U.S. SHALE OIL PRODUCTION AND WTI PRICES BEHAVIOUR

Manuel Monge, Luis A. Gil-Alana and Fernando Pérez de Gracia

University of Navarra, Pamplona, Spain

January 2017

The aim of this paper is to relate the shale oil revolution in the United States with WTI oil price behavior. Since the development of the combination of horizontal drilling techniques together with hydraulic fracturing in the 1970s, known as shale oil, oil markets have undergone a significant transformation with the unexpectedly strong rise in the United States production affecting oil prices. The goal of this paper is two-fold: first, we analyze the relationship of total United States crude oil production and WTI crude oil prices by studying its performance in the time-frequency domain applying wavelet tools for its resolution. Using wavelet methodologies, we observe a shift to higher frequencies of the wavelet coherency for the time period 2003-2009 and lower frequencies for the period 2009-2014. The results also indicate that during the period 2003-2009 the U.S. oil production and WTI oil prices time series are in phase; they move together, with total United States oil production leading. During the period 2009-2014 oil production and WTI oil prices time series are out of phase (negatively correlated), suggesting that oil production increases precede a decrease in WTI oil prices. In the second part of the paper and to give greater credibility to the results obtained through the wavelet transform, we analyze the behavior of WTI crude oil before and after the shale oil boom in the United States employing methodologies based on long run dependence. The results indicate that mean reversion takes place only for the data corresponding to the first subsample, ending at 2003. For the second subsample, as well as for the whole sample, lack of mean reversion is detected with orders of integration equal to or higher than 1 in all cases.

Keywords: Oil prices, oil production; wavelets; fractional integration

JEL Classification: C00, C22, E30, Q40, Q43

Correspondence author: Manuel Monge
University of Navarra
Faculty of Economics
E-31080 Pamplona
SPAIN

Phone: 00 34 948 425 625
Fax: 00 34 948 425 626
Email: mmonge.3@alumni.unav.es

1. Introduction

The production of shale oil[1] consists in horizontal drilling and the hydraulic fracturing of underground rock formations containing deposits of crude oil that are trapped within the rock. This process is used to extract crude oil that would be impossible to release through conventional drilling methods.

The boom production of shale oil in United States is an example of a technological change in a single industry in one country affecting international trade worldwide.

The evolution of United States oil production was in decline until the early 1970s. This trend was only briefly reversed by the development of the Alaskan oil fields in the late 1970s. According to Kilian (2016a), the expansion of United States shale oil takes place between 2003 and 2013 stimulated by high conventional crude oil prices resulting in this technology becoming competitive. According to Clements and Cummings (2016) the global oil market witnessed a period of stable and high prices with crude averaging over USD 110bbl between 2011 and mid-2014. During this period, there were geopolitical turbulences in Libya, Iraq and Iran that led to reduced exports, creating a market supply shortage. United States tight oil reduced the gap thanks to technological advances in horizontal funding techniques which were funded by high oil prices and low financing costs through credits. November 2008 was the reversal date of this trend in the United States. It can be shown that this reversal is largely due to U.S. fracking. Figure 1 plots the crude oil production at the Bakken, Eagle Ford, Haynesville, Marcellus, Niobrara, Permian and Uttica shale oil plays. According to the Energy Information Administration report (May 2016), the seven regions previously mentioned account for 92% of domestic shale oil production growth in United States.

[1] According to Mănescu and Nuño (2015) the definition provided by the Energy Information Administration (EIA) and the International Energy Agency (IEA) is that shale oil is the “tight oil” or “light, tight oil”. The term tight oil does not have a specific technical, scientific, or geologic definition. Tight oil refers to oil produced from very low-permeability shale, sandstone, and carbonate formations.

[Insert Figure 1 about here]

Figure 2 plots the crude oil production at the Bakken, Eagle Ford, Haynesville, Marcellus, Niobrara, Permian and Uttica shale oil and the WTI crude oil price behaviour during the period 2007-2016.

[Insert Figure 2 about here]

Following Kilian (2016a), the production of shale oil increased exponentially until growth became linear. In March 2014 the United States economy produced on average 8.2 millions of barrels/day (mbd) and imported 7.3 mbd. Of the total 15.5 mbd of crude oil, 3.6 mbd were produced from shale oil, accounting for half of the United States oil production and only a quarter of the total quantity of oil has been used by the United States economy.

Kilian (2016b) argues that the increase of shale oil production in the United States has displaced the Arab oil producing countries and their crude oil exports. This fact has occurred because United States refineries have increasingly exported refined products such as gasoline or diesel made from domestically produced crude oil.

The high oil prices in recent years made shale oil exploration economically viable. However, there were several factors which ensured that the shale oil revolution took place in the United States and not elsewhere. According to Alquist and Guenétte, (2014) there was a history of shale gas exploitation, legal incentives for land owners and an advanced oil production infrastructure. In the early 19th century, the technology necessary for shale oil exploration started to be developed and was perfected during the 1980s. Also, the United States has legal and institutional features that make the economic environment attractive for the extraction of unconventional oil, and have infrastructures that consist of a large number of state-of-the-art drilling rigs, an extensive pipeline network, and associated refineries.

According to Covert (2014) and related to the flow of shale oil, there is evidence of significant productivity gains in fracking that could further lower the cost of recovering shale oil and raise future estimates of recoverable shale oil.

From a different perspective, Sharenow and Worah (2013) showed that shale oil could provide a rebalancing global supply. The United States shale and eventually shale production globally, combined with production from Canada's oil sands, could potentially increase energy balances for the first time since the oil spikes of the 1970s, leading to the development of new reserve basins. Killian (2016a) argues that this has implications for the price of oil because the United States oil industry has been able to blend heavy crudes and shale oil in the right proportion to imitate mid-grade crude oil of type traditionally imported and refined along the Gulf Coast.

[Insert Figure 3 about here]

Figure 3 illustrates the total United States oil production and the behaviour of WTI crude oil prices.

It is common in the literature to utilize Fourier analysis to analyse the different relations at different frequencies, omitting the time information, despite it being difficult to identify structural changes with this type of analysis. In this paper we analyse the relationship of total United States oil production and WTI crude oil prices by studying its dynamic in the time-frequency domain through the application of wavelet tools for its resolution. To complement this and to give further credibility to the results obtained by the wavelet transform, we analyse the behaviour of WTI crude oil before and after the shale oil boom in the United States, employing methodologies based on long run dependence or long memory using the date break cited by the literature and by our wavelets results. In addition, we employ the methodology suggested in Gil-Alana (2008) to estimate breaks in the context of fractional integration.

Under the hypothesis that shale oil production affects WTI oil prices and according to Kilian (2016a) the evolution of the United States oil price is determined by the increases in shale oil production. The development of the United States refining, pipeline and rail infrastructure are important to understand and forecast the evolution of the domestic price of oil in the United States. There are others factors that Baumeister and Kilian (2016) mentioned such as oil supply shocks, demand shocks and shocks to oil price expectations. In this research we use the variables related with crude oil production in United States and WTI crude oil prices because we want to study how the overproduction related with the shale oil affects WTI behaviour.

