

SPATIAL CRUDE OIL PRODUCTION DIVERGENCE AND CRUDE OIL PRICE BEHAVIOUR IN THE UNITED STATES

Manuel Monge

Universidad Francisco de Vitoria, Spain

Luis Alberiko Gil-Alana

University of Navarra and Universidad Francisco de Vitoria, Spain

ABSTRACT

This paper deals with the analysis of (spatial) crude oil production divergence in the United States, paying particular attention to the domestic crude oil production between PADD 2 and PADD 3, which are the areas in which the bottleneck occurs and has a direct implication on the price of West Texas Intermediate (WTI). To this purpose, we use techniques based on fractional integration, fractional cointegration VAR (FCVAR) and wavelet analysis. Monthly data related to the oil production in the U.S. by regions (Anadarko, Appalachia, Bakken, Eagle Ford, Haynesville, Niobrara and Permian) from January 2007 to June 2020 are used. The results, using fractional integration and cointegration techniques, indicate that the time series analyzed are highly persistent and there is evidence of long run equilibrium relationships in some of the series. Finally, using wavelet analysis, we conclude that the most affected areas are Anadarko, Appalachia, Haynesville and Niobrara where an increase in the shale oil production is followed a decrease in WTI crude oil prices.

Keywords: Shale oil total production; WTI crude oil price; spatial crude oil price divergence; fractional integration; FCVAR model, wavelet analysis

JEL Classification: C22; Q41

Corresponding author: Prof. Manuel Monge
Universidad Francisco de Vitoria
Faculty of Law and Business
E-28223 Madrid
Spain

email: manuel.monge@ufv.es

* Luis A. Gil-Alana gratefully acknowledges financial support from the MINEIC-AEI-FEDER ECO2017-85503-R project from 'Ministerio de Economía, Industria y Competitividad' (MINEIC), 'Agencia Estatal de Investigación' (AEI) Spain and 'Fondo Europeo de Desarrollo Regional' (FEDER). Prof. Luis A. Gil-Alana and Dr. Manuel Monge also acknowledge financial support from internal Projects of the Universidad Francisco de Vitoria. Comments from the Editor and three anonymous reviewers are gratefully acknowledged.

1. Introduction

In the last century, a complex web of petroleum pipelines was the way to transport crude oil and refined petroleum products in the United States taking into account the cost-effective relationship. The United States, with more than 2.4 million miles of pipeline in which crude oil is transported, separated from the refined products, and flowing only in one direction, has the largest energy pipeline connection in the world. This web of crude oil pipelines is extensive with around 72,000 miles connecting regional markets with the crude oil lines. When access by sea is available, this option is used to transport oil cheaply. Other, more expensive ways of transporting crude oil and refined products are rail and trucks. These were used for areas with sparse demand and short distances.

The government of the United States, during the Second World War, divided the country into five Petroleum Administration for Defense Districts (PADDs), of which the Northeastern PADD has been divided into three sub-PADDs in order to aid in the planning and allocation of oil and refined products. In Figure 1, the different PADDs are shown where, according to Borenstein and Kellogg (2014), transport is relatively fluid but there is a potential risk of bottlenecks occurring.

[Insert Figures 1 and 2 here]

Borenstein and Kellogg (2014) claimed that PADDs 1, 2 and 3 were well integrated with one another, to the point that the primary crude oil pricing and the delivery futures contracts that traded on the New York Mercantile Exchange (NYMEX) were set in Oklahoma (PADD 2) because many of the pipes converge in Cushing, minimizing the risk due to minimal transportation constraints to major oil markets.

At the beginning of 2011, the connection in oil prices in PADD 2 and 3 changed when there was an increase in the production of crude oil due to the Bakken oil shale formation in North Dakota and the tar sands area of Alberta, Canada.

There are many factors behind the oversupply in Cushing, surpassing the capacity of the pipeline from there to the Gulf Coast, producing a spatial crude oil price divergence (see Borenstein and Kellogg, 2014). Among them we can mention the increase in crude oil production, the entire system designed to transport crude oil by pipeline (see Kilian, 2016)¹ and the lack of rail infrastructure or barge transportation capacity. Monge et al. (2017) analyzed the behaviour of shale oil production on the West Texas Intermediate (WTI) prices finding evidence that during the period 2009-2014, oil production and WTI crude oil prices time series were negatively correlated, suggesting that the increase in the production of oil produced a decrease in WTI crude oil prices. Also, they showed that for the time period from 2004 to 2016, the orders of integration of the series examined were higher than 1, implying lack of non-mean reversion behaviour.

The goal of this paper is to understand the behaviour of the spatial crude oil production divergence in the United States by regions (Anadarko, Appalachia, Bakken, Eagle Ford, Haynesville, Niobrara and Permian) that correspond to the areas PADD1, 2, 3, 4 and 5 (paying particular attention to domestic crude oil first purchase prices between PADD 2 and PADD 3, which is the area where the bottleneck occurs) and how this affects the price of crude oil.

In the literature there are plenty of articles dealing with the integration of energy markets. Examples are Adelman (1984), Weiner (1991, 1993), Serletis (1994), Rodriguez and Williams (1993, 1994), Gülen (1997, 1999), Gjolberg and Johansen (1999), De Vany and Walls (1999), Lin and Tamvakis (2001), Milonas and Henker (2001), Asche et al. (2002, 2003, 2006), Hammoudeh and Li (2004), Bachmeier and Griffin (2006), Bentzen (2007), Fattouh (2010), Chang et al. (2010), Wlazlowski et al. (2011) among many others.

¹ Kilian (2016) argued that the entire system is designed to transport imported crude oil from the Gulf Coast ports to the U.S. oil market hub in Cushing, Oklahoma instead of transporting all that oil to the refineries of the East Coast of the United States.

To the best of our knowledge, this is the first paper that analyzes the statistical properties of shale oil total production in the U.S. by regions in barrels per day (bbl/d) and the West Texas Intermediate (WTI) U.S. oil price, measuring the degree of persistence by using fractional integration techniques (Gil-Alana and Hualde, 2009 and Monge et al., 2017) in each of the specified locations. Also, we analyze the long-term relationships of the selected time series, focusing on the regions in which spatial crude oil price divergence occurs and using the fractional cointegration VAR (FCVAR) approach (Johansen and Nielsen, 2012). To conclude this study, we investigate possible structural changes caused by the oversupply and bottlenecks in the distribution, using methodologies based on wavelet transforms (Aguar-Conraria and Soares, 2014).

The rest of the paper is organized as follows. In Section 2 we describe the data used for this study. Section 3 presents the methodology applied in the paper. In Section 4 we discuss the main empirical results, while Section 5 concludes the paper.

2. Methodology

2.1. Fractional Integration

Fractional integration is a time series technique that allows for a fractional degree of differentiation. Given a time series, x_t , $t = 1, 2, \dots$, we say that it is integrated of order d , and denoted by $I(d)$ if its d -differences are stationary $I(0)$.

A series is integrated of order 0 or $I(0)$, also termed short memory, if the infinite sum of its autocovariances is finite, and within this category, we can include the case of uncorrelated series, e.g., a white noise process, but also, other models which are weakly dependent like those based on the stationary and invertible AutoRegressive Moving Average (ARMA)-type of models.

