Doctoral School of General and Quantitative Economics

COLLECTION OF THESES

Gábor Uliha

Oil and commodity markets' relationship with the macroeconomy

Ph.D. dissertation

Supervisor:

János Vincze, DSc
professor

Budapest, 2016

Department of Mathematical Economics and Economic Analysis

COLLECTION OF THESES

Gábor Uliha

Oil and commodity markets' relationship with the macroeconomy

Ph.D. dissertation

Supervisor:

János Vincze, Dsc
professor

Table of contents

Summary

The dissertation consists of three main chapters. First, I analyze the time-varying macroeconomic effects of the price of crude oil, and investigate why the 2000s' price rally didn't cause stagflation in the United States. After that, I use continuous wavelet transforms to detect the oil-macroeconomy relationship in an exporter country, namely Norway, and its net oil importer neighbour, Sweden. Finally, I test the robustness of short-term (1-3 days ahead) oil price forecasts' performance via econometric and machine learning tools.

In Chapter II I show that the price of oil is affected by many fundamental factors like geopolitical events, global economic growth, technological progress, market sentiment, foreign exchange market fluctuations and the cohesion of OPEC. I also present how increasing prices can affect the US macroeconomy. Higher inflation and unemployment rate, and lower GDP growth can be the result of direct and indirect effects as well. Direct effects contain rising input costs for production and shrinking disposable income for households. Indirect effects mean the so-called allocation, uncertainty and monetary policy effects (*Kilian* [2014]). These indirect (or asymmetric) effects can explain the lack of sky high economic growth and extremely low inflation rates after the oil price collapse of the 1980s and 1990s. But they can't explain the lack of stagflation during the price increase in the 2000s. Some researchers argue that the weakening effects of the oil price are due to technological progress (less oil is needed for one unit of GDP), more elastic labour markets, and the increased efficiency of monetary policy in the United States and several other developed economies (*Blanchard & Galí* [2007]). Others (like *Kilian* [2010]) support the theory that the standard *ceteris paribus* assumption is invalid when somebody analyzes the effects of oil price shocks because the macroeconomic effects differ due to the sources of the change in price: supply, aggregate demand and specific demand shocks have different consequences. I show in the thesis that both theories can be true at the same time: shocks matter, but during the last decades their macroeconomic effects changed dramatically. For doing this, I generate the shock series in a sample with monthly resolution (1973M1-2014M12) based on the VAR model of *Kilian* [2009], then I estimate the macroeconomic effects of the different types of shocks via Kalman-filter. The results suggest that the effects of aggregate and specific demand shocks diminished since the 1970s and 1980s, or at least were relatively modest in the first decade after the millennium, when aggregate demand shocks dominated the positive price trend (which were fuelled by rapid economic growth in Asia). That means, although source of price fluctuations really matters, the American economy became

more resistant against these shocks, so the two rival theories complement each other. Estimations also show that in the last few years there is some evidence for short-term growth acceleration and unemployment reduction after an oil price increase. This could be the consequence of the booming US shale oil industry. E.g. the recovery after the 2008 crisis got additional help from the positive specific demand shocks, which led to lower rates of unemployment and boosted economic growth. However, the price slump – started in mid-2014 – was also fuelled by specific demand shocks, which had significant short-term (maximum 1-2 quarters ahead) negative implications on the US economy by further reducing the already close to zero inflation rate and slowing down GDP expansion.

The aim of Chapter III is to present the co-movement in time and frequency space between the (Brent) oil price and some of the main macroeconomic aggregates (CPI inflation, industrial output and GDP growth) of Norway and Sweden in the 1970-2014 period by using the so-called continuous wavelet transform (CWT). I show that the oil-macroeconomy relationship can occur in different frequency bands, potentially with opposite signs at the same time, which justifies the usefulness of CWT as a complementary tool for time series analysis. It also becomes clear that the strength of the relationship in a given frequency often changes in time, and this phenomenon usually remains hidden by simple correlation analysis. I find that the reasons for time-varying co-movement is very similar to the US case: exogenous supply, aggregate demand and specific demand shocks show different patterns (wavelet coherency) with economic variables, and these patterns intensively changed both in time and frequency range. During the sample period the co-movement (wavelet coherency) between the real price of oil and industrial output/GDP growth became milder in Sweden, while Norwegian economic performance was more or less insensitive to oil price fluctuations both in the low (3-8 years) and high (1-3 years) frequency ranges. However, Norwegian variables showed higher coherency with the different shock series, although it remains true that the oil exporting economy managed to reach a low level of sensitivity in the case of shocks and price as well

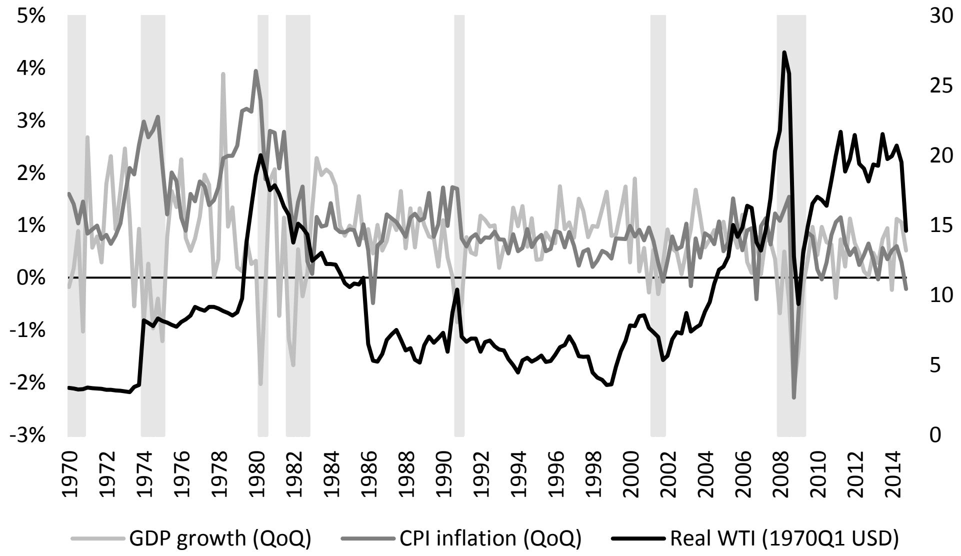
In Chapter IV I build several short-term forecasting models for predicting the nominal price of oil and test whether any of them can provide stable prediction accuracy through different time horizons (1, 2 or 3 days ahead), test sets (2011, 2012, 2013, 2014) and error indicators (MAE, RMSE, DA/HT, profit). In fact, oil price fluctuations and the relevant data generating process have long been researched by economists and market analysts, but the recent downward trend has put them and their predictability in the focus again. The thesis provides insights into the main oil price forecasting techniques and examines the precision of these real time predictions by using the WTI (West Texas Intermediate) spot price. It presents not only standard econometric tools (ARIMA, ARX, LPM, logit) but also some machine learning (SVM, NN) techniques. The benchmarks are simple heuristic models: random walk with and without drift. The results suggest the inclusion of oil futures prices and gasoline/heating oil spot prices in the model can reduce forecast errors, but the relative performance of these models is affected by the projection horizon, the test set and the way we measure prediction errors. In total, there is no universal model that could permanently give „*better than the benchmark*" forecast in every horizon/indicator pairs. Even if we focus only on one indicator, it sometimes happens that there is no method which can beat the benchmark every single year. When it can, then there is at least one year, when the difference is insignificant. Another conclusion shows higher robustness of performance in the case of econometric tools against machine learning models, however, the latter group has a great potential: some machine learning models (SVMs) could generate high profit rates, and the median prediction accuracy of machine learning techniques wasn't dominated by econometric models. A final remark: preliminary variable selection via LASSO and LASSOGLM couldn't eliminate the lack of robustness.

