Central European Journal of Economic Modelling and Econometrics

Crude Oil Price and Speculative Activity: A Cointegration Analysis

Robert Socha* and Piotr Wdowiński†

Submitted: 28.02.2018, Accepted: 26.08.2018

Abstract

The aim of the study is to discuss the relationship of the crude oil price, speculative activity and fundamental factors. An empirical study was conducted with a VEC model. Two cointegrating vectors were identified. The first vector represents the speculative activity. We argue that the number of short non-commercial positions increases with the crude oil stock and price, decreases with the higher number of long non-commercial positions. A positive trend of crude oil prices may be a signal for traders outside the industry to invest in the oil market, especially as access to information could be limited for them. The second vector represents the crude oil price under the fundamental approach. The results support the hypothesis that the crude oil price is dependent on futures trading. The higher is a number of commercial long positions, the greater is the pressure on crude oil price to increase.

Keywords: crude oil price, speculation, futures, cointegration, vector error correction model

JEL Classification: C32, C52, Q31

*University of Łódź; e-mail: robert.socha@uni.lodz.pl
†University of Łódź; e-mail: piotr.wdowinski@uni.lodz.pl

Introduction

An unexpected increase in crude oil price between 2004 and 2008 was an incentive for researchers to look for the factors that may influence the market. A potential influence on the oil trade can be associated with the higher activity of financial institutions. An easier and cheaper access to financial derivatives based on oil price supports the speculation with futures or options (so-called "paper barrels"). In a common sense, speculation refers to buying (or selling) a commodity with an expectation that its price will increase (or decrease) to make a profit. Hart and Kreps (1986) define a speculator as an agent who "buys with the intention of holding the commodity and then selling at a higher price at a later date". Kilian and Murphy (2014) define "speculation" as "buying crude oil not for current consumption but for future use". Moreover, they argue one should not consider this activity as an issue of oil futures market only. The existence of speculators is also expected on the spot market if they are present on the futures market. Undoubtedly, a variety of speculative activity may be more complex, especially if one considers agents trying to make a profit immediately or in a longer perspective or analyse speculation by taking into consideration different types of financial instruments. It is worth noticing that it is not necessary to buy crude oil futures or options directly to invest in this market. This possibility is also widely offered by investment funds (commodity index funds). Certainly, we have observed the transition from physical speculative transaction to purely virtual agreements and noted that the "virtualization" has reinforced a potential scale of investors' activity.
The aim of the present study is to discuss the relationship of the crude oil price, speculative activity on the NYMEX market and fundamental demand factors. We focus on the *third oil shock* only, for which the speculative activity has been considered as one of the possible factors for the changes in the crude oil price between 2002 and 2014.
The remainder of the paper is structured as follows. In the first section, we present a literature review. In the second section, trends on the spot and futures oil markets are introduced. The third section contains the methodology, model specification and data description. The fourth section provides empirical results of the VEC model. The conclusions are given in the final section.

1 Literature review

High crude oil prices observed in 2007 and 2008 have become a reason for discussion about the factors that influence this market (e.g. Hamilton, 2008; Hamilton, 2009; Kaufmann, 2011; Kilian, 2009; Dvir, Rogoff, 2013; Baumeister, Kilian, 2016b; Kim, Vera, 2018). Hamilton (2008, 2009) and Kilian (2009) point out the unexpected demand (especially from developing Asian economies) as the most important factor for the determination of crude oil price. Kaufmann (2011) concludes that non-fundamental elements, not directly related to oil production or consumption, should also be considered. One group refers to political and social disruptions, military conflicts, and natural disasters in the production area (Coleman, 2012; Breitenfellner *et al.*, 2009). There are several studies explaining changes in the oil price differently, as a result of higher trade of futures contracts for crude oil, low interest rates, growth of interest in so-called "alternative" investments or the downturn in the stock market (Kaufmann, 2011; Bencivenga *et al.*, 2012; Xiong, Tang, 2012; Hamilton, Wu, 2014; Gogolin, Kearney, 2016). The discussion can be summarized by the view of Diaz-Rainey *et al.* (2017) who indicate that "academic empirical studies have provided mixed and contradictory evidence". Kilian and Murphy (2014) argue that "there is no consensus in the academic literature on how to model the global market for crude oil".

With an increase in the volume of oil futures and options, researchers wonder if it has any impact on the spot price. This is defined in the literature as a "financialization" hypothesis. Broadly speaking, Fattouh *et al.* (2012) associate this concept with an "increasing acceptance of oil derivatives as a financial asset by a wide set of market participants including hedge funds, pension funds, insurance companies, and retail investors". Palley (2007) argues that this process must be inseparably associated with an increase in the significance of financial institutions over the real institutions operating in a specific industry. What is important, the role of futures trade in price determination has become more controversial when OPEC assigned the whole responsibility for the sharp oil price increase in 2007 to speculators (OPEC, 2007). As an effect OPEC observed that oil prices were "detached from market fundamentals – dynamics of supply and demand" (OPEC, 2008). The analyses of the futures market influence on the crude oil price are also complex and ambiguous. Below, we present the most relevant studies on the relationships between the crude oil price, speculative activity and fundamental factors (see Tables 10 and 11 for a detailed description of selected studies).

Fattouh *et al.* (2012) suggest that we should seriously doubt the influence of financial markets. The authors conclude that there are indications of the financialization of the oil futures market, but the role of speculation is still unclear. Knittel and Pindyck (2016) admit that one should not consider speculation as a factor that determined sharp changes in oil price in the period from 2004 to 2008. Similarly, one could not reject that it affected the oil price at all. Kilian and Murphy (2014) have found that strong and unexpected oil consumption caused a rapid increase in oil price in the period 2003-2008. It is important to notice that they investigated speculation on the spot market for crude oil and they resigned from including in their model any variable from the futures market. Alquist and Kilian (2010) find the negative value of the oil futures spread to be an indicator of possible fluctuations in the crude oil price. However, they show that the accuracy of prediction based on the price of oil futures is in most cases lower than that of a random walk. Lombardi and Van Robays (2011) claim that only about ten percent of oil price volatility is explained by futures trade. The rest is caused by oil supply shortage and stronger than expected oil demand.

On the contrary, Juvenal and Petrella (2011) argue that financial speculation is the second most significant source of oil price volatility. They find that the speculative shock is combined with positive price trends for other commodities than crude oil. This is an argument for further consideration of the role of investment in the commodity index funds. The main hypotheses of Juvenal and Petrella are confirmed by Beidas-Strom and Pescatori (2014). Xiong and Tang (2012) find that growth of investment in commodity index funds contributes to higher correlation of crude oil and other commodity prices. This theory can be explained on the basis of portfolio diversification, where risk of price volatility is shared between different assets. Kaufmann and Kolodziej (2013) conclude that "trader positions play an important role in price discovery" and "speculation could affect oil prices". As the speculation can be measured by the activity of investors on the commodities exchange (with the positions they hold), we can restrict ourselves to the definition of the U.S. Commodity Futures Trading Commission (CFTC) of "commercial" and "non-commercial" traders. According to CFTC a "commercial" trader on the futures market is a participant for hedging purpose, involved in a business activity that is hedged. "Non-commercial" trading refers to the group of participants not working directly on the oil market. We can suppose that the latter are interested in opening positions for extraordinary profit.

Möbert (2009) claims that it is not the number of speculative transactions but "the dispersion in beliefs" (measured as a difference between the number of non-commercial long and short positions) had an influence on the oil price. On the one hand, Haigh *et al.* (2005) reject the hypothesis that the activity of managed money traders in the U.S. futures market affects the crude oil price. On the other hand, Singleton (2013) shows that the activity of managed money traders and index investors might have a positive influence on returns in the oil futures market. Vansteenkiste (2011) argues that if the crude oil price is close to its trend level, only the "commercial" agents enter the market. When oil price shock is observed, the "non-commercial" traders increase their activity. Alquist and Gervais (2011) show that the changes in crude oil price are useful to predict the changes of non-commercial positions on the NYMEX market, but not the opposite way. Dées *et al.* (2008) indicate that prices of futures contracts for WTI are statistically significant for assessing the long-term price.

Diaz-Rainey, Roberts and Lont (2017) investigated whether the oil companies ("commercial" traders) had speculated just before the oil price peaked in 2008 (2004Q3-2008Q2). The authors employed their own measure of oil inventories (*Index of Scaled Physical Inventories*), which was calculated on the data from 15 publicly listed oil companies. They have found that some oil companies might participate in the speculative activity (British Petroleum, Royal Dutch Shell, Statoil, Total). In the same manner, Kolodziej *et al.* (2014) conclude that "large capital flows between equity and commodity markets could alter commodity prices beyond levels indicated by market fundamentals". They find that the daily returns of oil company stock (ConocoPhillips) are correlated with the daily returns of crude oil price. Zhang *et al.*

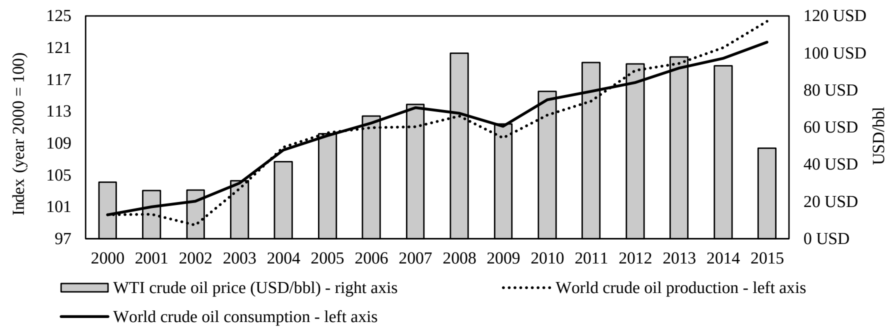
(2017) study whether the crude oil market volatility is influenced by the volatility of the stock market (the VIX and VSTOXX indexes). They argue that the oil market is related to the equity market and that the oil price reflects "not only its fundamentals, but also market risk aversion or investor sentiment".

The review of literature indicates that there are several arguments against and in favour of "financialization" hypothesis and that the crude oil price is likely to be determined by fundamental factors. What motivates the present study is that speculation may matter. Hence, the aim of the paper is to discuss the relationship of the crude oil price, speculative activity on the NYMEX market and fundamental factors. Our contribution to the literature is as follows. We combine three categories of variables: (a) speculation in futures and options (open positions in WTI futures and options), (b) physical speculation (crude oil stock in the North America), (c) demand (industrial production index). The motivation of the paper is to analyse whether the fluctuations in the crude oil price in 2002-2014 were caused by unexpected demand or speculative activity (Hamilton, 2009; Kilian, 2009; Kaufmann, 2011; Baumeister, Kilian, 2016b). Our approach does not only concern the relationship between price and open interest (as it is usually done in the literature), but also the perspective of investors and their expectations of the future price trends (long and short positions, commercial and non-commercial traders). Since we use the term of „speculation" in a broad sense, we can define three channels through which the speculative activity can affect the crude oil market: futures (and options) market, physical speculation and financial market. As the first two channels are included in our analysis, we put an effort to include in the model the financial variables or variables describing other commodities (i.e. U.S. dollar exchange rate, VIX index, S&P 500 and Dow Jones indexes, price of gold, S&P Goldman Sachs Commodity Index). Based on statistical criteria and conclusions of other studies, we finally include only the real effective U.S. dollar exchange rate.

2 Crude oil price, speculation and market fundamentals

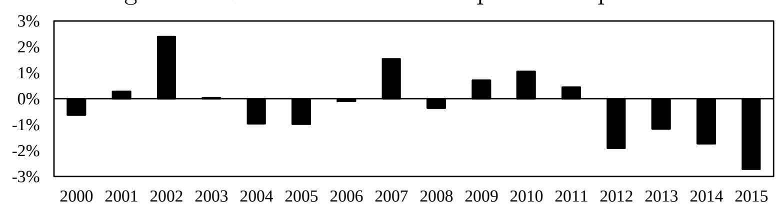
The discussion presented in the previous chapter must be completed with the analysis of trends on the physical and futures markets. As mentioned before, as long as speculation is expected to exist on the futures market, one should also expect the same on the spot market. This effect can be measured by the activity of investors on the commodity exchange (e.g., with positions they held) or by a physical stock of crude oil, changes in its consumption and production. Hence, it is important to look closer at the crude oil demand and supply in the period 2000-2015 (Figures 1 and 2).

Figure 1: World crude oil consumption, production and price during the period 2000-2015



The notation USD/bbl refers to U.S. dollar per barrel.
Source: Joint Organisations Data Initiative, U.S. Energy Information Administration.

Figure 2: Relation of oil consumption and production



Important to notice: the difference between oil production and consumption reflects the change in oil inventories and use of additional gasoline components.
Source: Joint Organisations Data Initiative, U.S. Energy Information Administration.

