A stochastic oil price model for optimal hedging and risk management

Teemu Pennanen* Luciane Sbaraini Bonatto†

February 15, 2022

Abstract

We develop a stochastic model for future monthly spot prices of the most important crude oils and refined products. The model is easy to calibrate to both historical data and users views even in the presence of negative prices which have been observed recently. This makes it particularly useful for risk management and design of optimal hedging strategies in incomplete market situations where perfect hedging may be impossible or prohibitively expensive to implement. We illustrate the model with optimization of hedging strategies for refinery margins in illiquid markets using a portfolio of 12 most liquid derivative contracts with 12 maturities traded on New York Mercantile Exchange (NYMEX) and Intercontinental Exchange (ICE).

Keywords: Oil prices, market risk, stochastic modeling, market incompleteness

1 Introduction

This paper develops a multivariate stochastic model to describe the dynamics of five oil spot prices which are the most important underlyings of the futures contracts traded on New York Mercantile Exchange (NYMEX) and Intercontinental Exchange (ICE): West Texas Intermediate crude (WTI), Brent crude, Reformulated Gasoline Blendstock (RBOB), NY Harbor Ultra Low Sulphur Diesel (HO), Low Sulphur Gasoil (Gasoil). WTI and Brent are the dominant crude oil futures contracts and the main benchmark prices for worldwide oil trading. The most important refined product futures are ultra-low sulphur diesel and gasoline that are settled by physical delivery in New York Harbor, the principal cash market in the USA, followed by gasoil futures, which are settled by delivery in barges in Amsterdam, Antwerp and Rotterdam, and it is used as the pricing reference for all distillate trading in Europe and other regions. These refined products are mainly transportation fuels and represent more than 40% of the global oil products demand.

*King's College London, Department of Mathematics, London WC2R 2LS, United Kingdom, *teemu.pennanen@kcl.ac.uk*

†King's College London, Department of Mathematics, London WC2R 2LS, United Kingdom, *luciane.sbaraini_bonatto@kcl.ac.uk*

Realistic descriptions of the uncertain price dynamics can be very useful for producers, suppliers, consumers as well as traders of oil derivatives markets. Stochastic models are essential in quantitative risk assessment and management as well as in refinery programming and in the construction of hedging strategies involving derivative contracts. Stochastic (statistical) models form the basis also for pricing and hedging of OTC (over-the-counter) derivatives as well as of production and storage facilities which fail the completeness assumptions behind risk neutral valuation principles.

This paper develops a model of the monthly values of the above five oil spot prices. The model captures the main statistical features and it is easy to calibrate to both user's views and historical data. The simple structure of the model allows for efficient numerical simulations which makes it well suited for practical applications requiring large sets of scenarios. Our model allows for the possibility of negative prices arising from extreme imbalances between supply and demand of oil. This became relevant during the drastic reduction of demand in the spring of 2020 due to Covid-19. The reduced demand and the difficulty of storing the excess oil resulted in negative prices in WTI.

The proposed model is not meant for forecasting but for describing the distribution of the five spot prices around a given forecast. The model captures both the user's views as well as the statistical dependencies over time and between the five products. Such dependencies are crucial in portfolio optimization and in pricing and hedging of derivative contracts. We illustrate the model with optimization of hedging strategies for refinery margins by using live derivatives quotes from NYMEX and ICE. We find that in the presence of bid-ask spreads and finite liquidity, it is not always possible or rational to execute a perfect hedge.

Various stochastic models for the spot and forward prices of crude oil can be found in the literature. Ornstein-Uhlenbeck stochastic process one-factor models as in (Schwartz, 1997) reflect the mean reverting price behavior of oil prices. (Schwartz, 1997) combined geometric brownian motion and mean reverting processes, extending the multifactor model of (Gibson and Schwartz, 1990) that was driven by a two-factor stochastic spot oil price model where spot price and convenience yield follow a mean-reverting process. (Cortazar and Schwartz, 2003) proposed a three-factor term structure model to describe the non-perfect correlation of all futures contract returns. Besides mean reverting spot prices, they modelled the convenience yield and the long-term equilibrium prices.

Most of the existing models, however, are based on the assumption of market completeness and the models are aimed for risk-neutral valuation. This is at odds with reality as the illiquidity present in the real quotes makes the market incomplete and nonlinear in terms of trading costs. The incompletenes makes perfect replication impossible and invalidates the risk-neutral valuation paradigm. The models developed in this paper describe price movements under the subjective measure (aka "real world measure", "statistical measure", "P-measure", ...) a traded uses to describe her views concerning the future market developments. The model is intended for risk management applications as well as for pricing and hedging of financial or physical contracts whose payouts depend on the five underlyings. "Calibration" to market quotes happens when derivative instruments are used in partial hedging of the cash-flows. This applies in both complete and incomplete markets. Together with the developed stochastic model, the derivative quotes thus determine the cost of hedging and the corresponding pricing of a contract.

The incorporation of user's views in a model is crucial when it is used for portfolio optimization and/or pricing and hedging in incomplete markets. Such views are largely driven by historical data but there may be other information that is not present in historical data but that is likely to influence future price development. Such additional information may come from e.g. equilibrium analysis of supply and demand that may be affected by politics or pandemics. The model developed in this paper allows for an easy incorporation of such views.

Our model captures many statistical features present in the historical time series such as seasonalities, mean reversion and dependencies among spot prices. Such features have been confirmed in the literature. The presence of seasonality was studied by (Kanamura et al., 2010) who found that the time spread trading profitability of heating oil and natural gas futures contracts are affected by seasonality. Also (Girma and Paulson, 1999) showed the existence of seasonality in crude oil spread and that trading seasonality is profitable. The tendency of oil prices to revert to their historical mean values is justified by the incentive that producers have to increase supply in a scenario where oil prices are very high, which combined with demand reduction can lead to a downward pressure on prices. The effect is the opposite when oil prices are low, where due to increased demand and lower supply, there is an upward pressure on prices. The mean reversion property has been used to construct profitable trading strategies. For example, (Kanamura et al., 2010) proposed a profit model for spread trading by studying the stochastic movements of the price spread of WTI crude, heating oil and natural gas futures and employing first hitting time density. They concluded that a mean-reverting model with long-term mean could be applied to spread trading and may be profitable. Also (Miffre and Rallis, 2007) studied the short and long-term mean reversion in commodity futures. They show that momentum strategies generate annual returns of over 9% when applied to commodity futures.

The approach developed in this paper differs from other papers where market incompleteness is addressed by the use of an arbitrary exogenous discount rate as described by (Guimarães Dias and Rocha, 1999). Moreover, It doesn't require that the spot markets consist of complete markets like in (Kanamura, 2015) who proposes a convenience yield-based pricing for commodity futures, which embeds the incompleteness of commodity futures markets in convenience yields.

The rest of this paper is organized as follows. We first present the oil market in Section 2. Section 3 discusses the stochastic modelling of spot prices and how we incorporate the main characteristics of the oil market and the user's views by modelling the forecast deviations. Section 4 addresses the estimation results, statistical residual analysis and copula estimation, followed by section 5 with the simulation results. Section 6 is a practical application to refinery hedging.

2 The market

The majority of oil derivatives traded on New York Mercantile Exchange (NYMEX) and Intercontinental Exchange (ICE) consists of futures contracts on the following two oil crudes and three oil products:

- 1.WTI: West Texas Intermediate Crude Oil

- 2.Brent: North Sea Brent Crude Oil
- 3.RBOB: Reformulated Gasoline Blendstock
- 4.HO: NY Harbor Ultra Low Sulphur Diesel
- 5.Gasoil: Low Sulphur Gasoil

All contracts have maturities in multiples of a month but have different daily settlement methodology and expiration dates with either financial or physical settlement at expiration. The daily settlement price methodology for the first 4 contracts is the volume-weighted average price of all trades executed between 14:28:00 and 14:30:00 New York time, the settlement period, rounded to the nearest tradable tick. The Gasoil contract has a different settlement period which is the volume-weighted average price of all trades executed between 16:28:00 and 16:30:00 London time. The expiration dates and the form of settlement at expiration of the future contracts are as follows:

- Brent: expires on the last trading day of the month having just financial settlement.
- WTI : expires on the 20th day of the month with physical delivery during the next forward month.
- Rbob: expires on the last trading day of the month with physical delivery from 7th to 30th next forward month.
- HO: expires on the last trading day of the month with physical delivery from 7th to 30th next forward month.
- Gasoil: expires on 12th with physical delivery from 16th to 31st of the same month.

We will study the settlement prices on the last trading day of each month for the five contracts that mature in the following month on Brent (ICE), WTI (NYMEX), HO (NYMEX), RBOB (NYMEX) and Gasoil (ICE). Figure 1 plots the historical prices for the five underlyings.

3 The model

Our aim is to create a model that is easy to calibrate and to simulate and that captures the statistical features as well as the user's views about future prices development. In Figure 1, we can observe that the prices are strongly dependent, present seasonalities and mean reverting behaviour.

