.457	0.376	0.523	0.711	0.581	0.718
Natural gas-Butane	0.657	0.464	0.632	0.593	0.430	0.594	0.805	0.771	0.923
Natural gas-Propane	0.643	0.509	0.677	0.600	0.484	0.650	0.805	0.714	0.853
Ethane-Propane	0.681	0.493	0.663	0.643	0.477	0.642	0.837	0.619	0.813
Ethane-Butane	0.574	0.401	0.559	0.527	0.369	0.519	0.733	0.619	0.783
Propane-Butane	0.762	0.556	0.731	0.728	0.532	0.709	0.899	0.733	0.882

Monthly data: 2002:01-2021:12 (T=240)

Alternatively, both Kendall's τ and Spearman's ρ statistics, which use the ranking of the data instead of the actual values of the data, can describe the nonlinear tail dependence structure. Specifically, Kendall's τ is as follows:

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

Spearman's ρ is as follows:

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

where X and Y are two random variables and F_1 and F_2 are the corresponding distribution functions.

The positive Kendall's τ and Spearman's ρ statistics in Table 4 indicate a positive relationship for all pairs of energy returns during the whole sample period as well as during the recessions. Moreover, the rank correlations are stronger during the recessions, indicating stronger comovements during economic contractions. The positive and high rank correlation values indicate that energy prices move together in the same direction during economic downturns. The highest Kendall's τ value and Spearman's ρ values are for the crude oil-natural gas pair, indicating that the probability of concordance in crude oil and natural gas price movements is significantly higher than the probability of discordance. The two measures of dependence are consistent with each other.

To explore the dependence structure in energy returns and the choice of the appropriate copula to use, we scatter plot all the pairs of ϵ_{it} and ϵ_{jt} $(i \neq j)$ in Figures 7, 8, 9, 10, 11, 12, 13, 14, 15 and 16. In general, the dots are mainly clustered in the center for all pairs. Moreover, it seems the dependence structures at the tails are symmetric. In other words, we do not observe the clustering of the dots in one tail obviously more sizeable than the clustering of the dots in the other tail in the scatter plots of all pairs.

We consider the four copula functions discussed in Section 4 — the Gaussian, Clayton, Gumbel, and Frank copulas. The estimates of the copula parameters are presented in Table 5. The copula parameters for all the pairs are highly statistically significant. They suggest that there is considerable bivariate dependence between the crude oil, natural gas, and hydrocarbon gas liquids markets over the sample period. The AIC values of the copula models are summarized in Table 6. The AIC clearly