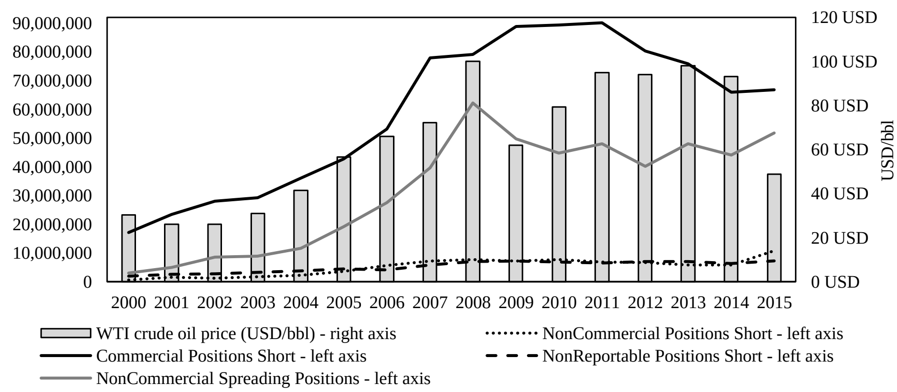
Between 2000 and 2004, demand for crude oil was higher than its production. In 2002 oil extraction was even lower than in 2000, although a significant part of world oil demand (2.4%) was met by the change in oil inventories. A temporary market contraction was an incentive for oil companies to increase production in 2003 and 2004 and it helped to achieve market stability. In 2007, an increase in oil production was stopped (0.1% year-to-year) but the same did not happen to oil consumption (increase of 1.7% year-to-year) and a shortage was observed (at the level of 1.6%). Crude oil price increased in 2008 to the highest level in history (147 USD per barrel). The financial crisis and weakness of the global economy contributed to the fall in oil consumption and crude oil price between 2008 and 2009 (the crude oil price dropped to 62 USD/bbl in 2009). In 2010 crude oil price reached the level of 100 USD/bbl. This was a result of a supply shortage. Oil production rose at the rate of about 1.5 p.p. lower than oil demand. In the period 2011-2014 oil prices fluctuated within the interval of 90-100 USD/bbl. From 2011 price stability was supported by higher oil production, especially by the exploitation of unconventional oil fields at the higher rate of capacity utilization. Firstly, higher level of oil price observed between 2004 and 2008 should be considered as a factor that encouraged to invest in capacity expansion, findings and development (F&D). Secondly, with the price of oil higher than 50 USD/bbl it is economically rational to exploit several categories of oil fields (i.e. tight oil, tar sands). On the other hand, increasing oil production was reflected in the depletion of crude oil price. Sellers had to accept a lower price if they wanted to place their production on the market. Furthermore, lower prices contributed to the reduction in the number of active oil rigs, due to a very low profitability of operations in unconventional oil fields.

The main role of the oil futures market is to hedge the risk in physical trading. However, this market became a popular field of interest for financial institutions when traditional stock or bond investments do not allow to achieve a positive rate of return. Xiong and Tang (2012) remind that the total value of investments in commodity index derivatives (made by individual investors) increased from 15 billion USD in 2004 to over 200 billion USD in 2008.

Figures 3 and 4 show the scale of WTI oil futures and options trade between 2000 and 2015. First of all, we have to notice that the interest in oil futures and options has changed dramatically for several categories of traders. As illustrated in Figure 3 the oil futures and options open interest in 2008 was almost seven times higher than in the year 2000. If we consider "short" transactions, the correlation between the oil price and the number of open positions is in the range from 0.87 to 0.92. Analyzing changes in the structure of oil traders it is worth mentioning that the share of non-commercial short positions is strictly correlated with the oil price. It is quite different if we consider other positions. With a higher share of commercial and non-reportable positions in an open interest, we observe a lower oil price. On the contrary, higher share of non-commercial spreading positions follows higher oil price. According to the U.S. Commodity Futures Trading Commission, "spreading" positions reflect "the extent to which each non-commercial trader holds equal long and short positions". These two positions are taken at the same time to make a profit from widening or narrowing the spread, not from the movement of the absolute prices. Gilbert (2008) reminds that speculation can be associated with "spread trading", because in this form there is no exposure to the price risk.

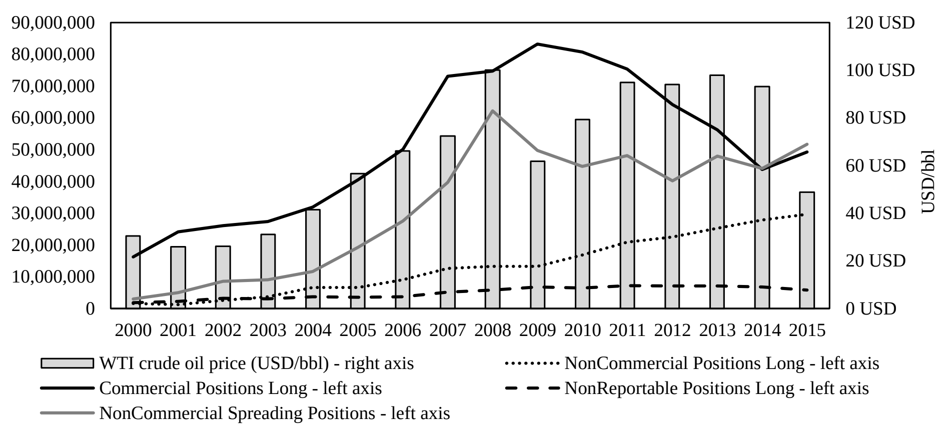
Figure 4 illustrates futures trade in categories of investors for the "long" positions. One significant difference is easily noticeable. Firstly, since 2004 we have observed an upward trend of trading with non-commercial long transactions. Secondly, the number of "long" commercial positions has fallen since 2009, but for "short" transactions this decline started in 2012. Thirdly, Figures 3 and 4 illustrate that commercial traders seem to be more flexible to the situation on the oil market while the situation of "long" non-commercial traders is more similar to "snowball effect", when an increase in the number of open positions seems to be an outcome of "popularity" of this investment form.

Figure 3: Crude oil price and oil futures and options open interest by category of trader (*short positions*) in the years 2000-2015 (data for WTI contracts on the NYMEX exchange)



Source: U.S. Commodity Futures Trading Commission, U.S. Energy Information Administration.

Figure 4: Crude oil price and oil futures and options open interest by category of trader (*long positions*) in the years 2000-2015 (data for WTI contracts on the NYMEX exchange)



Source: U.S. Commodity Futures Trading Commission, U.S. Energy Information Administration.

3 Methodology and data

At the first stage of empirical research, we analyse the stationarity of variables. For this purpose, we use ADF test (Welfe *et al.*, 2006). The results confirm that all variables contain a unit root, and then we apply Johansen cointegration analysis (Johansen, 1988; Lütkepohl, 2005; Kilian, Lütkepohl, 2017). This method starts with a $VAR(p)$ model:

$$x_t = \sum_{i=1}^{p} \Pi_i x_{t-i} + v + \mu_t \tag{1}$$

where:

x_t – vector of endogenous variables $(k \times 1)$,

v – vector of constant terms $(k \times 1)$,

Π_i – coefficient matrices $(k \times k)$,

μ_t – vector of error terms $(k \times 1)$, $\mu_t \sim N(0,\ \Sigma_\mu)$,

$t = 1, \ldots, T$.

The Johansen test for cointegration is sensitive to the lag length p, so the first step is to select an "optimal" number of lags. In practice, the selection of lag length is evaluated with the set of statistics (information criteria of Akaike, Schwarz, Hannan-Quinn, Final Prediction Error FPE, Likelihood Ratio statistic LR). The disadvantage of that approach is the high probability of achieving different results for particular criteria. For example, based on the Akaike criterion one may overestimate the lag length and with Schwarz or Hannan-Quinn criteria one can underestimate the lag length. Moreover, with selected lag length p we must test that the residuals of the VAR model are not autocorrelated or generally that they have appropriate stochastic properties. The next step is the determination of the cointegration rank with the Johansen test (e.g., Johansen, 1988; Johansen, Juselius, 1990; Lütkepohl, 2005; Majsterek, 2014). As long as we consider I(1) variables and assume that they are cointegrated, one may use a transformation of the model (1). This representation is known as the vector error correction (VEC) model given by:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \theta + \delta t + \zeta_t \tag{2}$$

where:

Π – matrix of total impact multipliers $(k \times k)$,

Γ_i – coefficient matrices $(k \times k)$,

θ – vector of constant terms $(k \times 1)$,

δ – coefficient vector $(k \times 1)$,

ζ_t – vector of error terms $(k \times 1)$, $\zeta_t \sim N(0, \Sigma_\zeta)$,

$\Pi = -\left(I_K - \sum_{i=1}^{p} \Pi_i\right)$

$\Gamma_i = -\sum_{s=i+1}^{p} \Pi_s$

$t = 1, \ldots, T$.

The k-dimensional process defined by (1) is cointegrated of order r if the rank of matrix Π fulfils the condition $0 < rk(\Pi) < k$. In this case one can decompose the matrix Π as:

$$\Pi = \alpha\beta^{'} \tag{3}$$

where:

α – matrix of adjustment (loading) parameters $(k \times r)$,

β – matrix of cointegrating vectors $(k \times r)$.

If the rank of matrix Π is zero, one should apply VAR$(p-1)$ model for Δx_t, and if $rk(\Pi) = k$, the process x_t has a stable VAR(p) representation.
It is worth noticing that the specification of the VEC model (2) has been extended by a deterministic trend and this factor was not included in the VAR model (1). Another important aspect is to test whether one should include a constant term and a deterministic trend in the VEC model. It is possible to decompose the vectors θ and δ in equation (2) as follows:

$$\theta = \alpha\gamma + \lambda \tag{4}$$

$$\delta = \alpha\kappa + \rho \tag{5}$$

and to consider a VEC model given by:

$$\Delta x_t = \alpha\left(\beta^{'} x_{t-1} + \gamma + \kappa t\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \lambda + \rho t + \zeta_t \tag{6}$$

where:

γ, λ – vectors of constant terms $(r \times 1)$ and $(k \times 1)$,

κ, ρ – coefficient vectors $(r \times 1)$ and $(k \times 1)$.

There are five cases of restricting deterministic terms in VEC model:

I model without a constant term ($\gamma = 0$, $\kappa = 0$, $\lambda = 0$, $\rho = 0$), II model with a restricted constant term ($\gamma \neq 0$, $\kappa = 0$, $\lambda = 0$, $\rho = 0$),

III model with an unrestricted constant term ($\gamma \neq 0$, $\kappa = 0$, $\lambda \neq 0$, $\rho = 0$),

IV model with a restricted linear trend ($\gamma \neq 0$, $\kappa \neq 0$, $\lambda \neq 0$, $\rho = 0$),

V model with an unrestricted linear trend ($\gamma \neq 0$, $\kappa \neq 0$, $\lambda \neq 0$, $\rho \neq 0$).

For the identification of proper restrictions, one may follow the "Pantula principle" (Pantula, 1989; Wdowiński, 2010). Generally speaking, one should consider all of the model parametrizations (I)-(V) for the possible cointegration ranks, starting with the hypothesis of no cointegration. This procedure is terminated with the null hypothesis not being rejected for the first time. It is worth pointing out that the first case (I) is omitted (Table 4), because then the variables should have zero mean in levels and in first differences. This case is observed very rarely in practice.
As we mentioned earlier, to select the proper VAR specification, it is necessary to check stochastic properties of the error term. Sometimes it requires to introduce additional deterministic variables (e.g. impulse dummy variables). In this case the asymptotic critical values of test statistics (Johansen, Juselius, 1990; Osterwald-Lenum, 1992) might be not applicable to test for a cointegration rank (Amisano, Giannini, 1997; Juselius, 2006; Lütkepohl, 2005; Nielsen 2004). Another problem with a statistical inference about the cointegration rank might be a small sample size (Kębłowski, 2013). There are two ways of handling this. One of them is the use of correction factors. The other way is to use the small sample critical values calculated with the bootstrap (Welfe *et al.*, 2006). We use the critical values computed with the bootstrap distribution (with the assumptions of sample size and additional deterministic components).
The next stage is the imposition of the long run restrictions. This can be specified with the following approach (Lütkepohl, 2005; Juselius, 2006):

$$\beta = [\beta_1, \ldots, \beta_r] = [H_1\Psi_1, \ldots, H_r\Psi_r] \tag{7}$$

where:

β_n – n-th cointegrating vector ($k \times 1$),

v_n – number of restrictions on β_n,

l_n – number of unrestricted parameters in the n-th cointegrating vector, $l_n = k - v_n$,

H_n – design matrices ($k \times l_n$),

Ψ_n – coefficient matrices ($l_n \times 1$),

$n = 1, \ldots, r$.