A series is said to be integrated of order d , or $I(d)$ if it can be represented as:

$$(1 - L)^d x_t = u_t, \quad t = 1, 2, \dots, \quad (1)$$

where L is the lag-operator, i.e., $L^k x_t = x_{t-k}$ and u_t is short memory or $I(0)$ process. Thus, for example, if u_t is $ARMA(p, q)$, we say then that x_t is an AutoRegressive Fractionally Integrated Moving Average, $ARFIMA(p, d, q)$ model. Clearly, if $d = 1$ in (1) we have the classical $ARIMA(p, 1, q)$ model, but as earlier mentioned, d can be any real value, including thus fractional numbers, and using a Binomial expansion,

$$(1 - L)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j L^j = 1 - dL + \frac{d(d-1)}{2} L^2 - \dots, \quad (2)$$

which is valid for any real value d , equation (1) can be expressed as:

$$x_t = d x_{t-1} - \frac{d(d-1)}{2} x_{t-2} + \dots + u_t. \quad (3)$$

Thus, the higher the value of d is, the higher the level of association between the observations is, and thus, the level of persistence in the data. In other words, the differencing parameter d can be taken as a measure of the degree of persistence of the data, and by allowing fractional values of d we permit a much higher degree of flexibility in the dynamic specification of the model. Moreover, this specification allows us to determine if shocks in the series will have transitory ($d < 1$) or permanent ($d \geq 1$) effects. These processes were introduced in the 80s by Granger (1980), Granger and Joyeux (1981) and Hosking (1981) but it was not until the late 90s that they become popular in the analysis of economic data. In the last ten years, there have appeared numerous papers dealing with energy issues and using fractional integration. Examples are among others those by Barros et al. (2012), Belbute and Pereira (2016), Solarin et al. (2018), Bozoklu et al. (2020), Gil-Alana et al. (2020a, b), etc.

2.2. Fractional Cointegrated VAR

A multivariate fractional cointegration model, named Fractionally Cointegrated Vector AutoRegressive (FCVAR) was proposed in Johansen (2008), and was later extended in Johansen and Nielsen (2010, 2012). In fact, it extends the classical Cointegrated Vector AutoRegressive (CVAR) model of Johansen (1996), allowing for series which are all integrated of order d and that cointegrate with order $d - b$, with positive b . An advantage of the FCVAR model is that it has the power to be used for stationary and nonstationary time series (Johansen and Nielsen, 2012 and Nielsen and Popiel, 2018).

From the non-fractional case CVAR model, where we assume that $Y_t, t = 1, \dots, T$ is a p -dimensional $I(1)$ time series, represented as:

$$\Delta Y_t = \alpha \beta' Y_{t-1} + \sum_{i=1}^k \Gamma_i \Delta Y_{t-i} + \varepsilon_t = \alpha \beta' L Y_t + \sum_{i=1}^k \Gamma_i \Delta L^i Y_t + \varepsilon_t, \quad (4)$$

we must replace the difference and lag operators by Δ^b and $L_b = 1 - \Delta^b$, in order to derive the FCVAR model. Then,

$$\Delta^b Y_t = \alpha \beta' L_b Y_t + \sum_{i=1}^k \Gamma_i \Delta L_b^i Y_t + \varepsilon_t, \quad (5)$$

which is applied to $Y_t = \Delta^{d-b} X_t$ such that

$$\Delta^d X_t = \alpha \beta' L_b \Delta^{d-b} X_t + \sum_{i=1}^k \Gamma_i \Delta^b L_b^i Y_t + \varepsilon_t \quad (6)$$

In the previous equation, a p -dimensional i.i.d. variable is represented by the term ε_t , with mean zero and variance-covariance matrix Ω . α and β are $p \times r$ matrices, where $0 \leq r \leq p$. The cointegrating relationships in the system are related to the columns of β . The parameters that govern the short-run behavior of the variables are indicated by Γ_i . The short-run dynamics of the system and the speed of adjustment responses to deviations from the equilibria are described by the coefficients in α . Finally, the parameter d is the order of differentiation of the observable time series and b refers to the level of reduction in the degree of integration of $\beta' X_t$. With this methodology, the FCVAR model, we can simultaneously model the long-run equilibria, the adjustment reactions to deviations from

the equilibria and the short-run dynamics of the system (see Johansen and Nielsen, 2012; Nielsen and Popiel, 2018).

Several empirical papers such as Jones, Nielsen and Popiel (2014), Baruník and Dvořáková (2015), Maciel, (2017), Aye et al. (2017), Dolatabadi et al. (2016, 2018), Gil-Alana and Carcel (2020), Gunay (2018), Poza and Monge (2020), Quineche (2020), and Caro et al. (2020) have been used for the estimation and testing of the FCVAR model, and Nielsen and Popiel (2018) reported the Matlab codes, which are available at the Nielsen webpage².

2.3. Wavelet Analysis

To analyse time series in the time-frequency domain and following Vacha and Baruník (2012), Aguiar-Conraria and Soares (2011, 2014), Dewandaru et al. (2016), Tiwari et al. (2016), Monge et al. (2017), Jammazi et al. (2017), and others we use the Continuous Wavelet Transform (CWT), applying wavelet coherency and wavelet phase-differences. This methodology is appropriate since stationarity is not a requirement; also, to find potential changes in its pattern, it is interesting to study the interaction of both the time and the frequency domains of the time series themselves.

The wavelet coherency is understood as a correlation of time series in a two-dimensional diagram that identifies hidden patterns or information in the domain of time and frequency. The $WT_x(a, \tau)$ of a time series $x(t)$ is obtained by projecting a mother wavelet ψ defined as:

$$WT_x(a, \tau) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{a}} \psi^* \left(\frac{t-\tau}{a} \right) dt, \quad (7)$$

where the wavelet coefficients of $x(t)$ is defined as $WT_x(a, \tau)$; a is the position of a wavelet in the frequency domain; the position in the time domain is defined by τ . Thus,

² <https://sites.google.com/view/mortennielsen/software?authuser=0>

the function of mapping the original time series in a function of the two previous parameters mentioned (τ and a) is obtained by the wavelet transform.

The type of mother wavelet chosen to carry out the analysis is the Morlet wavelet, which is a complex sine wave within a Gaussian envelope. This allows us to measure the synchronism between time series.³

The wavelet coherence is defined as:

$$WCO_{xy} = \frac{SO(WT_x(a,\tau)WT_y(a,\tau)^*)}{\sqrt{SO(|WT_x(a,\tau)|^2)SO(|WT_y(a,\tau)|^2)}}, \quad (8)$$

and it helps us to understand the interaction and the integration between the two time series.

The SO is a smoothing operator in time and scale that prevents the wavelet coherence from always being one for all times and scales (see Aguiar-Conraria et al. 2008 for details). Matlab computer programs for the calculation of estimators and test statistics in the CWT are provided in Aguiar-Conraria's website.⁴

Following the research line of Pinto et al. (2016) we know that the economic time series are an aggregation of components operating on different frequencies. According to Aguiar-Conraria and Soares (2014) the wavelet analysis performs the spectral characteristics of a time series as a function of a different scale of time, allowing us to localize the structural changes on time, the magnitude and the duration of them and to determine the synchronism between the time series employed.

³ See Aguiar-Conraria and Soares (2014) for the properties of this wavelet and for a more complete understanding of this procedure.