The contributions of the paper are twofold. First, we use a methodology based on a time-frequency technique that it is able to analyse the evolution of the different frequency components of the time series over time. Applying the wavelet transform it is possible to detect the evolution in time of the low frequency. This low frequency is related with the trend or long run component in the time series. Also, applying the wavelet transform we can detect the evolution over time of the high frequency components related to seasonality or the short run component, as well as the rapid changes in the time series[2]. We use wavelets to analyse the relationship between WTI oil prices and total oil production in the U.S. for the time period 2000-2016. Following Aguiar-Conraria and Soares (2011a,b), two tools are used to analyse the impact of crude oil production on the crude oil prices: the wavelet coherency and the wavelet phase-difference. The wavelet coherence is a localized correlation coefficient in the time-frequency space. The information on the delay between the oscillations of two time-series is the phase-difference. These concepts developed by Aguiar Conraria were previously examined in Naccache (2011), analysing the relationship between oil price

[2] Hogan and Lakey (2005) examined the relationship between time-frequency and time-scale (wavelets) methods.

and the economy. The analysis is performed in the time-frequency domain, using wavelet analysis. Following these authors, we use this methodology for three reasons. First, stationarity is not required in the wavelet analysis (in our case, oil prices are non-stationary). On the other hand, we can study how relations evolve between time and frequencies. And the last reason is related with the energy markets and the research by Kyrtsos et al. (2009). They argue that several energy markets display consistent non-linear dependencies. Second, in this paper we use some recently developed methods based on the concepts of long run dependence and long memory using fractional integration techniques (Gil-Alana and Hualde, 2009). The methodology used in the second part of the research is similar to the one applied in Monge et al. (2016). Fractional integration is more general than the standard methods that use exclusively integer orders of differentiation (i.e., AR(I)MA).

The rest of the paper is structured as follows. Section 2 reviews the oil supply and the implications on the WTI crude oil price behaviour in the U.S.. Section 3 presents the methodology applied in the paper. In Section 4 we discuss the main empirical results, while Section 5 concludes.

2. Is the shale oil supply behind the behaviour of WTI crude oil prices?

In the last decade, global oil markets have enjoyed a greater supply due to non-conventional sources of oil production in the United States and the Canadian Oil Sands combined with the production of biofuels.

According to the Bank of Canada (2015), if shale oil production were one-third of current oil production and the expected increase in global shale oil production were able to reach two-thirds of oil production, this increase could become uneconomical. In line with Benes et al. (2015), the cost of non-conventional oil production is likely to decline as new technologies will reduce the cost exploration and extraction.

Baumeister and Kilian (2016) basing their study on an alternative methodology, were unable to pin down the quantitative role of fracking. In their research they provide a strong evidence that a slowdown in the global demand for oil was a major contributor to this specific oil price decline, along with shocks to global oil production and oil price expectations.

Baffes et al. (2015) mentioned that the "unconventional" U.S. oil production differs from conventional ones, because they have a shorter life-cycle, around 2.5 - 3 years from the start until full extraction. Krane and Agerton (2015) and McCracken (2015) argue that oil supply from these sources tends to be more elastic to price changes than from conventional sources, even in the short term.

Another important factor mentioned by Baffes et al. (2015) is that OPEC does not have a legal clause on how to intervene when market conditions warrant, thus, allowing it to respond flexibly to changing circumstances. Also, these authors mentioned that changes in supply are due to the expansion of oil production in the United States, causing concerns regarding about supply disruptions to almost disappear. Furthermore, OPEC's policy has played a dominant role in explaining how the recent plunge in prices has been due to supply shocks.

To give consistency to this explanation, Arezki and Blanchard (2015) documented that demand related factors only contribute to 20-30 percent of the decline and the supply related factors, and OPEC's decision not to cut supplies were more important in driving the fall in oil prices. Also, Hamilton (2014) explains that only two-fifths of the decline in oil prices in the second half of 2014 was due to weak global demand. Baumeister and Kilian (2016) conclude in their research report that more than half of the oil price decline reflects the cumulative effects of earlier oil supply and demand shocks and, among the remaining half, the most influential shock was associated with the weakening global economy while positive oil supply shocks were limited between June and December 2014.

3. Methodology

3.1 Wavelet analysis

The wavelet transform offers localized frequency decomposition, providing information about frequency components. Wavelets have significant advantages over basic Fourier analysis when the series under study is stationary – see Gençay et al., (2002), Percival and Walden (2000) and Ramsey (2002). In our research, we use continuous wavelet analysis tools, mainly wavelet coherence, measuring the degree of local correlation between two-time series in the time-frequency domain, and the wavelet coherence phase differences.

3.1.1 The continuous wavelet transform

The continuous wavelet transform of a time series $x(t)$, with respect to the wavelet ψ, is a function $WT_x(a, \tau)$ defined as:

$$WT_x(a,\tau) = \int_{-\infty}^{+\infty} x(t)\psi_{a,\tau}^*(t)dt, \qquad (1)$$

where $WT_x(a,\tau)$ are the wavelet coefficients of $x(t)$ at a certain scale a and a shift τ, where,

$$\psi_{a,\tau}^* = \frac{1}{\sqrt{a}}\psi^*\left(\frac{t-\tau}{a}\right) \qquad (2)$$

is the complex conjugate of the wavelet function ψ. The parameter a is a scaling factor that controls the stretching factor of the wavelet and τ is a location parameter in time. Then, $WT_x(a,\tau)$ will be a matrix of time series. The scaling factor a is a positive real number that simply means stretching it (if $a > 1$), or compressing it (if $a < 1$). If a is positive, we assume that we are using an analytic or progressive wavelet, i.e., its Fourier transform is defined by the positive frequency axis, $\Psi(\omega) = 0$ when $\omega < 0$.

The lower the value of the scaling factor, the more higher frequency components are reflected in the continuous wavelet transform, thus we are dealing with the short-run components of the signal. As the scaling factor increases, we are dealing with lower frequency components of the time series, focussing on the long-run components. Then, the continuous wavelet transform is a multidimensional transform; from one-time series we obtain a matrix of time series that show different frequency components (depending on the scaling factor) of the original one.

If the wavelet function ψ is complex, then the wavelet transform $WT_x(a, \tau)$ will also be complex, with amplitude, $|WT_x(a, \tau)|$, and phase, $\phi_x(a, \tau)$. The real part of the wavelet transform, $\Re\mathrm{e}\{WT_x\}$, and its imaginary part, $\Im\mathrm{m}\{WT_x\}$ define the phase or phase-angle of the wavelet transform:

$$\phi_x = \mathrm{Arctan}\left(\frac{\Im\mathrm{m}\{WT_x\}}{\Re\mathrm{e}\{WT_x\}}\right). \tag{3}$$

The phase of a given time-series $x(t)$ is measured in radians, ranging from $-\pi/2$ to $+\pi/2$. Then, the phase is also a matrix containing the angle of each frequency component of the original time series. The phase will be used to extract conclusions of the synchronism between two time series, applying the wavelet coherency and the phase difference between time series (Aguiar-Conraria and Soares, 2011a,b and 2014).