1. The weakening macroeconomic effects of oil price shocks: the synthesis of two rival theories

1.1. Motivation

Figure 1 shows the United States' quarterly GDP growth and CPI inflation rates, and the real price of (WTI) oil. The oil crises of 1973-1974 and 1979 were followed by a period of stagflation (recession, deflation and increasing unemployment at the same time). In contrast, the rapidly rising price between 2002 and 2008 didn't cause such negative consequences, and until the Great Recession the US economy was in a good health with solid growth and low inflation. Some researchers found that the weakening effects of the oil price were due to technological progress (less oil was needed for one unit of GDP), more elastic labour markets, and the increased efficiency of monetary policy in the United States and several other developed economies (*Blanchard & Galí* [2007]). Others (like *Kilian* [2010]) supported the theory that the standard *ceteris paribus* assumption was invalid when somebody analyzed the effects of oil price shocks because the macroeconomic effects differed due to the sources of the change in price: supply, aggregate demand and specific demand shocks had different consequences. In Chapter II I show that both of the two rival theories can be true at once: shocks matter, but their macroeconomic effects have diminished over the decades.

Figure 1: US CPI inflation and GDP growth vs. the real price of WTI (1970Q1 - 2014Q4)



The WTI nominal price was deflated by the US implicit price index. The shaded areas show recessions (defined by the NBER). The source of the data is the Federal Reserve Economic Database (FRED).

1.2. Applied Methods

I identify the shocks through a VAR framework developed in *Kilian* [2009][1]:

$$A_0 z_t = \alpha + \sum_{i=1}^{24} A_i z_{t-i} + \varepsilon_t$$

Where *z* is the column vector of endogenous variables (namely – the dlog of – global crude oil production, real economic activity[2], and the logarithm of the real price of oil[3]), α contains constant terms, while ε is the vector of orthogonal and serially uncorrelated errors. I work with a monthly resolution sample from 1973M1 to 2014M12. I estimate the above structural form's reduced version:

$$z_t = \beta + \sum_{i=1}^{24} B_i z_{t-i} + e_t$$

Where $\beta = A_0^{-1}\alpha$, $B_i = A_0^{-1}A_i$ and $e_t = A_0^{-1}\varepsilon_t$. The shocks' identification strategy is the following:

$$e_t = \begin{bmatrix} e_t^{production} \\ e_t^{real\ activity} \\ e_t^{real\ price} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{production} \\ \varepsilon_t^{real\ activity} \\ \varepsilon_t^{real\ price} \end{bmatrix}$$

Under the first assumption producers can't react to changes in global activity and the real price of oil within a month. This is due to technological constraints and the role of uncertainty (how long the prices will stay higher/lower). The second assumption says real activity needs at least a month to change because of oil price fluctuations which is in line with empirical observations (and can be a consequence of contracts). Of course, prices can response „immediately" (within a month) for a shock in supply or activity. The three types of shocks are named via *Kilian*

[1] Details can be found in *Kilian* [2009] at page 1058.
[2] For further details see *Kilian* [2009] page 1056.
[3] In this chapter I use the refiner acquisition cost of imported crude oil (from the EIA database). Real prices are generated by the US CPI.

[2009]: supply (unexpected change in global crude oil production), aggregate demand (which affects not just oil but other commodity markets as well) and specific demand shocks (precautionary demand). I estimate the parameters by OLS, while for standard errors I use the bootstrap method of *Goncalves & Kilian* [2004] with 50 000 replications.

I estimate the macroeconomic effects of shocks by using the Kalman-filter technique[4]. Because of the resolution of some variables I work on a quarterly basis, so I convert all variables to this frequency. The monthly shocks are averaged for each quarter (1975Q1 – 2014Q4), while the macroeconomic aggregates (CPI, implicit price index, GDP, unemployment rate) are used as annualized quarterly growth rates. The form of measurement equations:

$$\Delta y_t = \delta_t + \sum_{i=0}^{12} \theta_{it} \hat{\xi}_{jt-i} + u_t$$

Where Δy_t is the annualized quarterly growth rate of the given macro variable, while $\hat{\xi}_{jt}$ is the average of type *j* monthly shocks (j = supply, aggregate demand, specific demand) in quarter *t*. The partial effects (θ_{it}) can be different for every quarter. The accumulated impulse response functions are created from these values, and will show the difference in the level of the macroeconomic variable. In the state equations, I assume that parameters follow a random walk process without drift:

$$\delta_t = \delta_{t-1} + \gamma_t$$
$$\theta_{it} = \theta_{it-1} + \rho_t$$

Where γ and ρ are Gaussian white noises with zero expected values. I use the smoothed versions of the estimated parameters.