The five prices are strongly related and follow each other in the long run. This suggests considering prices relative to each other. In our model, Brent is the reference price and we model the other products relatively to it. In order to quantify seasonalities, we computed the historical monthly medias for Brent and the other products ratios relative to Brent. Figure 2 shows the historical monthly log-medians of Brent and the relative prices from May 2008 to April 2020. We can observe seasonality in all 5 price series. Brent and WTI have they weakest values during the refinery maintenance period. Gasoline has stronger

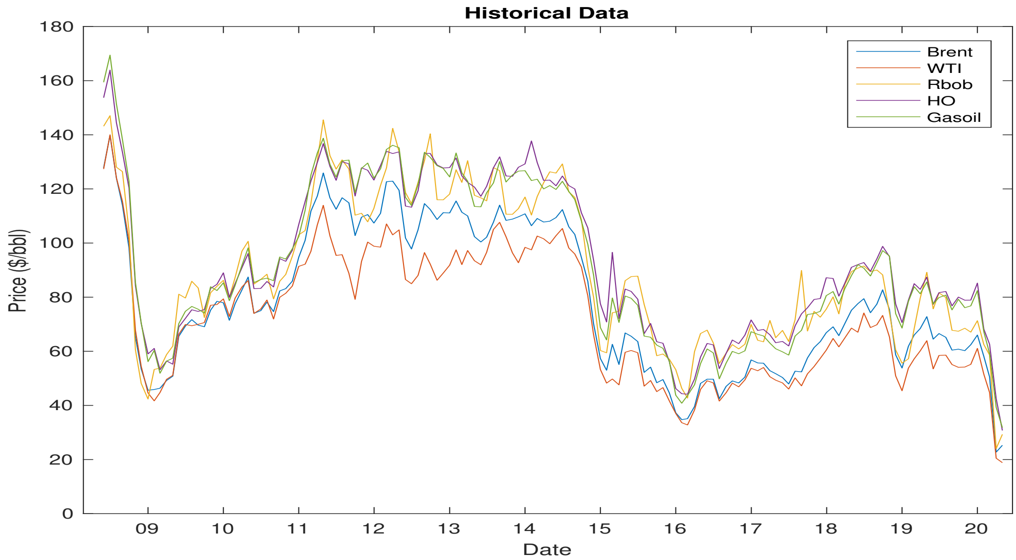


Figure 1: Historical end of the month nearest maturity prices

demand during summer driving season and also changes quality in March for a more expensive product that is consumed in summer, and in September for a cheaper quality consumed during winter. Those factors result in higher prices for the months of March until August and lower prices between September and February. The gasoil and diesel prices present seasonality as well due to additional demand for heating during the winter months.

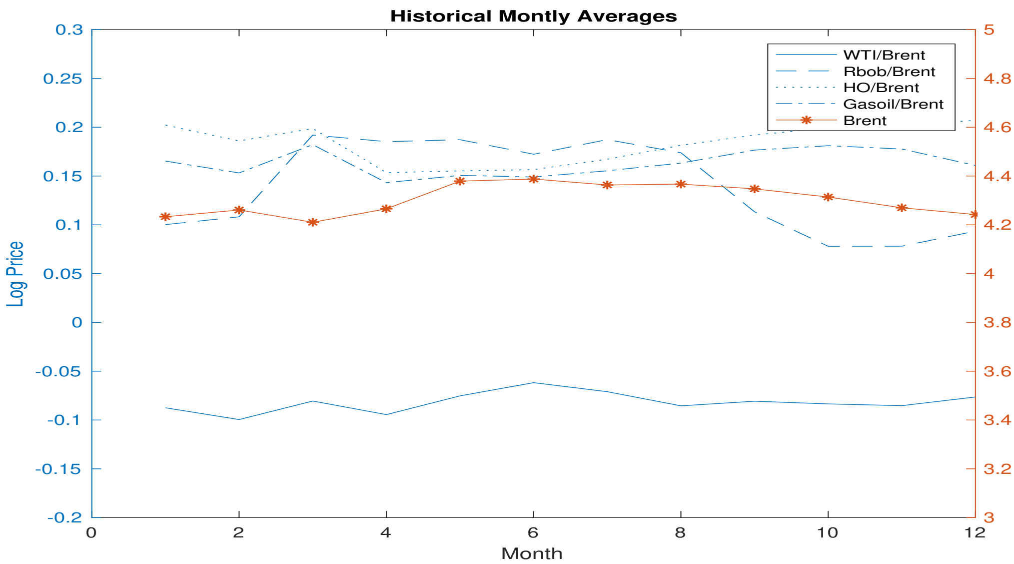


Figure 2: Seasonality of oil products.

The oil market has shown recently negative prices caused by extreme supply and demand imbalance created by COVID-19 crisis. Inventories in Cushing fill up quickly and led to negative prices for WTI in April 2020. At expiration, players who ended up net long WTI futures need to own storage to store the oil delivered. As oil futures trade at multiples of the physical oil production, non physical players had to close their positions before expiration and physical players didn't have the storage capacity and risk appetite to own crude. Clearly, short term extreme imbalances can lead to temporary negative prices but also questions the needed time for market to balance supply and demand. Producers could be able to balance the market by shutting down production, but need persistent prices below their breakeven marginal production cost as shutting down oil production is costly and could cause irreversible damages to the wells. Therefore, there is a finite lower bound where supply and demand balances which, depending on the market conditions, is not always zero.

We denote the lower bound of product i at time t by $B_{t,i}$ and the difference of the spot price from its lower bound by $S_{t,i}$. We model the development of the price vector $S_t = (S_{t,i})_{i=1}^5$ as a multivariate stochastic process $S = (S_t)_{t>0}$. Since S is positive, we start by taking logarithms. We develop a stochastic model for the logarithmic deviations

$$e^1 = \ln\ S_1 - \ln\ \bar{S}_1 \tag{1}$$

$$e^i = \ln \frac{S_i}{S_1} - \ln \frac{\bar{S}_i}{\bar{S}_1} \text{ for } i \neq 1, \tag{2}$$

where $\bar{S}_i$ is the monthly median of product i. Figure 3 illustrates the historical log-deviations assuming zero lower bound.

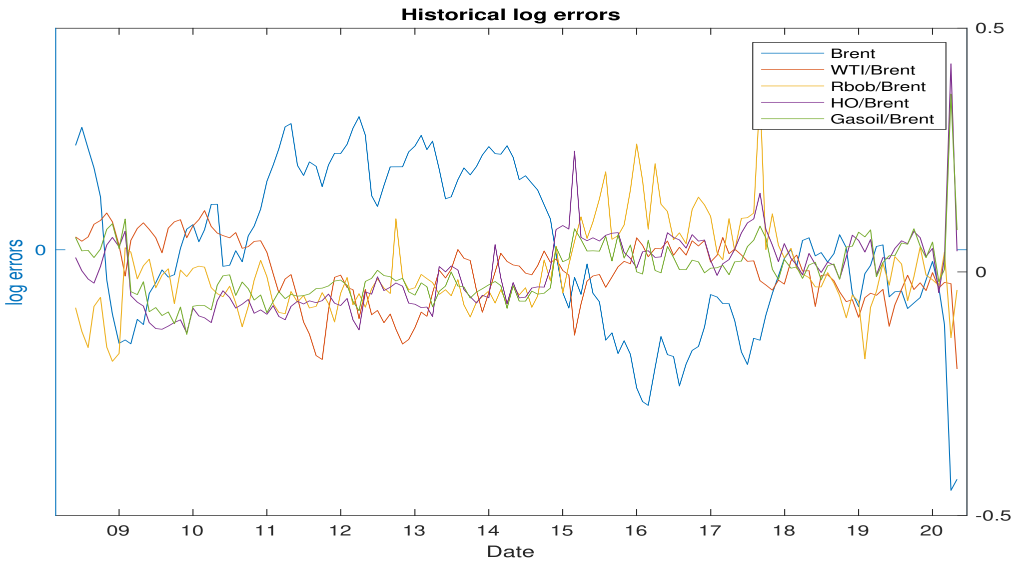


Figure 3: Historical log-deviations.

When modelling the future values of S, we can incorporate the trader's proprietary information in $\bar{S}$. One can interpret $\bar{S}$ as a forecast and then e can be viewed as random deviations from the forecasts. In practice, $\bar{S}$ represents views as well as any forecast for the future development of the medians based on, for example, technical analysis, machine learning or fundamental analysis such as equilibrium models for supply and demand. The user's deterministic predictive model incorporated in $\bar{S}$ should account for seasonalities and the relative value between the five underlings.

We model the monthly values of the five dimensional process $e = (e^i)_{i=1}^5$ with a linear stochastic difference equation of the form

$$\Delta e_t = Ae_{t-1}\Delta t + \varepsilon_t\sqrt{\Delta t}, \tag{3}$$

where $A \in \mathbb{R}^{5\times 5}$ and ε_t are iid $\mathbb{R}^5$-valued random vectors. The matrix A and the distribution of ε will be calibrated to historical data; see Section 4. The model can be viewed as a discrete time version of the stochastic differential equation

$$de_t = Ae_t dt + dW_t, \tag{4}$$

where W is a stationary martingale. Equation 3 implies that

$$e_t = (A\Delta t + I)^t e_0 + \sum_{s=1}^{t}(A\Delta t + I)^{t-s}\varepsilon_s\sqrt{\Delta t}, \tag{5}$$

where $e_0 = 0$. For e to be stationary, the matrix A needs to be such that $||A\Delta t + I|| < 1$. This implies that the stochastic shocks have transitory effects which result in a mean-reverting process.

The time series stationarity was tested by augmented Dickey-Fuller test for unit roots. Table 1 shows the results. With a null hypothesis of a unit root, all price series had p-values smaller than 0.05. This indicates we can reject the hypothesis of unit root and treat the time series as stationary.

Table 1: Time series test statistics for the regression $\Delta e_t = Ae_{t-1}$.

	Brent	WTI/Brent	RBOB/Brent	HO/Brent	Gasoil/Brent
ADF t-statistic	1	1	1	1	1
p-value	<0.01	< 0.00	< 0.00	< 0.00	< 0.00

4 Parameter estimation

This section presents the estimation results of the model. We use the nearest maturity end of the month historical data from May 2008 to April 2020; see Section 3.