⁴ <https://sites.google.com/site/aguiarconraria/joanasoares-wavelets>

3. Data

The data examined in this work comes from The Drilling Productivity Report from EIA.⁵ They correspond to the total number of drilling rigs in operation along with the estimates of drilling productivity and estimated changes in production from existing oil wells. They provide estimated changes in oil production for seven key regions in the U.S. Also, we use the West Texas Intermediate (WTI) from the Federal Reserve Bank of St. Louis⁶ for the crude oil prices in the U.S.. The monthly data analyzed covers the period from January 2007 to June 2020.

[Insert Figure 3 here]

The data used in the paper are presented in Figure 3; it displays the time series plots of U.S. shale oil productivity by area, showing the behaviour (similarities and differences) between the time series. We observe in the figure that shale oil production in the Permian region is more pronounced than in the other regions that show a softer behaviour.

4. Empirical results

4.1. Fractional integration and cointegration

We start by presenting the results based on the univariate analysis. For each series, we consider the following model,

$$y_t = \beta_0 + \beta_1 t + x_t; \quad (1 - L)^d x_t = u_t, \quad t = 0, 1, \dots, \quad (9)$$

where y_t refers to each of the observed time series; β_0 and β_1 are unknown parameters and referring, respectively, to the constant and to a linear time trend, while x_t is supposed to be $I(d)$; thus, d is the differencing parameter that we allow to be any real number and thus,

⁵ <https://www.eia.gov/petroleum/drilling/#tabs-summary-1>

⁶ <https://fred.stlouisfed.org/series/DCOILWTICO>.

potentially fractional, and the $I(0)$ error term u_t will adopt different forms across the presented tables. Thus, in Table 1 we suppose u_t is uncorrelated, following a white noise process; in Table 2 weak autocorrelation is permitted in the error term, and u_t follows here the exponential model of Bloomfield (1973); finally, in Table 3 and based on the monthly nature of the data, a seasonal monthly AR(1) process is imposed on u_t . We present in the three tables the estimated values of d (and the 95% confidence band of the non-rejection values of d using Robinson, 1994) under the three classical assumptions of no deterministic terms (second column in the tables), with an intercept (third column) and allowing for an intercept and a linear time trend (fourth column), marking in bold in the tables the selected cases according to these three standard specifications.

[Insert Table 1 about here]

Starting with the results based on white noise errors, (in Table 1) the first thing we observe is that the time trend is significant in 5 out of the 8 series examined, and the estimated values of d are equal to or higher than 1 in the majority of the cases. In fact, the only evidence of mean reversion, i.e., estimates of d significantly below 1 is found in the case of Andarko; the $I(1)$ hypothesis cannot be rejected for Haynesville and also for Oil Prices, while this hypothesis is decisively rejected in favour of $d > 1$ in the remaining cases.

Allowing for autocorrelation, we first present the results based on the model of Bloomfield (1973), in Table 2, and we see that the time trend is required in the cases of Bakken, Haynesville and Niobrara. Moreover, the hypothesis of mean reversion is now rejected in all cases, with all values of d being in the $I(1)$ and $I(d, d > 1)$ cases.

[Insert Tables 2 and 3 about here]

If the errors follow a seasonal AR(1) process, the results are fairly similar to those based on white noise errors, and d is equal to or higher than 1 in all except one single

case, again Anadarko. Thus, the results so far indicate that all series are highly persistent with shocks having permanent effects and only for Anadrako did we find a small degree of reversion to the mean with shocks being temporary, albeit with very long lasting effects.

[Insert Table 4 about here]

Extending the analysis to the multivariate case, we conduct the FCVAR model of Johansen (2008) testing for the existence of long run equilibrium relationships between each series and the oil prices in a vis-a-vis representation. The results are reported in Table 4, and they are extremely heterogeneous across the series. Thus, for two of the cases, Appalachia and Bakken, we cannot reject the hypothesis of $d = b$, this value being around 0.414 for Appalachia and 0.518 for Bakken; for other two cases, Anadarko and Haynesville there is no evidence of cointegration since the reduction in the degree of integration is almost null; finally, for the other three groups of series, there is a reduction of above 0.3 in the cointegrating long run relationship. We can conclude this section by saying that all the individual series are highly persistent, and evidence of long run equilibrium relationships between the crude oil production and oil prices are observed in some of the series, being very clear in the cases of Appalachia and Bakken, but also detectable for Eagle Ford, Niobrara and Permian. Thus, apparently there is a discrepancy between the univariate and the multivariate results, observing in the multivariate case a degree of mean reversion in the series of oil production examined (and in its relation with oil prices) that is not observed with the univariate results. This discrepancy may be explained by the presence of structural breaks which have not been detected in the data. Thus, a wavelet analysis is conducted in the following subsection.

4.2. Wavelet analysis

To see the possible presence of structural changes caused by oversupply and bottlenecks in the distribution, we use a multivariate wavelet analysis based on the time-frequency domain to estimate how shale oil production affects WTI crude oil prices in the U.S. at different frequencies and how they evolve over time.

In Figure 4, we do a preliminary analysis of each of the series that represents the production of shale oil in the U.S. On the left part of each plot we represent the monthly returns of the crude oil prices and the shale oil production. On the right, we plot the wavelet power spectra that represents for each moment and frequency, the intensity of the variance of the time series for each frequency of cyclical oscillations. According to Aguiar-Conraria and Soares (2014) the cone-of-influence (COI), represented by the black conic line, identifies the region where edge effects are important. The regions represented outside this line should be interpreted with caution. The low and high degree of variability is distinguished by a colour spectrum, ranging from dark blue to red, respectively. The local maxima in the power spectra is represented by a white line. The significance levels, 5% and 10%, are represented by black and grey contours, respectively.

[Insert Figure 4 about here]

In the case of WTI crude oil prices, the volatility is spread across the sample, but it is stronger at higher frequencies (short term). The red regions correspond to cycles of periods smaller than 32 months. Analyzing shale oil production in the U.S. we see that Bakken and Haynesville have one cycle. The first one corresponds to the beginning of the sample and Haynesville has a cycle that runs for virtually the entire sample. In both cases, the cycle lasts 12 months and it became apparent in the beginning of 2008. In the rest of the cases we find two or more cycles around the sample. We can highlight the case of Appalachia and Niobrara, where the cycles have 12 months and the cycles start in 2018 and 2017, respectively.

As it is difficult to discern any inter-relation between shale oil production by regions in the U.S. and WTI crude oil prices using previous analysis based on wavelet power spectra, Figure 5 shows a preliminary result that tells us when and at which frequencies the inter-relations are the strongest.

[Insert Figure 5 about here]

For this purpose we have calculated, first, the wavelet coherency and we identify that the main regions with statistically significant coherency are located at low frequencies (corresponding to cycles between 32 and 64 months where the regions show statistically significance coherence). The most important ones start around 2013 and are Anadarko, Appalachia, Haynesville and Niobrara where the coherency tells us how important and strong the relation is between the time series. Bakken starts around 2011.

Also, we can find the partial phase difference and the partial wavelet gain, which give us information about the magnitude of the impact that a shock in one variable has on the other. For the cases mentioned before, and looking at the 5% significance level, the phase difference is between π and $\frac{\pi}{2}$. This means that WTI and shale oil production, at these frequencies, display an anti-phase relation with shale oil production in these leading areas. Economically, that means that, at these frequencies, an increase in the shale oil production in the areas mentioned before is followed by a decrease in WTI crude oil prices, where the partial gain, interpreted as the modulus of the regression coefficient in the regression of shale oil production on WTI crude oil prices at each time and frequency, is stable at 0.04.