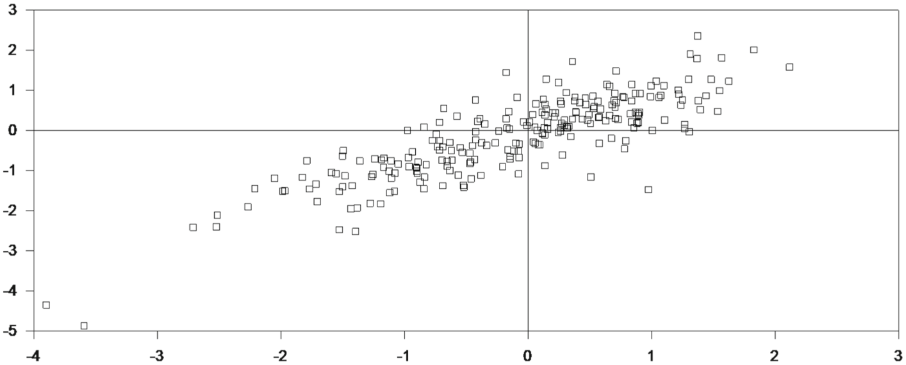

Fig. 7 Scatter plot of crude oil and natural gas

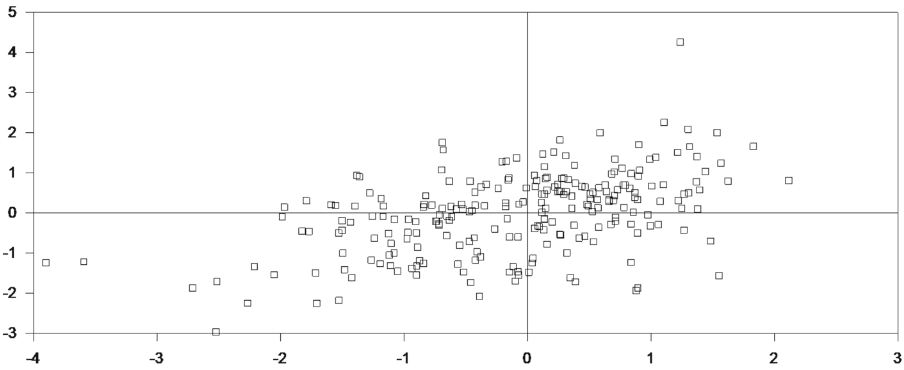

Fig. 8 Scatter plot of crude oil and ethane

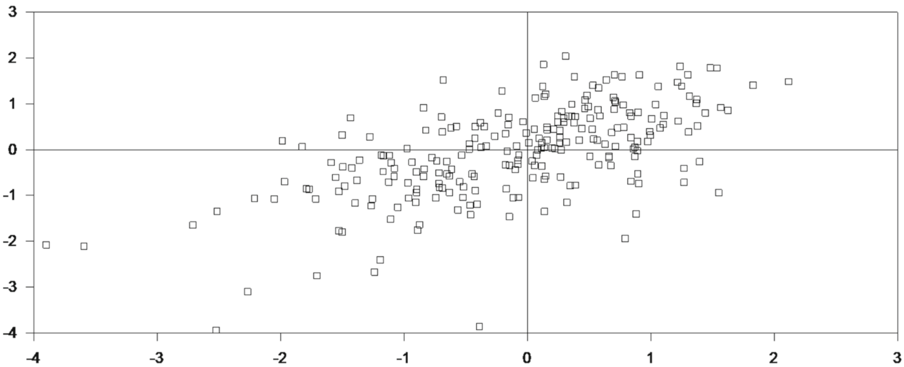

Fig. 9 Scatter plot of crude oil and propane

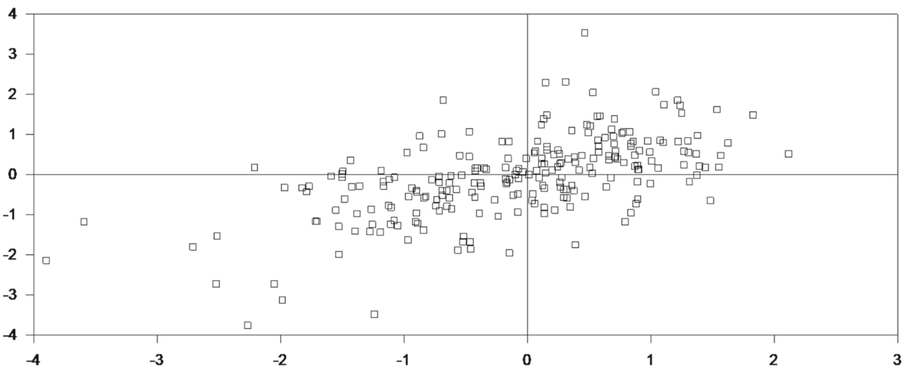

Fig. 10 Scatter plot of crude oil and butane

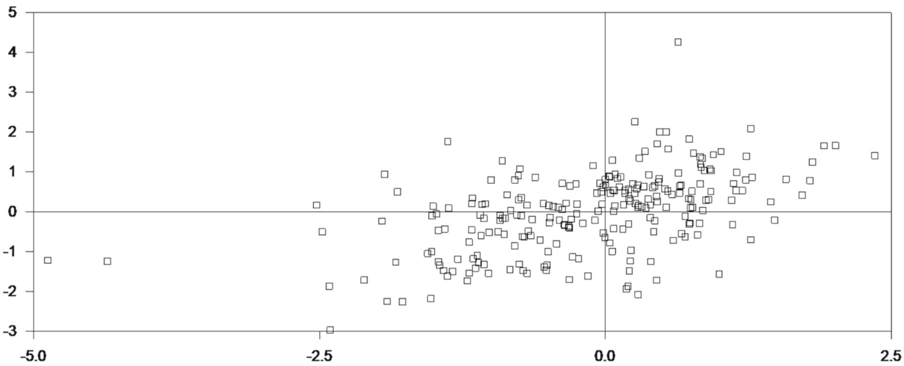

Fig. 11 Scatter plot of natural gas and ethane

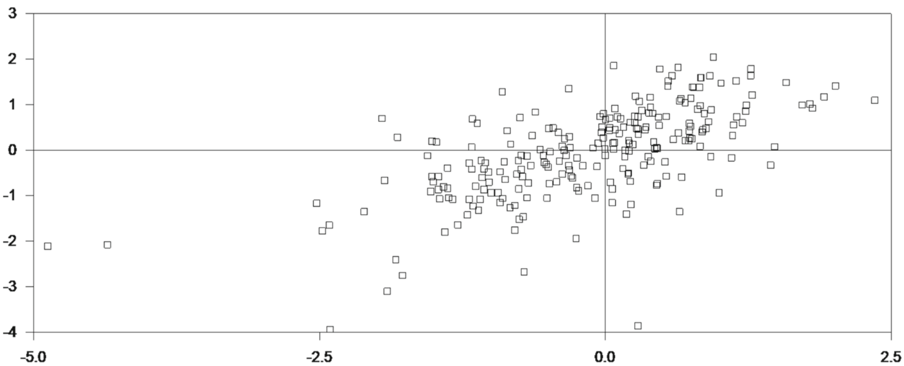

Fig. 12 Scatter plot of natural gas and propane

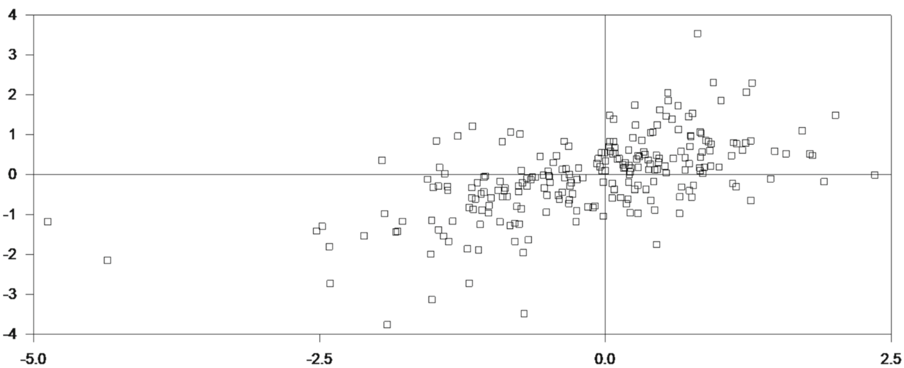