Under the cointegration rank r one may impose r^2 just-identifying restrictions on the cointegrating space and r on each cointegrating vector. It is worth mentioning that just-identifying restrictions are not tested (by imposing such restrictions the likelihood function does not change). In case of imposing additional restrictions (over-identifying) on the matrix of cointegration relationships β we have to apply the value of likelihood ratio test (LR), thus, it is possible to verify the assumptions based on economic theories. The restrictions are shown in Table 6. This stage is preceded by the identification of weakly exogenous variables for the long run parameters β (Engle *et al.*, 1983; Johansen, Juselius, 1990; Urbain, 1992). If the hypothesis that for the m-th variable $\alpha_{mn} = 0$ cannot be rejected ($m = 1, \ldots, k$ and $n = 1,\ \ldots,\ r$), the information contained in n-th cointegrating vector does not affect the dynamics of Δx_{mt}.

In the final stage of the empirical research, we analyze the impulse response function (IRF) and forecast error variance decomposition ($FEVD$). For the stationary process, represented by (1), one may determine Wold moving average (MA) representation (Lütkepohl, 2005):

$$x_t = \sum_{i=0}^{\infty} \phi_i \mu_{t-i} + u \tag{8}$$

where $u = \mathbb{E}\left(x_t\right)$, ϕ_i is a coefficient matrix, $\phi_0 = I_K$ and for $j = 1,\ 2,\ \ldots$ we can obtain (with $\Pi_j = 0$ for $j > p$):

$$\phi_i = \sum_{j=1}^{i} \phi_{i-j} \Pi_j \tag{9}$$

The $(m,\ n)$-th element of the matrix ϕ_i may be interpreted as a response of $x_{m,\ t+i}$ to a unit change in $x_{n,\ t}$ with constant past values of x_t. The impulse response analysis gives the possibility to observe not only to what extent an innovation in an equation of a particular variable affects another variable in the system, but also how long its effects are visible. One weakness of such an analysis is the correlation of residuals from different equations, therefore, it may be necessary to apply the Choleski decomposition to obtain uncorrelated innovations (Lütkepohl, 2005; Kilian, Lütkepohl, 2017). In practice, we need to decompose the covariance matrix of (1) as $\Sigma_\mu = A^{-1} A^{-1'}$, where A^{-1} represents a lower triangular matrix. In this way, we get the model (1) in its structural form (SVAR):

$$Ax_t = A \sum_{i=1}^{p} \Pi_i x_{t-i} + Av + \eta_t \tag{10}$$

where: $\eta_t = A\mu_t$, $\eta_t \sim (0,\ \Sigma_\eta)$, and the covariance matrix of the structural error terms is $\Sigma_\eta = A \Sigma_\mu A^{'} = I_K$.

With the Choleski decomposition the representation given by (8) can be rewritten as:

$$x_t = \sum_{i=0}^{\infty} \varphi_i \eta_{t-i} + u \tag{11}$$

where $\varphi_i = \phi_i A^{-1}$, $\varphi_0 = A^{-1}$.

For the nonstationary process represented by a VEC model (2), ϕ_i may be computed from the VAR model with integrated variables or from the levels version of the VEC model (Lütkepohl, Reimers, 1992; Lütkepohl, Krätzig, 2004). If the variables are integrated, as a result of this approach, permanent effects can be observed for some of the structural shocks. Further development of impulse response analysis for the nonstationary process x_t is an introduction of Beveridge-Nelson MA representation (Lütkepohl, Krätzig, 2004). In this way, we can obtain a structural VEC model (SVEC) and it is then possible to identify the transitory and permanent structural shocks.

One might wonder to what extent the variability of a particular variable is explained by each structural shock. The answer is an application of the forecast error variance decomposition, which is related to the impulse response functions (equation 11). In practice, we must compute the contribution of k-th shock to the h-step forecast error variance of the j-th variable (Lütkepohl, 2005; Kilian, Lütkepohl, 2017) given as:

$$\omega_{jk}(h) = \frac{\left(\varphi_{jk,0}^2 + \cdots + \varphi_{jk,h-1}^2\right)}{\sum_{k=1}^{K}\left(\varphi_{jk,\ 0}^2 + \cdots + \varphi_{jk,h-1}^2\right)} \tag{12}$$

Consequently, by analyzing all structural innovations we are able to find their contribution to the forecast error variance of each variable in the system.

3.1 Theoretical assumptions for model specification

Measuring the impact of speculative activity on the oil price requires understanding of the whole group of investors in their intention and their perspective of holding specific positions (short or long, hedging or purely speculative). There are several methods to measure the impact of futures contracts on the oil price. Vansteenkiste (2011) employs the speculative activity index T_t, described by Working (1960) as:

$$T_t = \begin{cases} 1+\frac{NCS_t}{HL_t+HS_t} & HL_t \le HS_t \\ 1+\frac{NCL_t}{HL_t+HS_t} & HL_t > HS_t \end{cases} \tag{13}$$

where for oil futures contracts:

NCS_t – number of non-commercial short positions,

NCL_t – number of non-commercial long positions,

HS_t – number of commercial (hedging) short positions,

HL_t – number of commercial (hedging) long positions.

The modification of this approach can be found in Hedegaard (2011):

$$T_t = \frac{NCL_t - NCS_t}{OI_t} \tag{14}$$

where:

OI_t – open interest for futures contracts.

We made several attempts to define such an indicator, when we wondered if the structure of investors is more significant for crude oil price than pure open interest. For example, we tried to include two activity indexes for commercial and non-commercial trading ($\frac{NCL_t+NCS_t}{2\ OI_t}$, $\frac{HL_t+HS_t}{2\ OI_t}$) and for short and long positions ($\frac{HL_t+NCL_t}{OI_t}$, $\frac{HS_t+NCS_t}{OI_t}$) or $NCL_t - NCS_t$ and $HL_t - HS_t$. We also modified Hedegaard's index as $\frac{|NCL_t-NCS_t|}{OI_t}$ to investigate the net share of non-commercial positions. In this specification, the surplus of long or short positions is assessed in the same way and it may be understood as how significant an activity of non-commercial agents is without giving consideration to the market in an upward or downward trend. There is a debate on whether the indexes are more reliable than nominal data. On the one hand, with a rapid increase in open interest we cannot capture changes in the perspective of different market participants, which is not difficult to notice when analyzing the structure of traders. On the other hand, even if the number of speculative transactions is at the highest level in history, the proposed index may not reflect this situation. Nevertheless, we find out that a nominal number of open positions is more important for price processes than the indexes described above. The same approach can be found in Kaufmann and Kolodziej (2013), who use the data for the number of positions in each group. For further empirical studies we have to remove two of the three variables: spreading transactions, commercial short and long positions. Including those variables in the model poses the risk of multicollinearity. Excluded variables are selected on the basis of statistical analysis (significance test) and with regard to its economic interpretation (as mentioned earlier, spreading transactions have different interpretation than open positions). It is worth noticing that the correlation coefficient between commercial short and long positions is about 0.96 in the sample, so the behaviour of investors (regardless of their perspective) has much in common.

Crude oil stock may be identified as an indicator of the future situation of the oil industry. An increase in crude oil stock may indicate a lower demand (or higher production) and this may result in a lower crude oil price. Dées *et al.* (2008) argue that "an increase in stocks reduces real oil price by reducing reliance on current production and thereby lowering the risk premium that is associated with a supply disruption". Fattouh *et al.* (2012) remind that it may be also "interpreted as a sign of speculative pressure in oil markets". The authors conclude that "speculators exogenously drive the futures price above its equilibrium level, which all else equal encourages new production and discourages consumption". As far as an increase in crude oil price cannot be exogenous to the oil market, the accommodation of oil stock in such a situation has to be explained on the basis of exogenous growth in world income and a limited capacity of oil production (Dvir, Rogoff, 2013) or as an effect of rising uncertainty about the future oil supply (Kilian, Murphy, 2014).

It is worth pointing out that the existence of speculators on the spot market not always has to be associated with changes in oil stock. From this point of view, we can consider an influence of the changes in crude oil stock on the number of open positions. As far as commercial investors are considered – by having perfect information about the oil market – positive influence of crude oil stock on the number of short commercial positions (or negative on the number of long positions) can be identified as an expectation of future supply-demand relationship (over-production in the oil market). An opposite relationship would indicate speculation (precautionary demand). Non-commercial investors have imperfect information and it is difficult for them to recognize if an increase in the crude oil stock is caused by the precautionary demand (uncertainty about future supply) or it is an indicator of future trends. The sign of this relationship may provide an answer if an increase in crude oil stock is still perceived as an indicator of a possible disequilibrium in the oil market.

The WTI price is denominated in U.S. dollar, the global trade of crude oil is, to a great extent, also denominated in this currency. In theory, with the U.S. dollar depreciation, the cost of crude oil purchases is lower in the non-dollar regions (in regional currencies), which may lead to an increase in demand and finally to the rise in crude oil price. For the oil suppliers (suppose an oil-exporting country with a currency pegged to U.S. dollar), their purchasing power decreases. Basher *et al.* (2012) argue that "since oil is denominated in dollars, in the face of a declining dollar, with the relative price of oil being set in equilibrium, the dollar price of oil must rise instantaneously". On the other hand, the appreciation of the U.S. dollar increases the income of oil-exporting countries and it may encourage higher production.

The point of departure for our research is general and (as we mentioned earlier) we consider a speculation in a broad sense. The final specification of an econometric model can be compared with the specification presented by Kaufmann and Kolodziej (2013). What is important, due to the use of monthly data, we are able to include a variable that represents the demand for commodities in the economy (industrial production index in the North America). Therefore, we can assess whether fundamental factors or speculation are more important in the determination of the crude oil price (Hamilton, 2009; Kaufmann, 2011; Kilian, 2009; Baumeister, Kilian, 2016b). Moreover, unlike Kaufmann and Kolodziej (2013), we do not include the oil futures prices in the oil spot price equation, what makes our interpretation different (Table 10). Finally, we use the following variables in the model:

$$x_t = \left[P_t,\ R_t, D_t,\ L_t^{COM},\ L_t^{NCOM},\ S_t^{NCOM},\ FX_t\right]' \tag{15}$$

where:

P_t – relative crude oil prices,

R_t – crude oil stock in the North America,

D_t – industrial production in the North America,

L_t^{COM} – WTI futures and options contracts: the number of commercial long positions,

L_t^{NCOM} – WTI futures and options contracts: the number of non-commercial long positions,

S_t^{NCOM} – WTI futures and options contracts: the number of non-commercial short positions,

FX_t – U.S. dollar real effective exchange rate.

An alternative way of capturing the influence of speculative activity on the crude oil spot market is to consider oil as an asset for alternative investments, which can be reflected by including a set of financial variables (gold price, volatility index VIX, S&P Goldman Sachs Commodity Index, Dow Jones, S&P500, the real exchange rate of the U.S. dollar) in an econometric model. Some of these variables were employed by Kolodziej *et al.* (2014), Andreasson *et al.* (2016), Gogolin, Kearney (2016) and Zhang *et al.* (2017). The main problem of such an approach is that we do not know to what extent soaring open interest of oil futures and options (Figures 3 and 4) is caused by: the capital flow (as a result of investment diversification), an increasing awareness of advanced technologies in the financial market (i.e. futures trading platforms) or the role of investment funds on the commodity market. The technology has released the possibility of futures trading for a very wide group of investors (even non-professional). In the empirical part of this study we have made several attempts at including such variables, but the only one that seems to be statistically significant is the U.S. dollar real effective exchange rate. One possible explanation is the data frequency. Kolodziej *et al.* (2014) and Zhang *et al.* (2017) base their studies on daily data and therefore their methodology also differs. Andreasson *et al.* (2016) analysed the causal relationships between futures returns of sixteen commodities (i.e. crude oil) and financial variables (S&P500, Currency Index), speculation measures (Working's T index) or the measures of uncertainty (VIX, Economic Policy Uncertainty Index). The authors have found the existence of a linear unidirectional causality from the speculative activity to the crude oil price for monthly data, but not for weekly data. Interestingly, when they considered a bidirectional linear causality for weekly data, the relationship among crude oil price and variables listed above has been proved in four cases at most and for monthly data in seven cases (over ten cases).