5. Concluding comments

We have examined in this article the spatial crude oil production divergence in the United States, paying special attention to domestic crude oil first purchase prices by regions

(Anadarko, Appalachia, Bakken, Eagle Ford, Haynesville, Niobrara and Permian) corresponding to the areas PADD1, 2, 3, 4 and 5. Of particular interest are the prices between PADD 2 and PADD 3, which are the areas where the bottleneck occurs.

For this purpose, first we have used techniques based on fractional integration, which are very appropriate to determine the nature of the shocks. Evidence of mean reversion is only found in the case of Anadarko. In all the other cases, the orders of integration are found to be equal to or higher than 1, implying high levels of persistence and permanency of shocks. Allowing autocorrelation, the hypothesis of mean reversion is rejected in all cases. If the errors follow a seasonal AR(1) process the results are fairly similar to those based on white noise errors, and d is equal to or higher than 1 in all areas except in Anadarko. In a multivariate context, using a fractional CVAR (FCVAR) approach, the results support the hypothesis of a long run equilibrium relationship, being very clear in the cases of Appalachia and Bakken, but also for Eagle Ford, Niobrara and Permian.

Finally, to see the possible presence of structural changes caused by oversupply and bottlenecks in the distribution and how the shale oil production affects WTI crude oil prices in the U.S. we use wavelet analysis. The coherency results, showing how important and strong the relation is between the time series indicate that the most important ones start around 2013 in Anadarko, Appalachia, Haynesville and Niobrara. Bakken starts around 2011. To see the magnitude of the impact that a shock in one variable has on the other we use the partial phase difference. At a 5% significance level, we can conclude by saying that an increase in the shale oil production is followed a decrease in WTI crude oil prices.

These results are in line with Monge et al. (2017), as their conclusions are also supported by our results. The evolution of the price of oil in the United States is

determined by the increase in shale oil production. The United States refining, pipeline and rail infrastructure play a fundamental role in price behaviour to understand and forecast the evolution of the domestic price of oil. This indicates that a bottleneck problem does not necessarily translate into a price increase, but rather that it would negatively affect the price of oil and cause it to persist over time, taking a long time to recover. This fact may have a negative effect on the profitability of oil-producing companies in the United States. Also, it might have implications in the economic policy of the country since the oil market represents 8% of the Gross Domestic Product (GDP). Thus, this research paper may be very helpful to institutions and companies that are exposed to crude oil market changes allowing a better understanding of the effects of spatial crude oil production divergence in the United States on the WTI crude oil price behaviour.

References

- Adelman, M. A. (1984). International oil agreements. *The Energy Journal*, 5(3).
- Aguiar-Conraria, L., Azevedo, N. and 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., and Soares, M. J. (2011). Oil and the macroeconomy: using wavelets to analyze old issues. *Empirical Economics*, 40, 645-655.
- Aguiar-Conraria, L., and Soares, M. J. (2014). The continuous wavelet transform: moving beyond uni- and bivariate analysis. *Journal of Economic Survey*, 28, 344-375.
- Adelman, M. A. (1984). International oil agreements. *The Energy Journal*, 5(3), 1-9.
- Asche, F., Osmundsen, P., & Tveterås, R. (2002). European market integration for gas? Volume flexibility and political risk. *Energy Economics*, 24(3), 249-265.
- Asche, F., Gjølberg, O., & Völker, T. (2003). Price relationships in the petroleum market: an analysis of crude oil and refined product prices. *Energy Economics*, 25(3), 289-301.
- Asche, F., Osmundsen, P., & Sandsmark, M. (2006). The UK market for natural gas, oil and electricity: are the prices decoupled?. *The Energy Journal*, 27-40.
- Aye, G. C., Carcel, H., Gil-Alana, L. A., Gupta, R. (2017). Does gold act as a hedge against inflation in the UK? Evidence from a fractional cointegration approach over 1257 to 2016. *Resources Policy*, 54, 53-57.
- Bachmeier, L. J., & Griffin, J. M. (2006). Testing for market integration crude oil, coal, and natural gas. *The Energy Journal*, 55-71.
- Barros, C.P., L.A. Gil-Alana and J.E. Payne (2012), Evidence of long memory behavior in U.S. renewable energy consumption, *Energy Policy* 41, 822-826.
- Baruník, J., Dvořáková, S. (2015). An empirical model of fractionally cointegrated daily high and low stock market prices. *Economic Modelling*, 45, 193-206.
- Belbute, J. and A.M. Ferreira, (2016). Does Final Energy Consumption in Portugal Exhibit Long Memory? A Fractional Integration Analysis. *Portuguese Economic Journal*. 15 (2), 59-77.B
- Bentzen, J. (2007). Does OPEC influence crude oil prices? Testing for co-movements and causality between regional crude oil prices. *Applied Economics*, 39(11), 1375-1385.
- Bloomfield, P. (1973), An exponential model in the spectrum of a scalar time series, *Biometrika* 60, 217-226.

- Borenstein, S., & Kellogg, R. (2014). The incidence of an oil glut: who benefits from cheap crude oil in the Midwest?. *The Energy Journal*, 15-33.
- Bozoklu, S., V. Yilanci and M.S. Gorus, (2020), Persistence in per capita energy consumption: a fractional integration approach with a Fourier function, *Energy Economics* 91, 104926.
- Caro, J. M. B., Golpe, A. A., Iglesias, J., & Vides, J. C. (2020). A new way of measuring the WTI–Brent spread. Globalization, shock persistence and common trends. *Energy Economics*, 85, 104546.
- Chang, C. L., McAleer, M., & Tansuchat, R. (2010). Analyzing and forecasting volatility spillovers, asymmetries and hedging in major oil markets. *Energy Economics*, 32(6), 1445-1455.
- De Vany, A. S., & Walls, W. D. (1999). Cointegration analysis of spot electricity prices: insights on transmission efficiency in the western US. *Energy Economics*, 21(5), 435-448.
- Dewandaru, G., Masih, R. and Masih, A. M. M. (2016). Contagion and interdependence across Asia-Pacific equity markets: An analysis based on multi-horizon discrete and continuous wavelet transformations. *International Review of Economics and Finance*, 43, 363-377.
- Dolatabadi, S., Nielsen, M. Ø., Xu, K. (2016), A fractionally cointegrated VAR model with deterministic trends and application to commodity futures markets, *Journal of Empirical Finance* 38, 623-629.
- Dolatabadi, S., Narayan, P. K., Nielsen, M. Ø., Xu, K. (2018). Economic significance of commodity return forecasts from the fractionally cointegrated VAR model. *Journal of Futures Markets*, 38(2), 219-242.
- Fattouh, B. (2010). The dynamics of crude oil price differentials. *Energy Economics*, 32(2), 334-342.
- Gil-Alana, L. A., & Hualde, J. (2009). Fractional integration and cointegration: an overview and an empirical application. In *Palgrave handbook of econometrics* (pp. 434-469). Palgrave Macmillan, London.
- Gil-Alana, L. A., Carcel, H. (2020). A fractional cointegration var analysis of exchange rate dynamics. *The North American Journal of Economics and Finance* 51, 100848
- Gil-Alana, L.A., M. Martin-Valmayor and P. Wanke (2020a), The relationship between energy consumption and prices. Evidence from futures and spot markets in Spain and Portugal, *Energy Strategy Review* 31, 100522.
- Gil-Alana, L.A., S. Sakiru and C. Lafuente (2020b), An investigation of long range reliance on shale oil and shale gas production in the US, *Energy* 195, 15.

