ORIGINAL PAPER

Portfolio stress testing applied to commodity futures

Florentina Paraschiv[1] · Stine Marie Reese[1] · Margrethe Ringkjøb Skjelstad[1]

Abstract

In this article, we construct a portfolio of commodity futures which mimics the Dow Jones Commodity Index and perform an extensive stress testing exercise with a focus on hybrid scenarios. The increased volume of investments in commodities as financial instruments over the last decades underline the importance of a more thorough framework for stress testing of related portfolios. Our study is the first to show comparatively the marginal impact of the model choice for portfolio components versus the marginal role of tail dependence on the portfolio profit and loss in stress testing exercises. We model the distribution of returns of portfolio components with an asymmetric AR-GARCH model combined with Extreme Value Theory for extreme tails, and employ multivariate copula functions to model the time-varying joint dependence structure. Our study reveals that indeed, for a realistic stress test, a special attention should be given to the tail risk in individual commodity returns as well as to tail correlations. We also draw conclusions about parameter risk persistent in stress testing exercises. Finally yet importantly, in line with Basel IIIb, the study highlights the importance of using forward-looking hybrid and hypothetical scenarios over historical scenarios.

Keywords Stress testing · Commodity futures · Risk measures · Extreme value theory · Copula functions

1 Introduction

Financial investments in commodities have grown rapidly over the last decades and became an important asset in portfolios of institutional investors such as pension funds, insurance companies, and hedge funds. The risk associated with weather, storage etc.

✉ Florentina Paraschiv
   florentina.paraschiv@ntnu.no

[1] NTNU Business School, Norwegian University of Science and Technology, 7491 Trondheim, Norway

 ⌀ Springer

led to the rise of commodity indices in the early 2000s, providing a hedge opportunity for commodity producers. The volumes of exchange-traded derivatives became 20 to 30 times higher than the physical production of many commodities (Silvennoinen and Thorp 2013; Paraschiv et al. 2015). The vivid interest in this asset class might be attributed to the perceived opinion that commodities show low correlation with traditional assets, and thus, provide diversification benefits in a mixed-asset portfolio (Bhardwaj et al. 2015; Gorton and Rouwenhorst 2006; Paraschiv et al. 2015). The empirical analyses in Silvennoinen and Thorp (2013) and Daskalaki and Skiadopoulos (2011) show increased integration of commodity and financial markets, with higher correlation, especially in bearish times (Cheung and Miu 2010). In Tang and Xiong (2012), it is shown that the increasing presence of index investors has exposed commodity prices to market-wide shocks, such as shocks to the world equity index, the US dollar exchange rate, and shocks to other commodities, such as oil. Furthermore, Adams and Glück (2015) show that the risk spillovers to commodities observed during the financial crisis were persistent over time and the volatility in commodity markets increased during the past decade (Tang and Xiong 2012; Basak and Pavlova 2016). These changes in commodity characteristics are often referred to as the financialization of commodity markets (Cheng and Xiong 2014) and lead to the need for an approach to measure and manage the associated risks in related financial investments.

A common tool for risk management is stress testing. European Banking Authority (2017, p. 28) points out in their new guidelines under development that "Institutions should ensure that the scenario analysis is a core part of their stress testing programme". Implementing stress testing is now mandatory for banks, due to regulations from Basel III formed in the post crisis environment (Basel Committee on Banking Supervision 2009).

An example in this sense is the study of Koliai (2016) which analyses existing risk models for stress testing purposes. The study presents a semi-parametric copula-GARCH risk model for equity indices, exchange rates and commodity prices, to perform stress testing on hypothetical portfolios, where the marginal distributions of returns are specified using EVT. Results show that different risk models produce significantly different results in terms of corresponding stress scenarios and impact on the portfolios.

The analysis in Paraschiv et al. (2015) is to our knowledge the only example of stress testing methodology applied to a portfolio of commodities which takes into account specific events that impacted this class of assets over time. It shows the importance of using forward-looking scenarios to enable the simulations of extreme quantiles, providing a better understanding of risk.

In this article, we apply stress testing techniques in line with the regulatory requirements from Basel III (Basel Committee on Banking Supervision 2017, 2018; Baudino et al. 2018) to a portfolio of commodity futures. The existing literature on stress testing of commodity portfolios is scarce, despite their popularity gained in practice. We update the analysis in Paraschiv et al. (2015), keeping the same procedure for constructing the stress portfolio as in the original study. However, we innovate in several directions. We extend the data set by including several new historical shocks, among which the oil price drop in 2014. Secondly, we enrich the spectrum of stress testing scenarios, focusing more on the forward-looking ones. The analysis in Paraschiv et al.

(2015) is limited to show the effects of a reoccurring financial crisis on the portfolio profit and loss. Our study shows the importance of combining historical estimations of model parameters with more flexible forward-looking scenario construction. Furthermore, this is the first study in the literature where we disentangle the effect of individual model components on the portfolio profit and loss.

For our portfolio construction we mimic the dynamics of Dow Jones Commodity Index. The DJCI is a broad commodity index consisting of 24 commodities in three major sectors: energy, metals and agriculture & livestocks. The weights are based on the traded volume, ensuring a liquid index.

For the marginal distributions of commodity returns, we use an asymmetric AR-GARCH process and model the tails by Extreme Value Theory. We further describe the joint dynamics of portfolio components by employing a copula function. Finally, we simulate the portfolio profit and loss distributions under different scenarios in a stress testing framework.

Our results are twofold. First we find that the simulated profit and loss distribution of the portfolio is highly sensitive to the choice of the modelling approach for the marginal distribution of portfolio components. In particular, a correct identification of tail risk is of great importance for the stress testing purpose. The marginal role of correlations/dependence structure among portfolio components seems, however, to have a less obvious impact for the stress test results. Secondly, we find the construction of hybrid scenarios to be a relevant tool to combine both historical information and the flexibility of forward looking approaches in line with the requirements from Basel III (Basel Committee on Banking Supervision 2009).

The remainder of this article is structured as follows. Section 2 offers an overview of the most relevant literature for our study. In Sect. 3 we will provide an introduction of our data, focusing on the characteristics of the portfolio. In Sect. 4 we show the theoretical background of the different methodologies applied. Section 5 shows details of the implementation of the methodology for our data set and of the simulation procedure. Finally, in Sect. 6 we will explain and apply stress testing and display our analysis.

2 Review of literature on stress testing

2.1 Regulatory requirements for Stress testing

As defined in Lopez (2005), stress testing is a risk management tool used to evaluate the potential impact on portfolio values of unlikely, although plausible events or movements in a set of financial variables. The recent financial crisis led the attention of banks and authorities to the insufficient methods of risk management and the need for more accurate stress testing became obvious, since financial institutions were not prepared to deal with the crisis. One main concern was that the scenario selection and simulation were carried out by separate units for each business line and for particular risk types (Basel Committee on Banking Supervision 2009). This indicates that the stress testing was isolated and did not provide a complete picture on the firm level.

Seemingly the most recent development in the methodology for stress testing of portfolios is the use of Extreme Value Theory (EVT) and copulas as input to the analysis. EVT was first introduced in Embrechts et al. (1997) to better model the tail distribution of risk factors. Extreme Value Theory focuses on shaping the tails rather than the whole distribution of returns, providing more rigorous estimates of risk for financial portfolios. In McNeil et al. (2015) the authors suggested using a combination of GARCH and EVT. This methodology is popular in recent literature, with the largest proportion of new studies focusing on stock markets or single commodities (Ghorbel and Souilmi 2014; Liu 2011; Wang et al. 2010; Aepli 2011).

2.2 Types of stress tests

Stress tests can be conducted with several methodologies. One can firstly differentiate between univariate and multivariate stress tests. Univariate stress tests aim at identifying the isolated influence of stressing or shocking one single risk factor of a portfolio (Aepli 2011, p. 4). This makes the univariate stress tests simple to apply, but very limited, since they do not take dependencies between the returns of portfolio components into account. Multivariate stress tests overcome this drawback. In Basel Committee on Banking Supervision (2009) we find a classification of stress test methodologies for financial institutions. One can consider different scenarios when running stress tests, historical, hypothetical and hybrid. The need for hypothetical scenarios was highlighted after the crisis, since risk managers mostly performed historical stress testing under Basel II (Basel Committee on Banking Supervision 2006). The European Banking Authority (2017, p. 28) pointed out that "the design of stress test scenarios should not only be based on historical events, but should also consider hypothetical scenarios, also based on non-historical events". Forward-looking scenarios are now required for European banks according to Basel Committee on Banking Supervision (2009). In Aepli (2011) the authors propose an extensive framework for complex stress testing for portfolios of futures that is in agreement with the regulations from Basel III formed in the post crisis environment.

2.2.1 Historical scenario

Historical scenarios are based on actual, realised data stemming from a historical episode of financial stress. This makes them realistic and easy to access. The profit and loss distribution in the historical scenario is simply given by the realised empirical distributions. In Lopez (2005), it is pointed out that historical scenarios are developed more fully than other scenarios since they reflect an actual stressed market environment that can be studied in great detail, therefore requiring fewer judgements by risk managers.

One major drawback with historical scenarios is the assumption that passed financial crises will reoccur with the same consequences on portfolio losses. This makes them unable to capture risks linked to new products that may have significant impact on the outcome of a crisis. The worst observed loss in the past might not reflect the worst possible outcome in the future. This drawback was proven to be essential in

the financial crisis of 2007 and resulted in the underestimation of the risk level and interaction between risks (Basel Committee on Banking Supervision 2009, p. 5).

Another drawback in historical scenarios is the sample size. Due to the limited number of observations, computing risk metrics in the higher confidence levels becomes problematic. This is a considerable drawback as the most extreme losses are of great interest in stress testing exercises.

2.2.2 Hypothetical scenario

Hypothetical scenarios are, unlike the historical scenarios, forward looking. Scenarios can be constructed in multiple ways, for example by shocking model parameters arbitrarily, based on own experiences of market movements. Hypothetical scenarios have the advantage of being more flexible and forward looking, making them more informative if conducted correctly. More focus on hypothetical stress testing scenarios allows the institution to be both well prepared for potential extreme unexpected outcomes, and lay the foundation to overcome these potential losses.

An extensive analysis has to be in place before constructing hypothetical scenarios, which can be both time consuming and difficult. In Basel Committee on Banking Supervision (2009, p. 5) it is pointed out that banks had implemented hypothetical scenarios prior to the financial crisis, but it was difficult for risk managers to obtain the support of the senior management, since the scenarios were extreme or innovative, and often were considered as implausible. Extremes that have not yet been experienced are often difficult to imagine and to be taken seriously.

In the financial regulatory frame from Basel Committee on Banking Supervision (2018) we note that during stress testing exercises consideration should be given to both historical and hypothetical events. This is to take into account new information and emerging risks in the foreseeable future. Furthermore, "when conducting stress tests it is important to be aware of the limitations of the scenarios (BIS 2018, p. 4). This also emphasizes the need to use several scenarios for a more correct result."

2.2.3 Hybrid scenario

Hybrid scenarios combine the knowledge found in historical scenarios with the flexibility of hypothetical scenarios, making them a suitable alternative in stress testing. Hybrid scenarios are also easier to implement than more extensive forward-looking scenarios, as they are anchored in actual experienced market conditions. Hybrid scenarios are constructed by using historical data during times of financial distress to calibrate the process of risk factor evolution, but allow extrapolation beyond experienced events.