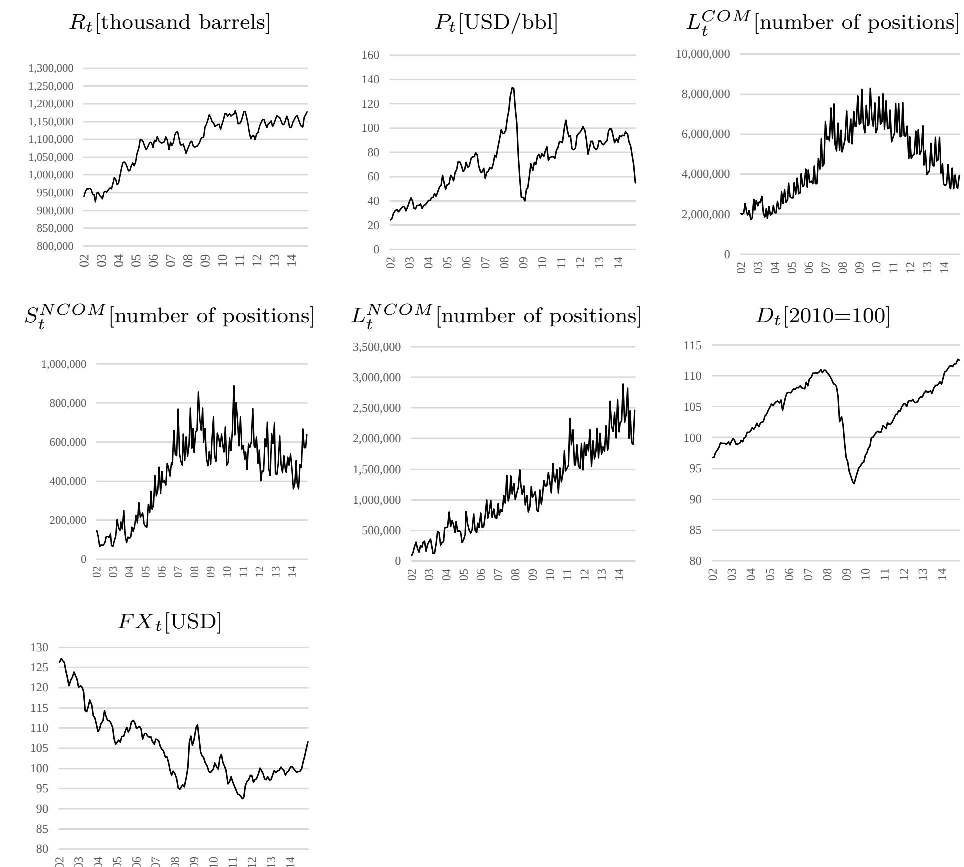
The other explanation might be that by building a model with the monthly data we are able to employ also variables like industrial production or crude oil stock in the North America. For example, in general the Dow Jones Index or the S&P 500 Index represent a "condition" of the U.S. economy (a leading indicator), so it may be close to the effect of including industrial production in the model. Moreover, Andreasson *et al.* (2016) analysed the existence of linear causal relationship from the S&P 500 or the VIX index to the crude oil price but their results were ambiguous (they confirmed the existence of such relationships at 10% significance level or only in one variant of the Granger causality test out of three considered).

3.2 Data description

The sample includes the period from January 2002 to December 2014. It is chosen with respect to the fundamental changes that were noticed on the crude oil market. The subject of our interest is the period of *the third oil shock*. This approach is motivated for the following reasons. The speculative activity is considered as one of the possible factors that caused a sudden volatility of crude oil price from 2002 to 2014. As we concluded (based on Figures 1 and 2), since 2015 we have observed a surplus of crude oil production that decreased the price (*shale revolution*). It can be expected that in the years 2002-2014 the price might be also determined by factors other than only oil supply and demand. Since 2015 we would rather expect that *the shale revolution* (increase in oil supply in the North America) was the most significant determinant of the crude oil price (Mănescu, Nuño, 2015; Baumeister, Kilian, 2016a). Monthly data for the spot price of WTI crude oil is obtained from U.S. Energy Information Administration (Figure 5). We use a relative oil price that is calculated as a relation of the WTI spot price to the U.S. CPI index from OECD database (Zaklan *et al.*, 2010; Coleman, 2012). Monthly data on futures trading activity is obtained from U.S. Commodity Futures Trading Commission and the data of crude oil stocks is from Joint Organisations Data Initiative (JODI). We consider the data for the NYMEX exchange (from futures-and-options combined report) to improve the comparability of data. We also analyze the physical crude oil stock in the North America. According to JODI, "stock level" should be understood as an amount of crude oil that is stored "within national territories (...), by importers, refiners, stock holding organisations and governments" (Joint Organisations Data Initiative, 2016). Monthly data for industrial production index in the North America (weighted average of U.S., Mexico, Canada) and real effective exchange rate of U.S. dollar is obtained from IMF Data. We decided to consider oil stocks and industrial production for the North America, because the physical oil market in this region is highly integrated (for example through pipelines or drilling infrastructure). All variables are in natural logarithms.

We can treat the oil market as global, although one might argue that we should not analyse only the WTI market. There are several reasons in favour of this assumption. First of all, speculation is mostly connected with the NYMEX market and WTI futures. The quality of data from the U.S. market is high. Secondly, we should analyse

Figure 5: Time series (before the seasonal adjustment)



Source: U.S. Energy Information Administration, OECD, U.S. Commodity Futures Trading Commission, Joint Organisations Data Initiative, IMF Data.

only the data from one region. For example, if we look at the price differential between WTI and Brent, we can simply understand that there are temporary, independent trends that determine regional market and in the sample from 2002 to 2014 the relationship between these prices has changed. Thirdly, the U.S. market is innovative if we consider supply trends and the production from unconventional sources. Our point of view is consistent with the arguments of Knittel and Pindyck (2016).

4 Empirical results

Table 1 presents the results of a unit root test. We calculate the ADF statistics for three possible specifications – without or with a constant and deterministic trend. All variables are integrated of order one (with significance level $\alpha = 0.05$). Empirical results are consistent with those discussed in the literature (Bencivenga *et al.*, 2012; Socha, 2013; Thomas *et al.*, 2010). Unit root results for U.S. dollar real effective exchange rate are in line with those presented in Amano, van Norden (1998) and Zhang (2013). Also, for North American industrial production we assume the existence of a unit root, although in the literature this is a subject of intense discussion (Krol, 1992; Kwiatkowski *et al.*, 1992; Perron, 1988; Zivot, Andrews, 1992) and there is some evidence that this variable is trend-stationary. Nowadays, Eksi *et al.* (2011) analysed this problem for recent data (between January 1997 and December 2008) from seven countries (including U.S. data). Our findings are consistent with them.

Table 1: Unit root analysis – ADF test results

Time series	test without constant		test with constant		test with constant and linear trend	
	H0: I(1), H1: I(0),	H0: I(2), H1: I(1),	H0: I(1), H1: I(0),	H0: I(2), H1: I(1),	H0: I(1), H1: I(0),	H0: I(2), H1: I(1),
P_t	0.21 (0.78)	−5.52 $(0.00)^a$	−2.42 (0.14)	−5.51 $(0.00)^a$	−1.98 (0.61)	−5.74 $(0.00)^a$
R_t	−1.76 (0.98)	−2.66 $(0.00)^a$	−1.71 (0.42)	−3.2 $(0.02)^b$	−1.95 (0.63)	−3.54 $(0.03)^b$
D_t	−1.11 (0.24)	−2.82 $(0.00)^a$	−1.93 (0.33)	−3.91 $(0.00)^a$	−2.44 (0.32)	−3.94 $(0.02)^b$
L_t^{COM}	0.35 (0.79)	−3.14 $(0.00)^a$	−1.14 (0.56)	−3.15 $(0.02)^b$	−0.32 (0.98)	−3.97 $(0.00)^a$
L_t^{NCOM}	2.29 (0.98)	−12.04 $(0.00)^a$	−2.03 (0.27)	−12.9 $(0.00)^a$	−2.57 (0.29)	−13.03 $(0.00)^a$
S_t^{NCOM}	1.46 (0.97)	−3.08 $(0.00)^a$	−2.41 (0.14)	−7.49 $(0.00)^a$	−1.61 (0.79)	−6.26 $(0.00)^a$
FX_t	−0.87 (0.33)	−8.36 $(0.00)^a$	−2.55 (0.11)	−8.39 $(0.00)^a$	−2.34 (0.41)	−8.66 $(0.00)^a$

The notation $^{a\ b\ c}$ means rejecting the null hypothesis with significance level, respectively: $\alpha = 0.01$, $\alpha = 0.05$, $\alpha = 0.1$. The null hypothesis (H_0) and an alternative hypothesis (H_1) were defined as: I(0) – stationarity, I(1) –integration of order 1, I(2) – integration of order 2. In brackets, the marginal significance level for test result is given.

The results of selecting a VAR lag order p are reported in Table 2. According to the information criteria one should choose between the lags 1, 3 or 4. This difference is a consequence of the weaknesses of individual criteria, for example, based on the AIC we would rather specify a high lag order, but with HQ we may underestimate the lag length. In small samples also the LR test can lead to different conclusions (Lütkepohl, 2005). In that case, we should pay more attention to the properties of VAR model residuals, especially to their autocorrelation. Hence, the number of lags required to achieve a proper specification of the VAR model is 3. Finally, we verify the properties of VAR(3) residuals with multivariate tests: White test for heteroscedasticity, Jarque-Bera test for normality, LM test for autocorrelation (Jarque, Bera, 1987; Doornik, Hansen, 1994; Urzúa, 1996; Lütkepohl, 2005). The results are reported in Table 3.

Table 2: Results of selecting the lag length in the VAR model

Lag	Criterion							
order	LR	FPE	AIC	SC	HQ	LM_1	LM_2	LM_4
1	3166.8	$3.16e-20$	-25.1	-19.4^*	-22.8	106.1	68.6	66.2*
2	123.2	$1.96e-20$	-25.7	-19.0	-23.0	100.7	58.4*	56.8
3#	152.8	$8.39e-21^*$	-26.6	-18.9	-23.5^*	54.7*	59.0	52.1
4	67.8*	$8.43e-21$	-26.7^*	-18.1	-23.2	58.8	54.2	56.5

Notation: LR – sequential modified LR test statistic, FPE – final prediction error, AIC – Akaike Information Criterion, SC – Schwarz Information Criterion, HQ – Hannan-Quinn Information Criterion, LM(k) – multivariate test for residual autocorrelation of ith order. In the LR and LM tests the significance level $\alpha = 0.05$ was assumed. Sign * means the lag order, which should be chosen with the specific criterion. Sign # indicates the selected lag order for the VAR model.

Table 3: Diagnostic tests of the VAR model – multivariate tests

Jarque-Bera test for residual normality	White test for residual heteroscedasticity
$JB = 19.37$ (0.15)	$WHITE = 1979.79$ (0.84)

Lagrange multiplier test for residual autocorrelation of the i-th order $LM(i)$

$LM_1 = 54.66$ (0.27)	$LM_2 = 58.43$ (0.17)	$LM_4 = 56.76$ (0.21)

The notation $^{a\ b\ c}$ means rejecting the null hypothesis with significance level, respectively: $\alpha = 0.01$, $\alpha = 0.05$, $\alpha = 0.1$. In brackets, the marginal significance level for test result is given.
Impulse dummy variables included in the VAR cover the following periods (presented for one-off events influencing the oil market during that period): Z_1 (value 1 for the period: 2002 September-October) – hurricane season in the Gulf of Mexico (hurricane Lili), oilrigs shutdown in this region, Z_2 (value 1 for the period: 2003 May, June and September) – hurricane Isabel in the Gulf of Mexico, Z_3 (value 1 for the period: 2005 March, April, May, September) – operational start of the Baku-Tbilisi-Ceyhan pipeline, lower forecasts of Asian economic growth, period after the hit of hurricanes Cindy, Katrina and Rita, Z_4 (value 1 for the period: 2006 April) – in April Nigerian supply disruptions, Z_5 (value 1 for the period: 2008 September-October) – a global economic recession, Lehman Brothers bank collapse, financial market crisis, hurricane Ike and Omar in North America mining areas, Z_6 (value 1 for the period: 2009 July) – low forecasts of economic growth for the most developed economies, lowest temperatures in North America for 15 years, Z_7 – (value 1 for the period: 2011 August-September) – Arab Spring, civil war in Libya (Battle of Tripoli), rumors of a decline in U.S. rating, Z_8 (value 1 for the period: 2012 January) – debt crisis in the euro area, expected lower economic growth in the United States and the euro area.

To analyse the long run relationships between variables we have started with an inference on the cointegration rank r with the trace test (Johansen, 1988; Johansen,

Table 4: Johansen trace test results and modulus of the largest roots of the companion matrix

Hypothesis: null and alternative	The trace test								Modulus of the largest roots of the companion matrix for the cointegration rank r	
	Model 2		Model 3		Model 4		Model 5			
	λ_{trace}	λ_{α}	λ_{trace}	λ_{α}	λ_{trace}	λ_{α}	λ_{trace}	λ_{α}	$r=2$	$r=3$
$H_0: r=0$, $H_1: r \geq 1$	**208.89**	**166.75**	**191.61**	**147.66**	**212.50**	**173.79**	**197.55**	**153.82**	1	1
$H_0: r=1$, $H_1: r \geq 2$	**140.03**	**127.14**	**122.89**	**109.07**	**136.39**	**130.59**	**121.76**	**119.46**	1	1
$H_0: r=2$, $H_1: r \geq 3$	**77.19**	**76.71**	*73.02*	*81.59*	86.48	96.45	72.78	89.71	1	1
$H_0: r=3$, $H_1: r \geq 4$	40.45	59.01	37.53	51.011	44.84	72.93	32.75	56.30	1	1
$H_0: r=4$, $H_1: r \geq 5$	16.67	37.54	14.37	30.51	21.66	39.95	14.72	33.30	1	0.87
$H_0: r=5$, $H_1: r \geq 6$	3.89	24.83	2.04	14.88	9.35	25.87	4.89	18.02	0.77	0.81
$H_0: r=6$, $H_1: r=7$	0.98	9.76	0.02	5.03	1.44	13.24	0.53	4.21	0.73	0.71

λ_{trace} – the trace test statistics, λ_{α} – critical value for the trace test calculated as a 0.95 quantile of bootstrap distribution of LR test statistics (with a given cointegration rank, sample size, number of endogenous variables, additional deterministic components). Bold values denote the rejection of the null hypothesis for an alternative hypothesis. Italicized values denote the decision of selecting VEC specification based on the Pantula principle. The bootstrap distribution was computed with Structural VAR software (version 0.45) by A. Warne (2008). For generating the bootstrap quantiles we used 999 replications. Part of the testing and estimation procedure was performed using JMulti 4.24 (Lütkepohl, Krätzig, 2004).