We chose the lower bound considering the lowest historical prices, put strikes traded in the market and the statistical improvements to the model such as normally distributed residuals. The lowest settlement prices for WTI was on the 20th April 2020 and the lowest put strike traded was -30 USD/bbl. Even though Brent hasn't traded negative prices, put zero strikes traded suggesting negative prices could be achieved. A lower bound of 80 USD/bbl makes WTI/Brent residuals normally distributed. For simplicity, we assume the same lower bound on all products. Figure 4 shows the transformed historical log-deviations.

$$B = \begin{bmatrix} Brent & WTI & RBOB & HO & Gasoil \\ 80 & 80 & 80 & 80 & 80 \end{bmatrix}$$

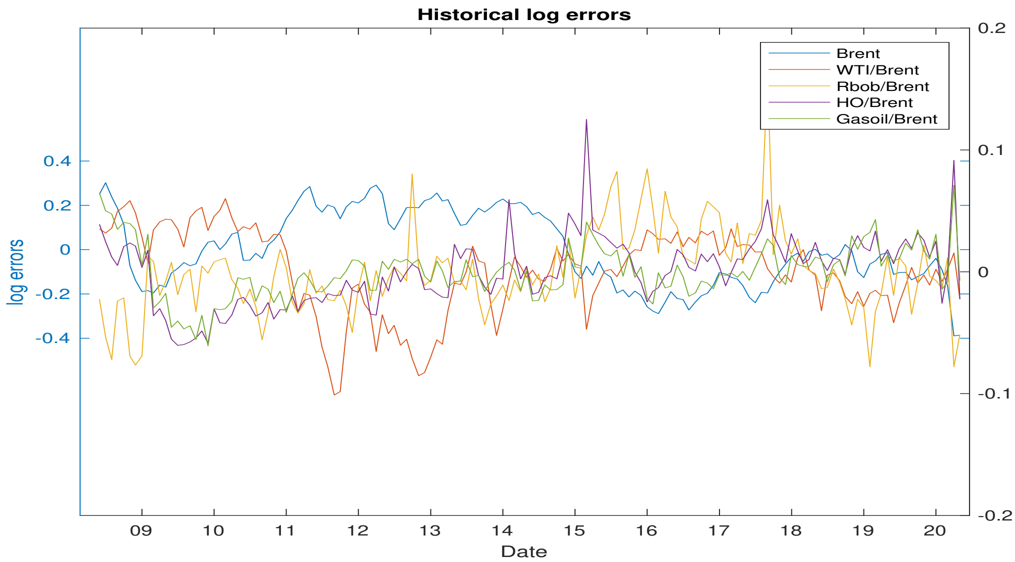


Figure 4: Historical log-deviations.

4.1 Regression results

The matrix A is estimated with OLS (Ordinary Least Square) using forward and backward stepwise regression (Draper and Smith, 1981). The stepwise method consist on either adding or deleting variables one at a time according to a specific criterion. The criterion chosen was the p-value for an F-test of the change in the sum of squared error by adding or removing the term.

In the forward procedure, we start with zero matrix A. We test row by row the addition of each term by using a chosen model fit criterion, adding the variable whose inclusion gives the most statistically significant improvement of the fit, and repeating this process until none improves the model to a statistically significant extent using F-test. The statistically significant improvement is measure by the F-ratio. In that sense, the variable considered for inclusion is the one giving the largest single degree of freedom F-ratio. The other variables are added to the equation if the group of variables are jointly significant. In this paper, the chosen criteria is when the corresponding p-values of the F test are less than the 0.01%.

Likewise, in the backward procedure, the terms are removed based on the p-value for the F-test of the change in the sum of squared error by removing the term. The estimated autoregression matrix A

$$
A = \begin{bmatrix}
Brent & WTI/Brent & RBOB/Brent & HO/Brent & Gasoil/Brent \\
-0.3792 & & & & \\
(0.1979) & & & & \\
 & -1.2156 & & & \\
 & (0.0060) & & & \\
-0.4970 & & -5.6886 & & \\
(0.0062) & & (0.0000) & & \\
 & & & -3.8477 & \\
 & & & (0.0000) & \\
 & & & & -4.1660 \\
 & & & & (0.0000)
\end{bmatrix},
$$

where the p-values of the respective t-test statistics are in parentheses and blanks correspond to zeros. The diagonal of A is negative for all products, with p-values of the respective t-test statistics below 0.05 except for Brent. The negative diagonal of A confirms the mean reversion behavior. For Brent prices there is evidence in the literature (Geman, 2009) (Schwartz, 1997) and (Pilipovic, 2007).

The main dependences are captured by data transformations in the equation (2), where Brent is the reference price and we model the other products relatively to it. As a consequence, just Rbob/Brent presented statistically improvement of the fit by including other predictors. Rbob/Brent shows dependence to Brent log deviations. It is natural to accept that higher Brent prices could lead to lower Rbob/Brent, or the contrary when there are lower Brent prices. The intercepts were removed from the regression equations since they were rejected with p-values higher than 0.50. Non-zero values would imply that $\lim_{t\to\infty} Ee_t \neq 0$ which would be inconsistent with the interpretation of e_t as a forecast deviation.

4.2 Residual analysis

In order to fully complete a stochastic model for the spot prices, we need to specify the distribution of the five-dimensional innovation terms. To this end, we analyze the historical values of the residuals obtained when estimating the parameter matrix A. The residual test results are reported in Table 2. The augmented Dickey-Fuller test confirms stationarity and absence of unit roots. We found no serial correlation except in Brent. Heteroscedasticity in WTI/Brent, RBOB/Brent, HO/Brent and Gasoil/Brent may be caused by the presence of outliers or skewness in the distribution which we can observe in kernel density plot of the residuals in Figure 5. Abosedra and Laopodis (1996) develop a univariate GARCH-model for crude oil prices to treat heteroscedasticity. The results below suggest a multivariate GARCH-models for our five-dimensional model but such models are often nontrivial to estimate and they are more cumbersome in simulation studies.

The volatility vector and correlation matrix are:

$$
\sigma = \begin{bmatrix}
Brent & WTI/Brent & RBOB/Brent & HO/Brent & Gasoil/Brent \\
0.1649 & 0.0518 & 0.0960 & 0.0714 & 0.0546
\end{bmatrix}
$$

Table 2: Residual test statistics for the regression $\Delta e_t = Ae_{t-1}$.

	Brent	WTI/Brent	RBOB/Brent	HO/Brent	Gasoil/Brent
Serial correlation (DW)	0.0002	0.9781	0.1448	0.1473	0.1130
Normality (JB)	0.0010	0.0526	0.0010	0.0010	0.0010
Heteroscedasticity (Arch)	0.6254	0.0351	0.0033	0.0001	0.0012
ADF t-statistic	1	1	1	1	1
p-value	<0.00	< 0.00	< 0.00	< 0.00	< 0.00

Notes. For serial correlation, normality and heteroskedasticity tests, the numbers are p-values of the statistics. Serial correlation is tested with the Durbin-Watson test of the null hypothesis that the residuals from a linear regression are uncorrelated. A small p-value suggests serial correlation. Normality is tested with the Jarque-Bera test, with null hypothesis of normality. A large p-value supports the normality assumption. The Arch test tests against heteroscedasticity, with the null hypothesis of homoscedasticity. A large p-value indicates absence of heteroscedasticity. Stationarity of the residuals is tested with the augmented Dickey-Fuller(ADF) tests where the number of lags selected was 4.

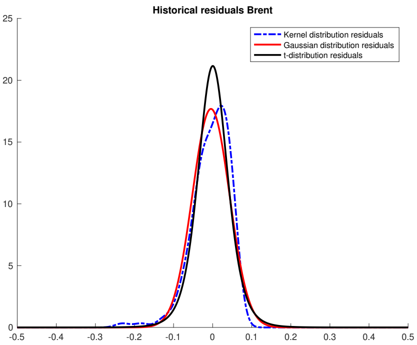
$$\rho = \begin{bmatrix} Brent & WTI/Brent & RBOB/Brent & HO/Brent & Gasoil/Brent \\ 1.0000 & -0.1199 & 0.1429 & -0.1995 & -0.2301 \\ & (0.1537) & (0.0887) & (0.0169) & (0.0057) \\ & 1.0000 & -0.0132 & 0.0184 & -0.0103 \\ & & (0.8754) & (0.8273) & (0.9031) \\ & & 1.0000 & 0.0652 & -0.0557 \\ & & & (0.4390) & (0.5088) \\ & & & 1.0000 & 0.7447 \\ & & & & (0.0000) \\ & & & & 1.0000 \end{bmatrix}$$

In risk management, one is often concerned with the tail of the distribution of losses, and often large losses in a portfolio are caused by simultaneous large moves in several components. The residual distributions impact the distribution of prices and the optimal portfolio allocation.

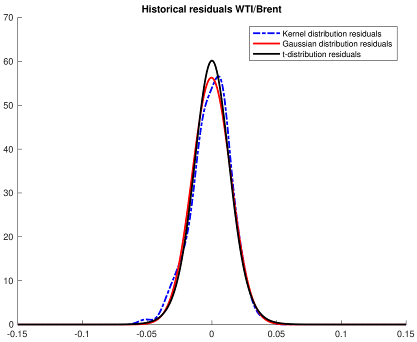
In the data analyzed, the residuals violated the Jarque-Bera test for normality for all series except WTI/Brent (see table 2). The heavy tails can also be observed in the Kernel density plot of the residuals in Figure 5. It suggests using another distribution to model the innovations.