Even though hybrid scenarios allow the construction of new possible scenarios, they are still somewhat backward looking in the sense that they do not fully explore the risk of shifting market conditions or risk associated with new products. However, Lopez (2005) points out that risk managers always face a trade-off between scenario realism and comprehensibility; that is, more fully developed scenarios generate results that are more difficult to interpret. The benefits from implementing hybrid scenarios should not be neglected as they balance this trade-off.

3 Data selection and description

3.1 Choice of commodity indices

Commodity indices have become quite popular in the last decades and several commodity indices have been developed, among which S&P Goldman Sachs Commodity Index (S&P GSCI) and the Dow Jones Commodity Index (DJCI).

The S&P GSCI consists of 24 commodities and the weights are based on trading volume. It is therefore often seen as a benchmark for investment performance in commodities. The trading volume in energy commodities is higher than any other commodity sector so this index is heavily based in energy (60% of the total weight in 2017 S&P Dow Jones Indices 2018). To get a more balanced portfolio across different commodity sectors the DJCI will be the focus in this article. It consists of 24 commodities divided into three major sectors: metals, energy and livestock & agriculture. The weights are based on the total volume traded, but unlike the S&P GSCI, the DJCI has constraints on total weight allocated in each sector and commodity. By not allowing any of the three sectors to obtain more than 35% of the weight, and no single commodity to constitute less than 2% or more than 17% of the total index, the DJCI becomes well diversified. These restrictions also provide continuity and high liquidity for potential investors. The weights are rebalanced annually. See S&P Dow Jones Indices (2017) for a detailed methodology.

To select the risk factors (commodities) for our analysis we apply the method introduced in Paraschiv et al. (2015). We take the ten commodities with the largest weights for 2017 in the DJCI, and form our test portfolio. This is done to get a more time efficient portfolio and to make the analysis more practical. The ten commodities add up to 76% of the DJCI, providing a good proxy for the movements of the entire index. To form our test portfolio we scale up the weights, proportionally to 100%. The weights of the ten commodities can be seen in Table 1. Our selection leaves us with three portfolio components in energy, three in metals and four in agriculture & livestocks.

3.2 Descriptive statistics

We extracted daily data from 1996 to 2017 from Thomson Reuters Eikon for continuous series of futures with approximately one year to maturity for the ten selected commodities. This leaves us with 5741 observations for each commodity. Details about the data extraction are found in Table 2.

As for our data, the rolling over was done by Eikon. Their methodology can be found in Thomson Reuters (2012). For monthly futures data roll over is done by jumping to the nearest future contract with a switch over following in the last trading day. They use the nearest contract month to form the first values of the continuous series, and when the contract expires, the next point of data is the next one year to maturity contract. They do not adjust for price differentials when adjusting the data, but we found this methodology to be sufficient for our analysis especially as our futures have one year to maturity.

Table 1 Portfolio weights scaled up from the weights in DJCI 2017. *Source*: S&P Dow Jones Indices (2016)

Commodity	Weight in index (%)	Weight in test portfolio (%)
Wheat	3.2	4.2
Corn	7.1	9.3
Soybeans	11.9	15.6
Live cattle	2.7	3.5
Copper	10.3	13.6
Gold	10.2	13.4
Aluminium	4.6	6.1
WTI	9.7	12.7
Brent	8.8	11.6
Natural gas	7.3	9.6
Sum	76.1	100

Figure 1 shows the historical price movements of commodities measured in a relative index value. We observe a general co-movement especially in the 2000s, when the commodity markets experienced a uniform rise in prices until the financial crisis.

We observe several structural breaks across commodities especially during the financial crisis in 2007–2009 that heavily affected commodity markets. The European Central Bank Delle Chiaie et al. (2017) brings evidence supporting that global activity has clear implications for the commodity markets. Their analysis shows that since the year 2000 the price drivers of oil have fundamentally changed, and during the time of the financial crisis global activity strongly affected the oil price. The acute drop in oil price in 2014 was driven by several factors, among which the increased supply of unconventional oil and a significant shift in OPEC policy (Baffes et al. 2015). What differentiates the price drop in 2014 from previous collapses in oil price is, according to Baffes et al. (2015), that the fluctuation could not be explained by a weakened demand or expansion of supply in isolation, but rather a combination of the two.

While during the financial crisis all commodity sectors were affected, the price drop in 2014 to a lesser degree showed spillover to non-energy sectors. This indicates the decoupling of oil price from other commodities in agriculture and metals. According to Erdős (2012) the co-integration of oil and natural gas ended in 2009 after an increase in shale gas production. We observe that the commodities in non-energy sectors in more recent years do not necessarily follow the oil price as closely as in the past decade, potentially affecting the dynamics of the commodity markets.

Table 3 shows the daily descriptive statistics of our time series. McNeil, Frey and Embrechts (2015, p. 117) present six stylized facts of financial returns that can be observed, which also apply to our data: (1) Return series are not i.i.d.; (2) Series of absolute or squared returns show profound serial correlation; (3) Conditional expected returns are close to zero; (4) Volatility appears to vary over time; (5) Return series are leptokurtic or heavy-tailed; (6) Extreme returns appear in clusters.

Table 2 Data extraction details. *Source*: Thomson Reuters Eikon

Commodity	Ticker	Classification	Units	Price quote	Exchange
WTI	CL	Energy	1000 barrels	USD/barrel	NYMEX
Brent	LCO	Energy	1000 barrels	USD/barrel	ICE
Natural gas	NG	Energy	10,000 million BTU	USD/MMBtu	NYMEX
Corn	C	Grains	5000 bushels	US cent/bushels	CBoT
Wheat	W	Grains	5000 bushels	US cent/bushels	CBoT
Soybeans	ISF	Grains	5000 bushels	US cent/bushels	CBoT
Live cattle	LC	Livestock	40,000 pounds	US cent/pound	CME
Gold	GC	Precious metal	100 ounces	USD/Troy oz	COMEX
Aluminium	MAL	Industrial metal	25 metric tonne	USD/MT	LME
Copper	HG	Industrial metal	25,000 pounce	USD/pounce	COMEX
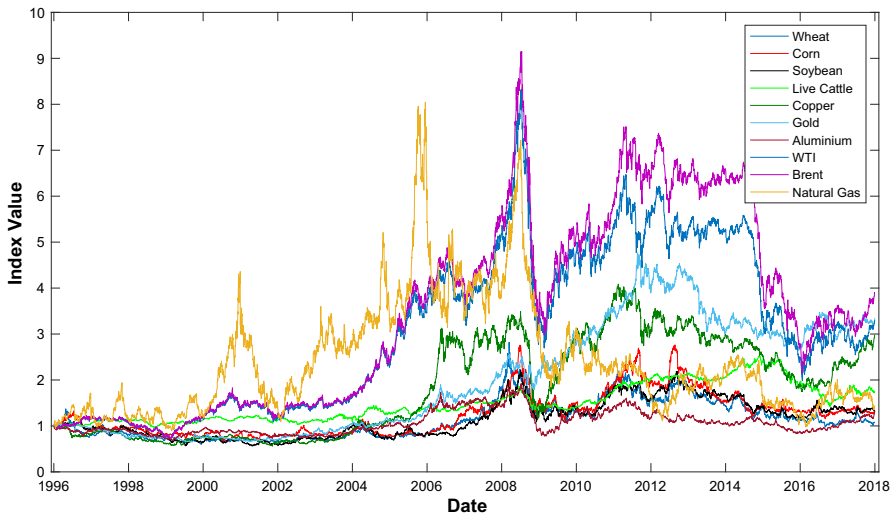
Fig. 1 Historical daily price movements from 1996 to 2017 for the ten commodities in relative value

We can also see graphically that the returns show autocorrelation, in line with the stylized fact (2) (see Fig. 2). From the probability plot we see that the returns follow the t distribution better than the Normal distribution, but the data deviates from the t distribution in the tails (see Fig. 3). This indicates heavy tailed returns. Additionally, ARCH-GARCH tests show evidence of conditional heteroscedasticity, in line with stylized fact (4). We performed the test for lags 5, 10, 15 and 20. We also tested for stationarity. Augmented Dickey-Fuller test, Phillips Perron test and KPSS test show that the returns of all commodities are stationary.

4 Methodology

4.1 Motivating the choice of modeling approach

As a result of the return characteristics for the ten commodities, as shown in Sect. 3, we model the conditional volatility with a GARCH process. A GARCH process can be extended in various ways, depending on the purpose. For commoditiy markets it has been shown that volatility tends to increase more after large negative returns than after large positive returns (Nyström and Skoglund 2002b). We therefore see it as an appropriate choice to extend to a GARCH-GJR model (Alexander 2008a), which includes a leverage parameter to capture this asymmetry.

Since the focus of our study is on stress testing, extreme returns are of special interest. We have shown the deviations from the normal and student t distribution for the returns, especially in the tails. Extreme Value Theory with the Peak over Threshold method has been employed in earlier studies (Aepli 2011; Paraschiv et al. 2015; Wang et al. 2010) showing a good modeling performance in shaping heavy tails. This is in