Fig. 13 Scatter plot of natural gas and butane

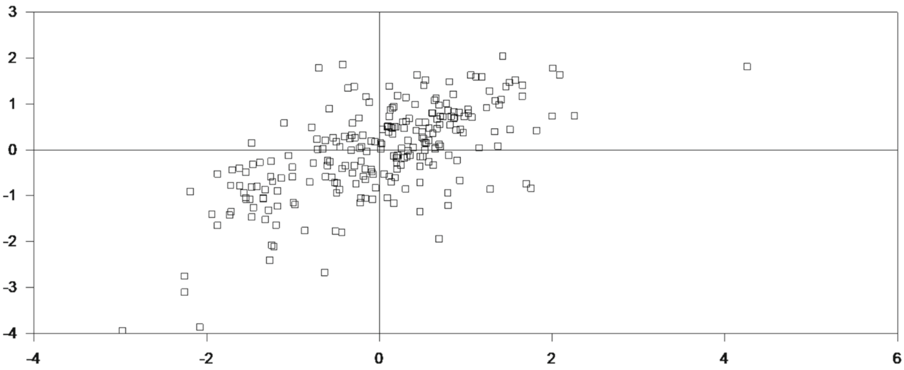

Fig. 14 Scatter plot of ethane and propane

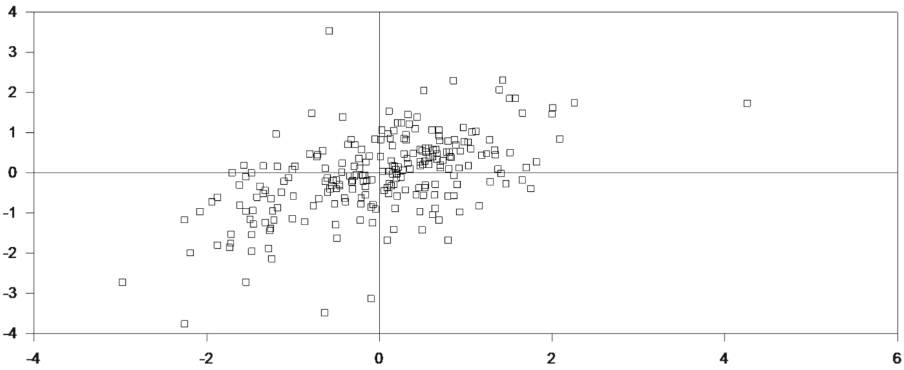

Fig. 15 Scatter plot of ethane and butane

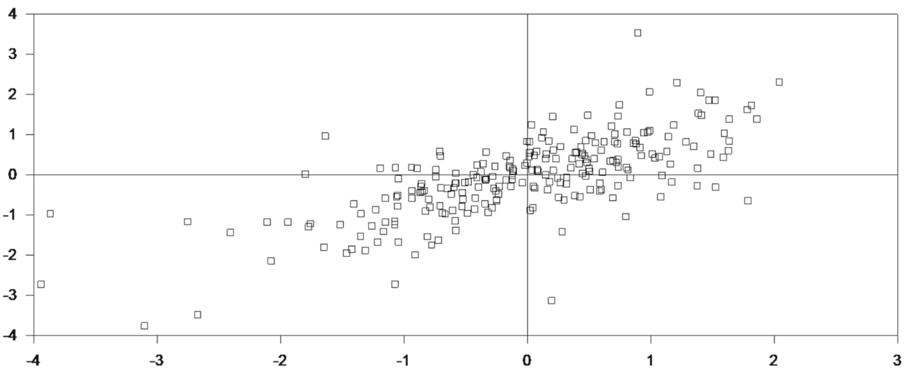

Fig. 16 Scatter plot of propane and butane

Table 5 Dependence parameter estimates of different copula models

	$\hat{\alpha}$			
Pair	Gaussian	Gumbel	Clayton	Frank
Crude oil-Natural gas	0.814 (0.000)	1.778 (0.000)	2.400 (0.000)	**8.611** (0.000)
Crude oil-Ethane	0.486 (0.000)	1.203 (0.000)	0.731 (0.000)	**3.245** (0.000)
Crude oil-Propane	0.610 (0.000)	1.320 (0.000)	1.072 (0.000)	**4.607** (0.000)
Crude oil-Butane	0.598 (0.000)	1.325 (0.000)	1.075 (0.000)	**4.457** (0.000)
Natural gas-Ethane	0.519 (0.000)	1.225 (0.000)	0.792 (0.000)	**3.686** (0.000)
Natural gas-Propane	0.654 (0.000)	1.362 (0.000)	1.201 (0.000)	**5.318** (0.000)
Natural gas-Butane	0.608 (0.000)	1.344 (0.000)	1.115 (0.000)	**4.572** (0.000)
Ethane-Propane	0.635 (0.000)	1.399 (0.000)	1.287 (0.000)	**5.045** (0.000)
Ethane-Butane	0.575 (0.000)	1.296 (0.000)	**0.976**(**0.000**)	3.865 (0.000)
Propane-Butane	0.716 (0.000)	1.502 (0.000)	1.586 (0.000)	**6.148** (0.000)