The wavelet or mother wavelet used to analyze the time series must satisfy certain technical conditions to provide effective time-frequency location properties (Daubechies, 1992). First, it has to be a function of finite energy, $\int_{-\infty}^{+\infty} \psi(t)dt = 0$. There are many different wavelet families, but the election of a certain wavelet will depend on the application itself.

Related to time localization properties, we can normalize the wavelet function so that $\int_{-\infty}^{+\infty} |\psi(t)|^2 dt = 1$. $|\psi(t)|^2$ defines a probability density function, and therefore we can obtain the mean, μ_ψ, and the standard deviation, σ_ψ, of this distribution. They are called the center and the radius of the wavelet, respectively. If we consider the Fourier transform of the mother wavelet, $\Psi(\omega)$, in a similar way we can calculate its mean and standard deviation, μ_Ψ and σ_Ψ.

These quantities define the Heisenberg box in the time-frequency plane: $[\mu_\psi - \sigma_\psi, \mu_\psi + \sigma_\psi] \times [\mu_\Psi - \sigma_\Psi, \mu_\Psi - \sigma_\Psi]$. We say that ψ is localized around the point (μ_ψ, μ_Ψ) of the time–frequency plane with an uncertainty given by $\sigma_\psi \sigma_\Psi$. In our context, the Heisenberg's uncertainty principle establishes that $\sigma_\psi \sigma_\Psi \geq 1/2$.

The Morlet wavelet,

$$\psi(t) = \pi^{-\frac{1}{4}} e^{i\omega_0 t} e^{-t^2/2} \tag{4}$$

is a complex valued wavelet, so we will be able to measure the synchronism between two-time series. This wavelet has optimal time–frequency concentration, in the sense that $\sigma_\psi \sigma_\Psi = 1/2$. Therefore, using this wavelet, we have the optimum trade off between time and frequency resolution. On the other hand, the Morlet can be considered as a wavelet (with finite energy, defined as before) when the frequency parameter ω_0 = 6. For this value of the Morlet wavelet, the wavelet scale, a, satisfies the inverse relation f ≈ $1/a$, as the rest of the most used mother wavelets.

3.1.2 Wavelet and cross wavelet power spectrum, and wavelet coherency

The wavelet power spectrum (WPS) or the scalogram of a time series $x(t)$, as it is called, is the squared amplitude of the wavelet transform, that is: $WPS_x(a,\tau) = |WT_x(a,\tau)|^2$. The wavelet power spectrum lets us know the distribution of the energy (spectral density) of a time-series across the two-dimensional time–frequency representation.

While the wavelet power spectrum shows the variance of a time-series in the time-frequency plane, the cross wavelet power spectrum (CWPS) of two time-series $x(t)$ and $y(t)$ shows the covariance between these time series in the time-frequency plane:

$$CWPS_{xy}(a,\tau) = \left|WT_x(a,\tau)WT_y(a,\tau)^*\right|, \tag{5}$$

where * represents the complex conjugate, as before.

Therefore, the complex wavelet coherency between two time series $x(t)$ and $y(t)$ is defined as the ratio of the cross-spectrum and the product of the power spectrum of both series:

$$WCO_{xy} = \frac{SO\left(WT_x(a,\tau)WT_y(a,\tau)^*\right)}{\sqrt{SO(|WT_x(a,\tau)|^2)SO\left(\left|WT_y(a,\tau)\right|^2\right)}}, \tag{6}$$

where *SO* is a smoothing operator in both time and scale. Without the smoothing operator, the wavelet coherency would be always one for all times and scales (see Aguiar-Conraria et al. (2008) for details).

As the WCO_{xy} is a matrix of complex time series, we can split it again into amplitude and phase, $WCO_{xy} = \left|WCO_{xy}\right|e^{i\phi_{xy}}$. The amplitude matrix is the wavelet coherency, WC_{xy} and the angle ϕ_{xy} is called the phase difference between both time series:

$$\phi_{xy} = \text{Arctan}\left(\frac{\Im\text{m}\{WCO_{xy}\}}{\Re\text{e}\{WCO_{xy}\}}\right), \tag{7}$$

ϕ_{xy} is the phase difference between time series $x(t)$ and $y(t)$, and tells us about the synchronism between those time series. ϕ_{xy} ranging from $-\pi$ to π.

On the one hand, if $\phi_{xy} = 0$ then both time series move in phase. This will mean that both time series increase or decrease their values at the same time. If

$\phi_{xy} \epsilon \left(-\frac{\pi}{2}, 0\right)$, they move in phase but the time series $x(t)$ is leading; if $\phi_{xy} \epsilon \left(0, \frac{\pi}{2}\right)$, the time series $y(t)$ is leading. Therefore, in these cases we can find that one time series anticipates the increase or decrease of the other one. On the other hand, a phase difference of π or $-\pi$ indicates an anti-phase relation, when one time series increases, the other one is decreasing in time. Finally, if $\phi_{xy} \epsilon \left(-\frac{\pi}{2}, -\pi\right)$, both time series are out of phase but $x(t)$ is leading; if $\phi_{xy} \epsilon \left(\frac{\pi}{2}, \pi\right)$, $y(t)$ is leading. In this case this means that one time series has a time delay with respect to the other.

3.1.3 Significance tests, Monte Carlo simulations

To check the statistical significance of the wavelet coherency, WC_{xy}, we rely on Monte Carlo simulations (Schreiber and Schmitz, 1996). We model each time series as an ARMA (p, q) process where $p = q = 1$, with no pre-conditions. Then we assess the statistical significance of the amplitude, not of the phase. The phase difference is not tested as there is no agreement in the scientific community about how to define the procedure. We should only take into account the phase difference when the amplitude of the wavelet coherency is statistically significant.

3.2 Fractional integration

We will also use techniques based on the concept of fractional integration, which means that number of differences required to render a series I(0) stationary may be a fractional value rather than an integer. A given time series *X(t), t = 1,2…* is said to follow an integrated of order *d* process (and denoted as *X(t)* ≈ *I(d)*) if

$$(1 - L)^d X(t) = U(t), \quad t = 1, 2, \ldots, \tag{8}$$

where d can be any real value, L is the lag-operator ($LX(t) = X(t-1)$) and $U(t)$ is I(0), defined for our purposes as a covariance stationary process with a spectral density function that is positive and finite at the zero frequency. Thus, $U(t)$ may display some type of time dependence of the weak form, i.e., the type of an AutoRegressive Moving Average (ARMA) form such that, for example, if $U(t)$ is ARMA (p, q), $X(t)$ is said to be ARFIMA (p, d, q).

Based on the specification in (8) different features can be observed depending on the value of d. Thus, if $d = 0$ in (8), $X(t) = U(t)$ and the process is said to be short memory or I(0). In this case, if $U(t)$ is ARMA, the autocorrelations decay exponentially fast. On the other hand, if $d > 0$ the process is said to be long memory, so-named due to the high degree of association between observations which are far distant in time. In this context, if $d < 0.5$ the process is still covariance stationary and the autocorrelations decay hyperbolically fast. As long as d is smaller than 1, the process is mean reverting with shocks disappearing in the long run, contrary to what happens with $d \geq 1$ where shocks are expected to be permanent, i.e. lasting forever.