Juselius, 1990). It has been found that among two maximum likelihood statistics proposed by Johansen (1988) for testing cointegration rank, the trace test seems to be more accurate (Majsterek, 2014). Moreover, at this stage we have also determined whether a constant or linear trend should be included in the VEC model (by carrying out the procedure of testing according to the "Pantula principle"). The results are presented in Table 4. We conclude that among the seven variables analysed we can identify two cointegrating vectors. Following the Pantula principle we consider different specification of the models, from most restrictive to the least restrictive and we select the option, when the null hypothesis is not rejected for the first time. That happens for the Model 3 (H_0: $r = 2$). Based on these results, we should include a constant term in the cointegrating relationship. The analysis of the largest roots of the companion matrix (for $r = 2$ and $r = 3$) confirms that we have correctly specified the cointegration rank. The analysis of adjustment matrix for an unrestricted VEC model provides a better understanding of how each cointegration relationship affects the changes of each variable. On the basis of these results, we should normalize the first cointegrating vector with respect to the number of non-commercial short run positions S_t^{NCOM} and the second to crude oil price P_t (Table 7). This approach provides a better insight into the activity of a group of investors with purely speculative intentions and the price process. In the next step, the identification of weakly exogenous variables for the long run parameters β is carried out based on the test proposed by Johansen and Juselius (1990). The results are presented in Table 5. With the cointegration rank $r = 2$, the number of zero-row restrictions should be at most 5 (Juselius, 2006). We cannot reject the null hypothesis of the long-run weak exogeneity for crude oil stock (R_t), industrial production in the North America (D_t), the number of long positions for commercial (L_t^{COM}) and non-commercial investors (L_t^{NCOM}). It is worth noticing that conditioning on the weakly exogenous variables has an impact on cointegration rank testing, because the asymptotic critical values become invalid. On the other hand, there are several advantages of analysing the partial system (conditional on weakly exogenous variables), e.g. higher stability of parameters (Juselius, 2006). In the case of our study, the cointegration rank does not change when we introduce a partial model.

To interpret the long run parameters within the framework of economic theories it is required to identify the cointegration vectors. In practice, it is justified to start with a set of just-identifying restrictions and then – if possible – with respect to the model properties, to add further over-identifying restrictions. The results are presented in Table 6.

The first cointegrating relationship provides an insight into the processes shaping short non-commercial open interest, the second one into the price process. In case of the second cointegrating vector, at the beginning, we considered the specification where crude oil prices were dependent on each kind of open positions included in the system. Based on the LR test statistics, it was allowed to accept this model structure, apart from the coefficients for non-commercial open positions (both short and long),

Table 5: Weak exogeneity tests

	Phase I		Phase II – testing with weakly exogenous R_t, L_t^{COM}, L_t^{NCOM}, D_t	
Variable	Likelihood ratio test statistics (LR)	Marginal significance level	Likelihood ratio test statistics (LR)	Marginal significance level
P_t	**16.44**	**0.00**^{a}	**32.16**	**0.00**^{a}
R_t	2.23	0.33	–	–
D_t	0.80	0.67	–	–
L_t^{COM}	0.74	0.69	–	–
L_t^{NCOM}	4.56	0.10	–	–
S_t^{NCOM}	**30.91**	**0.00**^{a}	**57.88**	**0.00**^{a}
FX_t	**10.02**	**0.00**^{a}	**23.81**	**0.00**^{a}

The notation $^{a\ b\ c}$ means rejecting the null hypothesis with significance level, respectively: $\alpha = 0.01$, $\alpha = 0.05$, $\alpha = 0.1$.

Table 6: Restrictions on the cointegrating vectors

Cointegrating vector	Coefficient						
	R_t	P_t	L_t^{COM}	S_t^{NCOM}	L_t^{NCOM}	D_t	FX_t
Cointegrating vectors restrictions							
β_1'	β_{11}	β_{12}	0	1	β_{15}	0	0
β_2'	0	1	β_{23}	β_{24}	β_{25}	β_{26}	β_{27}
LR test for imposing over-identifying restrictions: $LR = 11.81$ (0.46)							
Cointegrating vectors restrictions – final specification							
β_1'	β_{11}	β_{12}	0	1	β_{15}	0	0
β_2'	0	1	β_{23}	0	0	β_{26}	β_{27}
LR test for imposing over-identifying restrictions: $LR = 13.81$ (0.46)							

In brackets, the marginal significance level for test results is given.

which were statistically insignificant. Hence, we reduced the parameter space, setting previously mentioned coefficient values to zero. We assume that crude oil price is dependent on the industrial production and U.S. dollar real effective exchange rate, which is a reference to standard models of crude oil price or demand (Kaufmann *et al.*, 2007; Krichene, 2007). We also consider the restriction on $\beta_{11} = 0$ as this coefficient was statistically insignificant in all proposed specifications.

In Table 8, we present the diagnostic tests of the VEC model. First of all, we conclude on the basis of a multivariate Jarque-Bera test that the estimated residuals

Table 7: Estimation of the adjustment matrix (restricted VEC model)

Equation for a variable:	R_t	P_t	L_t^{COM}	S_t^{NCOM}	L_t^{NCOM}	D_t	FX_t
α_1'	–	–	–	-0.46 $(7.43)^a$	–	–	–
α_2'	–	-0.15 $(4.96)^a$	–	-0.69 $(5.19)^a$	–	–	-0.02 $(3.73)^a$

In brackets t-statistics are given. The notation $^{a\ b\ c}$ means rejecting the null hypothesis with significance level, respectively: $\alpha = 0.01$, $\alpha = 0.05$, $\alpha = 0.1$.

are normally distributed. The same conclusions can be drawn from results of the univariate Jarque-Bera test. The LM test (also known as the Breusch-Godfrey test) results indicate that there is no autocorrelation of residuals (for the orders 1, 3, 4 and 12). The result of multivariate White test for residual heteroscedasticity is also acceptable for the standard significance level.

Table 8: Diagnostic tests of the VEC model

Univariate tests							
Jarque-Bera test for residual normality							
Equation	1	2	3	4	5	6	7
	0.07 (0.96)	1.37 (0.50)	3.01 (0.22)	2.97 (0.23)	1.63 (0.44)	1.11 (0.57)	1.68 (0.43)

Multivariate tests	
Jarque-Bera test for residual normality	$JB = 11.84$ (0.62)
Lagrange multiplier test for residual autocorrelation (i-th order) $LM(i)$	$LM_1 = 56.99$ (0.20)
	$LM_2 = 56.47$ (0.22)
	$LM_3 = 59.32$ (0.12)
	$LM_4 = 62.67$ $(0.09)^c$
White test for residual heteroscedasticity	$WHITE = 1295$ (0.09)

The notation $^{a\ b\ c}$ means rejecting the null hypothesis with significance level, respectively: $\alpha = 0.01$, $\alpha = 0.05$, $\alpha = 0.1$. In brackets, the marginal significance level for test results is given.

The long-run parameter estimates are presented below – all coefficients are statistically significant (based on the t-statistics at the significance level $\alpha = 0.01$):

$$S_t^{NCOM} = 5.36\ R_t - 0.33\ L_t^{NCOM} + 1.67\ P_t - 64.30 \tag{16}$$

$$P_t = 0.57\ L_t^{COM} + 1.82\ D_t - 0.97\ FX_t - 0.05 \tag{17}$$

The signs of coefficients are mostly as anticipated. In Tables 10 and 11, the comparison of estimation results from our study and other research is presented. A direct comparison may be misleading because the authors did not use the variables in natural logarithms in all studies – especially if we consider the coefficient estimates (e.g. Kaufmann, Kolodziej, 2013).

First of all, it is worth noticing that short non-commercial open interest goes up with an increase in crude oil stock. As far as common knowledge is concerned, when demand for crude oil weakens and the stock level is built up, it means that prices may decline, and therefore the number of short positions might increase. A value of this coefficient, equal to 5.36, may seem overestimated (Novotný, 2012), especially if we consider the model with variables in the logarithm and we interpret this parameter as an elasticity. On average, in the sample period we have observed a change in crude oil stock of about 2.5% (year to year), while an increase in non-commercial open interest (short positions) has been much stronger, about 14.2%. One can notice that the adjustment of an open position is much easier on the futures market than with a physical stock. Changes in the oil stock are recognized as an indicator of physical market outlook that is why the reaction of financial investors can be much deeper. However, we cannot rule out that "non-commercial" investors gain access to imperfect information only and without the information available for commercial agents they are not able to notice the increase in crude oil stock as a precautionary demand (as a consequence of uncertainty about future supply disruptions or geopolitical tensions).

An influence of crude oil price on the number of short open positions is positive, what may be misleading as one could argue for a negative relationship. The downward trend of oil prices should encourage investors to sell their assets. If we consider the behavior of investors in the mature market, we would rather anticipate that the fall of crude oil prices leads to an increase in short open interest. Kaufmann and Kolodziej (2013) have found that the influence of crude oil price on the number of short open positions is negative, but this interpretation should be more complex, while their research is based on weekly data and they also included the oil futures price in the same equation (Table 10). Moreover, in their study the relationship between short open positions and the futures price is positive. Hence, we should pay close attention to Figures 3 and 4, where we can see that the trend of crude oil price from 2000 to 2008 led to a growth in popularity of investing in crude oil futures as many investors saw an opportunity to make a profit. The correlation between these two variables is also positive, for the whole monthly sample it is about 0.72. It is worth noticing that the share of short positions in a whole non-commercial open interest was higher in time of sudden oil price growth (over 35% from 2005 to 2009 and it was below 30% from 2011 to 2014). Barber *et al.* (2009) provide an explanation for such considerations based on the data from the U.S. stock market. The authors argue that the correlation between a short sale of assets and an increase in their price may be a result of a situation "where individual investors are providing liquidity to institutions which are selling overvalued stocks and buying undervalued stocks" and that "the trades of individual investors move prices away from fundamental value". The higher is oil price, the greater is the attractiveness of the oil market for non-commercial investors, despite the fact that they suffer from imperfect information about the crude oil market as they do not operate on this market on a daily basis. Andreasson *et al.* (2016) argue that "speculation volatility is much higher when prices pertain an upward pattern rather than gradually decline" and that "price declines do not attract much new agents with opposite price predictions". Therefore, we conclude that higher crude oil prices might be perceived as an attractor for speculative activity which is an explanation of a positive relationship, especially as we are talking about the market in the phase of strong development.

Furthermore, the higher is the number of long positions held by non-commercial agents, the lower is the interest in short positions. This gives a reason to suspect that the participation of non-commercial agents in the market is to some extent limited. Kaufmann and Kolodziej (2013) postulate that the elasticity between the number of long and short non-commercial open interests is nearly unity, which differs in our specification (Table 10). Nonetheless, we should remember that their research is based on weekly data. If the coefficient value had been equal to zero, we would have had no relationship between these variables. On the other hand, a discussed mechanism does not involve a full substitution of short and long positions, as the unit elasticity hypothesis for these variables was negatively verified ($LR = 51.34$ (0.00)).

The second cointegration equation represents the crude oil price. The outcome is similar to the standard models: the crude oil price is dependent on industrial production, as a factor representing the demand side of the pricing mechanism (Table 10). For the industrial production in the North America a parameter estimate is close to the results from Medlock (2013). On the contrary, in the study of Dvir and Rogoff (2013) the estimate of that parameter is even less than the one for the sample from 1975 to 2011. We have also included the U.S. dollar real effective exchange rate to investigate how the changes in a currency market are transferred to crude oil. In our empirical study, a negative relationship between the U.S. dollar real effective exchange rate and crude oil price is present, while a decline of the U.S. dollar leads to higher crude oil prices, but this effect is not one-to-one.