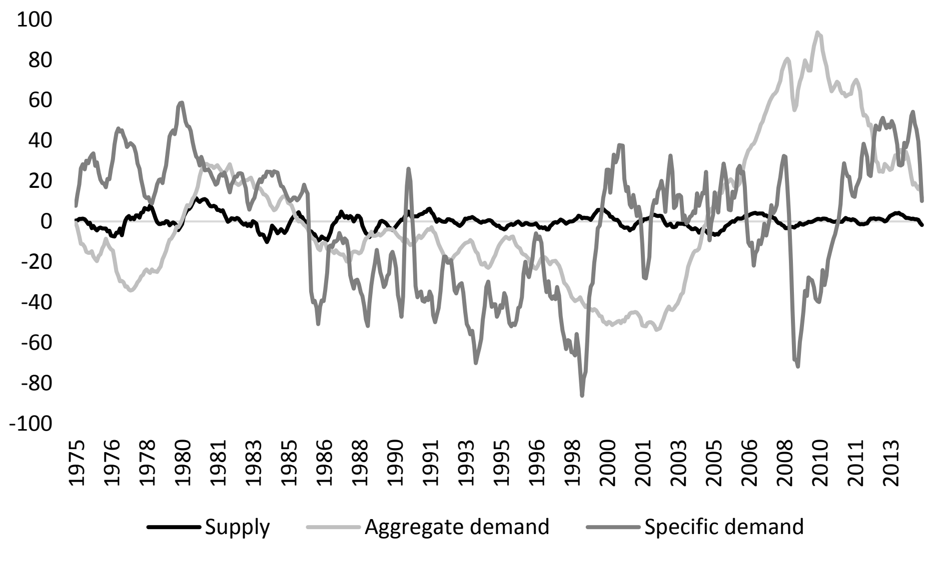
[4] For details see *Hamilton* [1994] Chapter 13.

1.3. Results

The chapter's main conclusions:

1. In line with the findings of *Kilian* [2009] the upward price trend of the 2000s was related to aggregate demand shocks, fuelled by rapid Asian economic growth (see Figure 2). My own calculations also show that after the break of trend (in 2008), specific demand shocks became the main drivers behind real oil price movements, however, aggregate demand shocks also meant an important factor. The price collapse of 2008 was mainly due to negative specific demand shocks, which occurred because of the expected deceleration of global economic growth and future oil demand. That was followed by a period of negative aggregate and positive specific demand shocks which finally caused a price rally (specific shocks dominated). The former was due to slowing economic growth, while the latter was the result of the Arab Spring and the belief that the crisis would soon be over and energy demand would accelerate. From mid-2014 specific shocks turned negative and not offset, but strengthened the aggregate shocks' downward pressure on price. The absolute impact of specific shocks dominated again.

Figure 2: The shocks' accumulated effect on the real price of oil (1975M1-2014M12)



Effect on 100*ln(rpo) where rpo is the real price of oil. Based on the VAR model of *Kilian* [2009].

2. During the last four decades supply shocks (unexpected falls in oil production) had marginal impact on the real price of oil, and – apart from some short periods – US CPI and GDP-deflator inflation were also unaffected by them. Until the early 1980s there

had been some evidence for slight recessionary effects on GDP growth, but later it diminished a lot. In the case of labour market I find the opposite: unexpected fall in production cause bigger spikes in unemployment rate now than it did some decades ago. Probably this is the result of the expanding US shale oil industry.

3. Aggregate demand shocks (unexpected demand growth for commodities) had had significant inflationary effects in the beginning of the sample period, but later the inflationary pressure became moderated. Results for the real economy are mixed. The impact on GDP undulated during the sample period, positive and negative trends followed each other, but during the upward price trend in the 2000s the recessionary pressure was historically weak, which helped avoid stagflation. The US labour market was also resistant against aggregate shocks, and after the millennium the unemployment rate responded by shrinking below its base level for the first two years after the shock had occurred.
4. Specific demand shocks are related to more favourable macroeconomic consequences nowadays than they were some decades ago. Their inflationary pressure diminished over time, while at the end of the sample a specific shock can significantly (and temporarily) reduce the unemployment rate. Similar to this, at these years the shocks also have some GDP acceleration effect for a quarter after occurrence, while formerly the divergence from the base path was insignificant.
5. In total, my results suggest the time-varying nature of oil market shocks: the main macroeconomic aggregates were almost unaffected by the 2000s' high prices because of two reasons. First, aggregate demand shocks caused the rising trend, second, the US economy became more resistant against this type of shocks. That means, both *Blanchard & Galí* [2007] and *Kilian* [2010] were correct: shocks matter, but their varied through time.
6. The results also enable us to analyze the recent oil price movements and their macroeconomic implications. After the Great Recession rising oil prices were rooted in positive specific demand shocks, which helped the recovery by boosting GDP growth and employment. On the other hand, the price collapse – started in mid-2014 – had some short-term (1-2 quarters ahead) recessionary and disinflationary effects.
7. There are some massive changes in the estimated effects and the accumulated impulse response functions during the sample's last 10 years. Partly, they can be the results of the crisis generated noise, but the US shale oil revolution also plays an important role.
8. The main conclusions remain valid under alternative model specifications.

2. Analyzing the oil-macroeconomy relationship in Sweden and Norway with the continuous wavelet transform

2.1 Motivation

The oil price collapse of 2014 caused a great drop in many producers' export and governmental income. Countries without enough financial reserves and diversified economies immediately suffered from falling revenues and deteriorating macroeconomic performance (*Arezki & Blanchard* [2014]). Many countries have long been known for high dependency on their oil sector, e.g. Saudi Arabia (*Aleisa & Dibooglu* [2004]), Venezuela (*Mendoza & Vera* [2010]), Russia (*Benedictow et al.* [2013]). However, there are some positive examples (e.g. Australia, Canada, Norway) of how to diversify the economy, handle oil related revenues and build an institutional environment, which is appropriate in reducing oil dependency and smoothing cyclical fluctuations of the economy (*Bjornland* [2000], *Bjornland & Thorsrud* [2014]). The goal of Chapter III is to detect the strength and time-varying nature of the co-movement between the oil price and some of the most important Norwegian macro aggregates (CPI inflation, growth rate of industrial output and gross domestic product). The analysis is executed by the so-called continuous wavelet transform (CWT) which enables the researcher to investigate the relationship in the time-frequency space. CWT is a useful addition for time series analysis and provides the opportunity to find formerly hidden context. I also estimate the indicators for Norway's neighbour, the oil importer Sweden, which is quite similar to Norway in many viewpoints. I try to find answers for the following research questions:

1. Does the CWT have any relevance in analyzing co-movement between oil price and macroeconomic variables?
2. Does the oil-macroeconomy relationship show a time-varying nature at the level of frequency bands?
3. If the answer is yes, then what can be said about the strength and the direction of the change?
4. Is it enough to analyze prices, or filtering out common factors – like global economic activity – would modify the conclusions? In other words: is it worth working with shocks instead of prices?
5. Was Norway successful in minimizing its dependence on oil?

2.2. Applied Methods

A comprehensive summary of the continuous wavelet transform and other indicators based on it can be found in *Aguiar-Conraira & Soares* [2014], who also developed the MATLAB toolbox for CWTs that I use for the research[5]. Now I focus just on the most important methodological elements that are necessary to understand the results.