We estimate the fractional differencing parameter d by means of both parametric and semiparametric techniques. In the parametric approach, we use the Whittle function in the frequency domain (Dahlhaus, 1989), while in the semiparametric case, we use a Gaussian semiparametric method that also uses the Whittle function on a band of frequencies that degenerates to zero (Robinson, 1995).

Also, based on the fact that long memory may be produced by the existence of breaks that have not been taken into account, we also conduct the methodology proposed in Gil-Alana (2008) that allows for breaks still in the context of fractional integration. In doing so, we provide evidence of a break at the end of 2003, which is consistent with Kilian (2016a), where he identifies that the expansion of United States shale oil starts after 2003 stimulated by the high prices of conventional crude oil, thereby making this technology competitive.

4. Empirical results

4.1 Data

The data examined in this work correspond to U.S. Crude Oil Production and WTI crude oil prices in the United States over the period 2000:01- 2016:03. The WTI crude oil prices data was obtained from the Federal Reserve Bank of St. Louis.[3] The database was in nominal prices (dollars as currency units), and we have deflated to real prices. We have used the Producer Price Index for All Commodities from the Federal Reserve Bank of St. Louis.[4] The base year to obtain the new crude oil prices is 2011.

Furthermore, we used monthly data of the U.S. Crude Oil Production (thousands barrels per day) over the period 2000:01- 2016:03 obtained from DataStream Database.

4.2 Empirical results

We first estimate the wavelet coherency between the monthly quantity of total U.S. oil production and the WTI crude oil prices. We rely on Monte Carlo simulations to test if the similitude of the wavelet coherency is statistically significant. We compute the complex wavelet coherence matrices between a surrogate for WTI crude oil prices and a surrogate for the total crude oil production. We do 1000 simulations modelling both time series as an ARMA (p, q) process, with no pre-conditions on p and q, with $p = q = 1$.

[3] Spot Oil Price: West Texas Intermediate, retrieved from FRED, Federal Reserve Bank of St. Louis. https://research.stlouisfed.org/fred2/series/OILPRICE/.

[4] US. Bureau of Labor Statistics, Producer Price Index for All Commodities, retrieved from FRED, Federal Reserve Bank of St. Louis https://research.stlouisfed.org/fred2/series/PPIACO/.

We estimate the wavelet coherency for frequencies corresponding to periods between 1.5 to 4 and 4.5 to 8 years.

[Insert Figure 4 about here]

In Figure 4 we display the empirical results. On the left panel (a) we have the wavelet coherency between oil production and WTI crude oil prices.[5] On the right hand side, we have the phase-differences: on the top (b) we have the phase-difference in the 1.5 - 4 year frequency band; at the bottom (c), we have the phase-difference in the 4.5 - 8 year frequency band. The regions surrounded by the black contour are the high coherency regions with significant values at 5%.

Analysing the wavelet coherency between oil production and WTI crude oil prices, we appreciate that the regions with higher coherency are between 2003 and 2014 corresponding to the wavelet scales of periods from the 1.5 to 7 year band. We focus our phase difference analysis on two frequency bands: 1.5 - 4 and 4.5 - 8 years.

To analyse the wavelet coherency graph, we have to focus on the regions of high coherency of the chart. In those regions we can observe the phase difference of the frequency band to extract some conclusions.

In the 1.5 - 4 year band, we identify a region of high coherency between 2003 and 2009, in the frequency bands between 2.5 and 3.5 years with a corresponding phase difference in this band between 0 and $\pi/2$. This result suggests that U.S. oil production and WTI oil prices time series are in phase, they move together, with total oil production leading.

We can find also a region of high coherency between 2009 to 2014 in the 4.5 - 8 year band, specifically between the 5 and 6 year frequency bands. The phase difference of that period stays between $-\pi/2$ and $-\pi$, suggesting that U.S. oil production and WTI

[5] Coherency ranges from blue (low coherency) to red (high coherency). The cone of influence is shown with a thick line, which is the region subject to border distortions.

oil prices time series are out of phase (negative correlated) with oil production oil leading. This suggests that the oil production increases precede a decrease on WTI oil prices. This coincides with Kilian (2016a) in the sense that the evolution of the U.S. price of oil is determined by increases in shale oil production.

From this wavelet coherency figure, we can observe a change across time in the common frequency bands between crude oil production and WTI; higher frequencies between the years 2003 and 2009 suggest that the influence in this period of time is a short-term relationship, reaching a maximum at the 3 year frequency band. This means that the crude oil production influence WTI oil prices faster that in preceding years. On the other hand, the relationship of both time series in the 2009 to 2014 period had a long term component, i.e., a lower frequency band of approximately between 5 to 6 years, suggesting a longer term impact of the crude oil production over the WTI crude oil prices.

Next we move to the long memory part of the paper, and taking into account that some authors argue that fractional integration (and even long memory, in a more general context) can be a spurious phenomenon caused by the presence of a structural break that had not been taken into account (Diebold and Inoue, 2001; Granger and Hyung, 2004) we perform first the approach suggested in Gil-Alana (2008), finding evidence in favour of a break at December 2003. Thus, we separate the whole dataset in two different subsamples, one from January 2000 to December 2003, and the second from January 2004 until the end of the sample. This is in line with the research conducted in Kilian (2016a) where he identifies that the expansion of U.S. shale oil starts after 2003 stimulated by high conventional crude oil prices, resulting in this technology becoming competitive.

Using fractional integration methods the results are presented across Tables 1 - 5. Tables 1 and 2 focus on a parametric approach, and the model considered is

$$y(t) = \beta_0 + \beta_1 t + X(t); \quad (1-L)^d X(t) = U(t), \ \ t = 1,2,..., \qquad (9)$$

Under the assumption of white noise errors (in Table 1) and Bloomfield's (1973) autocorrelated disturbances (in Table 2). The latter is a non-parametric approach of modeling I(0) processes that produce autocorrelations decaying exponentially as in the ARMA case. In both cases, we display the results of the estimates of d for the three standard cases examined in the literature of i) no deterministic terms (β_0 = β_1 = 0 a priori in (9)), ii) an intercept (β_0 unknown and β_1 = 0 a priori) and iii) an intercept with a linear trend (β_0 and β_1 unknown). We present the results for the two subsamples along with those corresponding to the whole dataset.

Starting with the results based on white noise errors (Table 1), we observe that in both cases of prices and production, the results for the whole sample are very similar to those corresponding to the second subsample; thus, for prices, the estimate of d is 1.25 for the whole sample and 1.28 for the data starting in 2004, and in both cases the unit root null hypothesis is rejected in favor of d > 1; however, for the data corresponding to the first subsample, the estimated value of d is about 0.81 and the unit root null cannot be rejected. Focusing now on production, the estimates of d are smaller than 1 in the three cases and the unit root null cannot be rejected in any of the three cases, however, once more, lower values are obtained for the data in the first subsample.