Table 3 Daily descriptive statistics of commodity returns for years 1996–2017

Commodity	Mean (%)	SD (%)	Max. (%)	Min. (%)	Skewness	Kurtosis	Jarque–Bera (p value)
Wheat	0.0014	1.53	8.15	− 12.17	− 0.11	6.86	3573 (0.001)
Corn	0.0046	1.45	9.74	− 14.48	− 0.22	8.19	6467 (0.001)
Soybeans	0.0049	1.36	7.04	− 8.11	− 0.30	6.48	2984 (0.001)
Live cattle	0.0098	0.72	6.61	− 6.89	− 0.68	12.98	24,210 (0.001)
Copper	0.019	1.55	11.41	− 11.26	− 0.14	7.80	5512 (0.001)
Gold	0.021	1.06	8.62	− 9.87	− 0.14	9.89	11,351 (0.001)
Aluminium	0.0046	1.13	6.37	− 7.60	− 0.25	6.09	2335 (0.001)
WTI	0.0208	1.62	10.00	− 9.19	− 0.17	6.33	2670 (0.001)
Brent	0.02367	1.58	9.26	− 9.26	− 0.11	6.00	2162 (0.001)
Natural gas	0.0064	2.75	21.64	− 31.12	− 0.08	10.30	12,716 (0.001)
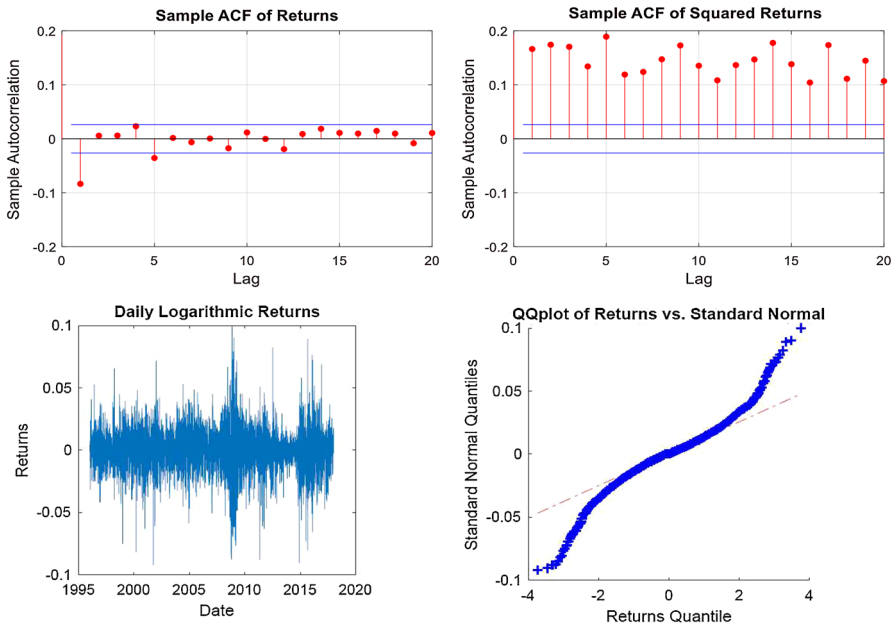
Fig. 2 Sample autocorrelation plot of the returns and squared returns for WTI, as well as the daily logarithmic returns and a quantile–quantile plot. Corresponding graphs for all commodities are available upon request
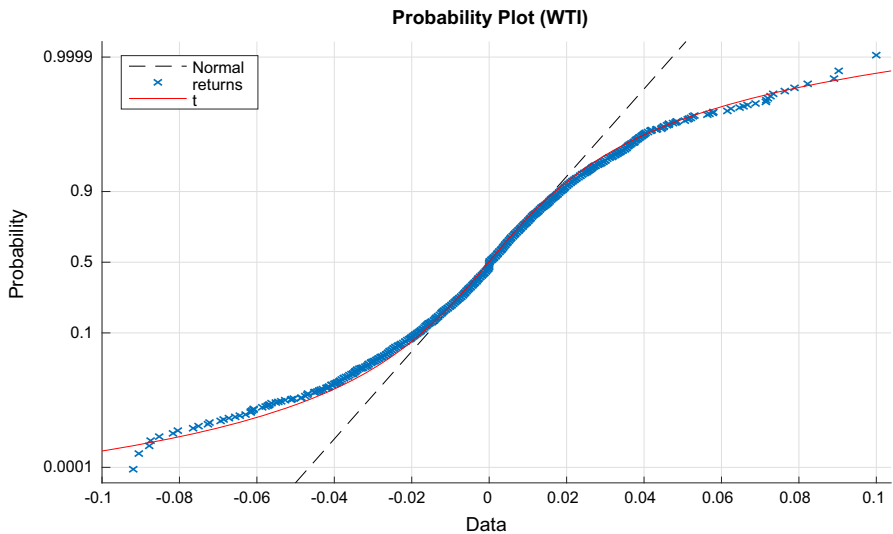
Fig. 3 Probability plot for WTI return versus standard normal and $-t$ distribution. Corresponding graphs for the other commodities are available upon request

line with the regulatory requirements for stress testing that point out the need to give a special attention to tail risk in asset returns.

Due to the common bust and boom cycles and co-integration of commodity markets, is important to model the dependence structure between returns in a realistic way. In Aepli et al. (2017) the authors show the importance of modelling time-variation and asymmetries in the dependence structure between the risk factors of a portfolio of commodity futures. In support of Basel III critics on over-reliance on historical correlations, the authors introduce multivariate dynamic copula models as a superior alternative. There exist a numerous amount of copulas to choose from, and the best choice is dependent on the aim of the analysis and the data. Analysis in Paraschiv et al. (2015), Aepli (2011) and McNeil et al. (2015) find the *t* copula to be superior over the Gaussian copula in the context of modelling multivariate financial return data. For our purpose, we therefore prefer a *t* copula over the more common Gaussian copula. The asymmetry of our data would probably be better modelled by an asymmetric copula, but as the *t* copula keeps the analysis tractable and allows a direct comparison across stress tests we find it suitable for this analysis. The subsequent subsections show the technical specification of the models employed.

4.2 GARCH

A simple autoregressive AR(p) process is a simple way to capture the autocorrelation between the individual commodity returns:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \tag{1}$$

where ϵ_t is i.i.d. with mean zero and variance σ^2.

The residuals from AR(q) models can be decomposed such that:

$$\epsilon_t = z_t \sigma_t \tag{2}$$

where z_t is i.i.d. with unit variance and σ_t is the conditional variance.

The generalized autocorrelation conditional heteroscedasticity model (GARCH) is then used to capture the volatility change and clustering of returns over time.

The symmetric normal GARCH assumes that the dynamic behaviour of the conditional variance is given by:

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \epsilon_t | I_{t-1} \sim N(0, \sigma_t^2). \tag{3}$$

The parameters of the GARCH model are estimated by maximising the value of the log likelihood function (see Alexander 2008a, p. 137).

Empirical evidence suggests that positive (negative) innovations to volatility correlate with negative (positive) innovations to returns (Nyström and Skoglund 2002a, p. 5). The rational behind is that negative impacts on returns have a tendency to increase

volatility, which is accounted for as the "leverage" effect (McNeil et al. 2015), since a fall in equity value causes a rise in the debt-to-equity ratio (leverage) of a company, therewith making the stock more volatile. The leverage effect is captured by extending the classical specification of the GARCH model (Eq. (3)) by an extra parameter, the leverage parameter λ. The GARCH-GJR can be written from the GARCH(1,1) model above, including the extra parameter (Alexander 2008a, p. 150):

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \lambda 1_{(\epsilon_{t-1}<0)} \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{4}$$

where the indicator function $1_{(\epsilon_t<0)} = 1$ if $\epsilon_t < 0$. Thus, in addition to the specification in Eq. (3), an additional shock to volatility for negative residuals is added to volatility to account for the asymmetry. The sign of λ is naturally expected to be positive.

In Nyström and Skoglund (2002a, pp. 10–12) the authors discuss which distribution should be assumed for the standardized residuals z_t for financial data, and find that using normal distribution as approximation for the high quantiles might lead to a significant underestimation. Related literature suggests the t distribution as an alternative distributional assumption, which might be more accurate in capturing fat tails, but unable to capture asymmetry. In Nyström and Skoglund (2002a), they use Extreme Value Theory to account for both the fat tails and the skewness and asymmetry of financial data. In this paper, we will apply this combined method.

4.3 Extreme value theory

Extreme value theory (EVT) is the study of improbable, but extreme events. EVT is more commonly used in weather and insurance, but has over the past decade become more popular also in financial studies. In Embrechts et al. (1997) the authors introduced a full framework for the analysis, and they argue that EVT should be given more attention in risk management for financial institutions. In McNeil et al. (2015) it is proposed a combination of GARCH-EVT models where the GARCH standardized residuals are used as input to EVT, since EVT requires the residuals to be i.i.d.

The theoretical framework for Extreme Value Theory is extensively shown in Nyström and Skoglund (2002b) and Embrechts et al. (1997).

The generalised Pareto distribution (GPD) is introduced for any $\xi \in \mathbf{R}$, $\beta \in \mathbf{R}_+$:

$$GP_{\xi,\beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)_+^{-\frac{1}{\xi}}, x \in \mathbf{R} \tag{5}$$

where $1/\xi$ is the tail index and x represent exceedances of standardized residuals z_t over the threshold that delimitates the extreme tail.

There are two practical methods in the literature for locating the threshold beyond which we define the tail of extreme values. The first is the block maxima method. In this approach we define blocks in the data, and then extract the maxima (maximum loss) in each block. There are several drawbacks of this approach. The local maxima in a block might not capture the actual maximas in the time series, and the second

and third maxima in a block might be of significance to the investor but will not be captured by the block maxima approach.

The second method, the Peak over threshold, focuses on the events that exceed a specified, high threshold. Here the observations over the threshold are asymptotically described by the generalised Pareto distribution. The Peak over threshold is the preferred method for practitioners as it makes better use of the data, and so we will use this approach. Determining the optimal threshold is challenging, and there are several methods which can be used. However, Nyström and Skoglund (2002b) argue that the threshold should be between 5 and 13% of the data.

$\tilde{\xi}$ and $\tilde{\beta}$ are parameter estimates of the generalised Pareto distribution. To estimate the parameters in the GPD we use maximum likelihood. This is the preferred method as it provides estimates of the parameters that are consistent and asymptotically normal as $n \to \infty$ given that $\xi > -1/2$. When using the maximum likelihood it is nearly invariant to the level of the threshold given that the threshold is within a reasonable limit.

Kernel smoothed interior

The data between the lower and upper tail thresholds are fitted by a Gaussian Kernel estimator. A kernel estimator is a function that derives a smooth curve from the observed data that is the best possible representation of the probability density.

4.4 Dependence structure

The GARCH-GJR-EVT process focuses on modelling the distribution of individual risk factors by modelling the conditional volatility, asymmetric adjustment and fat tails. However, this is done by modelling each risk factor in isolation, and it does not contain information about the dependence structure, which is a very important part of stress testing.

A copula allows for modelling the underlying joint distribution of two or more assets by only specifying the marginals. In Alexander (2008a), it is pointed out that one of the advantages of using a copula is that it isolates the dependence structure from the structure of the different marginal distributions. The theoretical background behind copulas was introduced in 1959 by Sklar. The use of copulas to measure dependence became more popular in the literature in the end of the 1990s, but only in the recent decade copula became a popular method employed in financial applications.

We will only focus on the theoretical background of the symmetric t copula. For a more detailed and complementary background of copulas we refer the reader to Alexander (2008a) or McNeil et al. (2015).

The multivariate t copula can be derived from the multivariate t distribution, and is defined as (Alexander 2008a, p. 268):

$$C_v(u_1, \ldots, u_n; \mathbf{\Sigma}) = \boldsymbol{t}_v(t_v^{-1}(u_1), \ldots, t_v^{-1}(u_n)), \qquad (6)$$

where $\boldsymbol{t}_v$ and t_v are multivariate and univariate Student t distribution functions. v is the degrees of freedom, and $\mathbf{\Sigma}$ is the correlation matrix.

5 Estimation procedure

In this section, we give an overview of technical details concerning the calibration of modeling approaches specified in Sect. 4 and show estimation results.

5.1 Application of the GARCH-GJR

To find the appropriate lag structure for the GARCH(p, q) process we estimate models with q and p ranging from 1 to 6. To select the best model for the data we perform Akaike (AIC) and Bayesian (BIC) information criteria (Box et al. 2015, p. 193). These criteria are the preferred ones for selecting the best GARCH fit for the data because it penalises models for additional parameters estimated.

$$AIC = -2(log\hat{L}) + 2NumParams$$
$$BIC = -2(log\hat{L}) + NumParams * log(n)$$

The process which minimises these criteria is considered to be the best specification. Table 4 shows the results from the AIC and BIC criterion tests for lags from 1 to 2. We tested up to 6 lags, but the results show insignificant parameters, and higher AIC and BIC criterion than the displayed models. The table indicates that the GARCH(1, 1) is the optimal choice overall, and we will continue with this specification. Similar results are found in Paraschiv et al. (2015) and Aepli (2011).

Table 5 displays the estimated GARCH-GJR parameters. The parameters closely align with previous empirical results for financial assets. Following Alexander (2008a, p. 137), the β is a measurement of the persistence in conditional volatility regardless of what happens in the market. Large β, above 0.9, indicates that high volatility following market stress will persist for a long time which is true for all of our commodities. α measures the reaction of conditional volatility to shocks in the market. The sum of the two parameters is the rate of convergence, for our risk factors the sum is close to 1, indicating high persistence and a relatively flat term structure of volatility forecasts. From the table we see that the estimated ARCH and GARCH coefficients, $\hat{\alpha}$ and $\hat{\beta}$, are significantly different from zero for all commodities. We found a significant leverage effect in the returns of oil products, copper and live cattle. Empirical evidence shows that after 2008 holding crude oil as financial asset gave higher returns than holding it as commodity, given the reduction in convenience yields and a change from backwardation to contango (see Kolodziej et al. 2014). We did not find evidence for the expected leverage effect in agricultural commodities and metals (excluding copper) though. In this case, negative impacts on returns are associated to negative shocks to volatility.