Monthly data: 2002:01-2021:12 (T=240). Numbers in parentheses are p-values. Numbers in bold font are corresponding to the lowest AIC values

selects the Frank copula for all cases except for the pair of ethane and butane. Notice that the Frank copula is able to capture the symmetric lower and upper tail dependence. The Frank copula suggests that all of the energy return pairs we study exhibit weakly positive and symmetric upper tail and lower tail dependence. The contagion effect intensity is different across the energy markets, with the highest between crude oil and natural gas markets. In the hydrocarbon gas liquids market, the dependence parameter between butane and propane is the highest among all the pairs, suggesting that extreme outcomes in the propane market are easier to extend to the butane market, and vice versa.

Table 6 AIC values of different copula functions

	AIC			
Pair	Gaussian	Gumbel	Clayton	Frank
Crude oil-Natural gas	4.537	2.101	−1.011	**−1.048**
Crude oil-Ethane	5.348	2.784	−0.210	**−0.242**
Crude oil-Propane	5.159	2.588	−0.388	**−0.439**
Crude oil-Butane	5.188	2.583	−0.391	**−0.422**
Natural gas-Ethane	5.305	2.762	−0.233	**−0.302**
Natural gas-Propane	5.068	2.546	−0.440	**−0.549**
Natural gas-Butane	5.167	2.563	−0.407	**−0.433**
Ethane-Propane	5.093	2.477	−0.492	**−0.502**
Ethane-Butane	5.217	2.627	**−0.334**	−0.326
Propane-Butane	4.896	2.372	−0.626	**−0.680**

Monthly data: 2002:01-2021:12 (T=240). Numbers in bold font are the minimums in each row

The statistically significant Frank copula parameter estimate indicates that in times of rare events, such as market crushes and large changes in energy market returns, as one energy return tends to reach its lower/upper limit, there is a high chance that the other energy return will be close to its lower/upper limit too. The existence of left (right) tail dependence implies a much higher downside (upside) risk in crude oil market investments than in the case of no-tail dependence. The high likelihood of extreme joint extreme values in the hydrocarbon gas liquids markets implies a higher value-at-risk than that of a joint normal distribution; as discussed in Poon et al. (2004), tail dependence measures the systematic risk in times of extreme market events.

We also explore the possible asymmetric tail dependence between returns of crude oil, natural gas, and hydrocarbon gas liquids using Clayton copula and Gumbel copula. Clayton copula can capture the lower tail dependence, and Gumbel copula can capture the upper tail dependence. Table 6 shows that Frank copula has the lowest AIC values for all pairs except for ethane and butane.

There is a theoretical literature on comovements of the energy markets. In general, the contagious movements can be explained by fear spillover, as in the financial markets, or by real links including production and trading. It has been argued, for example, that correlations between financial markets increase during market downturns as a consequence of investors facing greater uncertainty about the state of the economy. Asymmetric responses of agents to energy prices are also a possible cause of asymmetric dependence. Details on theoretical discussions of the comovements in the energy markets are beyond the scope of this paper.

7 Conclusion and policy implications

Frequent extreme events, such as financial crises, natural disasters, and global pandemics, suggest that the volatility and tail dependence in the crude oil, natural gas, and hydrocarbon gas liquids markets could remain an important feature in the energy markets landscape. This paper investigates the volatility and dependence structure in the crude oil, natural gas, and hydrocarbon gas liquids markets using copula GARCH-in-Mean models and monthly data over the period from January 2002 to December 2021. We first investigate the volatility of each energy return using ARMA GARCH-in-Mean models. The ARMA GARCH-in-Mean models show that volatility has a statistically significant positive effect on the returns of crude oil and natural gas, but a statistically significant negative effect on the returns of ethane. There is no statistically significant effect on the returns of propane and butane. The rank correlations show that the dependence among these markets increases during economic contractions indicating the potential contagious movements in the energy markets.