[Insert Tables 1 and 2 about here]

Looking now at the results based on autocorrelated errors, the values are substantially smaller, especially for those corresponding to the first subsample, but they are consistent with those presented above for the case of white noise errors, with lower degrees of integration during the first subsample. In fact, for prices, the estimate of d in the first subsample is found to be 0.39, and the hypothesis of mean reversion ($d < 1$) cannot be rejected in this case, although it is rejected in the second subsample and for the whole dataset. For production, the same result holds, and the estimate of d for the first subsample is even smaller (0.10) being close to 1 for the second subsample and the whole dataset.

Due to the disparity of the results depending on the specification of the error term, we also conduct the analysis with a semiparametric approach, where no functional form is imposed on the I(0) error term U(t).

Table 3 displays the results for the whole sample. We observe that the unit root is almost never rejected for prices, this hypothesis being rejected in favor of $d > 1$ for production with all bandwidth numbers.[6]

[Insert Tables 3 – 5 about here]

Table 4 focuses on the first subsample and here, evidence of mean reversion ($d < 1$) is obtained in many cases for prices and in all cases for production. Very different results are obtained in the results corresponding to the second subsample, support being found for the unit root or even $d > 1$ in the majority of the cases.

For the first subsample we conclude that there was a shock in the crude oil production and price will recover by itself over time with no need for a strong policy measures since the series will tend to revert to its trend sometime in the future. This behavior, however, is not observed with the data starting at 2004 or when the whole dataset is used. For these cases, shocks will have permanent effects and strong policy measures must be adopted to recover the original trends.

[6] We use a selected group of bandwidth numbers. The choice of the bandwidth clearly shows the trade-off between bias and variance: the asymptotic variance is decreasing with m while the bias is growing with m.

5. Concluding Remarks

In this research, we have analyzed the shale oil revolution and its effects on WTI oil price behavior. Since the development of the combination of horizontal drilling techniques together with hydraulic fracturing, known as shale oil, in the 1970s, , oil markets have undergone a significant transformation with the unexpectedly strong rise in US production affecting the oil prices.

It is very common to utilize methodologies based on Fourier analysis to analyse the different relations at different frequencies, omitting the time information, despite it being difficult to identify structural changes with this type of analysis. Hence we also use methodologies based on long run dependence or long memory processes.

First, we analyze the relationship of total U.S. crude oil production and WTI crude oil prices by studying its dynamics in the time-frequency domain applying wavelet tools for its resolution. Analyzing the wavelet coherency we appreciate that the regions with higher coherency are between 2003 and 2014 corresponding to the wavelet scales of periods from the 1.5 to 7 year band. We focus our phase difference analysis on two frequency bands: 1.5 - 4 and 4.5 - 8 years. Analyzing the regions of high coherency in the chart, we identify a region of high coherency in the 1.5 - 4 year band between 2003 and 2009, in frequency bands between 2.5 and 3.5 years with a corresponding phase difference in this band between 0 and $\pi/2$. This result suggests that U.S. oil production and WTI oil prices time series are in phase, they move together, with total oil production leading.

We can find also a region of high coherency between 2009 to 2014 in the 4.5 - 8 year band, specifically between the 5 and 6 year frequency bands. The phase difference of that period stays between $\pi/2$ and π, suggesting that oil production and WTI oil prices time series are out of phase (negatively correlated) with production oil leading.

This suggests that the oil production increases precede a decrease in WTI oil prices. This coincides with Kilian (2016a) in that the evolution of the U.S. price of oil is determined by the increases in shale oil production.

From this wavelet coherency figure, we can observe a change across time in the common frequency bands between crude oil production and WTI; higher frequencies between the years 2003 and 2009 suggest that the influence in this time period is a short-term relationship, reaching a maximum at the 3 year frequency band. This means that the U.S. crude oil production influences WTI oil prices faster than in the preceding years. On the other hand, the relationship of both time series in the 2009 to 2014 period had a long term component, i.e., a lower frequency band of approximately between 5 to 6 years, suggesting a longer term impact of the crude oil production over the WTI crude oil prices.

In the second part of the paper, we use fractional integration techniques to analyse the behavior of WTI crude oil before and after the shale oil boom in the U.S.. We chose the subsamples according to the methodology proposed in Gil-Alana (2008) for structural breaks in the context of fractional integration. The break date was found at the end of 2003, consistent with the results obtained in Kilian (2016a). The most notorious feature observed here is that mean reversion is detected in both production and prices series with the data ending at 2003. However, with data starting in 2004 or when the whole dataset is used, we notice the lack of mean reversion, with orders of integration equal to or higher than 1 in all cases.

Testing the hypothesis that shale oil production affects WTI oil prices, the evolution of the United States price of oil is determined by the increases in shale oil production. Also, the development of the United States refining, pipeline and rail infrastructure are important to understand and forecast the evolution of the domestic price of oil in the United States. There are others factors mentioned by Baumeister and Kilian (2016) such as oil supply shocks, demand shocks and shocks to oil price expectations. We have only taken into account the crude oil production in the United States and WTI crude oil prices because this research has mainly focused on how overproduction of shale oil affects WTI oil price behaviour. The influence of other variables using similar methodologies to those employed in this paper will be examined in future papers.

Acknowledgments

We would like to thank Aguiar-Conraria and Soares for providing us with the MATLAB wavelet package. Manuel Monge gratefully acknowledges financial support from the Asociacion de Amigos, University of Navarra. This paper is based on a chapter of Manuel Monge´s Ph.D. dissertation at the University of Navarra. Luis A. Gil-Alana gratefully acknowledges financial support from the Ministerio de Economía y Competitividad (ECO2014-55236). Fernando Perez de Gracia also acknowledges financial support from the Ministerio de Economía y Competitividad (ECO2014-55496).

References

Aguiar-Conraria, L., Azevedo, N., Soares, M. J. 2008. Using wavelets to decompose the time–frequency effects of monetary policy. Physica A: Statistical Mechanics and its Applications, 387, 2863-2878.

Aguiar-Conraria, L., Soares, M. J. 2011a. Oil and the macroeconomy: using wavelets to analyze old issues. Empirical Economics, 40, 645-655.

Aguiar-Conraria, L., Soares, M. J. 2011b. Business cycle synchronization and the euro: a wavelet analysis. Journal of Macroeconomics, 33, 477-489

Aguiar-Conraria, L., Soares, M. J. 2014. The continuous wavelet transform: moving beyond uni-and bivariate analysis. Journal of Economic Surveys, 28, 344-375.

Alquist, R., Guénette, J. D. 2014. A blessing in disguise: the implications of high global oil prices for the north american market. Energy Policy, 64, 49-57.

Arezki, R., Blanchard, O. 2015. Seven questions about the recent oil price slump. Energia, 36, 2-10.

Baffes, J., Kose, M. A., Ohnsorge, F., Stocker, M. 2015. The great plunge in oil prices: Causes, consequences, and policy responses. Policy Research Note No.1, World Bank.