Figure 4 displays the filtered residuals and the filtered conditional standard deviation of WTI, as given in Eq. (2). The other commodities show similar results, and are available upon request. We observe that the GARCH process models realistically the volatility clustering pattern in commodity returns.

To be able to apply EVT to the tails we need to standardise the filtered residuals from each return series. The standardised residuals are calculated by dividing the

Table 4 AIC & BIC criterion for *t* distributed residuals for GARCH process with various (p, q) lag structures

| Commodity	GARCH(p, q)							
	AIC criterion				BIC criterion			
	(1, 1)	(1, 2)	(2, 1)	(2, 2)	(1, 1)	(1, 2)	(2, 1)	(2, 2)
Wheat	*−32,810*	−32,809	−32,805	−32,808	*−32,764*	−32,749	−32,752	−32,741
Corn	*−33,868*	−33,865	−33,866	−33,863	*−33,821*	−33,805	−33,812	−33,796
Soybeans	*−34,272*	−34,271	−34,270	−34,268	*−34,225*	−34,211	−34,217	−34,202
Live cattle	*−42,571*	−42,569	−42,568	−42,569	*−42,524*	−42,509	−42,515	−42,503
Copper	*−33,360*	−33,357	−33,358	−33,357	*−33,313*	−33,297	−33,305	−33,290
Gold	*−37,650*	−37,649	−37,648	−37,644	*−37,603*	−37,589	−37,594	−37,578
Aluminium	*−36,445*	−36,444	−36,439	−36,446	*−36,399*	−36,384	−36,385	−36,379
WTI	*−32,432*	−32,428	−32,434	−32,431	*−32,385*	−32,368	−32,381	−32,364
Brent	*−32,637*	−32,634	−32,642	−32,638	*−32,591*	−32,574	−32,589	−32,572
Natural gas	*−25,884*	*−25,891*	−25,882	−25,889	*−25,838*	−25,831	−25,829	−25,823

Preferred model highlighted in italic

Table 5 Estimated GARCH-GJR(1, 1) parameters for the variance equation

| Commodity	$\hat{\omega}$ (SE)	$\hat{\alpha}$ (SE)	$\hat{\beta}$ (SE)	$\hat{\lambda}$ (SE)
Wheat	1.36e−06** (5.63e−07)	0.043*** (0.005)	0.966*** (0.004)	− 0.030*** (0.006)
Corn	1.43e−06** (5.80e−07)	0.097*** (0.010)	0.913*** (0.007)	− 0.021* (0.012)
Soybeans	2.36e−06** (7.30e−07)	0.071*** (0.009)	0.937*** (0.007)	− 0.039*** (0.009)
Live cattle	1.67e−06** (4.21e−07)	0.043*** (0.011)	0.907*** (0.009)	0.076*** (0.016)
Copper	1.05e−06** (5.30e−07)	0.034*** (0.006)	0.955*** (0.005)	0.017** (0.007)
Gold	3.80e−07 (3.01e−07)	0.056*** (0.007)	0.956*** (0.004)	− 0.024** (0.008)
Aluminium	8.35e−07** (4.13e−07)	0.053*** (0.007)	0.946*** (0.006)	− 0.007 (0.009)
WTI	6.96e−07 (5.19e−07)	0.034*** (0.006)	0.955*** (0.005)	0.019** (0.008)
Brent	8.07e−07* (5.30e−07)	0.036*** (0.006)	0.955*** (0.005)	0.015** (0.008)
Natural gas	1.28e−05*** (2.74e−06)	0.040*** (0.007)	0.944*** (0.008)	− 0.002 (0.008)

Standard errors (SE) in brackets. ω is the constant, α is the reaction, β is the persistence, λ is the leverage
****Significant on 1% level; **:significant on 5% level, *significant on 10% level
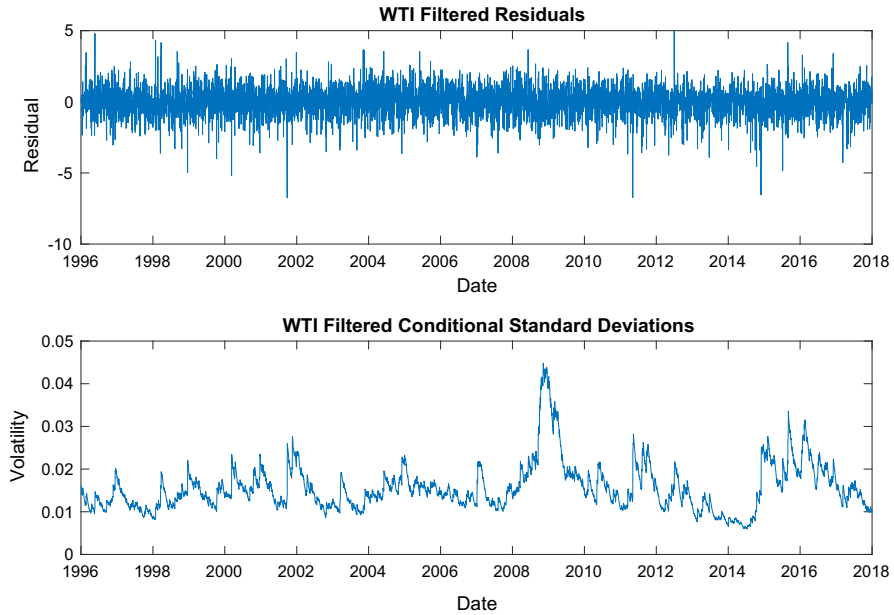
Fig. 4 Filtered residuals and filtered conditional standard deviation for WTI. Corresponding graphs for the other commodities are available upon request
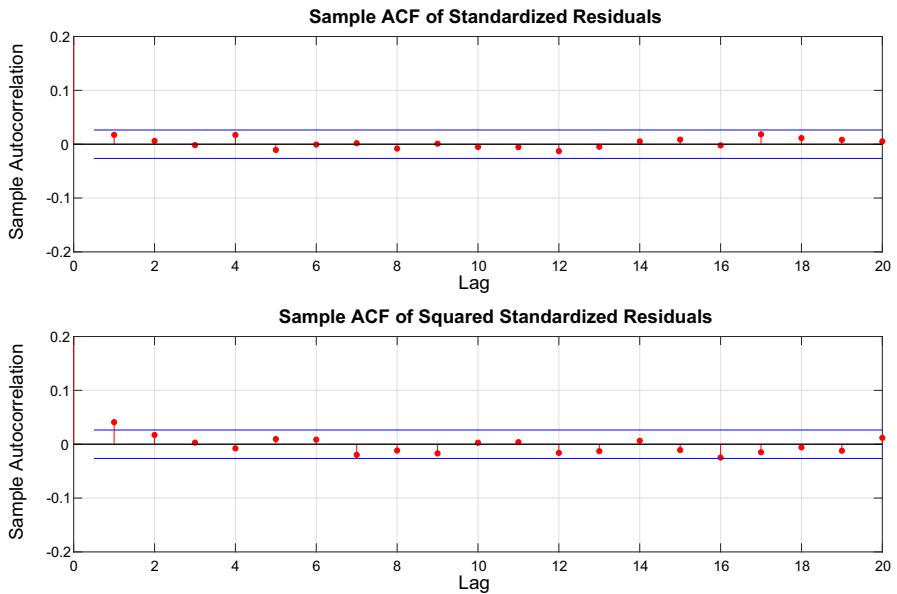
Fig. 5 Sample autocorrelation plot of standardised and squared standardised WTI residuals, showing that the residuals are now i.i.d. Corresponding figures for the other commodities are available upon request

Table 6 Comparison of upper tail parameters (ξ) for different thresholds

Commodity	Upper tail (ξ), Threshold = u			
	u = 7%	u = 10%	u = 11%	u = 12%
Wheat	0.0822	0.0223	0.0082	0.0063
Corn	0.1885	0.1775	0.1554	0.1320
Soybeans	0.0816	0.0504	0.0430	0.0356
Live cattle	0.2781	0.2854	0.2702	0.2714
Copper	− 0.0069	− 0.0290	− 0.0572	− 0.0555
Gold	0.2011	0.1450	0.1485	0.1413
Aluminium	− 0.0368	0.0022	0.0117	− 0.0278
WTI	0.0646	0.0101	− 0.0049	− 0.0158
Brent	− 0.0021	0.0086	− 0.0101	0.0006
Natural gas	0.1583	0.0728	0.0688	0.0871

The rest of the parameters from the different thresholds are found in Table 13

filtered residuals with the conditional variance $z_t = \frac{\epsilon}{\sigma_t}$ to obtain mean zero and unit variance. The standardised residuals are plotted in Fig. 5. We can now see graphically that the residuals are i.i.d. for WTI. The other commodities show similar results and are available upon request. The residuals are now applicable to be modelled by EVT.

5.2 Application of EVT

5.2.1 Estimation of the semi-parametric cumulative distribution functions

To locate the threshold, we have fitted the Generalized Pareto distribution (GPD) to the standardized residuals testing for parameter stability for different threshold values between 5% and 15%. This allows us to find a threshold where the tail indexes stabilise. In Table 6 we display the 7%, 10%, 11% and 12% thresholds upper tail index. The rest of the parameters from the different thresholds are located in Table 13. Notice that the tail index naturally becomes smaller as the threshold allows for more data in the extreme tail. However, generally speaking, the value of "ξ" observed at the 10% threshold has a similar value at subsequent thresholds.

We will therefore define the lower/upper tails in all commodity returns as starting at the 10% and 90% quantiles respectively. Previous studies (Aepli 2011; Paraschiv et al. 2015) have chosen a 10% threshold. We see the advantage of choosing a standard threshold as it gives us the opportunity to compare directly parameter estimates among commodity returns.

The next step is to fit the generalised Pareto distribution to the exceedances over threshold by using maximum likelihood. By optimising the log-likelihood function we estimate the tail indexes ξ and scale parameters β.

The estimated parameters for our risk factors are listed in Table 7. We have $\xi > 0$ for nine of the ten risk factors. This case coincides with the Fréchet distribution, which is characterized by a lower bound and gives an indication of an extreme tails in

Table 7 Maximum likelihood estimators for the generalized Pareto distribution parameters

Commodity	ξ		β	
	Upper tail	Lower tail	Upper tail	Lower tail
Wheat	0.0223	0.0822	0.6163	0.4942
Corn	0.1775	0.1561	0.5396	0.5081
Soybeans	0.0504	0.1014	0.5595	0.5924
Live cattle	0.2854	0.2509	0.4534	0.5704
Copper	− 0.0290	0.1178	0.5818	0.5654
Gold	0.1450	0.1003	0.5203	0.6121
Aluminium	0.0022	0.1003	0.5551	0.5197
WTI	0.0101	0.0926	0.5166	0.5805
Brent	0.0086	0.0664	0.5152	0.5965
Natural gas	0.0728	0.0714	0.6126	0.5191

Threshold (u) = 10%

Fig. 6 Generalized Pareto upper tail of the standardised residuals fitted versus empirical. Corresponding figures for the other commodities are available upon request
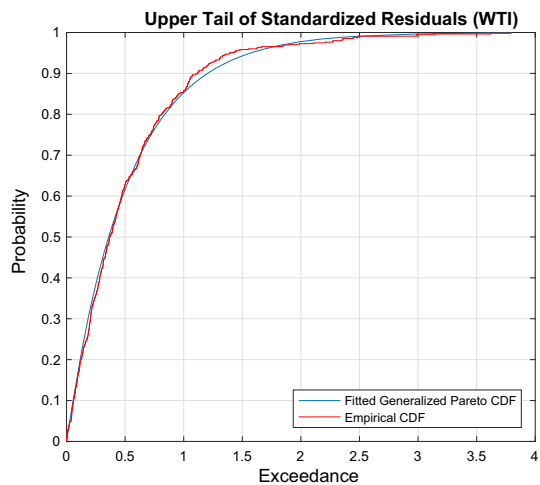

commodity returns. Only copper shows a negative sign for "ξ" (Weibull distribution, having an upper bound). The lower tails show a positive sign of ξ for all commodities. This suggests both fat upper and lower tails, and furthermore tail asymmetry. The findings are consistent with the theory and empirical results for financial time series (Nyström and Skoglund 2002b; Embrechts et al. 1997).