The lack of theoretical evidence on the dependence structure in the energy markets, and the observation of asymmetric comovements in these markets, motivate us to use a flexible dependence structure specification. Four copula models are used to examine the dependence of the marginals and the goodness of fit of different copulas is compared based on the AIC. By implementing the copula approach, we are able to capture the nonlinear, asymmetric tail dependence across the energy markets while allowing different univariate distributions for each energy market return. We find that the dependence intensity is different across the crude oil, natural gas, and hydrocarbon gas liquids markets. The Frank copula is the best copula to describe the bivariate dependence in all pairs of crude oil, natural gas, and hydrocarbon gas liquids returns, except for the pair of ethane and butane. It suggests that there is symmetric positive tail dependence for all of the considered pairs. The contagion effect is strongest between crude oil and natural gas. Within the hydrocarbon gas liquids markets, the dependence is strongest between propane and butane.

References

Aloui R, Ben A, Safouane M, Nguyen DK (2013) Conditional dependence structure between oil prices and exchange rates: A copula-GARCH approach. J Int Money Financ 32(C):719–738

Bollerslev T (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return. Rev Econ Stat 69(3):542–547

Chen X, Fan Y (2006) Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. J Econ 135:125–154

Clayton DG (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65:141–151

Dickey D, Fuller W (1981) Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49(4):1057–72

Efimova O, Serletis A (2014) Energy markets volatility modelling using GARCH. Energy Econ 43:264–273

Engle R (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50(4):987–1007

Ewing BT, Malikb F, Ozfidanc O (2002) Volatility transmission in the oil and natural gas markets. Energy Econ 24:525–538

Frank MJ (1979) On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. Aequationes Math 19:194–226

Gumbel EJ (1960) Distributions des valeurs extremes en plusieurs dimensions. Publications de lâ, Institute de Statistque de lâ Universite de Paris 9:171–173

Jahan S, Serletis A (2019) Business cycles and hydrocarbon gas liquids prices. J Econ Asymmetries 19(C):1–1

Jarque CM, Bera AK (1980) Efficient tests for normality, homoscedasticity, and serial independence of regression residuals. Econ Lett 6:255–259

Joe H. (2014) Dependence modeling with copulas. Chapman and Hall

Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? J Econ 54:159–178

Ning C (2010) Dependence structure between the equity market and the foreign exchange market: a copula approach. J Int Money Financ 29(5):743–759

Patton AJ (2006) Modelling asymmetric exchange rate dependence. Int Econ Rev 47:527–556

Phillips PCB, Perron P (1988) Testing for a Unit Root in Time Series Regression. Biometrika 75(2):335–346

Poon S, Rockinger M, Tawn J (2004) Extreme value dependence in financial markets: diagnostics, models, and financial implications. Rev Financ Stud 17(2):581–610

Reboredo J (2011) How do crude oil prices co-move?: a copula approach. Energy Econ 33(5):948–95

Rodriguez JC (2007) Measuring financial contagion: a copula approach. J Empir Finance, 14(3)

Serletis A, Xu L (2016) Volatility and a century of energy markets dynamics. Energy Econ 55(C):1–9

Sklar A (1973) Random variables, joint distributions, and copulas. Kybernetica 9:449–460

Tong BI, Wu C, Zhou C (2013) Modeling the co-movements between crude oil and refined petroleum markets. Energy Econo 40(C):882–897

Trivedi P, Zimmer D (2007) Copula modeling: an introduction for practitioners. Found Trends Econom 1(1):1–111

Wu C, Chung H, Chang Y (2012) The economic value of co-movement between oil price and exchange rate using copula-based GARCH models. Energy Econ 34(1):270–282

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.