- Gjolberg, O., & Johnsen, T. (1999). Risk management in the oil industry: can information on long-run equilibrium prices be utilized?. *Energy Economics*, 21(6), 517-527.
- Granger, C.W.J., (1980), Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics*, 14, 227-238.
- Granger, C.W.J., 1981, Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, 16, 121-130.
- Granger, C.W.J. and R. Joyeux, 1980, An introduction to long memory time series and fractional differencing, *Journal of Time Series Analysis* 1, 15-29.
- Gunay, S. (2018), Fractionally Cointegrated Vector Autoregression Model: Evaluation of High/Low and Close/Open Spreads for Precious Metals, *SAGE Journals* 8, 4.
- Hosking, J.R.M., 1981, Fractional differencing, *Biometrika*, 68, 168-176.
- Gülen, S. G. (1997). Regionalization in the world crude oil market. *The Energy Journal*, 109-126.
- Gülen, S. G. (1999). Regionalization in the world crude oil market: further evidence. *The Energy Journal*, 125-139.
- Hammoudeh, S., & Li, H. (2004). The impact of the Asian crisis on the behavior of US and international petroleum prices. *Energy Economics*, 26(1), 135-160.
- Jammazi, R., Ferrer, R., Jareño, F., and Shahzad, S. J. H. (2017). Time-varying causality between crude oil and stock markets: What can we learn from a multiscale perspective?. *International Review of Economics and Finance*, 49, 453-483.
- Johansen, S. (1996). *Likelihood-based inference in cointegrated vector autoregressive models*. New York, NY: Oxford University Press.
- Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory*, 24, 651-676.
- Johansen, S. and M. O. Nielsen (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* 158, 51-66.
- Johansen, S. and M.O. Nielsen (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* 80, 2667-2732.
- Johansen, S. and M. O. Nielsen (2016). The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models. *Econometric Theory* 32, 5, 1095-1139
- Jones, M. E., Nielsen, M. Ø., Popiel, M. K. (2014). A fractionally cointegrated VAR analysis of economic voting and political support. *Canadian Journal of Economics*, 47, 1078-1130.

- Kilian, L. (2016). The impact of the shale oil revolution on US oil and gasoline prices. *Review of Environmental Economics and Policy*, 10(2), 185-205.
- Lin, S. X., & Tamvakis, M. N. (2001). Spillover effects in energy futures markets. *Energy Economics*, 23(1), 43-56.
- Maciel, L. (2017). Technical analysis based on high and low stock prices forecasts: Evidence for Brazil using a fractionally cointegrated VAR model. *Empirical Economics*, 1-28.
- Milonas, N. T., & Henker, T. (2001). Price spread and convenience yield behaviour in the international oil market. *Applied Financial Economics*, 11(1), 23-36.
- Monge, M., Gil-Alana, L. A., & de Gracia, F. P. (2017). US shale oil production and WTI prices behaviour. *Energy*, 141, 12-19.
- Nielsen, M. Ø., Popiel, M. K. (2018). A Matlab program and user's guide for the fractionally cointegrated VAR model (Queen's Economics Department Working Paper No 1330). Ontario, Canada, K7L 3N6.
- Pinto AA, Gamba EA, Yannacopoulos AN, Herves-Beloso C. In: Trends in mathematical economics: dialogues between southern Europe and Latin America. Springer; 2016.
- Poza, C., & Monge, M. (2020). A Real Time Leading Economic Indicator based on Text Mining for the Spanish Economy. *Fractional Cointegration VAR and Continuous Wavelet Transform Analysis*. *International Economics*.
- Quineche, R. (2020), Consumption, Aggregate Wealth and Expected Stock Returns: An FCVAR Approach, *Journal of Time Series Econometrics* 13, 1, 21-42.
- Robinson, P.M. (1994) Efficient tests of nonstationary hypotheses, *Journal of the American Statistical Association* 89, 1420-1437.
- Rodriguez, A. E., & Williams, M. D. (1993). Is the world oil market 'one great pool'? A test. *Energy Studies Review*, 5(2), 121-130.
- Rodriguez, A. E., & Williams, M. D. (1994). The world oil market is "one great pool"? A response. *Energy Studies Review*, 5(3), 231-235.
- Serletis, A. (1994). A cointegration analysis of petroleum futures prices. *Energy Economics*, 16(2), 93-97.
- Solarin, S.A., L.A. Gil-Alana and U. AL-Muladi (2018), Stochastic convergence of renewable energy consumption in OECD countries: a fractional integration approach, *Environmental Science and Pollution Research* 25, 18, 17289-17299.
- Tiwari, A. K., Mutascu, M. I. and Albulescu, C. T. (2016). Continuous wavelet transform and rolling correlation of European stock markets. *International Review of Economics & Finance*, 42, 237-256.

Vacha, L. and Barunik, J. (2012). Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis. *Energy Economics*, 34, 241-247.

Wlzlowski, S., Hagströmer, B., & Giuliatti, M. (2011). Causality in crude oil prices. *Applied Economics*, 43(24), 3337-3347.

Weiner, R. J. (1991). Is the world oil market "one great pool"? *The Energy Journal*, 95-107.

Weiner, R. J. (1993). The world oil market is not one great pool: A reply to Rodriguez and Williams. *Energy Studies Review*, 5(3), 225-230.

Figure 1: Map of Petroleum Administration for Defense Districts (PADDs).
Source: EIA.