In Fig. 6 we display the empirical cumulative distribution function of the upper tail of the standardised residuals for WTI. The fitted distribution follows the empirical exceedances closely, and so the chosen distribution is well suited to estimate the tails for the commodities.

The last step is to combine the parametric generalized Pareto tails for each commodity with the corresponding Kernel smoothed interior to obtain the entire semi-parametric cumulative distribution function. Figure 7 displays the semi-parametric
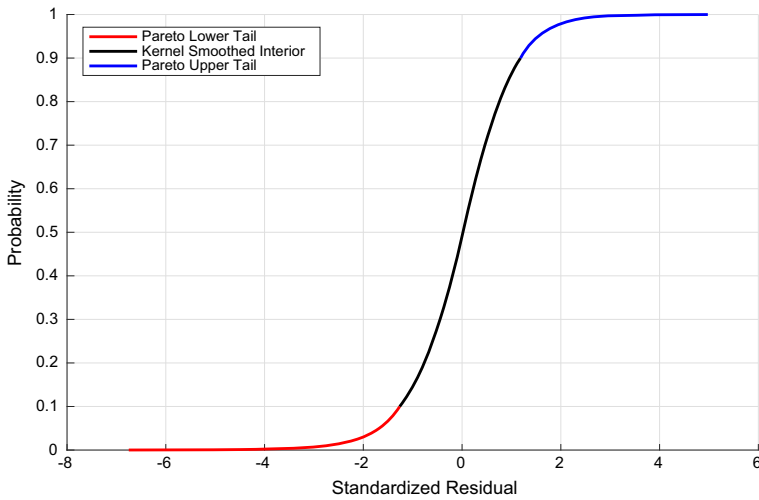
Fig. 7 Semi-parametric empirical cumulative distribution function of WTI. Corresponding figures for the other commodities are available upon request

empirical cumulative distribution function of WTI standardized residuals. The piecewise distribution object allows interpolation within the interior of the CDF, displayed in black, and extrapolation in each tail, displayed in red and blue for the lower and upper tail, respectively. The extrapolation allows for estimation of quantiles outside the historical record, and is therefore important for the stress testing exercise.

5.3 Application of the *t* copula and simulation steps

We fit a *t* copula to the standardized residuals of portfolio return series. Estimates of the degrees of freedom for the baseline scenario are given in Table 9. Given the parameters of the *t* copula (the correlation matrix Σ and the degrees of freedom parameter) we simulate jointly dependent portfolio returns. This is done by first simulating the corresponding dependent standardised residuals. We transform the dependent uniform variates obtained from fitting the *t* copula to standardised residuals through the inversion of the semi-parametric marginal cumulative distribution function of each risk factor. We therefore extrapolate into the generalized Pareto tails, and interpolate into the smoothed interior. This gives simulated standardized residuals consistent with those obtained from the GARCH-GJR(1, 1) filtering process described in Sect. 5.1. Residuals show no autocorrelation and are i.i.d. with unit variance.

5.4 Risk metrics

To compare the implications of various stress scenarios on the portfolio profit and loss profile, the use of Value at risk (VaR) and conditional value at risk (CVaR) is common. These risk metrics provide an indication of the quantile losses. VaR is the amount of

maximum potential loss at a given percentage. This risk metric is criticized because it is not coherent and ignores extreme values beyond the value at risk. CVaR corrects for the limitations of VaR (Alexander 2008b). For our analysis we include both risk metrics at various quantiles. This is in line with European Banking Authority (2017, p. 28): "The institutions should stress the identified risk factors using different degrees of severity as an important step in their analysis to reveal nonlinearities, threshold effects, i.e. critical values of risk factors beyond which stress responses accelerate".

5.5 Simulation steps

Based on the technical specifications given in Sect. 4, we simulate the profit and loss distribution of our portfolio of commodity returns at the end of the given time horizon as realistically as possible *without* and *with* the impact of stress. Generally speaking, in each type of stress scenario the technical simulation steps are outlined here:

1. To simulate jointly dependent equity returns with the parameters of the *t* copula one first of all has to generate the corresponding dependent standardized residuals. This is done by simulating **dependent uniform variates** based on the estimated degrees of freedom parameter and correlation matrix.
2. We transform them by inversion of the according share's semi-parametric marginal cumulative distribution function (Pareto tails and Gaussian kernel smoothed interior). The result are **standardized residuals** consistent with the ones obtained from the filtration of z_t in the GARCH model, namely i.i.d. with variance 1.
3. These simulated standardized residuals are then employed as the i.i.d. noise processes of the GARCH model. We simulate the asymmetric GARCH model to reestablish the heteroscedasticity and the autocorrelation of the original **commodity returns**. We use as seed for the GARCH model the last observed values of the data set and according volatilities.
4. The weights of the portfolio are held constant over the simulation horizon. We calculate the maximum simulated profit and loss (P&L distribution), the VaR (value at risk) and the expected shortfall (ES)

6 Stress testing and simulation results

In this section, we will perform stress tests on our portfolio of commodity futures.

By definition, stress testing is a risk management tool used to evaluate the potential impact on portfolio profit and loss profile of unlikely, although plausible historical or hypothetical events or movements in the portfolio risk factors. We will shock at one time various components of the model and assess which ones have the highest effect on the simulated profit and loss. Shocks are linked to stress scenarios as explained in this section.

For comparison purposes, we simulate a baseline scenario by calibrating the model on the entire data set and compare the profit and loss profile with others derived from stress scenarios as derived below. For each scenario we run 20,000 simulations over a 22 days horizon, which represent the average number of working days per

month. Note that the portfolio weights are held fixed over the risk horizon and that the simulation ignores any transaction costs required to re-balance the portfolio (the daily re-balancing process is assumed to be self-financing).

6.1 Stress test scenarios

We limit the study mostly to hybrid scenarios, where we calibrate the model for the risk factors or the copula to the restricted financial crisis data set versus the entire data set, and shock them simultaneously or one at a time. Besides hybrid, hypothetical scenarios could have been implemented. Examples could be a scenario with recession in China, which would decrease the demand for aluminium, oil, copper, soybeans and natural gas, and look at the change in dependence structure and volatility. Other scenarios could be natural disasters that affect crops or diseases that affect grains or livestocks. This is however out of the scope of this article. We refer the reader to Aepli (2011) for stress testing with hypothetical scenarios.

Our analysis consists of seven different scenarios. Underneath follows a brief description of the scenarios before the analysis is conducted.

Baseline scenario:
The baseline scenario is a default scenario simulation with the t copula and GARCH-GJR process calibrated on the entire data set, the historical time period from 1996 to 2017. None of the parameters are stressed in the baseline scenario. The baseline scenario is constructed to be a reference for normal times to assess the effect of stressing parameters compared to the steady state.

Historical scenario:
For the historical stress scenario we use the years 2007 and 2008 to observe the severity of losses in the financial crisis. This time period is known for high market stress with high return volatility and captures the simultaneous price drop during the financial crisis (see Sect. 3 for discussion). The historical scenario is a scenario that stems from the empirical distribution of returns during the financial crisis. We refer to the empirical profit and loss distribution as observed between 2007 and 2008. Unfortunately, due to the limited number of observations when resuming ourselves to observed returns, extreme loss quantiles are hard to estimate.

Hybrid scenarios:
Due to the limitations of the historical empirical scenario, we construct five hybrid scenarios. Hybrid scenarios allow extrapolation beyond realized returns, and are therefore appropriate to estimate extreme quantiles and events that have not yet occurred. The focus in our hybrid scenario construction is to examine which of the estimated parameters challenge mostly the portfolio profit and loss distribution in stress testing exercises. The parameters that changed between the different scenarios are the dependencies between risk factors, measured in Degrees of Freedom and correlations, and the GARCH-GJR coefficients. The parameters of the GARCH-GJR process and the t copula are here re-calibrated on the stress horizon, following the same procedure as in Sect. 4. The re-estimated tails from the generalized Pareto distribution parameters can be found in Table 14. To isolate the effect of various parameters, we compare scenar-

Table 8 Description of input parameters for simulation in hybrid scenarios

Scenario	Marginal distribution	Correlations	Degrees of freedom
Risk factor stress	Stress	Baseline	Baseline
Dependency stress	Baseline	Stress	Stress
Full stress	Stress	Stress	Stress
DoF shock	Baseline	Baseline	Stress
Risk factor stress without EVT	Stress	Baseline	Baseline

Baseline means the parameters from entire dataset 1996–2017 are used as input. Stress means the parameters are re-calibrated on our chosen time of financial distress, years 2007–2008

ios by mixing parameters from the baseline with those during the period of financial distress. The hybrid scenarios are described in Table 8.

Risk factor stress scenario aims to show the impact of stressing the model parameters describing the marginal distributions of the risk factors on the portfolio profit and loss distributions, without a change in the dependence between the risk factors.

Dependence stress scenario isolates the effect of stressing the dependence between the returns of portfolio components on the profit and loss distribution, without changing the parameters for the individual factors model (GARCH-GJR model).

Full stress scenario aims to simulate the effects of a recurring financial crisis on the portfolio. All model parameters refer to the financial crisis period

In the **degrees of freedom shock** we shock only the degrees of freedom of the copula, leaving all other parameters unchanged.

Risk factor stress **without EVT** highlights how the application of Extreme Value Theory to model the tails of portfolio components returns affect the profit and loss distribution of the portfolio. The risk factor distributions are here not modelled with EVT, but with a Student t distribution (see Sect. 4.1).