The continuous wavelet transform of the x(t) square integrable function[6]:

$$W_x(\tau, s) = \int_{-\infty}^{\infty} x(t)\psi^*_{\tau,s}(t)dt$$

Where $\psi_{\tau,s}(t)$ is the wavelet, while * is the complex conjugate. I also assume $s, \mathcal{T} \in \mathbb{R}, s \neq 0$. The wavelet comes from the so-called mother wavelet ($\psi(\cdot)$):

$$\psi_{\tau,s}(t) = s^{-0.5}\psi\left(\frac{t-\tau}{s}\right)$$

The scale parameter (s) shows the width of the window: low s implies narrow window, which is suitable for detecting high frequency processes, while high s is related to wide windows and low frequency bands[7]. A function is a mother (or analyzing) wavelet if it is a square integrable function ($\psi(\cdot) \in L^2(\mathbb{R})$) and satisfies the *admissibility condition*:

$$0 < \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|}{|\omega|} d\omega < \infty$$

[5] https://sites.google.com/site/aguiarconraria/joanasoares-wavelets/the-astoolbox

[6] In this chapter I work with analytic (or progressive) wavelets, which are complex functions and their Fourier transforms are zero for negative angular frequencies ($\Psi(\omega) = 0 \; for \; \forall\omega < 0$). Because of it the wavelet transform is calculated only for positive scale parameters ($s > 0$).

[7] Unfortunately endpoint problem appears here: the greater the scale parameter (the lower the frequency), the more data points are missing at the end and at the beginning of the sample period. In order to handle this drawback of the method, I follow the solution of *Aguiar-Conraira & Soares* [2014] and set the values of the missing data to zero. The affected areas will be marked on the charts.

Where Ψ is the Fourier transform of ψ. The most popular mother wavelet in economics is the so-called Morlet wavelet (*Goupillaud et al.* [1984])[8]:

$$\psi(t) = \frac{\pi^{-0.25} e^{-0.5t^2}}{\cos(\omega_0 t) - \mathrm{i} \cdot \sin(\omega_0 t)}$$

Where ω_0 controls the number of oscillations. Lower ω_0 results in better time but poorer frequency localization, while higher ω_0 improves the frequency, but worsens the time localization. In the thesis I use $\omega_0=6$ as a default setting, which is a widespread choice in the related literature[9].

To follow the co-movement of the two series not just in time but in different frequency bands, I use the wavelet coherency:

$$R_{x,y}(\tau, s) = \frac{\left|S\left(W_{x,y}(\tau, s)\right)\right|}{\left\{S[|W_x(\tau, s)|^2] S\left[\left|W_y(\tau, s)\right|^2\right]\right\}^{0.5}}$$

Where $R_{x,y}(\tau, s)$ and $W_{x,y}(\tau, s)$ are the wavelet coherency and cross-wavelet transform between the time series x and y. τ determines the time position, s denotes to scale, and *S(.)* is the smoothing function[10]. Definition of the cross-wavelet transform (*Hudgins et al.* [1993]):

$$W_{x,y}(\tau, s) = W_x(\tau, s) W_y^*(\tau, s)$$

The wavelet coherency is always between 0 and 1, where higher values indicate stronger co-movement (in the same or in the opposite direction). It's also possible to compute significance of coherencies[11] and to estimate the lead-lag structure by the phase difference[12]:

---

[8] It's worth mentioning that the Morlet wavelet doesn't satisfy the *admissibility condition*. However, choosing sufficiently large ω_0 makes it appropriate for empirical applications (*Foufoula-Georgiou & Kumar* [1994]).

[9] *Aguiar-Conraira & Soares* [2014] emphasized that in economic analysis they had met only with the Morlet wavelet and $\omega_0 \epsilon [5; 6]$. Choosing $\omega_0=6$ has the advantage of making it easier to interpret the results because the connection between scale and frequency simplifies to $f \approx 1/s$.

[10] Hamming window (*Harris* [1978]).

[11] I tested significance via *Aguiar–Conraira & Soares* [2014].

[12] From the definition one can easily understand the advantage of using complex wavelets. If the imaginary part would be always zero, then we couldn't work with phase and phase difference in the analysis.

$$\phi_{x,y}(\tau,s) = arctan\left(\frac{\Im\left(S\left(W_{x,y}(\tau,s)\right)\right)}{\Re\left(S\left(W_{x,y}(\tau,s)\right)\right)}\right)$$

Where $\Im(\cdot)$ denotes the imaginary, and $\Re(\cdot)$ the real part. The phase difference is between –π and π, and its value shows the type of co-movement (moving in the same or the opposite direction) and the lead-lag structure:

- $\phi_{x,y}(\tau,s) \in \left(-\pi; -\frac{\pi}{2}\right)$: out-of-phase, x leads
- $\phi_{x,y}(\tau,s) \in \left(-\frac{\pi}{2}; 0\right)$: in-phase, y leads
- $\phi_{x,y}(\tau,s) \in \left(0; \frac{\pi}{2}\right)$: in-phase, x leads
- $\phi_{x,y}(\tau,s) \in \left(\frac{\pi}{2}; \pi\right)$: out-of-phase, y leads

Of course high wavelet coherency doesn't mean causality, but it is possible to create indicators similar to partial correlation. I use the so-called partial wavelet coherency and partial phase difference that were introduced and described in *Aguiar–Conraira & Soares* [2014].

In the thesis I use three macroeconomic variables, namely monthly CPI inflation, monthly growth rate of industrial production and quarterly growth rate of GDP, for Sweden and Norway. Under the price of oil I mean the monthly averages of Brent crude, while real price is generated by deflating nominal values with US CPI inflation. I investigate the effect of shocks as well. The shock series are calculated the same way as in Chapter II, but this time I use the price of Brent instead of refinery acquisition costs for imported crude.

2.3. Results

Answers for the research questions:

1. The strength and the direction of co-movement between the price of oil and macroeconomic aggregates are not unified along the different frequency bands, even if we focus on a given point in time. E.g. it sometimes happens that the relationship is positive in one frequency and negative in an other one. Of course, this kind of connection can't be detected with a simple correlation analysis. This observation

suggests the usefulness of CWT techniques as an additional tool next to standard time series analysis.