4.3 Non-Gaussian innovations

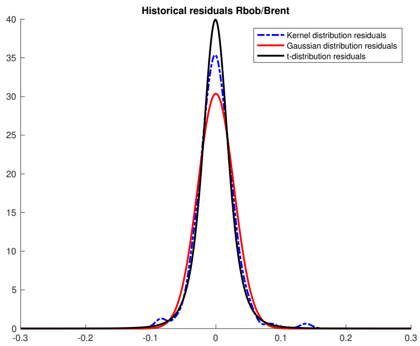
The main motivation to use the t-distribution is because the residuals are non-Gaussian. The t-distribution typically captures better the tail dependence and has gained popularity in quantitative risk management applications (Hofert, 2013). Specifically in the oil markets, (Mohammadi and Su, 2010) identified the presence of heavy tails in the residual distribution of crude oil spot prices, and (Reboredo, 2011)have found evidence of non-normality in the residuals marginals of different crude oil spreads using MA(1)-t-EGARCH(1,1). He concluded that in an event of market stress, crude oil prices tend to move together. (Fong and See, 2002) studied the temporal behaviour of volatility of daily returns on crude oil futures using regime switching models and comparing with a



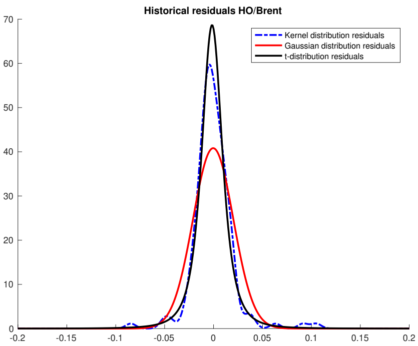
(a) Brent



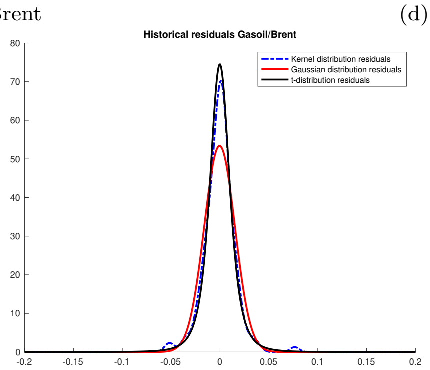
(b) WTI/Brent



(c) Rbob/Brent



(d) HO/Brent



(e) Gasoil/Brent

Figure 5: Kernel distribution of historical residuals together with Gaussian and t-distribution

Table 3: Monthly parameters of t-distribution for the residuals of the regression $\Delta e_t = Ae_{t-1}$.

	Brent	WTI/Brent	RBOB/Brent	HO/Brent	Gasoil/Brent
Location μ	0.003	0.0000	-0.0006	-0.0015	-0.0002
Scale σ	0.038	0.0137	0.0198	0.0111	0.0104
Degrees of freedom ν	6.1050	12.9991	4.0848	2.6283	3.3251

single regime GARCH, using for both models, student-t distribution to model the fatter tails in the return errors.

In our model, t-distribution was fitted to the historical data using maximum likelihood method and has the parameters listed in table 3. The use of t-distribution for modelling the innovations rises the need of a methodology to describe the dependancies between the random variables because the univariate t-distributed random vectors have different degrees of freedom and therefore the joint distribution would not be multivariate t (DeGroot, 2005; Monahan, 2011).

Such dependencies can be conveniently described by copulas which have been widely used in mathematical modeling of financial markets. Copulas provide an easy way to model and estimate the distribution of random vectors by treating the marginal distributions and the dependence structure separately. Copula is a multivariate probability distribution that links the marginal distributions together to form the joint distribution, being responsible for defining the dependency between the variables. The copula theory is based on Sklar's Theorem (Sklar, 1959) that proved any multivariate joint distribution can be described in terms of univariate marginal distribution functions and a copula.

We study two classes of copulas: the Gaussian and the t-Copula. The Gaussian copula is a multivariate probability distribution defined in the unit cube $[0, 1]^n$ (Embrechts et al., 2001) denoted by

$$C(u_1, ..., u_n) = N(N_1^{-1}(u_1), ..., N_n^{-1}(u_n)), \tag{6}$$

where N is the normal multivariate cumulative distribution function, $N_n^{-1}(u_n)$ is the uniform marginal distribution of each random variable n. In our model, the marginal distributions are obtained by fitting the t-distribution to each five residual historical data using maximum likelihood estimation (MLE). By applying the inverse cumulative distribution function, we transform the margins into uniform distributions which then can be fitted to a Gaussian copula. The Gaussian copula was fitted using MLE. The correlation matrix R obtained from data is:

$$R = \begin{bmatrix} Brent & WTI/Brent & RBOB/Brent & HO/Brent & Gasoil/Brent \\ 1.0000 & 0.1233 & 0.0909 & -0.1704 & -0.1450 \\ & 1.0000 & 0.0302 & 0.0715 & -0.0171 \\ & & 1.0000 & 0.1297 & -0.0162 \\ & & & 1.0000 & 0.6943 \\ & & & & 1.0000 \end{bmatrix}$$

The Gaussian copula has the property that its tail dependence is null, regardless of the correlation matrix. It means that there isn't correlation between the variables in the tails of the distributions. Tail dependence describes the limiting proportion that one marginal distribution exceeds a certain threshold given that the other margin has already exceeded it. On the other hand, the t-copula has positive tail dependence even when the random variables are uncorrelated. Therefore, the use of t-Copula has the advantage of increasing the probability it gives to simultaneous extremes in several dimensions.

A t-copula has uniform marginal distributions and correlation matrix, just as a Gaussian copula does. However, the difference relies in their dependence structure, even when applying the same correlation matrix. This is due to the fact that multivariate distributions are not uniquely defined by their marginal distributions, or by their correlations. Likewise the multivariate t-distribution, as the degrees of freedom is made larger, a t-copula approaches the corresponding Gaussian copula.

Following (Embrechts et al., 2001), if X has the stochastic representation

$$X =_d \mu + \frac{\sqrt{\nu}}{\sqrt{S}} Z, \tag{7}$$

where $\mu \in \mathbb{R}^5$, $S \sim \chi^2_\nu$ and $Z \sim \mathcal{N}_5(0, \sum)$ are independent. The copula of X can be written as

$$C^t_{\nu,R} = t^5_{\nu,R}(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_5)), \tag{8}$$

where $t^5_{\nu,R}$ is a multivariate t-distribution with probability density function given by

$$t^5_\nu(\mu, R) = \frac{\Gamma[\nu + 5)/2]}{\Gamma(\nu/2)\nu^{5/2}\pi^{5/2}|R|^{1/2}}[1 + \frac{1}{\nu}(x - \mu)^T R^{-1}(x - \mu)]^{-(\nu+5)/2}. \tag{9}$$

By using maximum likelihood method to fit a t-Copula, we obtain the the matrix R

$$R = \begin{bmatrix} Brent & WTI/Brent & RBOB/Brent & HO/Brent & Gasoil/Brent \\ 1.0000 & -0.1034 & 0.0388 & -0.1778 & -0.1494 \\ & 1.0000 & 0.0990 & 0.0967 & -0.0127 \\ & & 1.0000 & 0.1936 & 0.0127 \\ & & & 1.0000 & 0.7010 \\ & & & & 1.0000 \end{bmatrix}$$

The estimated t-copula degrees of freedom is $\nu = 7.6470$, which is considered low and indicates extreme price variations and tail dependence.

5 Simulation experiment

In order to study the behaviour of the calibrated spot price model, we generated 20,000 simulations 10 years into the future using antithetic sampling. For the median prices $\bar{S}$, we use the mid-price of the forward curve in 22/05/20 at

16:30pm for illustration. We observe the first 12 maturities market mid-prices and assume the following years will be kept constant.

The simulation results obtained with Gaussian innovations are illustrated in the Figure 7. The solid line represents the historical price series and 90% confidence interval with the median in the centre. The dashed line is a single simulated path. In order to illustrate the dependencies between the five prices more clearly, Figure 6 plots the single simulation of all 5 prices together to illustrate the dependencies. The simulated prices for the t-Copula Model are illustrated in the Figure 8.

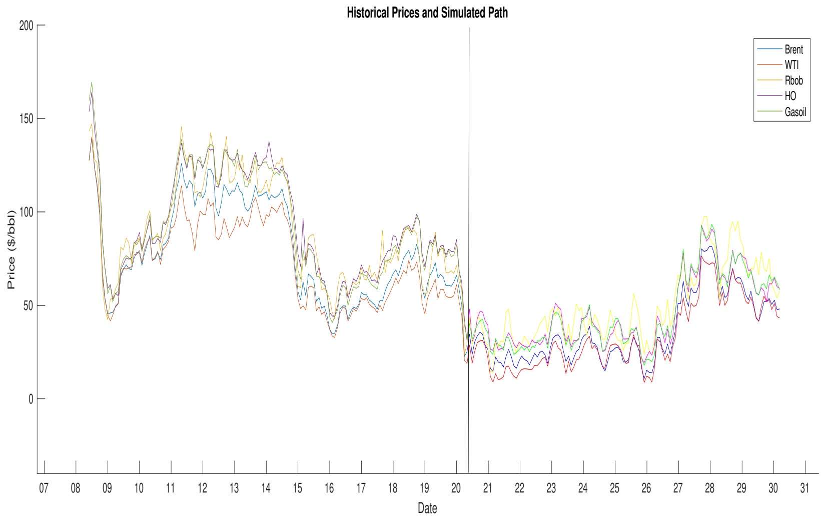
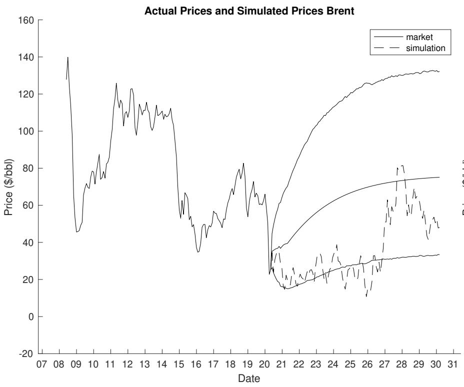


Figure 6: Historical end of the month nearest maturity prices and a simulated path

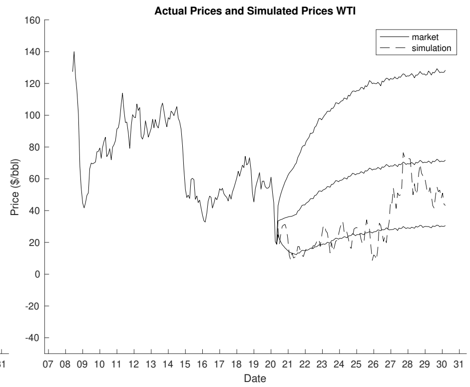
The simulations were implemented in Matlab on a Macintosh with 8 GB 3733 MHz LPDDR4X memory and 1.1 GHz Dual-Core Intel Core i3 processor. Computation 20,000 monthly scenarios 10 years into the future takes on average 0.4 seconds. This is fast enough for most risk management applications in practice. The model would be easy to implement also in C or Cuda wich should provide significant speedups over the current implementation.