6.2 Comparative analysis of simulated profit and loss distributions

6.2.1 Baseline scenario versus historical scenario

Figure 8 displays the simulated profit and loss (P&L) distribution for the returns in the baseline scenario versus the empirical distribution of P&L in the historical scenario. The simulated returns deviate in both the upper and lower tails. This can be further viewed in Table 9, where the maximum simulated loss is significantly larger for the historical returns than the simulated baseline, respectively. These results might be linked to the symmetry of the t copula. The baseline scenario represents normal market conditions while with the historical scenario we get an indication of its profit and loss profile assuming that a similar crisis will reoccur. This result highlights the importance of implementing forward-looking scenarios, both to simulate extreme returns in comparison to the baseline and to simulate beyond the profit and loss profile empirically observed.
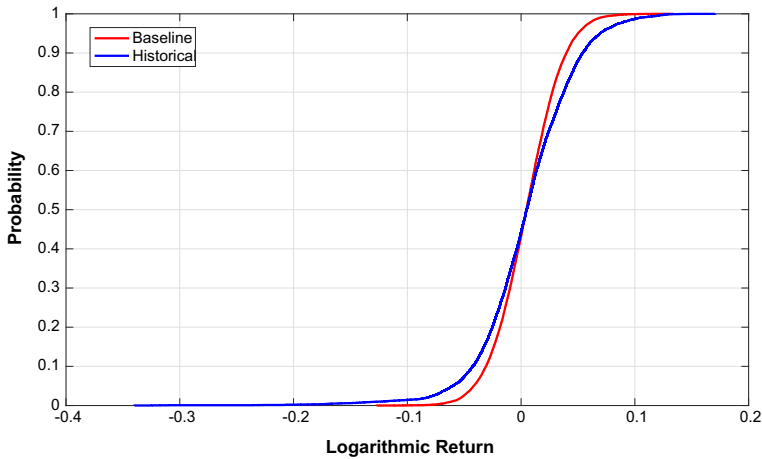
Fig. 8 Portfolio returns simulation, baseline versus historical scenario

Table 9 Simulation metrics for baseline scenario and historical scenario CDF

| Metric	Baseline	Historical scenario
Degrees of freedom	15.28	N/A
Max. simulated loss	− 12.72%	− 34.01%
Max. simulated gain	13.38%	17.08%
Simulated 90% VaR	− 3.09%	− 4.26%
Simulated 95% VaR	− 4.17%	− 6.01%
Simulated 99% VaR	− 6.13%	− 12.47%
Simulated 90% CVaR	− 4.48%	− 7.38%
Simulated 95% CVaR	− 5.39%	− 9.75%
Simulated 99% CVaR	− 7.30%	− 17.82%
Simulated 99.9% CVaR	− 10.10%	N/A
Simulated 99.99% CVaR	− 12.55%	N/A

The previous statement is further substantiated when we look at very high confidence levels displayed in Table 9. The historical scenario is limited to already experienced events so there are not enough observations in the data set to calculate the expected shortfall at very high confidence levels. This emphasizes the discussion about the scenarios in Sect. 2.1 and the drawback of using historical scenarios highlighted in Basel Committee on Banking Supervision (2009). In addition, the historical scenario neglects the dependence structure between the risk factors, which is highly relevant in stress testing. In European Banking Authority (2017, p. 24), it is stated that stress tests should take into account changes in correlations between risk types and risk factors and that correlations tend to increase during times of economic or financial distress. This statement and its implications for stress testing exercises will be further investigated in the next subsection where we analyse the hybrid scenarios.

Table 10 Risk metrics for hybrid scenarios

Metric	1 Risk factor stress	2 Dependency stress	3 Full stress	4 DoF shock	5 Without EVT
Degrees of freedom	15.28	13.78	13.78	13.78	15.28
Max. simulated loss	−36.40%	−18.67%	−45.90%	−16.53%	−34.24%
Max. simulated gain	31.93%	16.32%	42.89%	13.82%	32.38%
Simulated 90% VaR	−8.84%	−3.75%	−10.85%	−3.08%	−6.91%
Simulated 95% VaR	−11.91%	−5.05%	−14.53%	−4.12%	−9.78%
Simulated 99% VaR	−17.87%	−7.46%	−21.92%	−6.13%	−15.79%
Simulated 90% CVaR	−12.98%	−5.47%	−15.89%	−4.49%	−10.91%
Simulated 95% CVaR	−15.71%	−6.61%	−19.31%	−5.42%	−13.61%
Simulated 99% CVaR	−21.55%	−9.08%	−26.74%	−7.38%	−19.57%
Simulated 99.9% CVaR	−30.35%	−12.99%	−37.17%	−10.63%	−28.04%
Simulated 99.99% CVaR	−35.79%	−17.40%	−44.77%	−15.00%	−33.72%
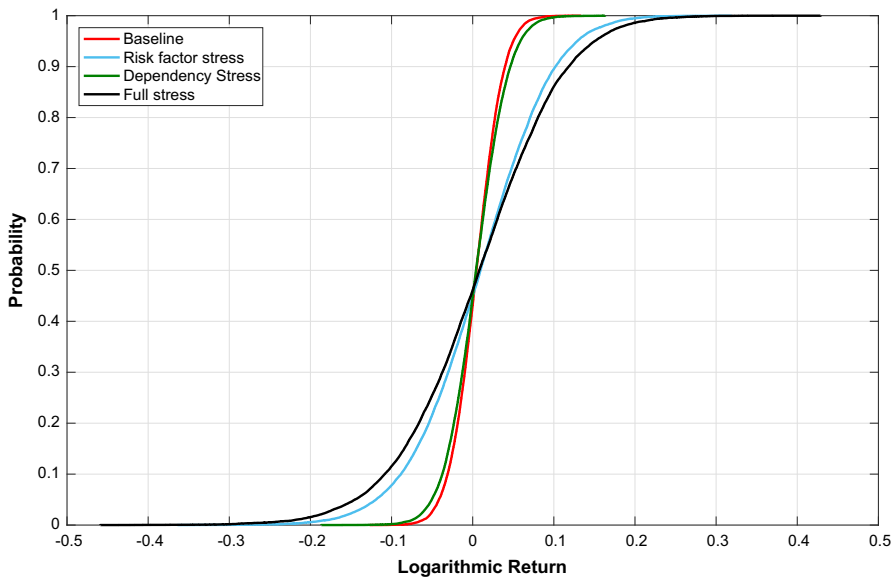
Fig. 9 Simulated one-month portfolio returns CDF for baseline versus hybrid scenarios: risk factor stress, dependency stress and full stress

6.2.2 Hybrid scenarios

Table 10 shows the risk metrics over the 5 sets of hybrid scenarios. The tail dependence for the simulated returns is measured in the degrees of freedom parameter from the t copula. From the entire data set the DoF is 15.28, while during the stressed period they shift to 13.78. Our decrease in DoF signals that the tail dependence in the commodity portfolio is increasing during times of stress. Lower degrees of freedom indicate a higher tendency of extreme events to occur jointly across risk factors (Paraschiv et al. 2015), which is in line with our simulation result.

Risk factor stress versus dependency stress

Figure 9 shows the baseline scenario, the scenario where we stress the dependencies between the risk factors, the full stress scenario and the scenario where the individual risk factors are stressed. Starting from the baseline we can see that by only stressing the dependencies, the simulation displays more severe losses (green vs. red). The correlation matrix and the decrease in DoF show that the dependencies between the risk factors increase in times of stress (see Table 11), which leads to larger simulated losses for the portfolio overall. However, by stressing only the GARCH-GJR-EVT parameters for the individual risk factors the effect on the portfolio P&L is even stronger (red vs. light blue). This result indicates that stressing the model parameters describing the marginal distributions of portfolio returns has a larger impact on the profit and loss distribution than stressing the dependencies between the risk factors. This shows that shocks in returns of portfolio components

Table 11 Correlation increases between baseline and stress scenario

Commodity	Wheat	Corn	Soybeans	Live cattle	Copper	Gold	Aluminium	WTI	Brent	Natural gas
Wheat	–	−0.0342	0.0085	0.1184	0.1634	0.1518	0.0993	0.1622	0.1688	0.0579
Corn	−0.0342	–	0.0562	0.1542	0.1561	0.1564	0.1569	**0.2181**	**0.2156**	0.1259
Soybeans	0.0085	0.0562	–	0.1797	0.1634	0.1851	0.1237	**0.2668**	**0.2614**	0.1246
Live cattle	0.1184	0.1542	0.1797	–	0.1342	0.1327	0.1160	0.1546	0.1552	0.1409
Copper	0.1634	0.1561	0.1634	0.1342	–	**0.2088**	0.0583	0.1615	0.1631	0.0925
Gold	0.1518	0.1564	0.1851	0.1327	**0.2088**	–	0.1596	**0.2597**	**0.2582**	0.1304
Aluminium	0.0993	0.1569	0.1237	0.1160	0.0583	0.1596	–	0.1573	0.1671	0.1023
WTI	0.1622	**0.2181**	**0.2668**	0.1546	0.1615	**0.2597**	0.1573	–	0.0750	0.1907
Brent	0.1688	**0.2156**	**0.2614**	0.1552	0.1631	**0.2582**	0.1671	0.0750	–	0.1959
Natural gas	0.0579	0.1259	0.1246	0.1409	0.0925	0.1304	0.1023	0.1907	0.1959	–

The correlation matrices from the Baseline and financial stress period are located in Tables 15 and 16. (Italic is correlation increase larger than 0.15. Bold is correlation increase larger than 0.20)

is of higher impact on the profit and loss than the shifts in their dependence structure and correlations.

We compare further the mentioned scenarios with the full stress scenario. Naturally this stress scenario simulates the largest tail losses since both the dependencies and the individual parameters are stressed (black line). Comparing the risk metrics in Table 10 we see that the risk factor stress scenario simulates the second largest losses, after the full stress scenario, which substantiates the previous result.

The full stress scenario in Paraschiv et al. (2015) gives more severe losses overall. Our study replicated the methodology for modeling the marginal distributions of portfolio components and dependence structure, allowing for a direct comparison of results. The difference between the results might be explained by: (i) The difference in weights of the test portfolio where our study uses weights from 2017 while Paraschiv et al. (2015) use the weights from 2013. (ii) Our extended data sample. We include the years 1996–1998, and 2011–2017 beyond the original data set. (iii) Differences might be due partially to the randomness in the scenario generation.

In Paraschiv et al. (2015) natural gas makes 15.11% of the portfolio, while for our portfolio it is 9.6%. From the descriptive statistics natural gas is by far the most volatile commodity and we observe that natural gas has performed poorly over the last decades compared to most of the other commodities. Several structural breaks in natural gas prices are also included in our data set, examples being the supply shortfall in Libya 2011 and the Russian export stop in 2012 (Nick and Thoenes 2014). Weighting more the natural gas the portfolio might therefore be one of the main reasons of the more severe simulated loss in the original study.

Soybean is the second commodity with the most deviating weight from Paraschiv et al. (2015). In our portfolio soybean makes 15.66% of the total weight, in comparison to 6.89% in Paraschiv et al. (2015). Over our time period soybean returns showed low volatility. We expect that the increased allocation in soybean in our portfolio provides the same consequences as the down-scaling of natural gas.
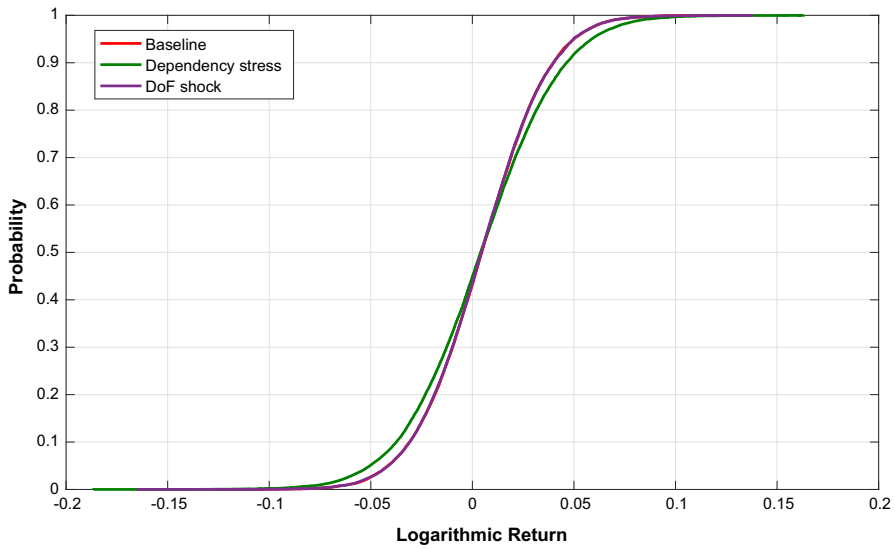
DoF Shock Versus Dependency Stress

In Fig. 10 we compare the scenario where both correlation and DoF are stressed (green line), with the scenario where only the DoF are shocked from 15.28 for the baseline to 13.78 for the financial crisis (purple line). For both scenarios the parameters of the GARCH-GJR-EVT model are calibrated on the entire data set. By doing so we can discuss the impact of correlations as a driver of losses in isolation.