2. The value and significance of the wavelet coherency varies over time and frequencies. The changing relationship can be the result of smaller dependence on crude oil (*Blanchard & Galí* [2007]) or just the consequence of different types of shocks behind the price fluctuations (*Kilian* [2010]). It seems again that both theories are true at the same time: wavelet coherencies between exogenous shocks and economic variables change over time and frequency and the type of the shock also matters.
3. During the sample period the wavelet coherency between the real price of oil and industrial output/GDP growth became smaller in Sweden, while Norwegian economic performance was more or less insensitive to oil price fluctuations both in the low (3-8 years) and high (1-3 years) frequency bands (see Figure 3). The Swedish GDP growth showed a strong and significant co-movement with the oil price until the mid 1990s in the low (3-8 years) frequency range. The results show the two series were in phase and GDP was leading (the phase difference is between 0 and $\pi/2$). The latter result suggests common factors (namely the global economic activity) played an important role during the period: it's not likely that booming Swedish economy could cause higher oil prices on its own, it's more likely that favourable global economic environment boosted the demand for Swedish export products and crude oil at once. Similar patterns can be found in the case of inflation. The wavelet coherencies in Sweden lost their significance in the 3-8 years frequency bands, although in the 2000s they somewhat strengthened again, but now in the higher (1-3 years) ranges.
4. Partial analysis (i.e. filtering out the effect of global economic activity) highlights the importance of common factors. Partial wavelet coherencies suggest weaker (and less significant) relationship between the price of oil and the macro variables than it seemed whit simple (not partial) wavelet coherencies. It also becomes clear that the co-movement should be analyzed not on the price but on the shock level (see conclusion No. 2).
5. Macroeconomic performance shows weaker connection with the price of oil in Norway than in Sweden where the relationship has become milder over the decades, while in Norway it was detectable only temporarily. On the other hand, if someone looks at the wavelet coherencies between shocks and macro variables she will realize that in the case of the real economy (industrial production and GDP) the co-movement was somewhat

stronger in the oil exporting economy (see Figure 4). Nonetheless, Norway still seems to be an economy highly independent from oil prices or oil market shocks.

Figure 3: Wavelet coherencies and phase differencies between the quarterly growth rates of GDP and the real price of oil in Sweden and Norway[13]

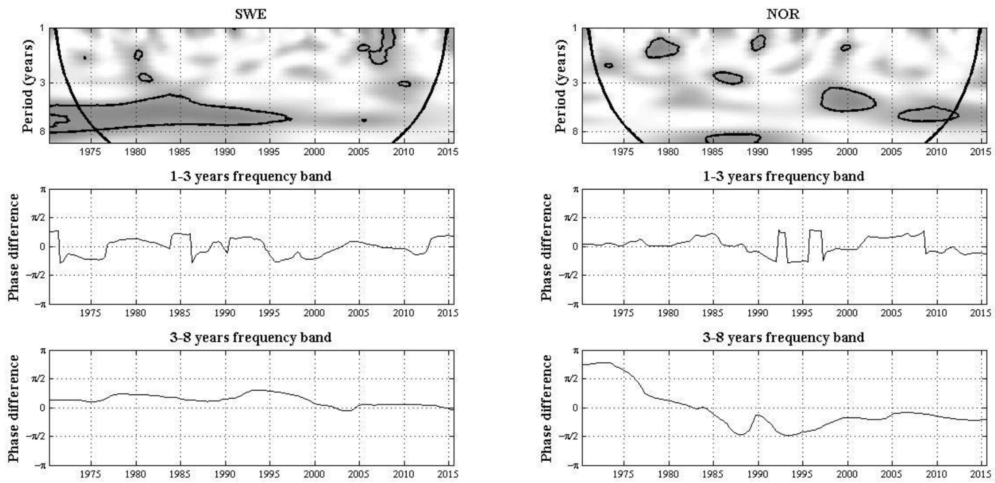
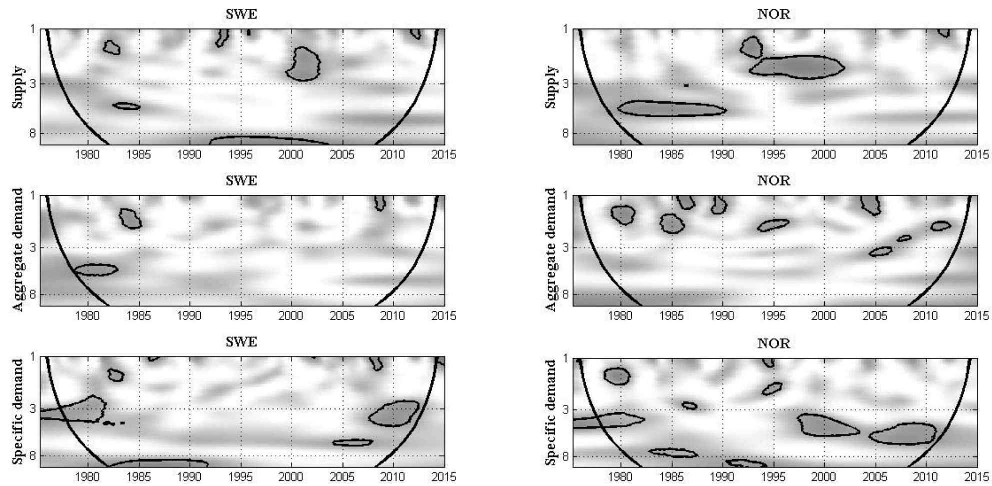


Figure 4: Wavelet coherencies between quarterly GDP growth rates and oil shocks in Sweden and Norway



[13] On the upper panel the horizontal axes show the time, the vertical axes show the frequency (in years). Dark colours denotes to high, light to low wavelet coherency. The cone of influence, which indicates the region affected by edge effects, is shown with a thick black line. The black contour designates the 5% significance level based on an ARMA(1,1) null. The lower panels show the phase differences.

3. Performance stability of short-run oil price forecasts

3.1 Motivation

Due to its role in transportation and macroeconomic performance crude oil market is closely watched by traders, analysts and economists, and there is significant demand for reliable price forecasts both in the short and the long run (i.e. from days to decades). Nevertheless, the wide range of predictions and the never-ending list of positive and negative risks highlight the complex nature of forecasting. *Hamilton* [2009] argues that the real price of oil can't be projected even on a monthly or quarterly basis, while *Baumeister & Kilian* [2013] find evidence against it. Although in the case of monthly/quarterly price forecasts there are examples for comparing dozens of models in one study (*Alquist & Kilian* [2010], *Alquist et al.* [2011]), I didn't find similarly exhausting attempts for short-run (i.e. 1-3 days ahead) projections. Another common problem in the „short-run" literature is that some very promising results were born (e.g. *Haidar & Kulkarni* [2009], *Shabri & Samsudin* [2014], and *Yu et al.* [2008]), which produced high direction accuracy, but not with real time predictions. That makes really hard to compare the different results to each other and their robustness in time: they use only a few models, the test sets are different across articles, and sometimes the predictions indirectly use future information. The aim of Chapter IV is to test the short-run (1-3 days ahead) prediction performance and the robustness of performance on a wide range of models. As I know these models haven't been tested in the same study yet. The research questions:

1. Does the inclusion of oil futures and gasoline/heating oil spot prices improve the prediction accuracy?
2. Which models show the best performance in forecasting the price and/or the direction of the change?
3. What can be said about the performance stability across different forecast horizons? Is it true that the best day-ahead predictors are among the best 2/3 days ahead forecasters?
4. What can be said about the performance stability across different test samples? Is it true that the best predictors for the year 2011 are among the best forecasters in 2012, 2013 and 2014?
5. In total, is it possible to create a framework, which manages to continuously give more accurate forecasts than a simple heuristic model?