The last factor included in this equation is the number of commercial long positions on the WTI futures. A relatively high value of its coefficient may be interesting as well as the failure of finding insignificant the number of non-commercial long positions. It seems that the higher is the number of commercial long positions, the greater is the pressure on the price to increase, so it can be considered as a growth incentive in the long-run. An influence of futures open interest on the crude oil price is a point of the biggest differences between our research and other studies. The sign of this relationship is as anticipated, but in empirical works there are noticeable disparities between particular coefficient values. For weekly data, the estimates are between 0.004 and 0.017, and for monthly data 0.04. Fratzscher *et al.* (2014) use the SVAR and they indicate that the effect of 1% positive shock in the open interest leads to a decrease of 0.61% in oil price. Merino, Albacete (2010) show that the increase in the number of long positions leads to an even stronger reaction of the oil price.

The last observation should be considered in greater detail. With the long position an investor expects that the price of an asset will rise and he will make a profit by selling it in the future. Hart and Kreps (1986) remind that "buying cheap and selling dear" strategy leads to price stabilization. When we buy cheap in a period t and sell dear in $t+1$, we put additional downward pressure on the prices in $t+1$, which may lead to their decrease in $t+2$ (and again we buy cheaper in $t+2$). Such an understanding of this activity refers to the "storage" demand or in other words – speculative demand – and it may be considered as stabilizing the market. In a common sense, this point of view can be odd, as far as we very often feel that generally "speculation" and "destabilization" are the same. There are clearly counter examples. Baumol (1957) argues that "speculation involves purchases during the upswing and sale during the downswing". The author adds, "but it (refers to speculation, storage demand) must also have a destabilizing influence in accelerating both upward and downward movements because speculative sales occur when prices are falling, and purchases are made when prices have begun to rise". Hence, we should rather ask whether this affects a frequency or an amplitude of price fluctuations? Therefore, based on empirical results, we cannot claim that commercial long open positions are stabilizing or destabilizing crude oil prices. Our view is consistent with an opinion presented by Vansteenkiste (2011), who argues that if the crude oil price is close to its trend, only commercial agents enter the market. The long run perspective for our study should be related to the long-term trend level.

4.1 Robustness of the results

In Figure 6 we present the results of recursive estimates of long run parameters. We argue that the estimates are stable. When modelling the crude oil market, one could consider to use the Brent price instead of the WTI price. We verified whether the results are robust to this change in the model specification:

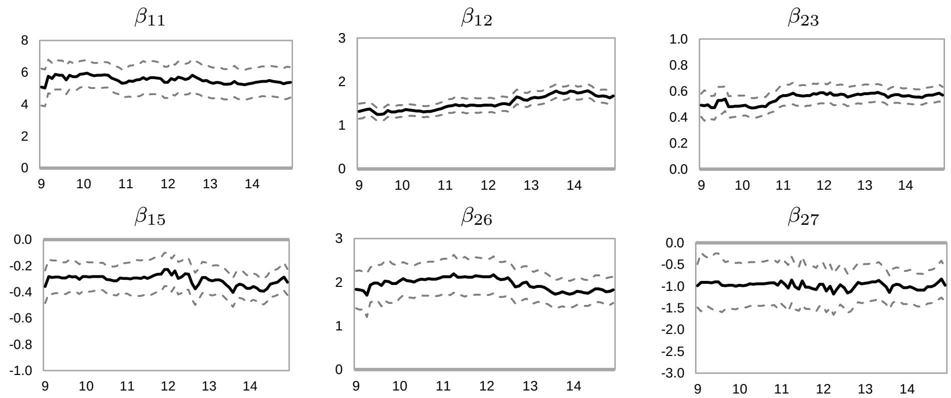
$$S_t^{NCOM} = 5.41\ R_t - 0.32\ L_t^{NCOM} + 1.48\ P_t - 64.20 \tag{18}$$

$$P_t = 0.62\ L_t^{COM} + 2.10\ D_t - 1.18\ FX_t - 0.04 \tag{19}$$

Firstly, the influence of crude oil price on speculative activity of non-commercial traders is weaker. Secondly, in the equation of the Brent price the estimates are higher than for the WTI price, especially for the industrial production (D_t) and the U.S. dollar real exchange rate (FX_t). This may be the consequence of a shift in a differential between these two benchmarks and that to some extent they are determined by different trends (Fattouh, 2007; Kaufmann, 2011). However, a comparison of the results in equations (16)-(17) and (18)-(19) leads to the conclusion that the differences are acceptable.

One should expect that in the period of fundamental changes in oil supply and demand, the role of speculative activity may be less significant. It is noticeable that *the shale oil revolution* in 2015 had an impact on the price of oil (Mănescu, Nuño,

Figure 6: Recursive estimation of long run coefficients



Recursive estimation (± one standard error) was prepared for the period from January 2009 to December 2014.

2015; Zhang *et al.*, 2017). To make a reference to this notice we verified how long-run parameter estimates might change if the sample was extended to the period of shale revolution (from January 2002 to December 2016). The results are presented below:

$$S_t^{NCOM} = 6.00\ R_t - 0.33\ L_t^{NCOM} + 1.62\ P_t - 73.23 \tag{20}$$

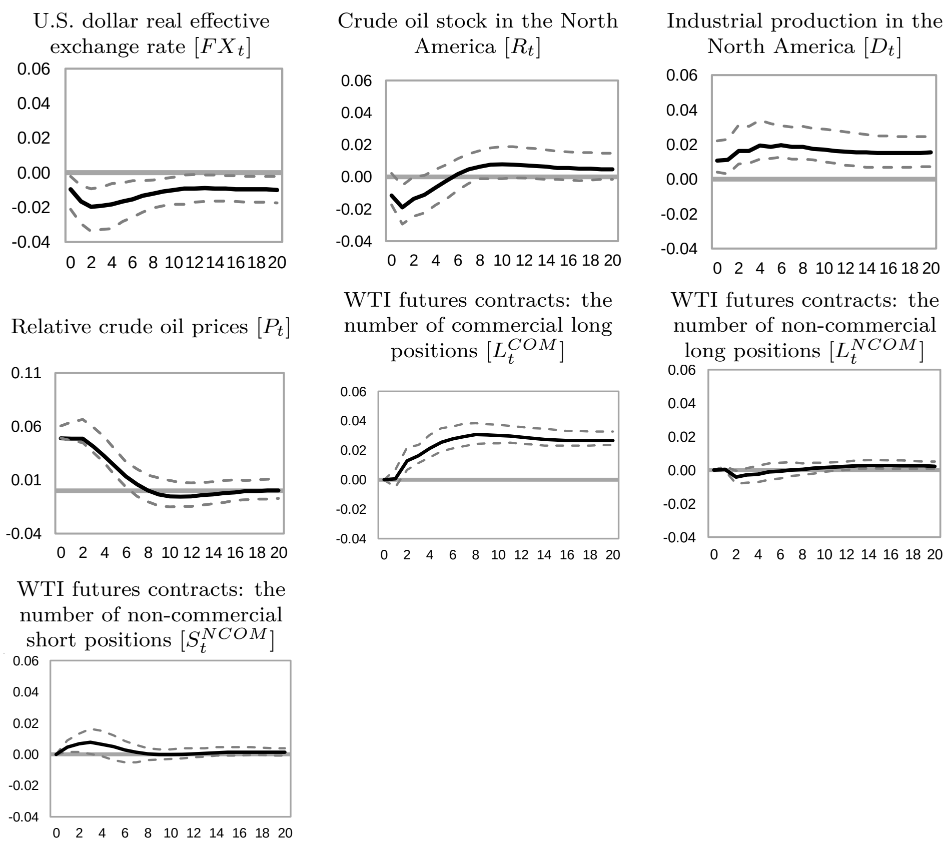
$$P_t = 0.41\ L_t^{COM} + 1.90\ D_t - 1.10\ FX_t - 5.94 \tag{21}$$

There are two main conclusions. Firstly, it is important to notice that in the longer sample there are still two cointegrating vectors. The structure of the model remains stable. The first cointegrating vector represents the speculative activity of non-commercial traders, but the second one relates to the price of crude oil. Secondly, the parameter estimates are not different to a large extent from the results of our base model (16)-(17). We can conclude, however, that the role of market fundamentals increases and the speculative activity has a lower impact on the crude oil price.

In the equation for speculative activity, the parameter for crude oil stock (β_{11}) equals 6.00 and the highest value in recursive estimation was 5.95 (5.36 in our base model). Direct influence of crude oil price on the number of short positions of non-commercial traders (β_{12}) is slightly lower (1.62) than in the base model (1.67). Based on these results, we may argue that investors pay more attention to the movement of crude oil stock (or production) as an indicator of the future market situation, than to current trends. Similar situation is visible in the second cointegrating vector representing the crude oil price. The role of commercial open interest (β_{23}) declines and this estimate for the extended sample (0.41) is lower than the minimum in the recursive estimation (0.47). On the contrary, crude oil price is determined to a greater extent by the demand for commodities and the U.S. dollar real effective exchange rate, the factors that can be seen as crucial for oil production and international trade. Based

on this discussion, we conclude that the role of speculative activity decreases as the fundamental factors become more important.
In the literature review, we have shown that the influence of speculative activity on the crude oil price can be measured differently. One can consider to include the variables representing the financial market. In the model extended by the Dow Jones index, the estimate of β_{26} increases significantly (3.94). If we include the VIX index, the highest difference of estimates in comparison to equations (16)-(17) is observed for industrial production (β_{26}= 2.21). Other parameters change by less than 10%. The changes to the specification of the model do not change the number of cointegrating relations.

Figure 7: Responses of WTI crude oil price to structural shocks



In the impulse response analysis, one standard deviation shocks are considered (Lütkepohl, 2005). In each figure, the solid line represents the mean of impulse response function and the dashed line indicates the bootstrap confidence interval (Hall, 1992, Lütkepohl, Krätzig, 2004) as 0.025 and 0.975 quantiles of bootstrap distribution (999 bootstrap replications).

4.2 Impulse response analysis

Impulse response analysis improves understanding of how the dynamic system responds to exogenous shocks. This analysis facilitates the investigation of the impact of shocks in the crude oil market, which is strongly exposed to various types of exogenous disruptions. The point of departure for this analysis is a recursive identification of structural shocks (Choleski decomposition), which is made in relation to other studies of the crude oil market (Kilian, 2009; Kilian, Murphy, 2014; and for cointegrated variables – Ratti, Vespignani, 2012).
The first shock can be defined as the currency market innovations. In the literature, there are different views on a question of influence of the U.S. dollar on the crude oil price and vice versa (Breitenfellner, Cuaresma, 2008; Fratzscher *et al.*, 2014). As we consider the crude oil market, not the U.S. economy, we decide not to allow the crude oil price to contemporaneously influence the U.S. dollar real effective exchange rate. This assumption is supported by the results of other studies (Su *et al.*, 2016; Taghizadeh-Hesary, Yoshino, 2015). The second structural shock represents the U.S. aggregated demand innovation. This assumption can be explained on the basis of slow adjustment of the commodity demand to the price movement (for example, due to technological limitations of immediate substitution). The next shock corresponds to the innovations in the crude oil stock, which may be interpreted as physical speculation shock. We suppose that it can be caused by anticipated oil production disruptions, for example due to the announcements of decision makers (at OPEC conferences). The fourth shock refers to innovations specific to the oil market – precisely, to sudden exogenous events that lie behind the fundamental factors of oil production, consumption, macroeconomic environment and also speculative activity, for example natural disasters (Deepwater Horizon oil spill), military conflicts (Libyan Civil War, sanctions against Iran), social unrests in the oil territories (strikes in Nigeria, Tunisia).
Unlike the delay in adjusting the fundamental variables like oil demand or stock, one can expect that financial variables respond smoothly to the exogenous innovations. The fifth shock represents the financialization disruptions and it can be determined as a change in the activity of investors *within* the oil industry on the futures and options market. We suppose that oil companies are better informed that firms outside the industry and their open interest do not react to shock in the positions of investment funds, banks or brokers within the same month. The sixth shock one can consider as a representation of the current situation on the financial market that is why it can be interpreted as an attractiveness of investing in the oil market for investors *outside* the industry. Exogenous distortions on the market of other commodities (gold, copper), bonds or stocks may lead to an increase or decrease in the activity of market participants on the futures and options market of crude oil. Last shock represents the purely financial speculation of agents *outside* the oil industry. The number of non-commercial short positions is the last element of the identification scheme and we suppose that any exogenous shock

to other variables is immediately reflected in the activity of purely speculative agents.