6 An application to refinery hedging

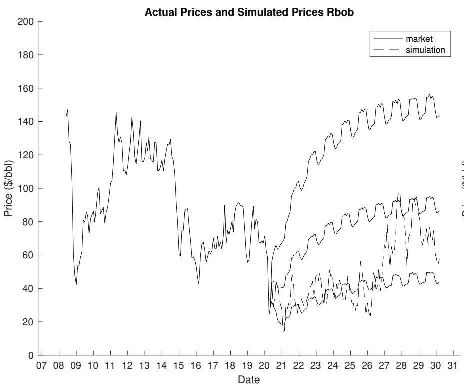
This section illustrates how the stochastic models developed in the previous sections can be used in risk management and the optimization of hedging strategies using futures contracts on the five underlying products. Building on the stochastic model, we build an optimization model that looks for optimal buy-and-hold portfolios on exchange-traded futures contracts so as to achieve the best possible terminal wealth distribution according to a user specified risk measure. The optimization model has several practical applications. As an illustration, we will consider a problem of hedging the cashflows generated by a (rudimentary simplification) of an oil refinery.



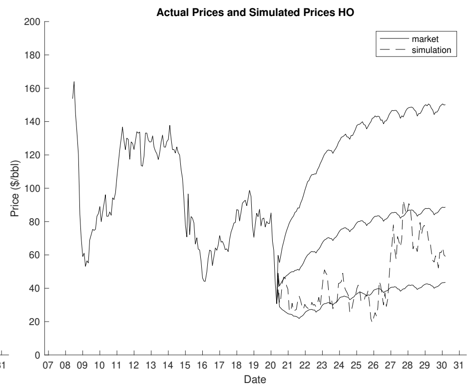
(a) Brent



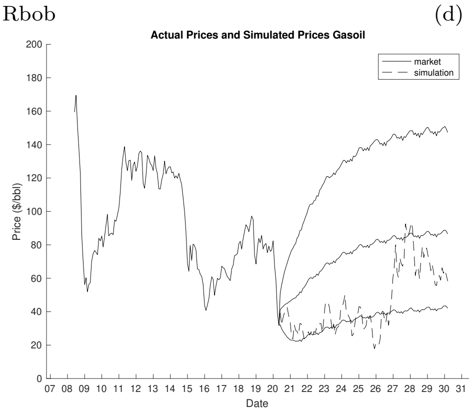
(b) WTI



(c) Rbob

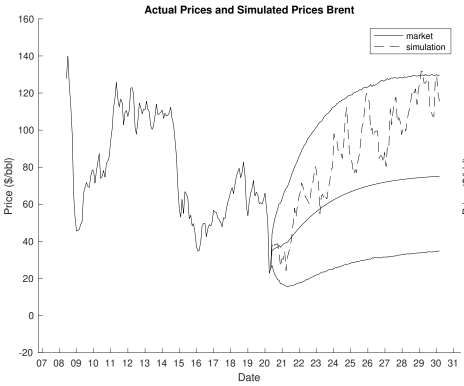


(d) HO

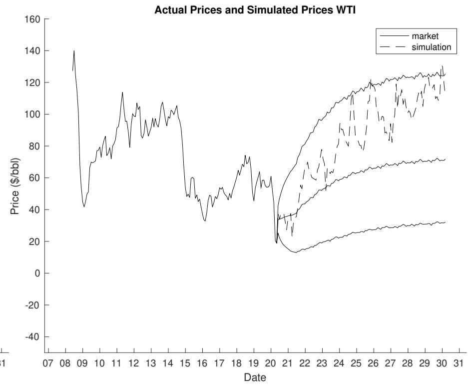


(e) Gasoil

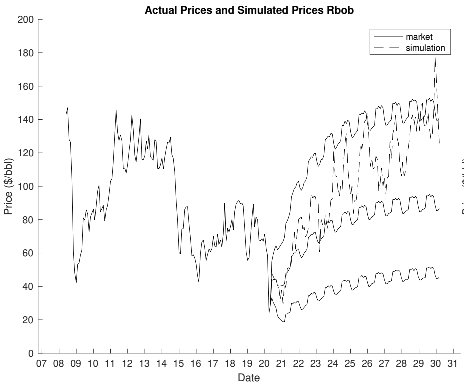
Figure 7: Actual and simulated prices obtained with Gaussian innovations



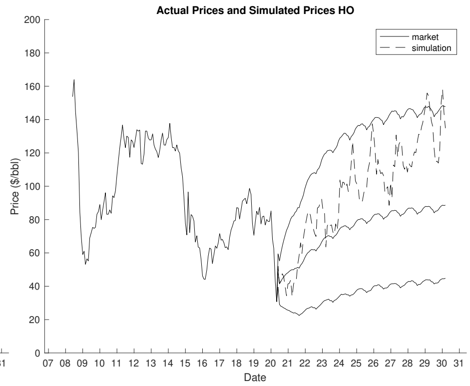
(a) Brent



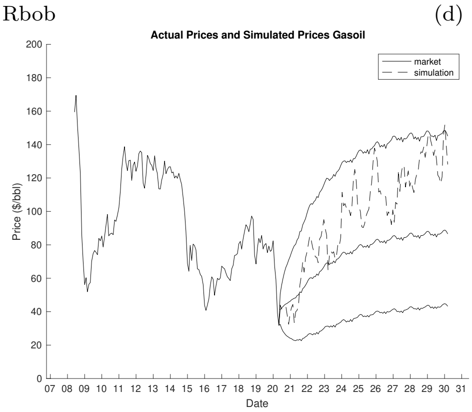
(b) WTI



(c) Rbob



(d) HO



(e) Gasoil

Figure 8: Actual and simulated prices obtained with t-copula model for innovations

In incomplete markets, optimal hedging strategies depend on the distributional properties of the future prices. This was the main motivation for building a model that allows user's views to be incorporated directly as the future medians. The dependencies across the five underlying prices are crucial when trading derivative contracts on all of them. The importance of the dependencies is pronounced in the case of crack-spread contracts whose payoffs are functions of price differences. Moreover, having a reasonable description of the dynamic properties of the price processes is essential when taking positions in multiple maturities. In the examples below, we use futures contracts for 12 different maturities. With 12 different futures contracts for each maturity, we end up optimizing over 144 different contracts.

Refineries are exposed to the price risk of both the inputs and outputs, i.e., the cost of the crude intake and the revenue from selling the products. This market risk is usually managed by hedging programs using futures, swaps and options. The refinery margin hedging program is not always simple to execute as, depending on the refinery capacity, it may require significant volumes to be traded in the market. Some refineries use financial institutions to manage this risk and therefore pay a premium for it.

We approach this problem with stochastic optimization. This involves simulating the refinery margin and finding the optimal hedge subject to the available liquidity. The model is useful for both the risk management department that needs to execute the hedge of the production plan as well as for financial institutions or market makers that buy the risk from the refinery by swapping fix for floating cash-flows and need to execute the hedge in the market.

For illustration, we consider a topping refinery, which has a particularly simple configuration. It has a basic crude oil distillation and support operations. It can separate the crude oils into refined products, but cannot modify the natural yield patterns. That means that this type of refinery doesn't have the units to convert less valuable oil products into transportation fuels.

We assume that the refinery is located in the United States with capacity to process 100 thousand barrels a day. Its output is 50% gasoline and 50% diesel. The refinery processes WTI crude and supply gasoline and 80% of the diesel production to the domestic market. The refinery exports diesel to Europe and therefore sell it at the gasoil reference price. The volume exported is set to one cargo per month which would be equivalent to 300 thousand barrels.

For simplicity, we assume the refinery will produce fixed volumes per month, although this assumption would be easy to change to a flexible monthly production plan. It follows that the monthly refinery margin is given by

$$C_t = 1500 S_{t,rbob} + 1200 S_{t,ho} + 300 S_{t,gasoil} - 3000 S_{t,wti}. \tag{10}$$

Figure 9 illustrates the historical refinery margin together with the median and 90% confidence band computed with our model. The shape of the simulated refinery margins shows the seasonality expected during March to September due to gasoline's change in quality and driving season demand as illustrated in the Figure 2.

If we assume perfectly liquid cash, the monthly margins can be rolled until the end of the year resulting in the terminal net margin of

$$C_T = \sum_{t=1}^{T=12} C_t * (1+r)^{(T-t)} \tag{11}$$

where r is the monthly interest rate. Figure 10 shows the histogram of the simulated end of the year unhedged refinery margin using our spot price model with Gaussian innovations.