As the above questions show I use an inductive approach, I have no fixed theory to test, and the final conclusions are valid only for the group of models and period of time that I apply in the thesis. From these unique observations one can only get conjectures and not general conslusions.

3.2. Applied Methods

Table 1 contains the main information about the methodologies used in Chapter IV. Heuristic models (random walk with and without drift) serve as benchmarks for evaluation, while predictions are executed via time series analysis and machine learning techniques.

Table 1: Prediction models

Method	Type	Prediction variable	Model calibration	References
Random walk	heuristic	price, direction	daily	-
ARIMA	econometric	price, direction	daily/yearly	*Hamilton* [1994]
ARX[14]	econometric	price, direction	daily	*Wang, Jain* [2003]
LPM[15]**/logit**	econometric	direction	daily	*Wooldridge* [2012]
Neural network	machine learning	price, direction	yearly	*Kriesel* [2007]
SVM[16]	machine learning	direction	yearly	*Cristianini, Shawe és Taylor* [2000]

The daily frequency sample starts at the beginning of 2002 and ends November 2014, the test set is 2011-2014. Data are coming from the Energy Information Administration's database. The included variables:

- WTI crude spot price (USD/bbl, 2002.01.02-2014.11.04.)
- New York Harbor Conventional gasoline spot price (USD/gallon, 2002.01.02-2014.11.04.)
- New York Harbor Heating Oil No. 2 spot price (USD/gallon, 2002.01.02-2014.11.04.)

[14] Autoregressive exogenous model.
[15] Linear probability model.
[16] Support vector machine.

- WTI futures prices (Contract 1, 2, 3 and 4, USD/bbl, 2002.01.02-2014.11.04.)

In the case of gasoline and heating oil I use the 1 barrel = 42 gallons relationship to convert the variables into the same unit.

The models have the following structure:

- target variable: transformation of the spot WTI price (dlog(WTI) or its standardized value, or max{0,sign[dlog(WTI)]})
- predictors: autoregressive terms (1-20 day delays), last 1-20 values of dlog predictor prices (or their standardized forms) and/or spreads (e.g. $\log(P_{gasoline})-\log(P_{WTI})$)

I always use nine combinations of the six predictor variables:

- Contract 1, 2, 3 and 4 separately and together (5 cases)
- gasoline and heating oil separately and together (3 cases)
- futures Contract 1, 2, 3, 4 and gasoline/heating oil spot together (1 case)

In total, there are 1604 models (projections) for the 1 day ahead direction forecast, while there are 404 models for everything else (1-3 days ahead price forecast and 2-3 days ahead direction forecast). This is the consequence of the LPM, logit, SVM and neural network models built only for 1 day ahead direction forecasts.

I estimate models and make projections in the following structure:

- the first element of the training set is always 2002.01.02[17]
- the last element of the training set is 2010.12.31 during the first estimation, then it always shifts by one trading day (econometric models excluding ARIMA) or a year (ARIMA and machine learning tools)
- in the case of SVM the start of test sets are 2011.01.03, 2012.01.03, 2013.01.02 and 2014.01.02, while they end at 2011.12.30, 2012.12.31, 2013.12.31 and 2014.10.31[18] respectively
- in the case of neural networks I follow the same structure as in the previous point, but I use the last 12 months of the training set as a validation set

[17] Dates denote trading days.

[18] In fact, these are the beginning of the last forecasts, e.g. a prediction which starts at 2014.10.31 will give projections for 2014.10.31, 2014.11.03 and 2014.1104.

The evaluation happens in yearly resolution from January 2011, so during 2011-2013 I have 252 forecasts for each period, while for 2014 I generate 211 projections. The performance is measured by several indicators, namely RMSE (root mean squared error), MAE (mean absolute error), DA/HT (direction accuracy/hit rate), and a profit indicator for measuring the effectiveness of trading. The latter measures accumulated profits when decisions are made based on 1 day ahead direction forecasts:

$$PROFIT_t = I(\widehat{p_t} > p_{t-1}) \cdot \left[(1 + PROFIT_{t-1}) \cdot \frac{p_t}{p_{t-1}} - 1\right] + [1 - I(\widehat{p_t} > p_{t-1})] \cdot PROFIT_{t-1}$$

I use the following benchmarks for evaluation:

- RMSE: random walk without drift
- MAE: random walk with drift
- DA/HT: random walk with drift, which always indicates growing prices, but in 2014 I use the stricter 50% limit to avoid the below 50% benchmark
- profit: „*buy and hold”* strategy, but for 2012 and 2014 I use the stricter limit of 0% in order to avoid negative benchmark values

Unfortunately the possible number of combinations of model types, variables and lag structure is so large that there is no chance for testing all of them. However, LASSO (*Least Absolute Selection and Shrinkage Operator*) and LASSOGLM methods with 10-fold cross-validation are fine for preliminary input variable selection[19]. After estimating the formerly described models and evaluating their performance, I also make an experiment to detect whether the preliminary variable selection ensures the robustness of prediction accuracy. For forecasting prices I use OLS and neural network (NN), for forecasting directions I make projections via LPM, logit, SVM and neural network models. Parameters of NN models are refreshed once a year, SVM is calibrated on a daily and yearly frequency as well, while all other models are reestimated every day.

[19] For details of the LASSO methodology see *Tibshirani* [1996].