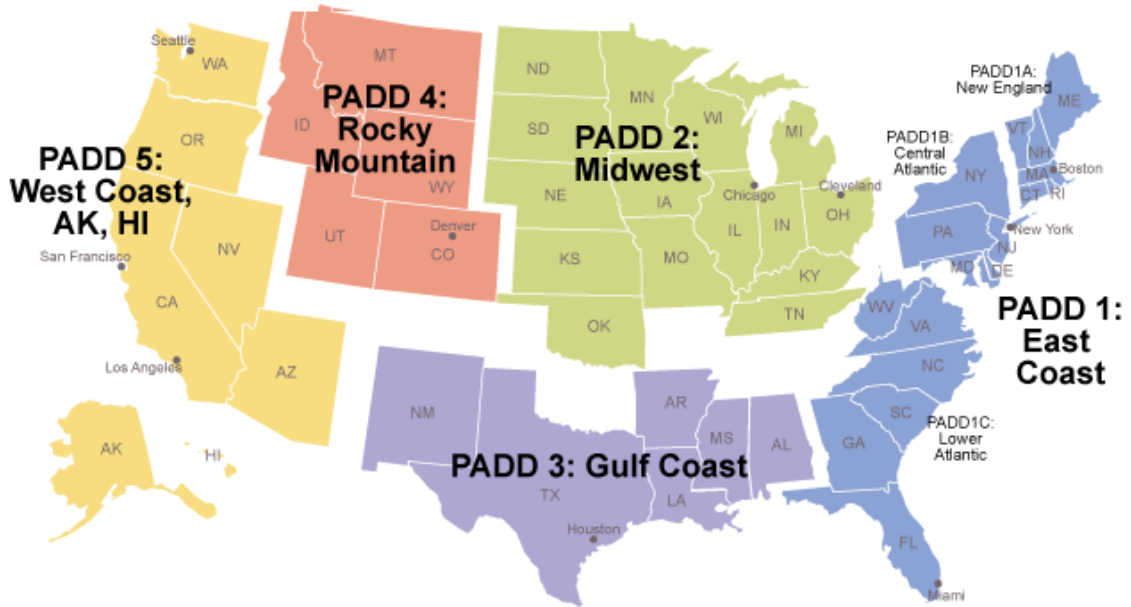
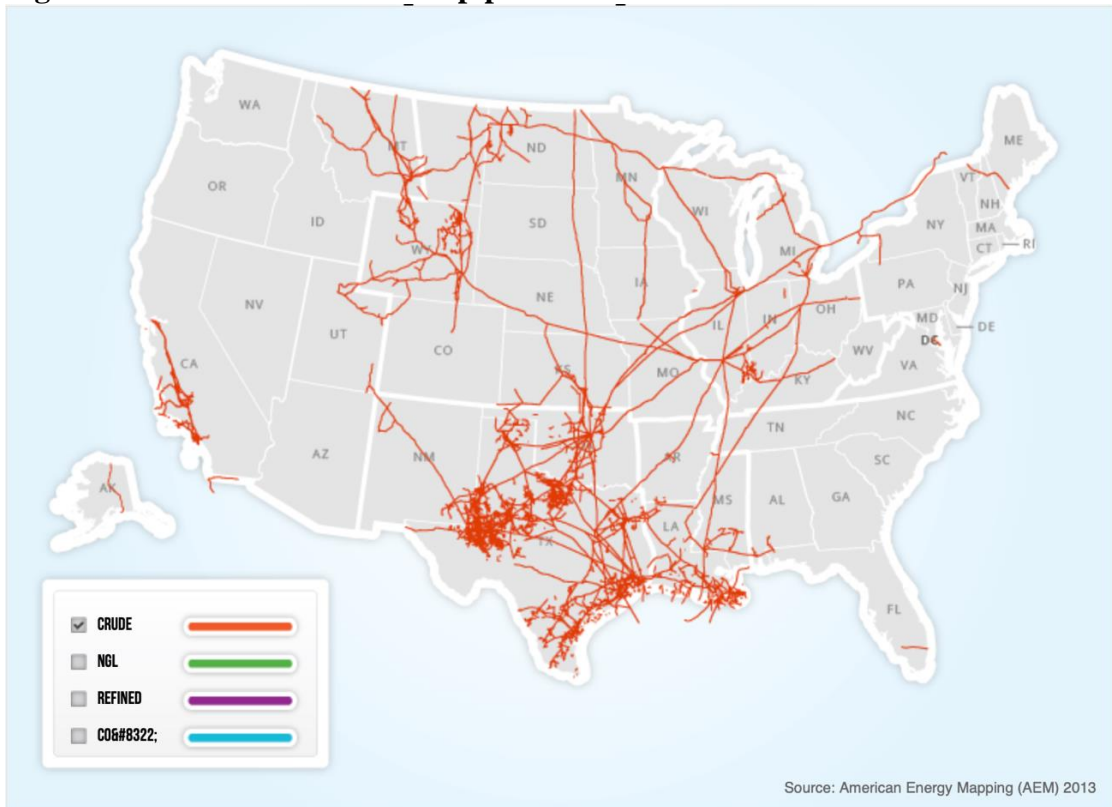


Figure 2: United States crude oil pipelines.