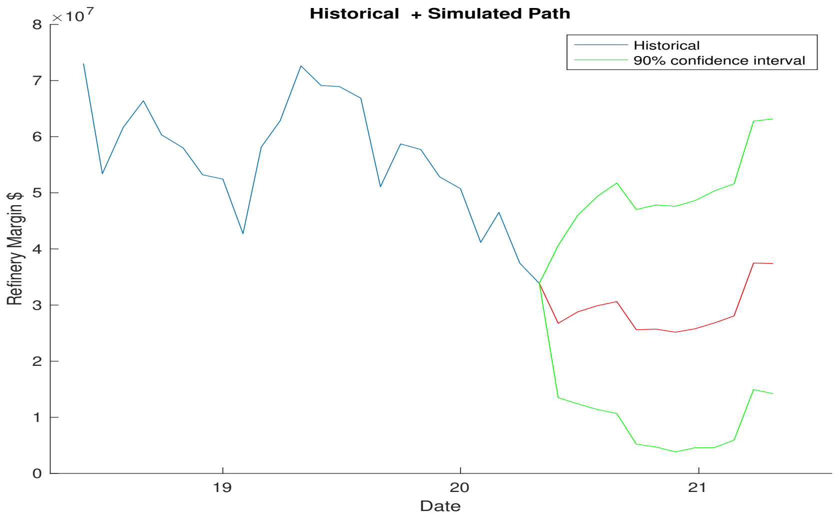


Figure 9: Historical and simulated monthly values of the refinery margin. For the simulation, the plot gives the medians and 90% confidence intervals for the monthly values.

6.1 Hedging instruments

Derivatives instruments can be used to hedge the uncertainty in the refinery margin. We focus on the most liquid derivatives contracts whose payouts only depend on the spot prices. For every maturity, there are 12 derivatives contracts to trade: 5 outright futures on the underlying products and 7 futures on the spreads: HO-WTI crack, HO-Brent crack, RBOB-WTI crack, RBOB-Brent crack, Gasoil-Brent crack, WTI-Brent spread, RBOB-HO spread.

The crack spreads essentially represent the production margin for refineries. The WTI-Brent spread is an important differential to compute the crude export economics for USA and Europe. RBOB-HO spread indicates which product a refinery should maximize production. In practice many refiners use futures contracts or other derivatives to hedge their price risk.

The payouts of long and short positions, respectively, in a futures contract with maturity t are given by

$$c_t^{l,k} = X_t^k - F_t^a \tag{12}$$

$$c_t^{s,k} = F_t^b - X_t^k, \tag{13}$$

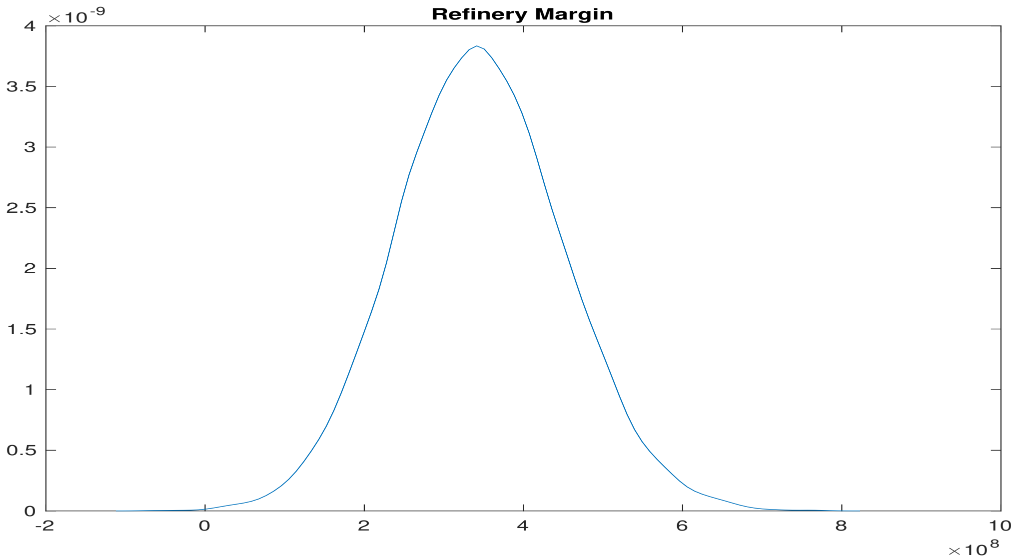
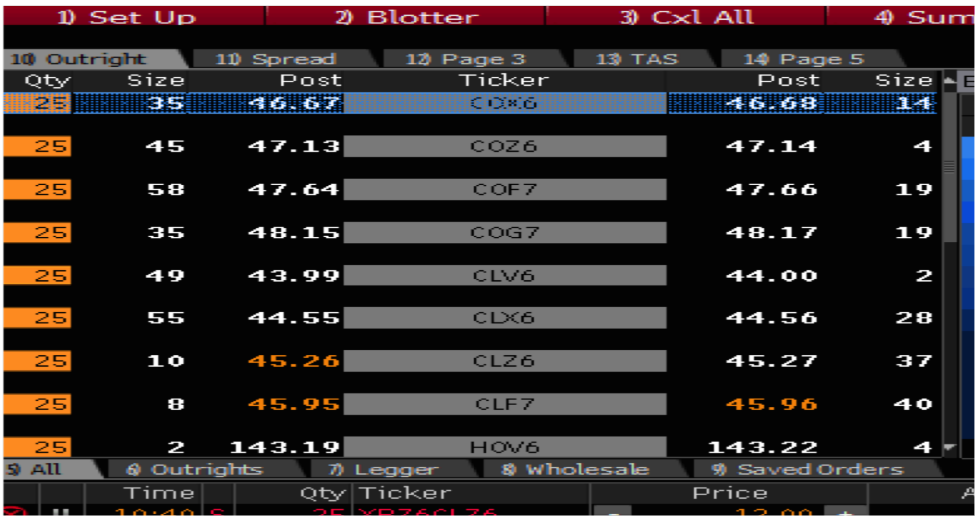


Figure 10: Kernel density topping refinery margin

where for outright contracts $X_t^k = S_{t,i}$ and spread contracts $X_t^k = S_{t,i} - S_{t,j}$ Here F_t^b, F_t^a are the bid and ask market quotes. Both bid and ask prices come with finite quantities q_a, q_b. Figure 11 is a screenshot from Bloomberg of the Brent forward prices from November to January contracts showing the market bid and ask prices in USD/barrel and the volume available to trade at this specific price. The volume is listed as a size in lots of 1,000 barrels.



Qty	Size	Post	Ticker	Post	Size
25	35	46.67	COX6	46.68	14
25	45	47.13	COZ6	47.14	4
25	58	47.64	COF7	47.66	19
25	35	48.15	COG7	48.17	19
25	49	43.99	CLV6	44.00	2
25	55	44.55	CLX6	44.56	28
25	10	45.26	CLZ6	45.27	37
25	8	45.95	CLF7	45.96	40
25	2	143.19	HOV6	143.22	4

Figure 11: Market bid and ask futures prices and volume - Bloomberg

6.2 Optimal hedging

We will use futures contracts to hedge the refinery margin. Each maturity t has a set K of contracts, where for each contract k, we denote the payouts of long on short positions $c_t^{k,l}$ and $c_t^{k,s}$, respectively. The hedged margin process w is given by

$$w_t = R_t w_{t-1} + \sum_{k \in K} (x_t^{k,l} c_t^{k,l}) + \sum_{k \in K} (x_t^{k,s} c_t^{k,s}) + C_t, \tag{14}$$

where $x_t^{k,l}$ and $x_t^{k,s}$ are the sizes of the long and short position in a derivative contract $k \in K$ with maturity t. Our aim is to find derivative positions such that the the distribution of the terminal net wealth w_T is optimal in terms of a user specified risk measure.

If there were no quantity constraints, the refinery margin would be easy to hedge perfectly. The refinery margin in equation (11) could be perfectly hedged, for example, by selling 1500 RBOB contracts, 1200 HO contracts and 300 Gasoil contracts and buying 3000 WTI contracts for each maturity. As the spread contracts are a linear combination of the outright contracts, the perfect hedge is not unique. Indeed, the refinery margin could be perfectly hedged also by using spread contracts, selling 1500 RBOB/WTI crack spreads, selling 1200 HO/WTI crack spreads, selling 300 Gasoil/Brent crack spreads and buying 300 WTI/Brent spreads. In practice, however, it is not always possible to execute the perfect hedge, since there isn't always enough volume available to trade in the forward maturities. Even if it would be possible to execute the perfect hedge, it is not necessarily "optimal". The optimality of a hedge depends on market quotes as well as subjective factors such as views on the future prices and risk preferences. The views will be described by the model developed in Section 3.

A classical way to describe risk preferences are mean-variance and expected utility. Because the spot prices are unsymmetrically distributed, the probability distribution of the payouts in equation (12) is not symmetric. In this situation, variance is not a good measure of risk as it penalizes the right tail of the distribution as much as the left.

We describe the risk preferences with the expected utility function. As usual, we assume the utility function to be increasing and concave. We can write the optimal hedging problem as

$$\begin{aligned} \text{maximize} \quad & Eu(w_T) \quad \text{over} \quad x^l, x^s \in R^{144} \\ \text{subject to} \quad & 0 \leq x_t^{l,k} \leq q_t^{a,k} \\ & 0 \leq x_t^{s,k} \leq q_t^{b,k}, \end{aligned}$$

where E denotes the expectation and the terminal wealth w_T is given by equation (14). This is a 288 dimensional convex optimization problem with 144 variables describing long and short positions. The optimal portfolio depends on the agent views described by the probabilistic description of S, risk preferences described by u, financial position described by the initial wealth w_0 and the stochastic margin C. The fact that our optimization problem is convex it is an important feature, as convexity makes numerical optimization easier; see e.g. Nesterov (2018); Ben-Tal and Nemirovski (2001).

6.3 Optimization results

Our goal is to find the optimal hedge for the production plan described in Section 6 above for the 12 forward months given the bid and ask market quotes F_t^b, F_t^a with finite quantities q_a, q_b. We use the market quotes on 22/05/20 at 16:30pm from Bloomberg and optimize the portfolio with respect to the exponential utility function

$$u(w_T) = -\frac{1}{\delta}e^{(\delta w_T)},$$

where δ is the risk aversion. We use a numerical quadrature to compute an approximation of expected utility

$$Eu(w_T) \approx \sum_{n=1}^{N} \frac{1}{N} u(w_T^n),$$

where N is the number of scenarios drawn from the stochastic model from Section 3. In the numerical experiments below, we use Monte Carlo approximation with antithetic sampling. This results in a finite dimensional convex optimization problem which we then solve with the interior-point solver of Mosek (ApS, 2019). For 20,000 scenarios it takes on average 8 seconds to find a solution on a Macintosh with 8 GB 3733 MHz LPDDR4X memory and 1.1 GHz Dual-Core Intel Core i3 processor.