From Fig. 10 we observe that a small shock to degrees of freedom does not provide a significant stress scenario. The baseline scenario and the DoF shock scenario do not deviate much from each other (red vs. purple), although the DoF shock scenario simulates larger extreme losses in the lower quantiles (see Table 10). Furthermore, we see that the scenario where both the correlations between the risk factors as well as the DoF are shocked displays the largest simulated loss. This indicates that shocking the DoF in isolation is a limitation in a stress testing exercise.

For more forward-looking hypothetical scenarios the implementation of more severe shocks to DoF might be of interest. We therefore tested by including a set of hypothetical scenarios where several more substantial downward changes to DoF

Fig. 10 Simulated one-month portfolio returns CDF for baseline versus hybrid scenarios: dependency stress and DoF shock

Table 12 Risk metrics from different DoF shock scenarios

| Metric	DoF					
	5	7	10	13.78	15.28	17
Max. simulated loss	− 16.72%	− 17.25%	− 13.85%	− 16.53%	− 12.72%	− 15.88%
Max. simulated gain	15.28%	15.47%	14.05%	13.82%	13.38%	16.77%
Simulated 90% VaR	− 3.08%	− 3.11%	− 3.12%	− 3.08%	− 3.09%	− 3.11%
Simulated 95% VaR	− 4.12%	− 4.18%	− 4.20%	− 4.12%	− 4.17%	− 4.19%
Simulated 99% VaR	− 6.18%	− 6.32%	− 6.31%	− 6.13%	− 6.13%	− 6.15%
Simulated 90% CVaR	− 4.50%	− 4.55%	− 4.55%	− 4.49%	− 4.48%	− 4.53%
Simulated 95% CVaR	− 5.46%	− 5.52%	− 5.50%	− 5.42%	− 5.39%	− 5.47%
Simulated 99% CVaR	− 7.57%	− 7.64%	− 7.48%	− 7.38%	− 7.30%	− 7.47%
Simulated 99.9% CVaR	− 11.14%	− 11.28%	− 10.36%	− 10.63%	− 10.10%	− 10.54%
Simulated 99.99% CVaR	− 14.44%	− 17.13%	− 12.81%	− 15.00%	− 12.55%	− 14.20%

DoF 15.28 is from the baseline (1996–2017), and DoF 13.78 is from the time of financial stress (2007–2008). The other DoF are hypothetical shocks

are included. The results can be found in Table 12. We used the degrees of freedom from the financial crisis period and also the one corresponding to the baseline scenario and then we added more extreme values at both ends. For our data we found that extreme shocks to DoF yield no substantial increase in simulated tail losses.

Impact of EVT

In Fig. 11 we display two hybrid scenarios to highlight the importance of implementing EVT for modelling extremely large return changes of portfolio components before
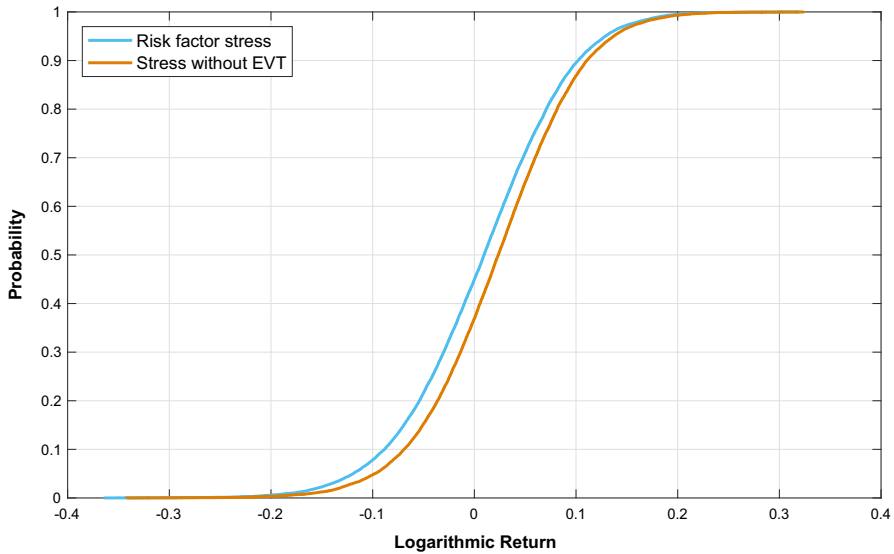
Fig. 11 Simulated one-month portfolio returns CDF for baseline and hybrid scenarios: risk factor stress and scenario without modelling with EVT

running the actual stress test. For both scenarios the correlation matrix and the DoF parameter are calibrated on the entire data set, so the difference between them comes from how the individual risk factors are modelled. In the risk factor stress scenario the tail distributions are modelled with EVT where the tail indexes are calibrated on the financial crisis data, and the other scenario with a Student t distribution. One can see that the scenario where EVT is implemented estimates more severe losses, where simulated 99.99% CVaR is -35.79% in comparison to -33.72% for the scenario without EVT. Overall, the profit and loss distribution in the stress test excluding EVT is shifted to the right. Applying EVT strengthens the accuracy and understanding of the most extreme, potential losses. In light of this, we conclude that the risk is potentially underestimated when the individual risk factor distributions disregard extreme events (Embrechts et al. 1997).

7 Conclusion

In this study, we update the analysis in Paraschiv et al. (2015) with a more extensive data set, and a more detailed focus on stress testing. In particular, hybrid and hypothetical scenarios are explored, in line with the regulatory requirements for stress testing calling for forward looking scenarios. Our stress testing exercise are based on rearranging arbitrarily shocks linked to specific extreme events or time to reveal the importance of correlations, tail correlations, or extreme movements in portfolio components on the profit and loss distribution. This is the first study in the literature that clearly illustrates the marginal impact of the model assumed for the individual portfolio components versus the marginal role of tail dependence and correlations on the portfolio risk profile.

We mimic the DJCI, by forming a portfolio of ten commodities. We use a GARCH-GJR approach to model stylized facts observed in commodity return data, and

implement Extreme Value Theory to model the tails accurately. To account for the dependence structure we apply a *t* copula. We then stress test the portfolio with different scenarios, examining the drivers of the profit and loss distribution.

Our study revealed three main results. First, we bring empirical evidence showing the importance of hybrid (forward-looking) scenarios for comprehensive stress testing. In addition, we show the value added of forward looking over historical scenarios and show numerically the drawbacks of the latter. We confirm the stress testing requirements from Basel III accordingly to which different stress testing approaches cannot be used in isolation, but combined, for a comprehensive picture. Our second finding is that, before implementing a stress test, a special attention should be given to an accurate model identification for the evolution of returns of portfolio components and dependence structures. In addition, our third finding enhanced the previous findings in Paraschiv et al. (2015) by disentangling the effects of stressing at one time the model parameters for the individual portfolio components versus their correlations and tail dependence. We found clear evidence that the first accounts more than the latter while stress testing the portfolio profit and loss profile. At the same time, our analysis represents an integration of the "model risk" concept into stress testing exercises, highly relevant for portfolio managers. Special attention should be given to extreme tails, in line with the regulatory frame on stress testing.

Our analysis is bounded by the number of stress scenarios and simulations based on the random number generator. Stress scenarios display tendencies, and the numbers generated cannot be transferred directly to risk management. On the other hand, the simulations can form expectations and contribute to an overall understanding of stress testing for capital requirements. For further analysis it would be interesting to update our analysis with asymmetric copulas to better capture the dependence structure. In addition, a more extensive use of hypothetical shocks to a commodity portfolio would be of interest.

Appendix

See Tables 13, 14, 15 and 16.

Table 13 Comparison of ML estimators for the GPD parameters for different thresholds (u) for years 1996–2017

Commodity	ξ		β	
	Upper tail	Lower tail	Upper tail	Lower tail
u = 7%				
Wheat	0.082	0.114	0.566	0.477
Corn	0.188	0.233	0.564	0.456
Soybeans	0.082	0.071	0.540	0.645
Live cattle	0.278	0.313	0.516	0.560
Copper	− 0.007	0.116	0.551	0.591
Gold	0.201	0.124	0.488	0.609
Aluminum	− 0.037	0.148	0.592	0.494
WTI	0.065	0.075	0.466	0.624
Brent	− 0.002	0.109	0.527	0.562
Natural gas	0.158	0.123	0.531	0.473
u = 10%				
Wheat	0.022	0.082	0.616	0.494
Corn	0.178	0.156	0.540	0.508
Soybeans	0.050	0.101	0.559	0.592
Live cattle	0.285	0.251	0.453	0.570
Copper	− 0.029	0.118	0.582	0.565
Gold	0.145	0.100	0.520	0.612
Aluminum	0.002	0.100	0.555	0.520
WTI	0.010	0.093	0.517	0.581
Brent	0.009	0.066	0.515	0.597
Natural gas	0.073	0.071	0.613	0.519
u = 11%				
Wheat	0.008	0.092	0.632	0.480
Corn	0.155	0.152	0.557	0.504
Soybeans	0.043	0.098	0.564	0.591
Live cattle	0.270	0.253	0.454	0.553
Copper	− 0.057	0.102	0.620	0.577
Gold	0.148	0.103	0.507	0.602
Aluminum	0.012	0.089	0.546	0.528
WTI	− 0.005	0.098	0.534	0.568
Brent	− 0.010	0.063	0.535	0.597
Natural gas	0.069	0.062	0.614	0.529
u = 12%				
Wheat	0.006	0.075	0.634	0.493
Corn	0.132	0.138	0.580	0.514

Table 13 continued

Commodity	ξ		β	
	Upper tail	Lower tail	Upper tail	Lower tail
Soybens	0.036	0.096	0.571	0.588
Live cattle	0.271	0.266	0.440	0.525
Copper	− 0.056	0.097	0.621	0.578
Gold	0.141	0.073	0.508	0.636
Aluminum	− 0.028	0.070	0.587	0.549
WTI	− 0.016	0.077	0.547	0.588
Brent	0.001	0.044	0.523	0.621
Natural gas	0.087	0.060	0.586	0.528

Table 14 Recalibrated Maximum Likelihood estimators for the generalized Pareto distribution parameters for the time of financial distress, years 2007–2008