Source: American Energy Mapping (AEM) 2013

Figure 3: U.S. shale oil productivity

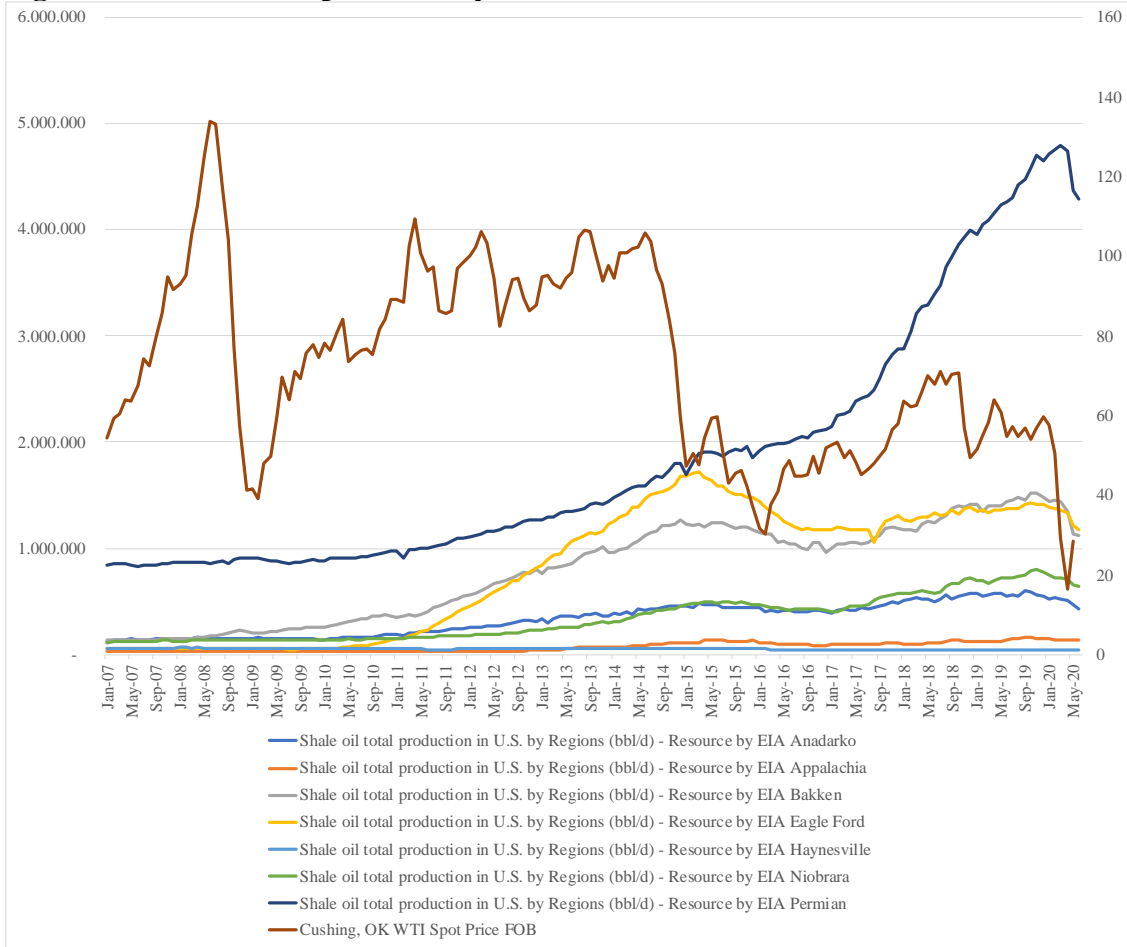


Figure 4. Time series and Wavelet Power Spectrum plots

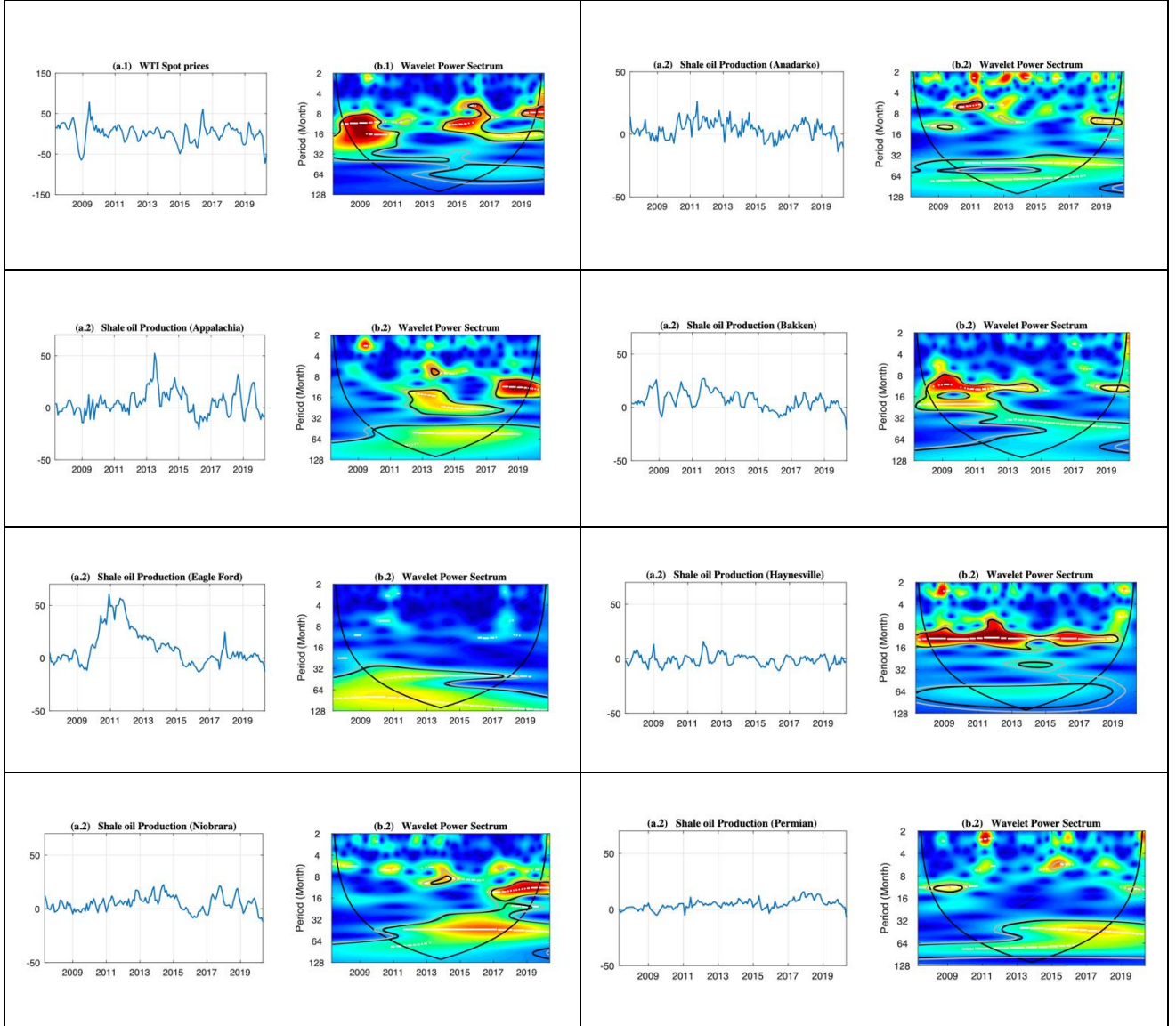
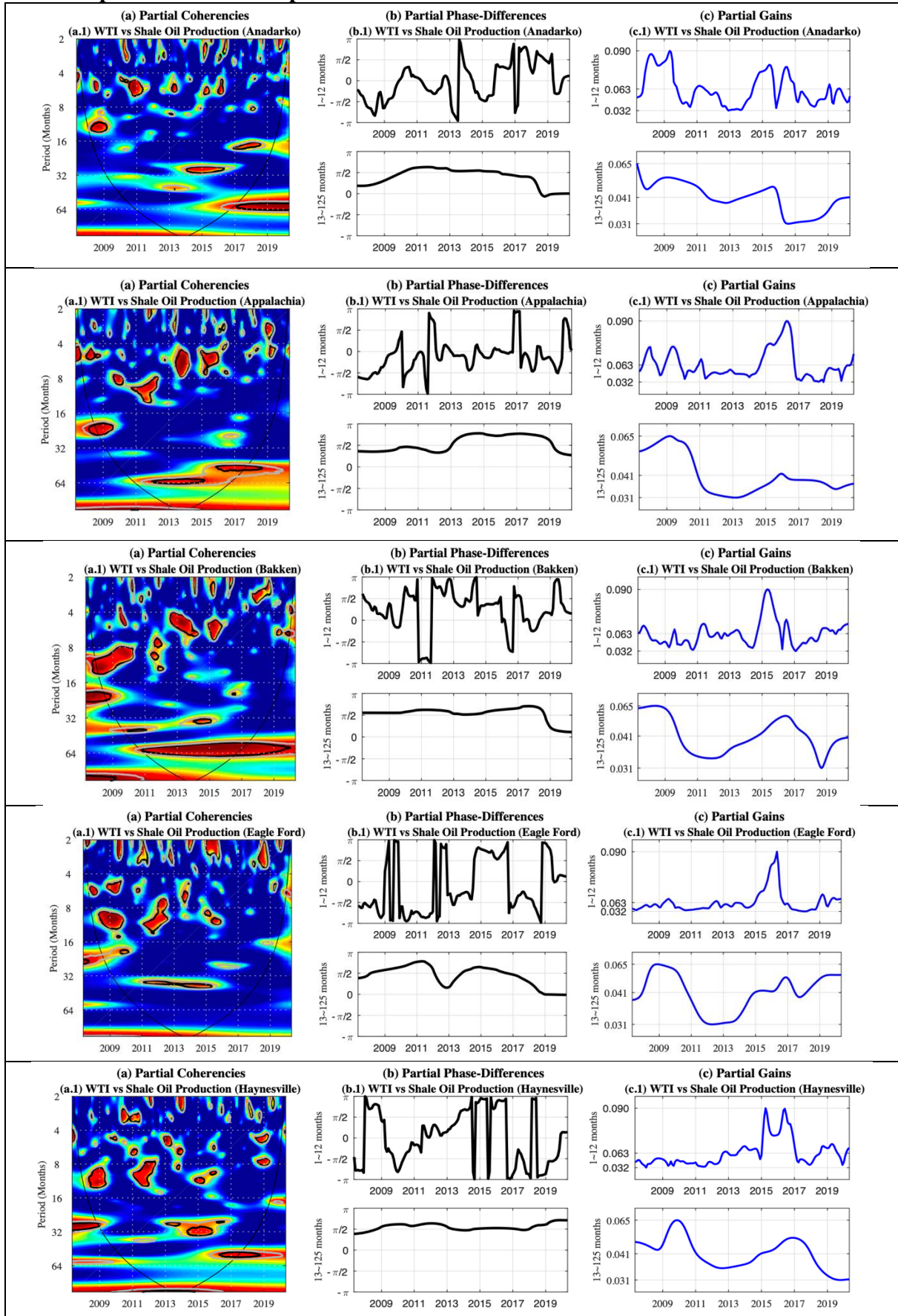


Figure 5. Wavelet Coherency, phase-differences and wavelet gain between WTI crude oil price and shale oil production in U.S.