Figure 12 plots the distribution of the unhedged margin together with with the distributions of two optimally hedged margins. The first optimization was done with risk aversion $\delta = 1$ and the second one with $\delta = 90$. The optimization reduces the risk significantly and, as expected, higher risk aversion leads to smaller risk. Figure 13 compares the optimal hedge portfolios for the two risk aversions. The higher risk aversion trades mostly contracts that offset the refinery margin risk. At a lower risk aversion, the trader keeps a significantly larger long position in gasoil and gasoline crack spread contracts.

A major advantage of our stochastic model is the easy incorporation of user's views. For example, at the beginning of the Covid-19 pandemic, oil companies revised down their long-term outlook for oil prices in view of the expected lower global economic growth. Such a views can be easily incorporated in the model. To illustrate the effects on optimal portfolios, we change the median for HO 12th maturity from 49.50 $/bbl to 44.50 $/bbl and reoptimize the hedge. Figure 14 illustrates the optimal portfolio for this scenario using risk aversion $\delta = 1$. Lower expected prices for HO means that the trader would sell HO and lock in the crack margin at the current forward curve prices. Given the change of views, the long position in HO is mostly transferred to Gasoil because of its strongest correlation with HO.

Another important feature of our model is the statistical dependencies between the spot prices. Such connections can be exploited e.g. in construction of optimal hedging strategies under finite liquidity. To illustrate, we increase the refinery production capacity to 1000 kbbl/d and reoptimize the hedge. In this case, the finite liquidity available at the best quotes impose constraints on the optimal hedge. When the available liquidity in a given contract is exhausted, the optimal hedge starts taking positions in positively correlated contracts. Figure 15 shows the optimal portfolio for this situation using $\delta = 1$. As expected,

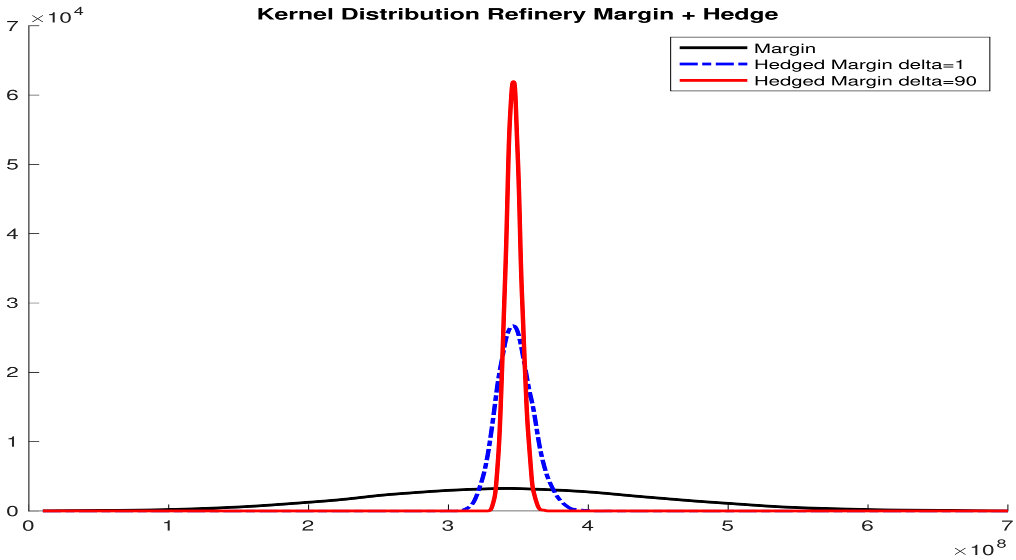


Figure 12: Distributions of the unhedged refinery margin together withe those obtained with optimal hedges corresponding to risk aversions of 1 and 90

the overall volume traded is increased and for contracts bound by the finite liquidity, the model chooses nearby maturities of the same underlying contracts to hedge the exposure.

The above simple experiments illustrate how statistical properties of a spot price model affect optimized hedging strategies. This was the main motivation for building a stochastic model that captures the main dependencies and allows for the incorporation of user's views into the model. When used for purposes of portfolio optimization, it is also important that the model is easy and fast to sample from. Such features become even more important in pricing and hedging of derivative contracts under finite liquidity and market incompleteness. This will be studied in subsequent papers.

References

Salah S. Abosedra and Nikiforos T. Laopodis. Stochastic behavior of crude oil prices: A garch investigation. *The Journal of Energy and Development*, 21(2):283–291, 1996. ISSN 03614476. URL `http://www.jstor.org/stable/24808808`.

MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019. URL `http://docs.mosek.com/9.0/toolbox/index.html`.

Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications.* SIAM, 2001.

Gonzalo Cortazar and Eduardo S Schwartz. Implementing a stochastic model for oil futures prices. *Energy Economics*, 25(3):215–238, 2003.

Morris H. DeGroot. *Some Special Multivariate Distributions*, pages 48–66. John Wiley & Sons, 2005. ISBN 9780471729006. doi: 10.1002/0471729000.ch5. URL `http://dx.doi.org/10.1002/0471729000.ch5`.

N. Draper and H. Smith. *Applied Regression Analysis*, volume 2. John Wiley & Sons, 1981.

Paul Embrechts, Filip Lindskog, and Alexander McNeil. Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*, 2001.

Wai Mun Fong and Kim Hock See. A markov switching model of the conditional volatility of crude oil futures prices. *Energy Economics*, 24(1):71–95, 2002.

Helyette Geman. *Commodities and commodity derivatives: modeling and pricing for agriculturals, metals and energy.* John Wiley & Sons, 2009.

Rajna Gibson and Eduardo S Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *The Journal of Finance*, 45(3):959–976, 1990.

Paul Berhanu Girma and Albert S Paulson. Risk arbitrage opportunities in petroleum futures spreads. *Journal of Futures Markets*, 19(8):931–955, 1999.

Marco Antonio Guimarães Dias and Katia Rocha. Petroleum concessions with extendible options: investment timing and value using mean reversion and jump processes for oil prices. *Institute for Applied Economic Research Working Paper*, (620), 1999.

Marius Hofert. On sampling from the multivariate t distribution. *R Journal*, 5 (2), 2013.

Takashi Kanamura. The role of incompleteness in commodity futures markets. *Frontiers in Applied Mathematics and Statistics*, 1: 11, 2015. ISSN 2297-4687. doi: 10.3389/fams.2015.00011. URL `https://www.frontiersin.org/article/10.3389/fams.2015.00011`.

Takashi Kanamura, Svetlozar T Rachev, and Frank J Fabozzi. A profit model for spread trading with an application to energy futures. *The Journal of Trading*, 5(1):48–62, 2010.

Joëlle Miffre and Georgios Rallis. Momentum strategies in commodity futures markets. *Journal of Banking & Finance*, 31(6):1863–1886, 2007.

Hassan Mohammadi and Lixian Su. International evidence on crude oil price dynamics: Applications of arima-garch models. *Energy Economics*, 32(5): 1001–1008, 2010.

John F. Monahan. *Numerical Methods of Statistics.* Cambridge, 2011.

Yurii Nesterov. *Lectures on Convex Optimization.* Springer Publishing Company, Incorporated, 2nd edition, 2018. ISBN 3319915770.

Dragana Pilipovic. *Energy risk: Valuing and managing energy derivatives.* McGraw Hill Professional, 2007.

Juan C Reboredo. How do crude oil prices co-move?: A copula approach. *Energy Economics*, 33(5):948–955, 2011.

Eduardo S Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance*, 52(3):923–973, 1997.