Bank of Canada. 2015. Monetary policy report January 2015. Bank of Canada.

Baumeister, C., Kilian, L. 2016. Understanding the decline in the price of oil since June 2014. Journal of the Association of Environmental and Resource Economists, 3, 131-158.

Benes, J., Chauvet, M., Kamenik, O., Kumhof, M., Laxton, D., Murula, S., J. Selody. 2015. The future of oil: geology versus technology. International Journal of Forecasting, 31, 207-221.

Clements, M., Cummings, A. 2016. U.S. shale: the saviour and the scourge of the oil market. SYZ Wealth Management.

Covert, T. R. 2014. Experiential and social learning in firms: the case of hydraulic fracturing in the bakken shale (Job Market Paper). Working Paper.

Cuestas, J. C., Gil-Alana, L. A. 2016. Testing for long memory in the presence of non-linear deterministic trends with Chebyshev polynomials. Studies in non-linear dynamics and econometrics, 20, 57-74.

Dahlhaus, R. 1989. Efficient parameter estimation for self-similar process, Annals of Statistics, 17, 1749-1766.

Daubechies, I. 1992. Ten lectures on wavelets. Society for Industrial and Applied Mathematics, Philadelphia.

Diebold, F. X., Inoue, A. 2001. Long memory and regime switching, Journal of Econometrics, 105, 131-159.

Gençay, R., Selçuk, F., Whitcher, B. 2002. An introduction to wavelets and other filtering methods in finance and economics. Academic Press.

Gil-Alana, L. A. 2008. Fractional integration and structural breaks at unknown periods of time. Journal of Time Series Analysis, 29, 163-185.

Gil-Alana, L. A., Hualde, J. 2009. Fractional integration and cointegration. An overview with an empirical application. The Palgrave Handbook of Applied Econometrics, 2, 434-472.

Granger, C.W.J., Hyung, N., 2004. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. Journal of Empirical Finance, 11, 399-421.

Hamilton, J. 2014. Oil prices as an indicator of global economic conditions. Econbrowser.

Hogan J., Lakey, J. 2005. Time-frequency and time-scale methods: adaptive decompositions, uncertainty principles, and sampling. Birkhäuser, Boston.

Kilian, L. 2016a. The impact of the shale oil revolution on U.S. oil and gas prices. Review of Environmental Economics and Policy. Forthcoming.

Kilian, L. 2016b. The impact of the fracking boom on Arab oil producers. Working Paper.

Krane, J., Agerton, M. 2015. Effects of low oil prices on U.S. shale production: OPEC calls the tune and shale swings. Baker Institute Research Paper.

Kyrtsos, C., Malliaris, A. G., Serletis, A. 2009. Energy sector pricing: on the role of neglected nonlinearity. Energy Economics, 31, 492-502.

McCracken, R. 2015. Shale oil's response to prices may call for industry re-evaluation. Energy Economist, Platts.

Monge, M., Gil-Alana, L. A., Perez de Gracia, F. 2016. Crude oil price behavior before and after military conflicts and geopolitical events. Under evaluation in Energy.

Naccache, T. 2011. Oil price cycles and wavelets. Energy Economics, 33, 338-352.

Percival, D. B., Walden, A.T. 2000. Wavelet methods for time series analysis. Cambridge University Press.

Ramsey, J. B. 2002. Wavelets in economics and finance: past and future. Studies in Nonlinear Dynamics and Econometrics, 6, 1-28.

Robinson, P. M. 1995. Gaussian semi-parametric estimation of long range dependence.

Annals of Statistics, 23, 1630-1661.

Schreiber, T., Schmitz, A. 1996. Improved surrogate data for nonlinearity tests. Physical Review Letters, 77, 635-638.

Sharenow, G. E., Worah, M. P. 2013. Shale oil: a deep dive into implications for the global economy and commodity investors.<http://www.pimco.com/ en/insights /pages /shale-oil-a-deep-dive-into-implications-for-the-global-economy-and-commodity-investors.aspx>.