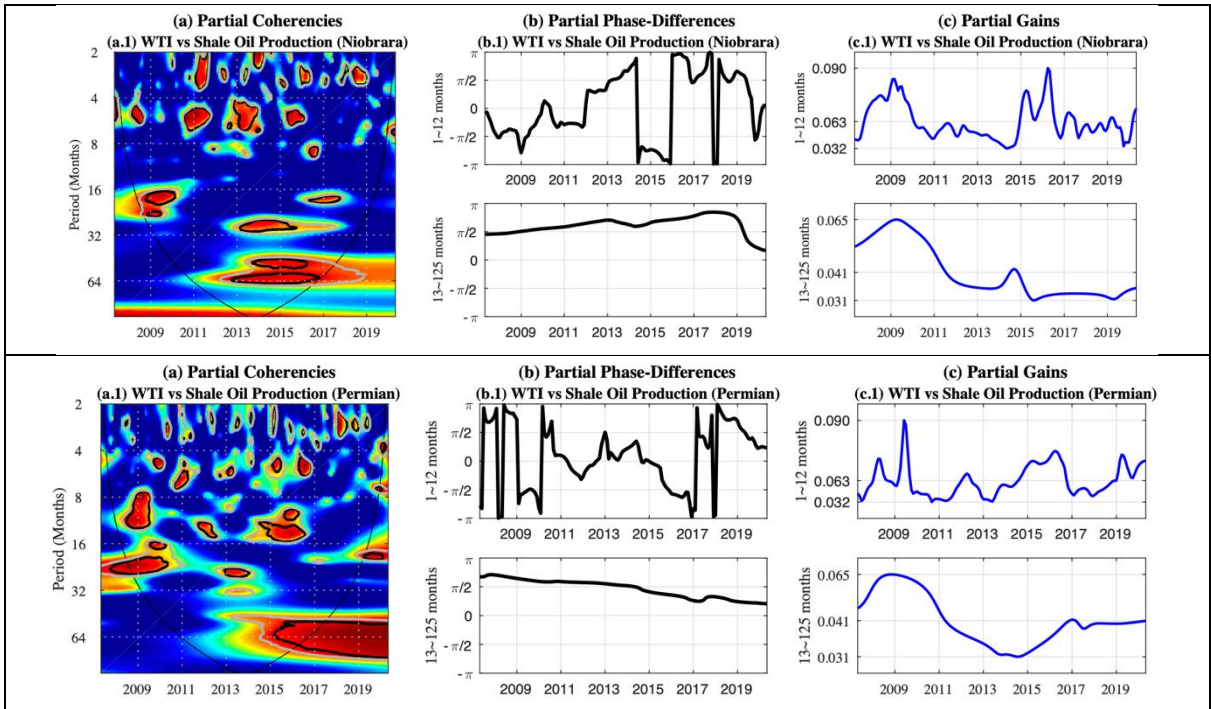


Table 1: Estimated values of the differencing parameter d: White noise residuals

	No regressors	An intercept	A time trend
Anadarko	0.96 (0.86, 1.09)	0.91 (0.84, 0.98)	0.90 (0.83, 0.98)*
Appalachia	0.97 (0.87, 1.10)	1.09 (1.01, 1.19)	1.09 (1.01, 1.19)
Bakken	0.95 (0.85, 1.08)	1.24 (1.15, 1.36)	1.23 (1.14, 1.35)
Eagle Ford	0.97 (0.86, 1.10)	1.41 (1.35, 1.49)	1.41 (1.35, 1.49)
Haynesville	0.97 (0.88, 1.11)	0.95 (0.85, 1.08)	0.95 (0.85, 1.08)
Niobrara	0.96 (0.86, 1.10)	1.17 (1.08, 1.30)	1.18 (1.08, 1.30)
Permian	0.97 (0.87, 1.10)	1.07 (1.01, 1.16)	1.08 (1.01, 1.18)
Oil prices	1.04 (0.91, 1.23)	1.15 (0.99, 1.39)	1.15 (0.99, 1.38)

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

Table 2: Estimated values of the differencing parameter d: Autocorrelated (Bloomfield) residuals

	No regressors	An intercept	A time trend
Anadarko	0.92 (0.74, 1.13)	1.12 (0.99, 1.30)	1.12 (0.99, 1.30)
Appalachia	0.93 (0.74, 1.17)	1.21 (1.06, 1.41)	1.22 (1.06, 1.42)
Bakken	0.88 (0.73, 1.09)	1.23 (1.08, 1.41)	1.21 (1.06, 1.41)
Eagle Ford	0.90 (0.70, 1.12)	1.66 (1.54, 1.81)	1.67 (1.54, 1.85)
Haynesville	0.94 (0.78, 1.16)	0.94 (0.78, 1.22)	0.94 (0.77, 1.23)
Niobrara	0.90 (0.74, 1.15)	1.12 (0.97, 1.31)	1.13 (0.97, 1.34)
Permian	0.91 (0.76, 1.15)	1.22 (1.09, 1.43)	1.27 (1.13, 1.46)
Oil prices	0.77 (0.64, 0.96)	0.79 (0.61, 1.07)	0.80 (0.62, 1.07)

Table 3: Estimated values of the differencing parameter d: Seasonal monthly AR(1) residuals

Series	No regressors	An intercept	A time trend
Anadarko	0.96 (0.85, 1.10)	0.89 (0.83, 0.98)	0.89 (0.82, 0.98)*
Appalachia	0.97 (0.86, 1.10)	1.07 (0.99, 1.18)	1.08 (0.99, 1.18)
Bakken	0.95 (0.84, 1.08)	1.23 (1.13, 1.35)	1.22 (1.13, 1.34)
Eagle Ford	0.97 (0.86, 1.10)	1.40 (1.33, 1.49)	1.40 (1.33, 1.49)
Haynesville	0.97 (0.86, 1.11)	0.91 (0.81, 1.05)	0.91 (0.80, 1.05)
Niobrara	0.96 (0.86, 1.10)	1.16 (1.06, 1.28)	1.16 (1.06, 1.28)
Permian	0.97 (0.85, 1.10)	1.07 (1.00, 1.16)	1.07 (0.99, 1.18)
Oil prices	1.0e4 (0.91, 1.23)	1.15 (0.99, 1.38)	1.15 (0.99, 1.38)

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

Table 4: FCVAR results

Series	d	b	μ_1	μ_2	Γ_1	Γ_2
Anadarko	0.894	0.010	56.401	126398.04	13781.81	28756598.99
Appalachia	0.414	0.414	57.929	28272.341	0.172	73.549
Bakken	0.518	0.518	54.805	136273.457	0.033	279.164
Eagle Ford	0.971	0.687	62.465	51390.41	-0.285	905.37
Haynesville	1.412	0.010	56.382	59957.63	-70368.42	21979520.57
Niobrara	1.361	0.914	56.024	111845.25	-0.256	204.72
Permian	0.634	0.366	55.574	859239.32	0.006	100.48