Closer to one great pool? Evidence from structural breaks in oil price differentials *

Michael Plante[†] Grant Strickler [‡]

August 6, 2019

Please do not distribute or cite without permission from author.

Abstract

We show that the oil market has become closer to "one great pool," in the sense that price differentials between crude oils of different qualities have generally become smaller over time. We document, in particular, that many of these quality differentials experienced a major structural break in or around 2008, after which there was a marked reduction in their means and, in many cases, volatilities. Differentials between residual fuel oil, a low-quality fuel, and higher-valued products, such as gasoline and diesel, experienced similar breaks during the same time period. Several factors explain these shifts, including a growing ability of the global refinery sector to process lower-quality crude oil and the U.S. shale boom, which has unexpectedly boosted the supply of high-quality crude oil. Differentials between crude oils of similar quality in general did not experience breaks in or around 2008, although we do find evidence of breaks at other times.

Keywords: crude oil price differentials, oil, structural breaks, refining

JEL Codes: Q40, C22

*This work has benefited from comments by participants at The Society for Nonlinear Dynamics and Econometrics 2019 conference, the 42nd International Association for Energy Economics International Conference, and the Commodity and Energy Markets Association Annual Meeting 2019. We also thank Julie Harris and John Powell at the U.S. Energy Information Administration for useful discussions on refineries and U.S. refining data, as well as Mark Agerton, Alexander Chudhik, Reinhard Ellwanger, Kei-Mu Yi and Mine Yucel for helpful comments. Any errors or omissions are our own. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

[†]Corresponding author, Federal Reserve Bank of Dallas, 2200 N Pearl St, Dallas, TX 75201, michael.plante@dal.frb.org.

[‡]Federal Reserve Bank of Dallas, 2200 N Pearl St, Dallas, TX 75201, grant.strickler@dal.frb.org.

1 Introduction

In 1984 Morris Adelman famously wrote, “The world oil market, like the world ocean, is one great pool.” If this were literally true, it would imply that all crude streams would be perfectly substitutable for one another in the refining process. We would then expect to see generally small price differentials between different crude streams over the long-run, reflecting primarily transportation costs. In reality, crude oil can have a wide range of physical properties that play an important role in how one crude stream is priced relative to another, and one often observes large price differentials between crude streams of different qualities.

These quality differentials are important for many oil market participants. For refiners, they can affect profitability and influence investment decisions about specific equipment, such as cokers, that could improve the profitability of processing lower grades of crude.[1] Oil producers and fiscal authorities are concerned about these differentials because of the impacts they can have on revenues earned from producing or taxing certain types of oil.[2] They can also affect a government’s choice of the benchmark used to set official selling prices.[3] Finally, for academics, analysts and others interested in understanding the upstream and downstream oil markets, these differentials provide important signals about how supply and demand conditions change over time for one type of crude relative to others.

In a certain sense, these quality differentials reflect limits on the global refining sector’s ability to treat various crude streams as substitutes for one another when it comes to transforming them into the valuable petroleum products that consumers desire. In this paper, our question of interest is whether the average values of crude oil quality differentials have declined over time. That is, can we find evidence that crude oils of different types may have become more substitutable for one another and that the oil market has become closer to “one great pool”?

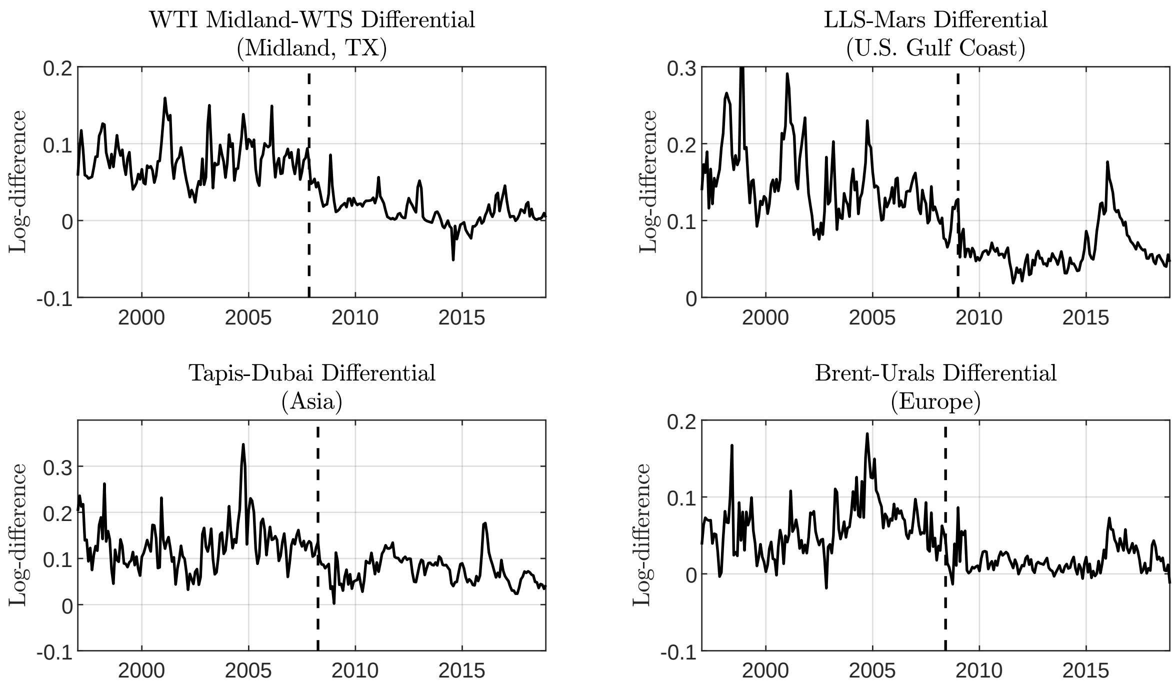
To answer this question, we construct price differentials between numerous crude oils of various types and then test whether these differentials have experienced shifts in their means using a structural breakpoint test. While it is well known in the industry and literature that changing market conditions can cause short-run variations in quality differentials, little has been said about whether they have been affected by structural breaks that have more

[1] This topic has received attention from trade press and market analysts since at least the early 2000s. More recently, the shale boom and IMO 2020 have renewed interest in these issues. See, for example, Evans and Mowler (2002), PIW (2005), Piotrowksi (2009), Tuttle (2019) and recent analysis on IMO 2020 by the U.S. Energy Information Administration and the International Energy Agency.