Commodity	ξ		β	
	Upper tail	Lower tail	Upper tail	Lower tail
Wheat	0.0452	0.0021	0.4820	0.6991
Corn	− 0.1141	− 0.0896	0.4715	0.5985
Soybeans	− 0.0309	− 0.1642	0.5149	0.7738
Live cattle	0.2708	0.2140	0.4478	0.5416
Copper	− 0.5240	− 0.0681	0.9359	0.5994
Gold	0.1864	− 0.0017	0.3939	0.7062
Aluminium	0.1976	− 0.1715	0.4346	0.6775
WTI	− 0.1542	− 0.1232	0.5696	0.5557
Brent	− 0.1876	− 0.1805	0.5822	0.5868
Natural gas	0.1348	0.0107	0.5633	0.5841

Threshold: 10%

Table 15 Correlation matrix for the baseline, years 1996–2017

Commodity	Wheat	Corn	Soybeans	Live cattle	Copper	Gold	Aluminium	WTI	Brent	Natural gas
Wheat	1.0000	**0.6217**	**0.4768**	0.1259	0.1251	0.1221	0.1048	0.1468	0.1393	0.0793
Corn	**0.6217**	1.0000	**0.6215**	*0.1508*	*0.1556*	0.1393	0.1316	*0.1689*	*0.1634*	0.1071
Soybeans	**0.4768**	**0.6215**	1.0000	0.1462	**0.2035**	*0.1633*	*0.1676*	*0.1995*	*0.1971*	0.1144
Live cattle	0.1259	*0.1508*	0.1462	1.0000	0.0983	0.0450	0.0907	0.1095	0.1035	0.0493
Copper	0.1251	*0.1556*	**0.2035**	0.0450	1.0000	**0.2637**	**0.5829**	**0.2641**	**0.2496**	0.0668
Gold	0.1221	0.1393	*0.1633*	0.0450	**0.2637**	1.0000	**0.2280**	**0.2057**	*0.1979*	0.0735
Aluminium	0.1048	0.1316	*0.1676*	0.0907	**0.5829**	**0.2280**	1.0000	**0.2156**	**0.2023**	0.0714
WTI	0.1468	*0.1689*	*0.1995*	0.1095	**0.2641**	**0.2057**	**0.2156**	1.0000	**0.9083**	**0.2566**
Brent	0.1393	*0.1634*	*0.1971*	0.1035	**0.2496**	*0.1979*	**0.2023**	**0.9083**	1.0000	**0.2368**
Natural gas	0.0793	0.1071	0.1144	0.0493	0.0668	0.0735	0.0714	**0.2566**	**0.2368**	1.0000

Italic is correlation between 0.15 and 0.19. Bold is correlation larger than 0.20

Table 16 Correlation matrix for scenarios re-calibrated on time of financial distress, years 2007–2008

Commodity	Wheat	Corn	Soybeans	Live cattle	Copper	Gold	Aluminium	WTI	Brent	Natural gas
Wheat	1.0000	0.5875	0.4854	0.2443	0.2885	0.2738	0.2041	0.3090	0.3080	0.1372
Corn	0.5875	1.0000	0.6777	0.3050	0.3117	0.2956	0.2885	0.3870	0.3790	0.2330
Soybeans	0.4854	0.6777	1.0000	0.3259	0.3670	0.3484	0.2913	0.4663	0.4586	0.2391
Live cattle	0.2443	0.3050	0.3259	1.0000	0.2326	*0.1777*	0.2067	0.2641	0.2587	*0.1902*
Copper	0.2885	0.3117	0.3670	0.2326	1.0000	0.4725	0.6412	0.4256	0.4127	*0.1594*
Gold	0.2738	0.2956	0.3484	*0.1777*	0.4725	1.0000	0.3876	0.4654	0.4561	0.2038
Aluminium	0.2041	0.2885	0.2913	0.2067	0.6412	0.3876	1.0000	0.3729	0.3694	*0.1737*
WTI	0.3090	0.3870	0.4663	0.2641	0.4256	0.4654	0.3729	1.0000	0.9833	0.4473
Brent	0.3080	0.3790	0.4586	0.2587	0.4127	0.4561	0.3694	0.9833	1.0000	0.4326
Natural gas	0.1372	0.2330	0.2391	*0.1902*	*0.1594*	0.2038	*0.1737*	0.4473	0.4326	1.0000

Italic is correlation between 0.15 and 0.19. Bold is correlation larger than 0.20

References

Adams Z, Glück T (2015) Financialization in commodity markets: a passing trend or the new normal? J Bank Finance 60:93–111. https://doi.org/10.1016/j.jbankfin.2015.07.008

Aepli M (2011) On the design of stress tests. Master's thesis. University of St. Gallen

Aepli MD, Füss R, Henriksen TES, Paraschiv F (2017) Modeling the multivariate dynamic dependence structure of commodity futures portfolios. J Commod Mark 6:66–87. https://doi.org/10.1016/j.jcomm.2017.05.002

Alexander C (2008a) Market risk analysis: practical financial econometrics, vol 2. Wiley, England

Alexander C (2008b) Market risk analysis: value at risk models, vol 4. Wiley, England

Baffes J, Kose MA, Ohnsorge F, Stocker M (2015) The great plunge in oil prices: causes, consequences and policy responses, (policy research note). World Bank Group. https://openknowledge.worldbank.org/discover?query=the%20great%20plunge. Accessed 26 Apr 2018

Basak S, Pavlova A (2016) A model of financialization of commodities. J Finance 71(4):1511–1556. https://doi.org/10.1111/jofi.12408

Basel Committee on Banking Supervision (2006) International convergence of capital measurement and capital standards. Technical report, Bank for International Settlements. https://www.bis.org/publ/bcbs128.htm. Accessed 11 Apr 2018

Basel Committee on Banking Supervision (2009) Principles for sound stress testing practices and supervision. Technical report. Bank for International Settlements. https://www.bis.org/publ/bcbs155.htm. Accessed 11 Apr 2018

Basel Committee on Banking Supervision (2017) Supervisory and bank stress testing: range of practices. Technical report. Bank for International Settlements. https://www.bis.org/bcbs/publ/d427.htm. Accessed 20 Apr 2018

Basel Committee on Banking Supervision (2018) Stress testing principles. Technical report. Bank for International Settlements. ISBN 978-92-9259-207-3 (online)

Baudino P, Goetschmann R, Henry J, Taniguchi K, Zhu W (2018) Fsi insights on policy implementation no 12. stress-testing banks—a comparative analysis. Technical report, Financial Stability Institute. ISSN 2522-2481 (print)

Bhardwaj G, Gorton G, Rouwenhorst GK (2015) Facts and fantasies about commodity futures ten years later, (working paper). National Bureau of Economic Research, Cambridge. https://doi.org/10.3386/w21243

Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis: forecasting and control. Wiley, Hoboken

Cheng I-H, Xiong W (2014) Financialization of commodity markets. Annu Rev Financ Econ 6(1):419–441. https://doi.org/10.1146/annurev-financial-110613-034432

Cheung CS, Miu P (2010) Diversification benefits of commodity futures. J Int Financ Mark Inst Money 20(5):451–474. https://doi.org/10.1016/j.intfin.2010.06.003

Daskalaki C, Skiadopoulos G (2011) Should investors include commodities in their portfolios after all? New evidence. J Bank Finance 35(10):2606–2626. https://doi.org/10.1016/j.jbankfin.2011.02.022

Delle Chiaie S, Ferrara L, Giannone D (2017) Common factors of commodity prices, (working paper). European Central Bank, Frankfurt am Main. https://doi.org/10.2866/116302

Embrechts P, Mikosch T, Klüppelberg C (1997) Modelling extremal events: for insurance and finance. Springer, London

Erdős P (2012) Have oil and gas prices got separated? Energies 49:707–718. https://doi.org/10.1016/j.enpol.2012.07.022

European Banking Authority (2017) Draft guidelines on institution's stress testing, (consultation paper). European Banking Authority, London. http://www.eba.europa.eu/regulation-and-policy/supervisory-review-and-evaluation-srep-and-pillar-2/guidelines-on-stress-testing-and-supervisory-stress-testing. Accessed 26 Apr 2018

Ghorbel A, Souilmi S (2014) Risk measurement in commodities markets using conditional extreme value theory. Int J Econom Financ Manag 2(4):188–205. https://doi.org/10.12691/ijefm-2-5-4

Gorton G, Rouwenhorst KG (2006) Facts and fantasies about commodity futures. Financ Anal J 62(2):47–68. https://doi.org/10.2469/faj.v62.n2.4083

Koliai L (2016) Extreme risk modeling: An evt-pair-copulas approach for financial stress tests. J Bank Finance 70:1–22. https://doi.org/10.1016/j.jbankfin.2016.02.004

Kolodziej M, Kaufmann RK, Kulatilaka N, Bicchetti D, Maystre N (2014) Crude oil: commodity or financial asset? Energy Econ 46:216–223

Liu J (2011) Extreme value theory and copula theory: a risk management application with energy futures. PhD thesis, University of Victoria

Lopez JA (2005) Stress tests: useful complements to financial risk models. RBSF Economic Letter, pp 1–4. https://www.frbsf.org/economic--research/files/el2005--14.pdf. Accessed 20 Mar 2018

McNeil AJ, Frey R, Embrechts P (2015) Quantitative risk management: concepts, techniques and tools. Princeton University Press, Princeton

Nick S, Thoenes S (2014) What drives natural gas prices?—A structural VAR approach. Energy Econ 45:517–527. https://doi.org/10.1016/j.eneco.2014.08.010

Nyström K, Skoglund J (2002a) A framework for scenario-based risk management Swedbank, Group Financial Risk Control, pp 1–56. http://gloria--mundi.com/UploadFile/2010--2/knjs.pdf. Accessed 01 Mar 2018

Nyström K, Skoglund J (2002b) Univariate extreme value theory, GARCH and measures of risk. Swedbank, Group Financial Risk Control, pp 1–27. http://gloria--mundi.com/UploadFile/2010--2/knjsug.pdf. Accessed 14 Feb 2018

Paraschiv F, Mudry P-A, Andries AM (2015) Stress-testing for portfolios of commodity futures. Econ Model 50:9–18. https://doi.org/10.1016/j.econmod.2015.06.005

Silvennoinen A, Thorp S (2013) Financialization, crisis and commodity correlation dynamics. J Int Financ Mark Inst Money 24:42–65. https://doi.org/10.1016/j.intfin.2012.11.007

S&P Dow Jones Indices (2016) S&P Dow Jones Indices announces 2017 weights for the Dow Jones Commodity Index. Press release. S&P Global. https://us.spindices.com/search/?query=2017+weights&Search=Go&Search=Go. Accessed 20 Jan 2018

S&P Dow Jones Indices (2017) Dow Jones Commodity Index Methodology, Index methodology. S&P Global. https://us.spindices.com/documents/methodologies/methodology-dj-commodity-index.pdf. Accessed 12 Feb 2018

S&P Dow Jones Indices (2018) S&P GSCI Methodology, Index methodology. S&P Global. https://us.spindices.com/indices/commodities/sp-gsci. Accessed 12 Feb 2018

Tang K, Xiong W (2012) Index investment and the financialization of commodities. Financ Anal J 68(6):54–74. https://doi.org/10.2469/faj.v68.n6.5

Thomson Reuters (2012) Thomson Reuters Indices: continuous commodity total return indices. Methodology report. Thomson Reuters. https://www.thomsonreuters.com. Accessed 25 Jan 2018

Wang Z-R, Chen X-H, Jin Y-B, Zhou Y-J (2010) Estimating risk of foreign exchange portfolio: using VaR and CVaR based on GARCH–EVT-Copula model. Physica A 389(21):4918–4928. https://doi.org/10.1016/j.physa.2010.07.012