Appendix

Figure 1. Oil production of selected U.S. shale plays

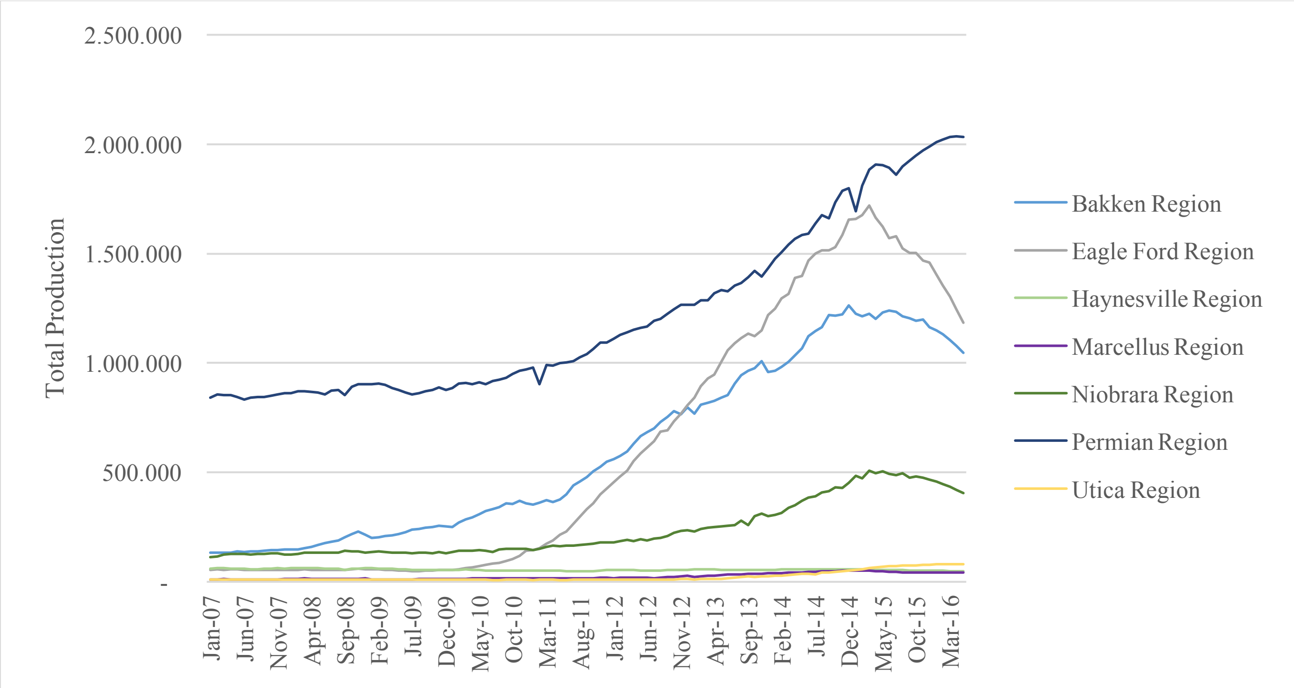
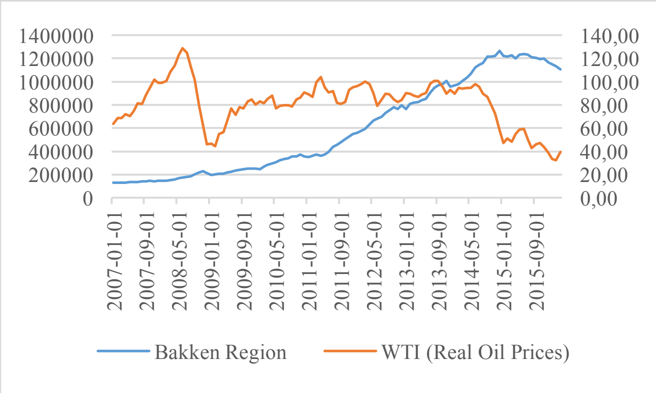
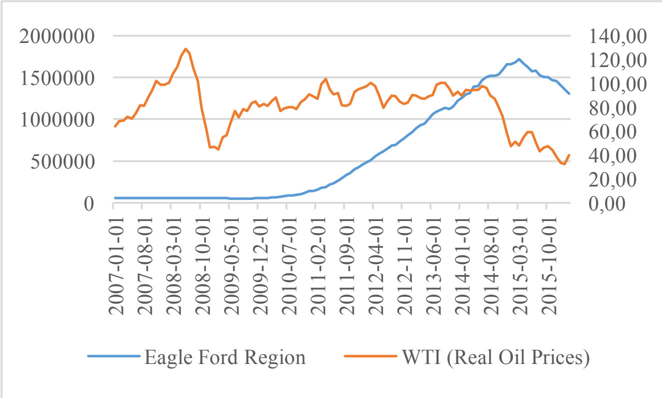
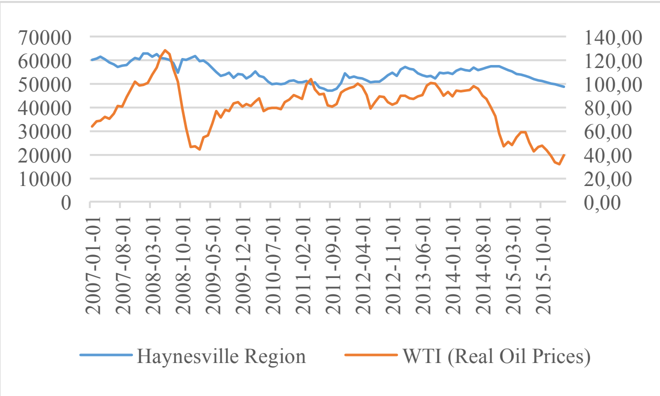
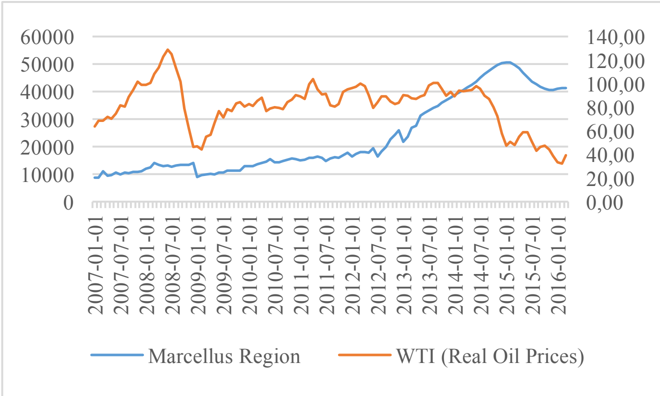


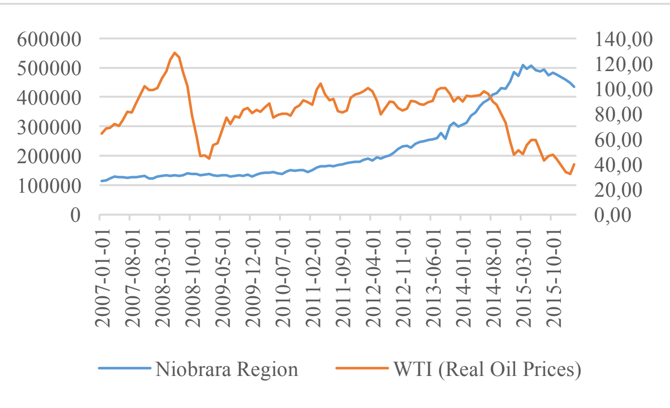
Figure 2. US shale oil production by region and the behaviour of WTI crude oil prices.

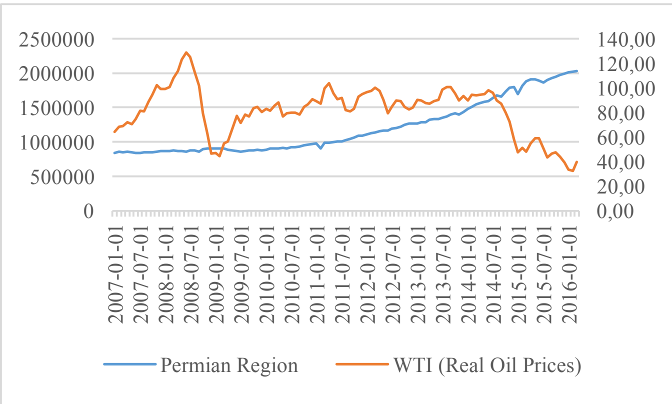












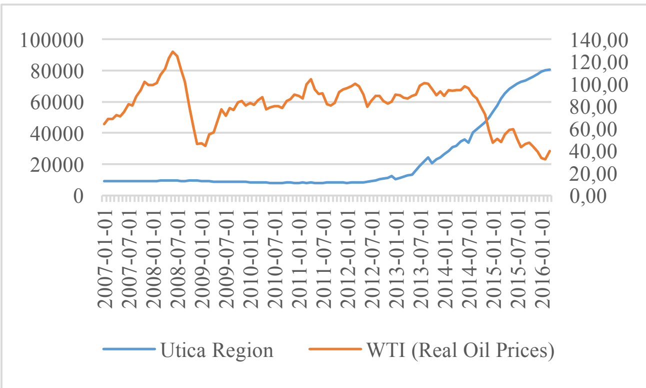


Figure 3. Total U.S. oil production and the behaviour of WTI crude oil prices.