A. Sklar. Fonctions de rpartition n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

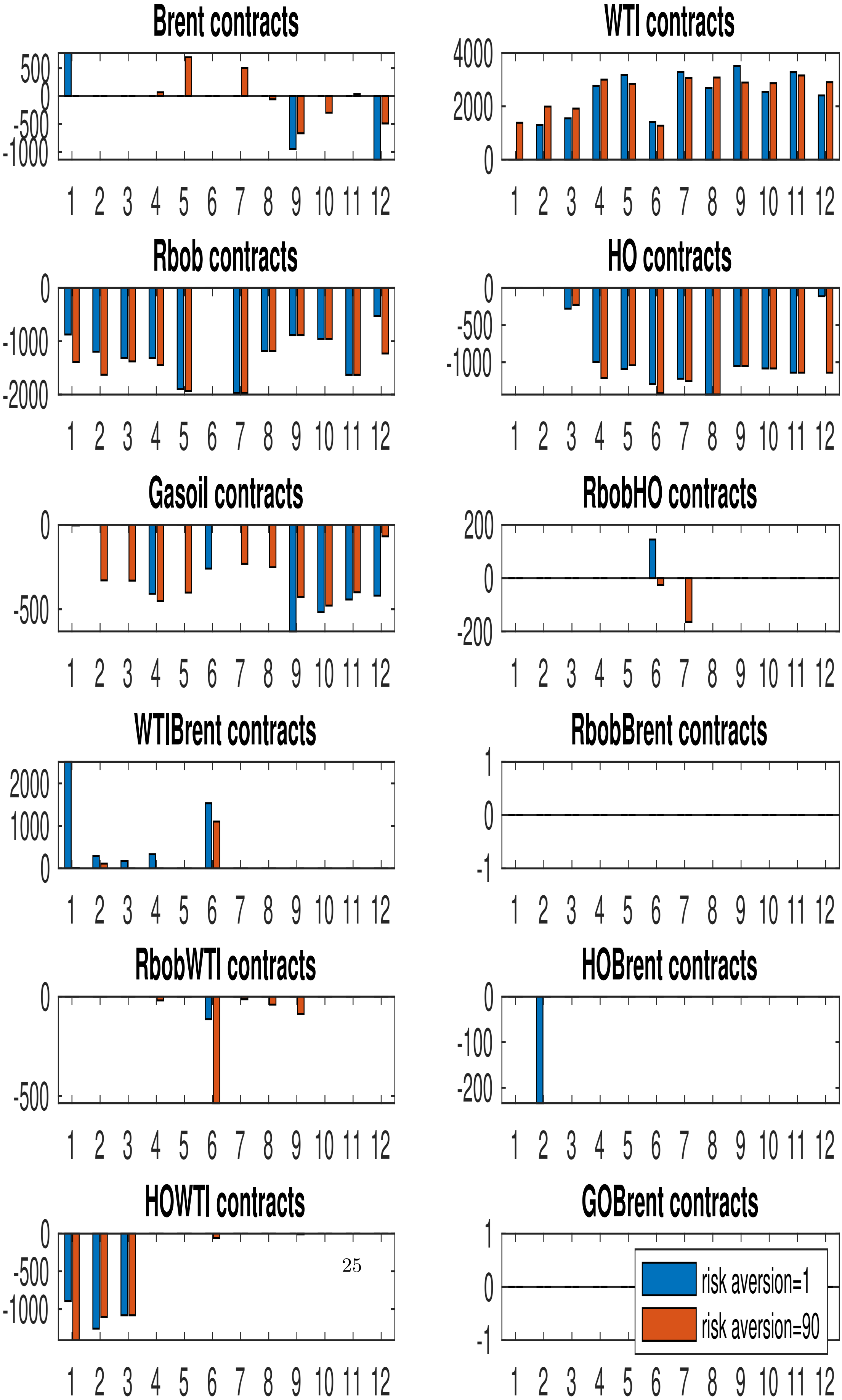


Figure 13: Optimal hedge using exponential utility with $\delta = 90$ and $\delta = 1$

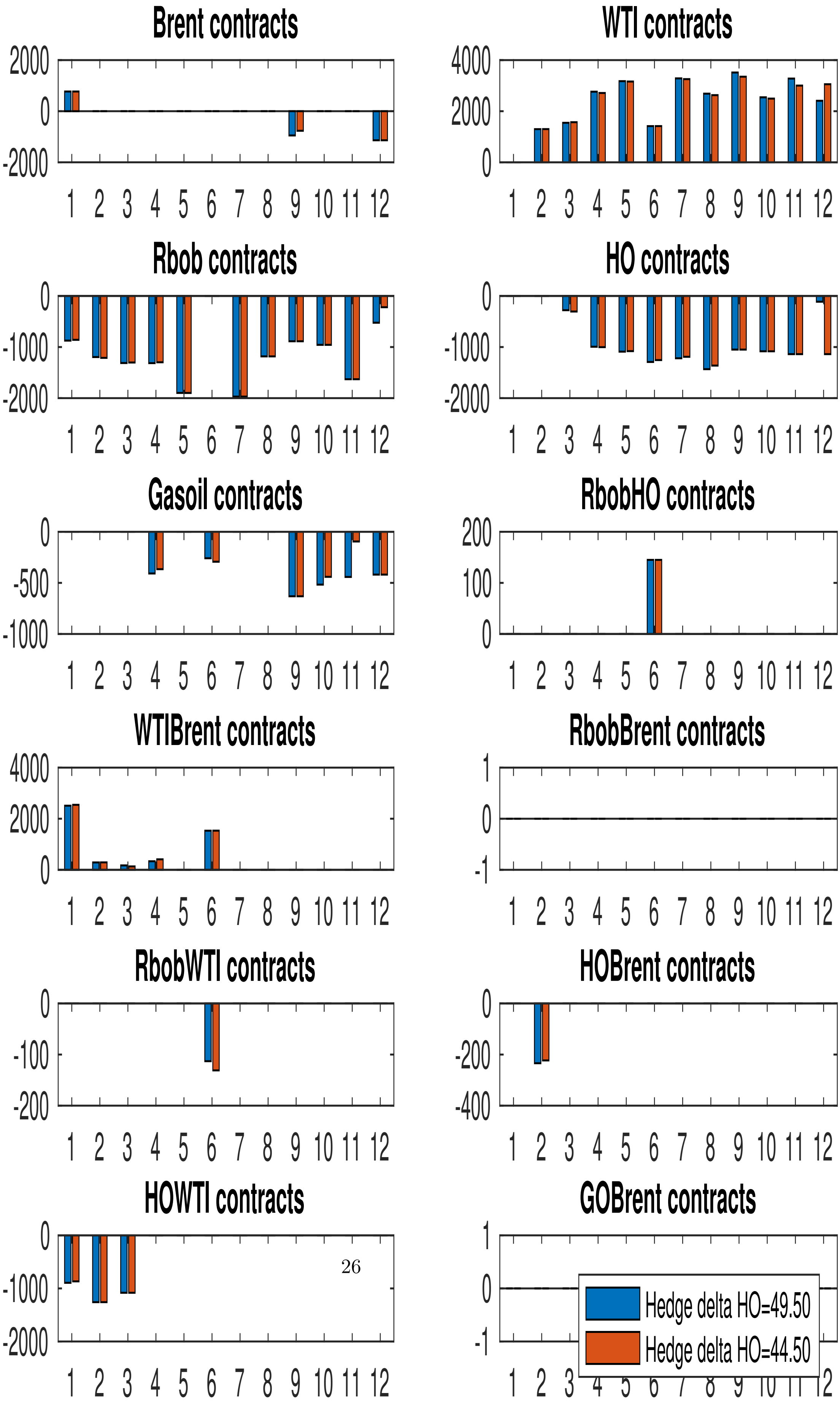


Figure 14: Optimal hedge scenario by changing the median for HO 12th maturity from 49.5 $/bbl to 44.5 $/bbl

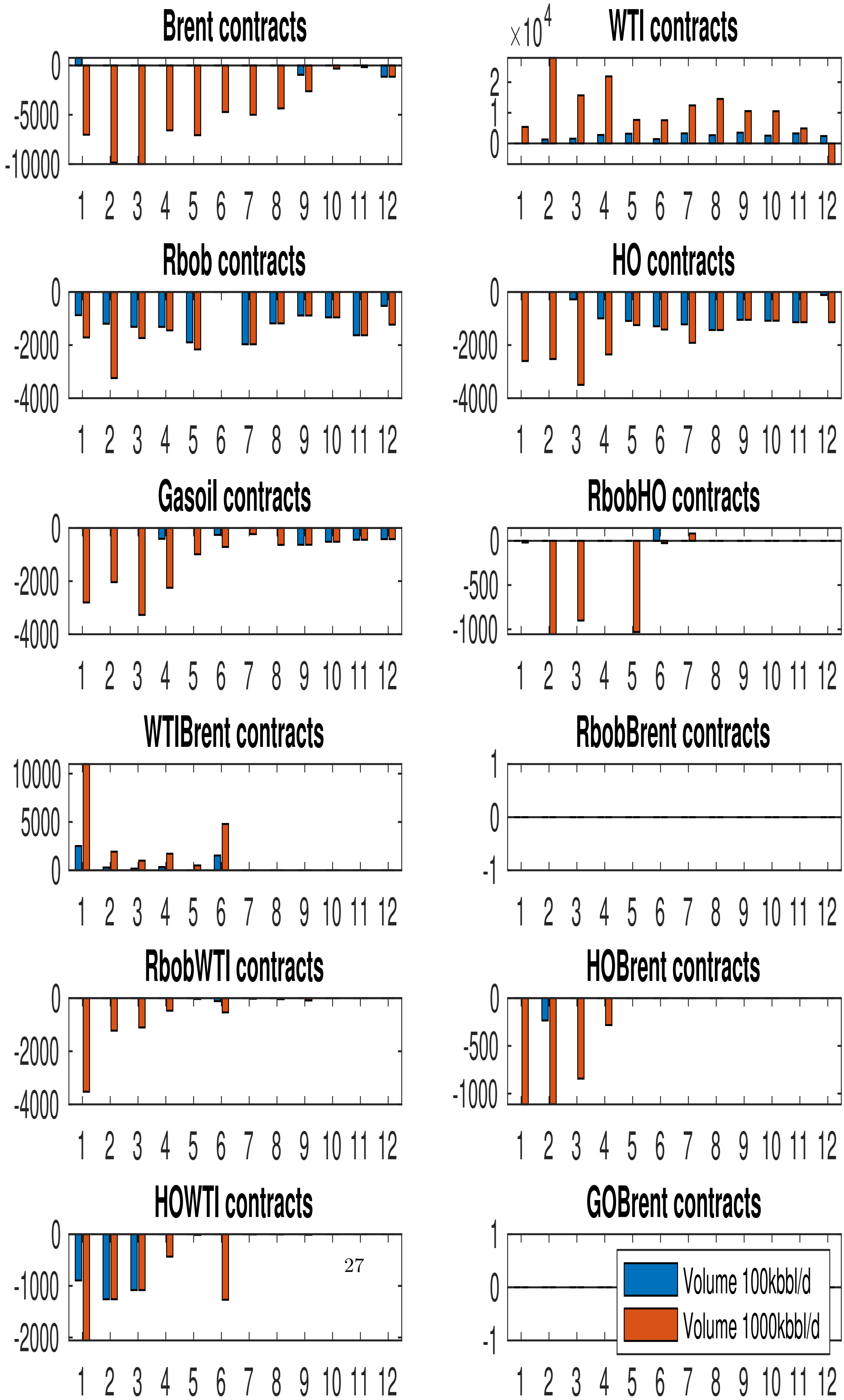


Figure 15: Optimal hedge for refinery capacity 1000 kbbl/d