Table 9: Forecast error variance decomposition for WTI crude oil price

	Variable						
Horizon	FX_t	D_t	R_t	P_t	L_t^{COM}	L_t^{NCOM}	S_t^{NCOM}
1	0.03	0.07	0.05	0.85	0.00	0.00	0.00
2	0.05	0.06	0.09	0.79	0.00	0.00	0.01
3	0.06	0.08	0.08	0.75	0.01	0.00	0.01
4	0.07	0.09	0.07	0.72	0.03	0.00	0.02
5	0.07	0.10	0.07	0.68	0.05	0.00	0.03
6	0.08	0.11	0.07	0.64	0.08	0.00	0.03
7	0.08	0.12	0.06	0.60	0.11	0.00	0.03
8	0.08	0.13	0.06	0.56	0.14	0.00	0.03
9	0.09	0.13	0.05	0.53	0.17	0.00	0.03
10	0.09	0.13	0.05	0.50	0.20	0.00	0.03
11	0.09	0.14	0.05	0.47	0.23	0.00	0.03
12	0.09	0.14	0.05	0.45	0.25	0.00	0.02
16	0.09	0.15	0.04	0.38	0.33	0.00	0.02
20	0.09	0.15	0.04	0.33	0.37	0.00	0.02

The results of impulse response analysis are presented in Figure 7. In relation to the aim of this study, in this part we focus on the response of the crude oil price. In addition to the mean of impulse response functions, the confidence intervals are also shown. With wide intervals, the interpretation of impulse response functions may be more uncertain. Serwa, Wdowiński (2017) postulate that the impulse response function can be considered statistically significant if its mean is at least two standard deviations above or below zero. This problem applies to the response of WTI oil price to an impulse in the number of non-commercial positions, both long and short, seems to be insignificant and the reaction is temporary and very weak. Physical speculation shock is reflected in a decrease in crude oil price, but the negative reaction is also temporary and it lasts only six months. Moreover, the response to this shock is significant only in the first month. Response of crude oil price to an innovation in the currency market is negative, which is consistent with the estimates of long-run equation and in the horizon of twenty months the effect of this shock is permanent. The highest value of impulse response function is visible very fast, in the first month after the shock occurred. This supports the hypothesis that the crude oil price can be correlated with the U.S. dollar and that changes on the currency market lead to instantaneous reaction of participants on the oil market and price adjustment.

Table 10: VEC estimation results (cointegrating vectors) and comparison to other research

	Method	R_t	P_t	L_t^{COM}	S_t^{COM}	L_t^{NCOM}	S_t^{NCOM}	P_t^{F1}	D_t	FX_t	Period	Other variables included in an equation	Data frequency
Estimation results (cointegration equations, VEC)													
	VEC (CE_t^1)	-5.36^a	-1.67^a	–	–	0.33^a	1	–	–	–	1/2002-12/2014	–	Monthly
	VEC (CE_t^2)	–	1	-0.57^a	–	–	–	–	-1.82^a	0.97^a	1/2002-12/2014	–	Monthly
Estimation results from other research													
Dvir, Rogoff [2013]	VEC (spec. 1)	1	3.98^a	–	–	–	–	–	3.58^a	–	5/1931-12/1972	Oil production	Monthly
	VEC (spec. 2)	1	-0.10^b	–	–	–	–	–	-0.65^a	–	1/1975-12/2011		Monthly
Dées, Gasteuil, Kaufmann, Mann [2008]	ECM	2.06^{b*}	1	–	–	–	–	$-3.25a\dagger$	–	–	Q1/1986-Q4/2006	U.S. refinery utilization, OPEC spare capacity utilization	Quarterly
Merino, Albacete [2010]	VEC	11.17^b	1	–	–	-1.97^a	–	–	–	–	1/1999-12/2007	OPEC surplus of production capacity	Monthly
Kaufmann, Kołodziej [2013]	CVAR (spec. 1)	0.002^b	1	–	–	–	0.004^b	-1.004^a	–	–	1/1997-12/2010	Interest rate (3-month or 6-month T-Bill rate), number of WTI spreading positions on the NYMEX	Weekly
	CVAR (spec. 2)	0.019^b	1	–	0.017^b	–	–	-1.016^a	–	–	1/1997-12/2010		Weekly
	CVAR (spec. 3)	–	–	0.04^a	–	1	-1.04^a	–	–	–	1/1997-12/2010		Weekly
Novotný [2012]	OLS	$-0.74^{b\#}$	1	–	–	–	–	–	-5.31^a	0.03^a	1/1994-9/2010	Interest rate (3-month), U.S. refinery capacity utilization	Monthly
Zhang [2013]	ECM	–	1	–	–	–	–	–	–	0.79^n	1/1973-6/2010	–	Monthly

Variable P_t^{F1} is the futures price of WTI crude oil (the near-month contract). ***Important to notice:*** in the above comparison, it should be noted that there are minor differences in the definitions of variables used in particular research (e.g. 'FX' may denote US Dollar REER, US Dollar NEER or USD/EUR exchange rate, depending on the individual model specification). The notation $^{a\ b\ c}$ means rejecting the null hypothesis with significance level, respectively: $\alpha = 0.01$, $\alpha = 0.05$, $\alpha = 0.1$, ^{n} means "non-reported". Symbol * means days of stocks consumption, # – lagged (one period), † – the difference between the prices of 4-month and near-month contracts for WTI.

Table 11: Comparison to other research cont.

	Estimation results (cointegration equations, VEC)								
	Method	R_t	P_t	OI_t	D_t	FX_t	Period	Other variables included in an equation	Data frequency
Medlock [2013]	GMM (spec. 1)	0.064^a	1	-0.26^a	0.58^a	0.29^a	Q1/1986-Q1/2013	China demand, Non-OECD demand, Drilling cost index, OPEC spare capacity	Quarterly
	GMM (spec. 1)	0.059^a	1	-0.19^a	-1.51^a	0.30^a	Q1/1986-Q1/2013		Quarterly
Fratzscher, Schneider. Van Robays [2014]	SVAR	–	1	-0.61^b	–	0.3^a	2/01/2001-19/10/2012	U.S. stock return, U.S. short interest rate, VIX index	Daily
Bencivenga, D'Ecclesia. Triulzi [2012]	VEC	–	1	0.04^a	–	-2.80^a	1993-2009	Interest rate, Gold spot price, U.S. oil import	Monthly

Important to notice: in the above comparison, it should be noted that there are minor differences in the definitions of variables used in particular research (e.g. 'FX' may denote US Dollar REER, US Dollar NEER or USD/EUR exchange rate, depending on the individual model specification). The notation $^{a\ b\ c}$ means rejecting the null hypothesis with significance level, respectively: $\alpha = 0.01$, $\alpha = 0.05$, $\alpha = 0.1$, ^{n} means "non-reported". Symbol * means days of stocks consumption, # – lagged (one period), † – the difference between the prices of 4-month and near-month contracts for WTI.

For disruption in aggregate demand in North America we can observe an increase in crude oil price. This effect expires very slowly, but it is not an upsurge in price. These results are consistent with conclusions from studies of Hamilton (2008) and Kilian (2009). The authors point out that an unexpected demand in the years 2007-2008 was an incentive to a sudden price increase in the oil market. The highest value of the impulse response function is observed after six months. A surplus of oil demand leads to the release of crude oil stock which allows to spread an increase in crude oil price over time.

An adjustment of the crude oil price to a shock in the activity of commercial traders (long positions) is also significant. Moreover, the results indicate that this effect is permanent in the horizon of twenty months, also the reaction of the crude oil price is not instantaneous. To complete the analysis of factors that affect crude oil price, we should pay attention to Table 9, where we present the forecast error variance decomposition for WTI oil price. According to Juvenal, Petrella (2011) and Beidas-Strom, Pescatori (2014) the demand shock corresponding to the speculative activity is the second most important factor for crude oil price variability. Based on the results, about 15% of the 1-step-ahead forecast error variance of crude oil price corresponds to the innovations in crude oil stock, industrial production and U.S. dollar. This situation changes for the long-term forecast, when about 25% of the 12-month forecast error is accounted for by financialization shock and additional 2% by purely speculative activity. It is worth noticing that with such a forecast horizon, we observe an increase in the role of aggregated demand; still, it becomes the second most important innovation accounted for WTI forecast error variance (besides crude oil price itself). About 9% of the variability of WTI oil price can be attributed to the shock in the U.S. dollar market.

When we look again at the response of the crude oil price to an impulse in the number of long commercial positions, we can notice that it takes about eight months to achieve the maximum value of impulse response function. Based on these results, we argue that with a longer horizon an increase in the role of commercial speculative activity may be considered. A few months after the shock on the market has occurred, we should pay greater attention to the perception of investors, than to the fundamental factors (this situation might be an opportunity to create a speculative bubble).

5 Conclusions

High crude oil prices observed in 2007 and 2008 have become a motivation for the discussion of factors that influence the oil market (Hamilton, 2008, 2009; Kaufmann, 2011; Kilian, 2009). One group of potential determinants refers to a speculative activity. The main objective of the article is to analyze the long-run relationship of the crude oil price, speculation (both in the futures and spot markets) and fundamental factors. An empirical study was conducted with the VEC model with two cointegrating vectors. To interpret the long-run parameters within the framework

of economic theories we considered the model with over-identifying restrictions.
The first equation represents the speculative activity. Based on the analysis of the adjustment matrix for an unrestricted VEC model, the equation is normalized to the short non-commercial open interest. We argue that short non-commercial open interest increases with the crude oil stock and decreases with the higher number of long positions held by non-commercial agents. The first relationship can be perceived in the supply-demand framework. However, it is not mutually exclusive with the hypothesis that "non-commercial" investors gain access only to imperfect information. Agents outside the oil industry are "on thin ice" as access to information for them is costly and limited. An influence of crude oil price on the number of these open positions is positive, what may be misleading is that one could anticipate a negative relationship. This effect may occur because the higher crude oil prices could be perceived as an indicator of the market attractiveness. Positive trend of crude oil prices may be a signal for traders outside the industry to invest in that market, as long as we understand a non-commercial investor as the one who is able to follow also short-run trends and market corrections. To sum up, our specification seems appropriate provided that the investors' activity is based on a long-run situation.
The second equation represents market fundamentals and it is normalized to the crude oil price. Generally, the results of the study support the hypothesis that the crude oil price is dependent on futures trading. The higher is commercial open interest (long positions), the greater is the pressure on crude oil price growth. It is worth noticing that we failed with the number of non-commercial positions (long and short). We conclude that the crude oil price is dependent on industrial production, as a factor representing the demand side of the pricing mechanism, and the real effective exchange rate of U.S. dollar. The international trade of crude oil is denominated in the U.S. dollar which has an impact on decisions of importers and oil producers.
The results discussed above are generally robust to the change in model specification and the time span. The extended sample, which covered the period of shale revolution, is also investigated. We argue that the impact of speculative activity on the crude oil price declines as fundamental factors (i.e. demand for commodities) become more important.
The results of impulse response function and forecast error variance decomposition provide further conclusions about the dynamics of the system. Initially, the aggregate demand on commodities, the U.S. dollar or crude oil stock are more important for the determination of the crude oil price and its variability. However, with a longer horizon, the activity of commercial traders (long positions) becomes most important. This may be a reason to presume that after a shock in the oil price has occurred, from a certain point of time, the price has been detached from the fundamentals, which gives us a reason to conclude that we should pay more attention to the outlook of the futures market.
It is worth mentioning that the problem of financialization can be considered based on three aspects. Firstly, the structure of participants on the oil futures market may

be analyzed. Secondly, we should also consider changes in the behavior of financial institutions (e.g. positions they held). Thirdly, one can look at the activity of the financial sector not only from the perspective of commodities trading in the financial market. Significant shifts should be identified if we consider the volume of transactions on the whole futures market. The latter is not considered in this study, it may be further analyzed in other research. In addition, it would be a challenge to cover the OTC trade on the crude oil derivatives or the data on commodity Exchange-Traded Funds (ETF). It would be also interesting to investigate the views presented by Vansteenkiste (2011) and to include in the system deviations of crude oil price from its trend level, instead of spot prices. It can also be investigated whether the crude oil price adjusts asymmetrically to the deviation from the long-run equilibrium and short-run changes (Leszkiewicz-Kędzior, Welfe, 2014; Welfe, Karp, 2017).

References

[1] Alquist R., Gervais O., (2011), The Role of Financial Speculation in Driving the Price of Crude Oil, *Bank of Canada Discussion Paper* no. 2011(6).

[2] Alquist R., Kilian L., (2010), What do we learn from the price of crude oil futures? *Journal of Applied Econometrics* 25, 539–573.

[3] Amano R.A., van Norden S., (1998), Oil price and the rise and fall of the US real exchange rate, *Journal of International Money and Finance* 17, 299–316.

[4] Amisano G., Giannini C., (1997), *Topics in Structural VAR Econometrics*, Springer, Berlin.

[5] Andreasson P., Bekiros S., Nguyen D.K., Uddin G.S., (2016), Impact of speculation and economic uncertainty on commodity markets, *International Review of Financial Analysis* 43, 115–127.

[6] Barber B.M., Odean T., Zhu N., (2009), Do Retail Trades Move Markets? *The Review of Financial Studies* 22, 151–186.

[7] Basher S.A., Haug A.A., Sadorsky P., (2012), Oil Prices, Exchange Rates and Emerging Stock Markets, *Energy Economics* 34, 227–240.

[8] Baumeister C., Kilian L., (2016a), Understanding the Decline in the Price of Oil since June 2014, *Journal of the Association of Environmental and Resource Economists* 3, 131–158.

[9] Baumeister C., Kilian L., (2016b), Forty Years of Oil Price Fluctuations: Why the Price of Oil May Still Surprise Us, *Journal of Economic Perspectives* 30, 139–160.