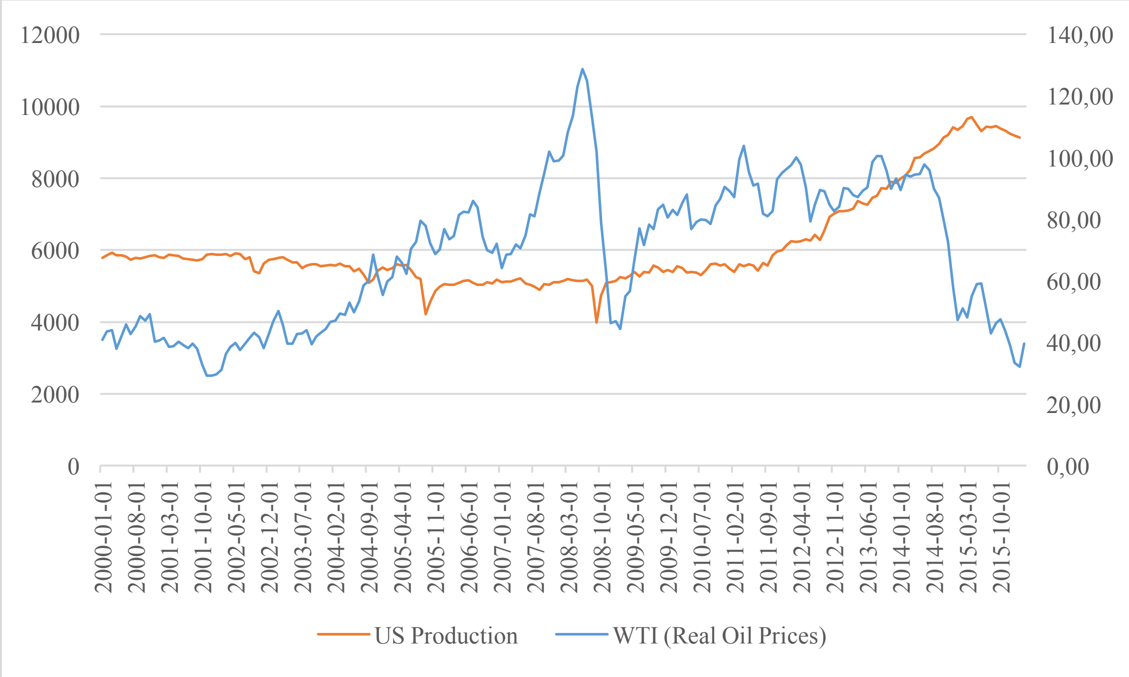
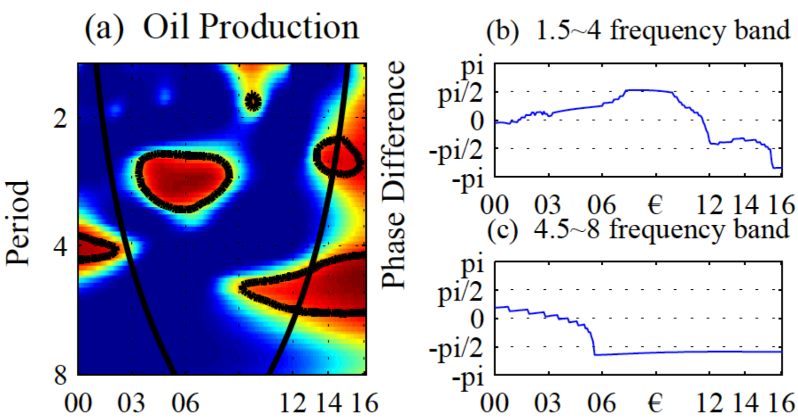


Figure 4: Wavelet Coherency and phase-difference between Total U.S. Oil Production and WTI crude oil price.



On the left: Wavelet Coherency between Total U.S. Oil Production and WTI. On the right: Phase-difference between Total U.S. Oil Production and WTI at 1.5~4 year (top) and 4.5~8 year (bottom) frequency bands

Table 1: Estimates of d based on white noise disturbances

Series	No det. terms	An intercept	A linear time trend
Prices (total)	1.17 (1.04, 1.34)	**1.25 (1.12, 1.42)**	1.25 (1.12, 1.42)
Prices (1st subsample)	0.97 (0.78, 1.27)	**0.81 (0.60, 1.19)**	0.81 (0.60, 1.19)
Prices (2nd subsample)	1.16 (1.03, 1.35)	**1.28 (1.13, 1.47)**	1.28 (1.13, 1.47)
Production (total)	0.96 (0.88, 1.07)	**0.97 (0.91, 1.05)**	0.97 (0.91, 1.05)
Prod. (1st subsample)	0.93 (0.75, 1.21)	**0.73 (0.47, 1.18)**	0.70 (0.36, 1.18)
Prod. (2nd subsample)	0.93 (0.83, 1.07)	**0.96 (0.89, 1.05)**	0.95 (0.88, 1.05)

Table 2: Estimates of d based on autocorrelated (Bloomfield) disturbances

Series	No det. terms	An intercept	A linear time trend
Prices (total)	0.84 (0.64, 1.15)	**0.96 (0.79, 1.25)**	0.96 (0.79, 1.25)
Prices (1st subsample)	0.71 (0.33, 1.16)	**0.39 (0.06, 0.79)**	0.39 (0.06, 0.79)
Prices (2nd subsample)	0.91 (0.69, 1.24)	**0.96 (0.74, 1.34)**	0.96 (0.74, 1.33)
Production (total)	0.94 (0.81, 1.11)	**1.01 (0.93, 1.12)**	1.01 (0.92, 1.13)
Prod. (1st subsample)	0.78 (0.35, 1.23)	**0.10 (-0.23, 0.53)**	-0.24 (-0.68, 0.43)
Prod. (2nd subsample)	0.87 (0.72, 1.08)	**0.99 (0.89, 1.13)**	0.99 (0.88, 1.15)

Table 3: Estimates of d using a semiparametric method for the whole dataset

m	Prices	Production	Lower 95%	Upper 95%
8	0.842	1.500**	0.709	1.290
9	0.778	1.500**	0.725	1.274
10	0.787	1.500**	0.739	1.260
11	0.772	1.378**	0.752	1.247
12	0.802	1.348**	0.762	1.237
13	0.790	1.375**	0.771	1.228
14	0.778*	1.363**	0.780	1.219
15	0.823	1.280**	0.787	1.212

*: Evidence of mean reversion at the 5% level; ** Evidence of d > 1 at the 5% level.

Table 4: Estimates of d using a semiparametric method for the first subsample

m	Prices	Production	Lower 95%	Upper 95%
5	0.630*	0.202*	0.632	1.367
6	0.800	0.361*	0.664	1.335
7	0.950	0.277*	0.689	1.310
8	0.773	0.410*	0.709	1.290
9	0.673*	0.488*	0.725	1.274
10	0.655*	0.500*	0.739	1.260
11	0.720*	0.500*	0.752	1.247
12	0.745*	0.500*	0.762	1.237

*: Evidence of mean reversion at the 5% level.

Table 5: Estimates of d using a semiparametric method for the second subsample

m	Prices	Production	Lower 95%	Upper 95%
6	0.554*	1.500**	0.664	1.335
7	0.679*	1.500**	0.689	1.310
8	0.766	1.487**	0.709	1.290
9	0.742	1.349**	0.725	1.274
10	0.811	1.418**	0.739	1.260
11	0.784	1.284**	0.752	1.247
12	0.775	1.150	0.762	1.237
13	0.781	1.095	0.771	1.228

*: Evidence of mean reversion at the 5% level; ** Evidence of d > 1 at the 5% level.