3.3. Results

Chapter IV ends with the following conclusions:

1. The inclusion of WTI futures and gasoline/heating oil spot prices are highly recommended, regardless of the test set, the forecast horizon and the evaluation indicator. When we focus solely on RMSE, then futures spreads are recommended, while in other cases both the futures and the spot prices can improve prediction accuracy.
2. Econometric models provided more robust performance than machine learning techniques (i.e. in the group of „better than the benchmark" models econometric methods were overrepresented and remained in that group with greater – although still low – probability in the case of other test sets/forecast horizons). Nonetheless, one must be careful with comparing them. On the one hand, the parameters of machine learning models were calibrated just once a year, and on the other hand parameters of the learning algorithms' were not optimized, they ran next to „widely used" settings, which can be suboptimal in the current situation. However, SVM generated the highest profits during the whole sample (2011-2014) and the median performance of machine learning models were not worse than econometric ones' (see Table 2).
3. Results show the group of „better than the benchmark" models varies through test sets and forecast horizons. There are only a few (sometimes zero) models which can beat the benchmark every year. Considering the not too long test interval (2011-2014), it can easily happen that these methods wouldn't be able to outperform the simple benchmark predictions on a longer test set. It also has a high probability because there is no model which can significantly outperform the benchmark in each year.
4. From the former point it also becomes clear that although it seems possible to create a model with better forecast ability than a simple heuristic rule has, its robustness will always be questionable and there is no guarantee that a model, which could outperform the benchmark for a long time, will be as good in the coming months or years.
5. Results in Table 3 suggest preliminary variable selection via LASSO doesn't help reaching better point predictions. Regardless of the model (OLS or NN) price forecasts are inaccurate, sometimes significantly poorer than the benchmark. In the case of direction accuracy it is recommended using LASSOGLM instead of simple LASSO[20]. Apart from the SVM it is true for almost every model/forecast horizon pairs that the hit

[20] LASSOGLM uses the (0/1) direction of the change in price and not the price or its *dlog* transformation.

rate is bigger with LASSOGLM than with LASSO. Even if it doesn't exist, the difference remains marginal. In most of the cases the methods were able to outperform the „*buy and hold*" strategy's -11,9% profit indicator, but they managed to go sensibly above zero only with LASSOGLM. Primarily the good 1 day ahead forecasters, namely the LPM, logit, and neural network models did a good job with a rate of 33-37%.

6. However, robustness through different test sets is not provided by the LASSO method: none of the models generated forecasts that were significantly better than the ones created by simple heuristic rules. Therefore, using the LASSO doesn't change the main conclusions.

Table 2: Machine learning models' relative performance against econometric (time series) tools

		2011	**2012**	**2013**	**2014**
RMSE	*1-day*	***99,66%***	***99,14%***	***99,61%***	100,69%
	2-day	100,08%	100,64%	***99,83%***	100,10%
	3-day	100,01%	100,20%	100,12%	***99,41%***
MAE	*1-day*	***99,75%***	***99,01%***	***99,39%***	100,84%
	2-day	100,45%	100,30%	***99,89%***	100,88%
	3-day	100,06%	100,13%	100,27%	***99,39%***
DA	*1-day*	99,22%	***100,78%***	95,56%	95,68%
	2-day	94,43%	98,12%	***102,11%***	99,04%
	3-day	96,37%	***103,41%***	***102,46%***	***104,00%***
profit		***0,71%***	-0,19%	-4,99%	-4,84%

The table shows the ratio of median values of the error indicators in the case of RMSE, MAE and DA, and the difference between the medians in the case of profit. Cases are highlighted if machine learning models provide better forecasts.

Table 3: Forecast accuracy and profit indicator on the whole sample (2011-2014)

		RMSE			MAE			DA			PROFIT
		D1	**D2**	**D3**	**D1**	**D2**	**D3**	**D1**	**D2**	**D3**	
LASSO	**OLS**	1,55	2,18	2,65	1,15***	1,62***	2,00***	49,3%	50,6%	50,6%	-15,7%
	NN	1,56	2,18	2,66	1,16***	1,61***	1,99**	50,9%	49,9%	51,8%	**-6,7%**
	LPM							51,6%	49,2%	**52,4%**	**-1,5%**
	logit							51,3%	49,9%	**53,1%**	**-11,2%**
	NN_binary							51,7%	50,3%	**52,6%**	**0,3%**
	SVM_Y							49,7%	50,1%	52,0%	-31,1%
	SVM_D							49,1%	49,9%	51,6%	-19,4%
LASSOGLM	**LPM**							**53,6%***	51,1%	**52,5%**	**34,1%**
	logit							**53,7%***	50,4%	52,0%	**36,7%**
	NN_binary							**54,4%****	50,5%	**52,5%**	**33,0%**
	SVM_Y							49,8%	48,1%	47,4%*	**7,9%**
	SVM_D							49,6%	46,8%**	49,5%	-12,8%

Worse than the benchmark forecasts (where the related test's null hypothesis can be rejected) are highlighted with dark gray, while light gray cells denote better than the benchmark predictions. In this table I don't use 0% (for profit) and 50% (for DA/HT) constraints of the benchmarks. Asterisks denote p-values (* p<0.1, ** p<0.05, *** p<0.01). The applied tests: RMSE – *Clark & West* [2006], MAE – *Diebold & Mariano* [1995], DA – *Pesaran & Timmermann* [1992]. Abbreviations in the table: OLS (ordinary least squares), NN (neural network), LPM (linear probability model), logit (logistic regression), NN_binary (neural network with binary – 0/1 – target variable), SVM_Y (yearly calibrated support vector machine), SVM_D (daily calibrated support vector machine), DX (X days ahead forecast).

References

Aguiar-Conraira, L. - Soares, M. J. [2014]: *The Continuous Wavelet Transform: moving beyond uni- and bivariate analysis*. Journal of Economic Surveys. 28(2), pp. 344-375. DOI: http://dx.doi.org/10.1111/joes.12012

Aleisa, E. - Dibooglu, S. [2004]: *Oil Prices, Terms of Trade Shocks, and Macroeconomic Fluctuations in Saudi Arabia*. Contemporary Economic Policy. 22(1), pp. 50-62. DOI: http://dx.doi.org/10.1093/cep/byh005

Alquist, R. - Kilian, L. [2010]: *What do we learn from the price of crude oil futures?* Journal of Applied Econometrics. 25(4), pp. 539-573. DOI: http://dx.doi.org/10.1002/jae.1159

Alquist, R. - Kilian, L. - Vigfusson, R. J. [2011]: *Forecasting the Price of Oil*. C.E.P.R. Discussion Paper. No. 8388. DOI: http://dx.doi.org/10.2139/ssrn.1911194

Arezki, R. - Blanchard, O. [2014]: *Seven Questions about the Recent Oil Price Slump*. IMFdirect - The IMF Blog, December 22, 2014.

Baumeister, C. - Kilian, L. [2013]: *Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach*. Bank of Canada Working Paper. No. 2013-28. Ottawa.

Baumeister, C. - Kilian, L. - Zhou, X. [2013]: *Are Product Spreads Useful for Forecasting? An Empirical Evaluation of the Verleger Hypothesis*. Bank of Canada Working Paper. No. 2013-25. Ottawa.