[10] Baumol W., (1957), Speculation, Profitability and Stability, *The Review of Economics and Statistics* 39, 263–271.

[11] Beidas-Strom S., Pescatori A., (2014), Oil Price Volatility and the Role of Speculation, *IMF Working Paper* no. WP/14/218.

[12] Bencivenga C., D'Ecclesia R.L., Triulzi U., (2012), Oil prices and the financial crisis, *Review of Managerial Science* 6, 227–238.

[13] Breitenfellner A., Cuaresma J.C., (2008), Crude Oil Prices and the Euro-Dollar Exchange Rate: A Forecasting Exercise, *University of Innsbruck Working Papers in Economics and Statistics* no. 2008-08.

[14] Breitenfellner A., Cuaresma J.C., Keppel C., (2009), Determinants of Crude Oil Prices: Supply, Demand, Cartel or Speculation? *Monetary Policy & the Economy* 2009(4), 111-136.

[15] Coleman L., (2012), Explaining crude oil prices using fundamental measures, *Energy Policy* 40, 318–324.

[16] Dées S., Gasteuil A., Kaufmann R., Mann M., (2008), Assessing the factors behind oil price changes, *European Central Bank. Working Paper* no. 855.

[17] Diaz-Rainey I., Roberts H., Lont D.H., (2017), Crude inventory accounting and speculation in the physical oil market, *Energy Economics* 66, 508–522.

[18] Doornik J., Hansen H., (1994), An Omnibus Test for Univariate and Multivariate Normality, *Unpublished Manuscript*, Nuffield College.

[19] Dvir E., Rogoff K., (2013), Demand Effects and Speculation in Oil Markets: Theory and Evidence, presented at: *IMF Seminars „Understanding Commodity Market Fluctuations"*, Washington D.C., 20–21 March, available at: `http://www.imf.org/external/np/seminars/eng/2012/commodity/pdf/dvir.pdf`.

[20] Eksi I.H., Izgi B.B., Senturk M., (2011), Reconsidering the Relationship between Oil Prices and Industrial Production: Testing for Cointegration in some of the OECD Countries, *Eurasian Journal of Business and Economics* 4, 1–12.

[21] Engle R.F., Hendry D., Richard J.F., (1983), Exogeneity, *Econometrica* 51, 287–300.

[22] Fattouh B., (2007), WTI benchmark temporarily breaks down: it is really a big deal, *Middle East Economic Survey* 50.

[23] Fattouh B., Kilian L., Mahadeva L., (2012), The Role of Speculation in Oil Markets: What Have We Learned So Far? *CEPR Discussion Paper* no. DP8916.

[24] Fratzscher M., Schneider D., Van Robays I., (2014), Oil prices, exchange rates and asset prices, *European Central Bank. Working Paper* no. 1689.

[25] Gilbert Ch.L., (2008), Commodity Speculation and Commodity Investment, *University of Trento Department of Economics Working Papers.*

[26] Gogolin F., Kearney F., (2016), Does speculation impact what factors determine oil futures prices? *Economics Letters* 144, 119–122.

[27] Haigh M.S., Hranaiova J., Overdahl J., (2005), Price Volatility, Liquidity Provision and the Role of Managed Money Traders in Energy Futures Markets, *U.S. Commodity Futures Trading Commission, Working Paper*, Washington DC.

[28] Hall P., (1992), *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.

[29] Hamilton J.D., (2008), Understanding Crude Oil Prices, *The Energy Journal* 30, 179–206.

[30] Hamilton J.D., (2009), Causes and Consequences of the Oil Shock of 2007-08, *Brookings Papers on Economic Activity* 40, 215–261.

[31] Hamilton J.D., Wu J.C., (2014), Effects of Index-fund investing on commodity futures prices, *National Bureau of Economic Research Working Paper* no. 19892.

[32] Hart O.D., Kreps D.M., (1986), Price Destabilizing Speculation, *Journal of Political Economy* 94, 927–952.

[33] Hedegaard E., (2011), How Margins are Set and Affect Asset Prices, Job Market Working Paper, available at: https://pdfs.semanticscholar.org/828d/73fcea38e4bae7161abf4369b41185c616db.pdf?_ga=2.232486870.857145122.1509793810-2917677.1509793810.

[34] Jarque C.M., Bera A.K., (1987), A test for normality of observations and regression residuals, *International Statistical Review* 55, 163–172.

[35] Johansen S., (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231–254.

[36] Johansen S., Juselius K., (1990), Maximum Likelihood Estimation and Inference on Cointegration – With Applications to the Demand for Money, *Oxford Bulletin of Economics and Statistics* 52, 169–210.

[37] Joint Organisations Data Initiative, (2016), Jodi Oil Manual, available at https://www.jodidata.org/_resources/files/downloads/manuals/jodioil-2nd-manual.pdf.

[38] Juselius K., (2006), *The Cointegrated VAR Model: Methodology and Applications, Advanced Texts in Econometrics*, Oxford University Press, Oxford.

[39] Juvenal L., Petrella I., (2011), Speculation in the Oil Market, *Federal Reserve Bank of St. Louis, Research Division Working Paper* no. 2011-027A.

[40] Kaufmann R., (2011), The role of market fundamentals and speculation in recent price changes for crude oil, *Energy Policy* 39, 105–115.

[41] Kaufmann R., Dées S., Karadeloglou P., Sanchez M., (2007), Modelling the world oil market: Assessment of a quarterly econometric model, *Energy Policy* 35, 178–191.

[42] Kaufmann R., Kolodziej M., (2013), The Role of Trader Positions in Spot and Futures Prices for WTI, *Energy Economics* 40, 176–182.

[43] Kębłowski P., (2013), Właściwości wybranych metod małopróbkowego wnioskowania o rzędzie kointegracji, *Przegląd Statystyczny* 60, 163–185.

[44] Kilian L., (2009), Not All Oil Price Shocks Are Alike: Disentangling, Demand and Supply Shocks in the Crude Oil Market, *The American Economic Review* 99, 1053–1069.

[45] Kilian L., Lütkepohl H., (2017), *Structural Vector Autoregressive Analysis*, Cambridge University Press, Cambridge.

[46] Kilian L., Murphy D., (2014), The Role of Inventories and Speculative Trading in The Global Market for Crude Oil, *Journal of Applied Econometrics* 28, 454–478.

[47] Kim G., Vera D., (2018), Recent drivers of the real oil price: Revisiting and extending Kilian's (2009) findings, *Energy Economics*, doi.org/10.1016/j.eneco.2017.12.020.

[48] Knittel C., Pindyck R., (2016), The Simple Economics of Commodity Price Speculation, *American Economic Journal: Macroeconomics* 8, 1–27.

[49] Kolodziej M., Kaufmann R.K., Kulatilaka N., Bicchetti D., Maystre N., (2014), Crude oil: Commodity or financial asset? *Energy Economics* 46, 216–223.

[50] Krichene N., (2007), An Oil and Gas Model, *IMF Working Paper* no. WP/07/135.

[51] Krol R., (1992), Trends, Random Walks and Persistence: An Empirical Study of Disaggregated U.S. Industrial Production, *The Review of Economics and Statistics* 74, 154–159.

[52] Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y., (1992), Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* 54, 159–178.

[53] Leszkiewicz-Kędzior K., Welfe A., (2014), Asymmetric Price Adjustments in the Fuel Market, *Central European Journal of Economic Modelling and Econometric*s 6, 105–127.

[54] Lombardi M.J., Van Robays I., (2011), Do financial investors destabilize the oil price? *European Central Bank Working Paper Series* no. 1346.

[55] Lütkepohl H., (2005), *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.

[56] Lütkepohl H., Krätzig H., (2004), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge.

[57] Lütkepohl H., Reimers H.E., (1992), Impulse response analysis of cointegrated systems, *Journal of Economic Dynamics and Control* 16, 53–74.

[58] Majsterek M., (2014), Modelowanie systemów skointegrowanych. Aspekty teoretyczne, *Bank i Kredyt* 45, 433–466.

[59] Maňescu C.B., Nuño G., (2015), Quantitative effects of the shale oil revolution, *Energy Policy* 86, 855–866.

[60] Medlock K., (2013), Speculation, Fundamentals, and the Price of Crude Oil, *The James A. Baker III Institute for Public Policy*, *Centre of Energy Studies Working Paper*.

[61] Merino A., Albacete R., (2010), Econometric modelling for short-term oil price forecasting, *OPEC Energy Review* 34, 25–41.

[62] Möbert J., (2009), Dispersion in beliefs among speculators as a determinant of crude oil prices, *Deutsche Bank Research Working Paper* no. 32.

[63] Nielsen H.B., (2004), Cointegration Analysis in the Presence of Outliers, *The Econometrics Journal* 7, 249–271.

[64] Novotný F., (2012), The Link Between the Brent Crude Oil Price and the US Dollar Exchange Rate, *Prague Economic Papers* 2012(2), 220–232.

[65] OPEC, (2007), OPEC Bulletin 01/2007, Organization of the Petroleum Exporting Countries, Vienna.

[66] OPEC, (2008), OPEC Bulletin 01/2008, Organization of the Petroleum Exporting Countries, Vienna.

[67] Osterwald-Lenum M., (1992), A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics, *Oxford Bulletin of Economics and Statistics* 54, 461–472.

[68] Palley T., (2007), Financialization: What It Is and Why It Matters, *The Levy Economics Institute, Working Paper* no. 525.

[69] Pantula G., (1989), Testing for Unit Roots in Time Series Data, *Econometric Theory* 5, 256–271.

[70] Perron P., (1988), The Great Crash, The Oil Price Shock, And The Unit Root Hypothesis, *Econometric Research Program, Research Memorandum* no. 338.

[71] Ratti R., Vespignani J.L., (2012), Crude Oil Prices and Liquidity, the BRIC and G3 countries, *MPRA Paper* no. 44049.

[72] Serwa D., Wdowiński P., (2017), Modeling Macro-Financial Linkages: Combined Impulse Response Functions in SVAR Models, *Central European Journal of Economic Modelling and Econometrics* 9, 321–355.

[73] Singleton K.J., (2013), Investor Flows and the 2008 Boom/Bust in Oil Prices, *Management Science* 60, 300–318.

[74] Socha R., (2013), Analiza relacji cen wybranych gatunków ropy naftowej, *Polityka Energetyczna – Energy Policy Journal* 16, 17–31.

[75] Su X., Zhu H., You W., Ren Y., (2016), Heterogeneous effects of oil shocks on exchange rates: evidence from a quantile regression approach, *Springerplus* 5.

[76] Taghizadeh-Hesary F., Yoshino N., (2015), Macroeconomic Effect of Oil Price Fluctuations on Emerging and Developed Economies in a Model Incorporating Monetary Variables, *Asian Development Bank Institute Working Papers* no. 546.

[77] Thomas A., Mühleisen M., Pant M., (2010), Peaks, Spikes, and Barrels: Modeling Sharp Movements in Oil Prices, *IMF Working Paper, Strategy, Policy, and Review Department* no. WP/10/186.

[78] Urbain J.P., (1992), On weak exogeneity in error correction models, *Oxford Bulletin of Economics and Statistics* 54, 188–203.

[79] Urzúa C. M., (1996), On the correct use of omnibus tests for normality, *Economics Letters* 53, 247–251.

[80] Vansteenkiste I., (2011), What is driving oil futures prices? Fundamentals versus speculation, *European Central Bank, Working Paper Series* no. 1371.

[81] Warne A., (2008), *Structural VAR (version 0.45)*, available at: `http://www.texlips.net/svar/`.

[82] Wdowiński P., (2010), *Modele kursów walutowych.*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

[83] Welfe A., Karp P., (2017), Makroekonometryczny miesięczny model gospodarki Polski WM-1, *Gospodarka Narodowa* 28, 5–38.

[84] Welfe A., Karp P., Kębłowski P., (2006), Mechanizmy makroekonomiczne w gospodarce polskiej, Analiza ekonometryczna, Wyd. Uniwersytetu Łódzkiego, Łódź.

[85] Working H., (1960), Speculation on Hedging Markets, *Food Research Institute Studies* 1, 185–220.

[86] Xiong W., Tang K., (2012), Index Investment and the Financialization of Commodities, *Financial Analysts Journal* 68, 54–74.

[87] Zaklan A., Zachmann G., Neumann A., (2010), The dynamics of global crude oil production, *DIW Berlin Discussion Papers* no. 1075.

[88] Zhang Y.-J., Chevallier J., Guesmi K., (2017), "De-financialization" of commodities? Evidence from stock, crude oil and natural gas markets, *Energy Economics* 68, 228–239.

[89] Zhang Y., (2013), The Links between the Price of Oil and the Value of US Dollar, *International Journal of Energy Economics and Policy* 3, 341–351.

[90] Zivot E., Andrews D.W.K., (1992), Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis, *Journal of Business & Economic Statistics* 10, 251–270.