Benedictow, A. - Fjærtoft, D. - Løfsnæs, O. [2013]: *Oil dependency of the Russian economy: an econometric analysis*. Economic Modelling. 32(1), pp. 400-428. DOI: http://dx.doi.org/10.1016/j.econmod.2013.02.016

Bjornland, H. C. [2000]: *The dynamic effects of aggregate demand, supply and oil price shocks - A comparative study*. The Manchester School. 68(5), pp. 578-607. DOI: http://dx.doi.org/10.1111/1467-9957.00220

Bjornland, H. C. - Thorsrud, L. A. [2014]: *What is the effect of an oil price decrease on the Norwegian economy?* Working Paper. http://home.bi.no/a0310125/BT_OilPrice_2014.pdf

Blanchard, O. J. - Galí, J. [2007]: *The Macroeconomic Effects of Oil Price Shocks: Why are the 2000s so different from the 1970s?* NBER Chapters, in: International Dimensions of Monetary Policy, pp. 373-421. DOI: http://dx.doi.org/10.3386/w13368

Clark, T.E. - West, K. D. [2006]: *Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis*. Journal of Econometrics. 135(1-2), pp. 155–186. DOI: http://dx.doi.org/10.1016/j.jeconom.2005.07.014

Cristianini, N. - Shawe-Taylor, J. [2000]: *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. First Edition*. Cambridge University Press. Cambridge. DOI: http://dx.doi.org/10.1017/cbo9780511801389

Diebold, F. X. - Mariano, R. S. [1995]: *Comparing predictive accuracy*. Journal of Business and Economic Statistics. 13(3), pp. 253–263. DOI: http://dx.doi.org/10.1080/07350015.1995.10524599

Foufoula-Georgiou, E. - Kumar, P. [1994]: *Wavelet Analysis in Geophysics: An Introduction*. Wavelet Analysis and Its Applications. 4, pp. 1-43. DOI: http://dx.doi.org/10.1016/b978-0-08-052087-2.50007-4

Goncalves, S. - Kilian, L. [2004]: *Bootstrapping autoregressions with conditional heteroskedasticity of unknown form*. Journal of Econometrics, Elsevier. 123(1), pp. 89-120. DOI: http://dx.doi.org/10.1016/j.jeconom.2003.10.030

Goupillaud, P. - Grossman, A. - Morlet, J. [1984]: *Cycle-octave and related transforms in seismic signal analysis*. Geoexploration. 23(1), pp. 85-102. DOI: http://dx.doi.org/10.1016/0016-7142(84)90025-5

Haidar, I. - Kulkarni, S. [2009]: *Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices*. International Journal of Computer Science and Information Security. 2(1), pp. 81-88.

Hamilton, J. D. [1994]: *Time series analysis*. Princeton University Press. Princeton.

Hamilton, J. D. [2009b]: *Understanding Crude Oil Prices*. The Energy Journal. 30(2), pp. 179-206. DOI: http://dx.doi.org/10.5547/issn0195-6574-ej-vol30-no2-9

Harris, F. J. [1978]: *On the use of windows for harmonic analysis with the discrete Fourier transform*. Proceedings of the IEEE. 66(1), pp. 51-83. DOI: http://dx.doi.org/10.1109/proc.1978.10837

Hudgins, L. - Friehe, C. - Mayer, M. [1993]: *Wavelet transforms and atmospheric turbulence*. Physics Review Letters. 71(20), pp. 3279-3282. DOI: http://dx.doi.org/10.1103/physrevlett.71.3279

Kilian, L. [2009]: *Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market*. American Economic Review. 99(3), pp. 1053-1069. DOI: http://dx.doi.org/10.1257/aer.99.3.1053

Kilian, L. [2010]: *Oil Price Shocks, Monetary Policy and Stagflation*. RBA Annual Conference Volume, in: Renée Fry & Callum Jones & Christopher Kent (ed.), Inflation in an Era of Relative Price Shocks Reserve Bank of Australia.

Kilian, L. [2014]: *Oil Price Shocks: Causes and Consequences*. Annual Review of Resource Economics. 6(1), pp. 133–154. DOI: http://dx.doi.org/10.1146/annurev-resource-083013-114701

Kriesel, D. [2007]: *A Brief Introduction to Neural Networks*. http://www.dkriesel.com

Mendoza, O. - Vera, D. [2010]: *The Asymmetric Effects of Oil Shocks on an Oil-exporting Economy*. Latin American Journal of Economics. 47(135), pp. 3-13. DOI: http://dx.doi.org/10.4067/s0717-68212010000100001

Pesaran, M. H. - Timmermann, A. [1992]: *A simple nonparametric test of predictive performance*. Journal of Business and Economic Statistics. 10(4), pp. 461-465. DOI: http://dx.doi.org/10.2307/1391822

Shabri, A. - Samsudin, R. [2014]: *Daily Crude Oil Price Forecasting Using Hybridizing Wavelet and Artificial Neural Network Model*. Mathematical Problems in Engineering. DOI: http://dx.doi.org/10.1155/2014/201402

Tibshirani, R. [1996]: *Regression Shrinkage and Selection via the Lasso*. Journal of the American Statistical Association. Series B. 58(1), pp. 267-288.

Wang, G. C. S. - Jain, C. L. [2003]: *Regression analysis – Modeling and Forecasting*. Graceway Publishing Company. Great Neck.

Wooldridge, J. M. [2012]: *Introductory Econometrics: A Modern Approach. Fifth Edition*. South-Western Cengage Learning. Mason.

Yu, L. - Wang, S.Y. - Lai, K.K. [2008]: *Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm*. Energy Economics. 30(5), pp. 2623–2635. DOI: http://dx.doi.org/10.1016/j.eneco.2008.05.003

List of publications

Peer Reviewed Journal Publications

Uliha, G. [2012a]: *Az olajár makrogazdasági hatásai: mitől lett egyre gyengébb?* Külgazdaság. 56(5-6), pp. 78-92.

Uliha G. [2015]: *Rövid távú olajár-előrejelzések teljesítményének stabilitása.* Statisztikai Szemle. 93(3), pp. 189-224.

Uliha, G. [2016a]: *Az olajár és a makrogazdaság kapcsolatának elemzése folytonos wavelet transzformáció segítségével.* Forthcoming in Statisztikai Szemle. 94(5)

Working Papers

Uliha, G. [2012b]: *Az olajár gyengülő makrogazdasági hatásai: két versengő elmélet szintézise.* VIII. KDI konferencia

Uliha G. [2014]: *Olajár előrejelzések összehasonlítása.* X. KDI konferencia

Work in Progress

Uliha, G. [2016b]: *Az olajár gyengülő makrogazdasági hatásai: két versengő elmélet szintézise.*