[2] For example, Khrennikova and Mazneva (2018) and DiPaola (2019).

[3] See, for example, Kemp (2009) on the 2009 Saudi Aramco decision to switch its benchmark from West Texas Intermediate Crude to the Argus Sour Crude Index.

Figure 1.1: Oil price differentials in four areas



Notes: Figure plots log (percent) differentials between a higher and lower grade crude oil from 1997 to 2018 in four areas of the world, calculated as $\ln P_{high} - \ln P_{low}$. West Texas Intermediate (WTI), Louisiana Light Sweet (LLS), Tapis and Brent are light, sweet crudes. West Texas Sour (WTS) is light, sour, while Dubai, Mars and Urals are medium, sour crudes.

permanently changed their average levels.

To provide some motivation for our interest in structural breaks, we plot in Figure 1.1 examples of differentials between higher and lower grade crudes for four parts of the world: Midland, TX; the U.S. Gulf Coast; Europe; and Asia. These are log-differentials using monthly data from 1997 to 2018, considering the price of the high-quality crude relative to a lower-quality one. Visually, there is strong evidence of at least one break in the means of these differentials, occurring sometime around 2007 or 2008. The vertical lines denote the breakpoints as determined by a more formal statistical test. Visual inspection of other quality-related differentials, not shown here, strongly suggests the existence of structural breaks in many of their means, as well.

Our first contribution is to document more rigorously and systematically the extent to which differentials between crude oils of different types have experienced structural breaks in their means. Using the sequential breakpoint test of Bai (1997), we find that almost all of the differentials we look at have experienced at least one break in their mean. In particular, a large number of these quality differentials—25 out of 27 cases to be exact—experienced a significant break around 2007 and 2008. We then investigate how the means have shifted over time and find that most quality differentials have narrowed. In many cases, particularly those that experienced a break around 2008, a major drop in volatility accompanied the reduction in the mean. After the break, the means and volatilities are often half of their pre-break levels. Related to these findings, we also show that a differentials between residual fuel oil, a low-quality fuel produced in greater abundance in low-grade crude oil, and higher-valued petroleum products have narrowed significantly following a set of breaks that occurred around 2008.

Price differentials between different types of oil reflect the limits to arbitrage that exist across crude oil quality. A major decline in their average values suggests dramatically fewer opportunities than existed in the 1990s and parts of the 2000s for arbitrage involving substitution of high and low-quality crudes. We discuss several possible explanations for why this has happened. One factor is the continued global buildup of more complex refineries, which have the ability to increase the production of high-value petroleum products, such as gasoline and diesel, from low-grade crude oils. Another is the U.S. shale oil boom, which has unexpectedly boosted the supply of high-grade, light crude oil. This has reduced, on the margin, the need for more complex refineries to process low-grade crude oils. These factors have apparently been strong enough to reduce the differentials, despite the fact that over time changes in environmental regulations and falling demand for residual fuel oil should have led to wider differentials.

Regarding the timing of the cluster of breaks, we find some evidence that the Great Recession played a role in this outcome. We first present some evidence that utilization of refining capacity to upgrade lower-quality crude oils was at relatively high levels prior to the Recession. We then show that there were some significant capacity additions during the Recession, which occurred during a period of falling demand for petroleum products, due to the Recession itself. Together, these two forces sharply reduced utilization rates for upgrading capacity in 2008 and 2009. Although demand for high-valued petroleum products, such as gasoline and diesel, has continued to grow since then, additional upgrading capacity appears to have been sufficient to meet incremental demand, keeping utilization rates at more modest levels and quality differentials at relatively low levels.

Finally, we also investigated whether oil price differentials between crudes of the same type, for example, two light, sweet crude oils, experienced a similar set of breaks, particularly around 2008. If that were true, it would suggest a broader change in the oil market not necessarily connected to quality. Overall, we do not find any evidence for this. We do, however, find that differentials between similar-type crude oils have experienced their own set of breaks. Many appear connected to changing market conditions in the United States, occurring either in the mid-2000s or after 2010, and affecting numerous differentials related to crude oils in the U.S. Gulf Coast, particularly light, sweet crude oil. A modest contribution on our part is to show that these breaks are more prevalent than previously documented in the literature.

We note that our work is connected with previous research papers, such as Weiner (1991), Sauer (1994), Ripple and Wilamoski (1995), Gülen (1997), Gülen (1999), and Bachmeier and Griffin (2006) that have considered to what extent Adelman's statement holds true. Those works have mainly looked at the degree to which oil prices move together across space and time, often using cointegration models. Or, to elaborate on the Adelman's metaphor, these works ask if there is a disturbance in one part of the pool that generates waves (price movements), do the waves spread out and affect other parts of the pool?

In our work, we consider the idea that long-run price differences also tell us something about how close the oil market is being "one great pool" but from the quality perspective. Or, elaborating on the metaphor again, to what extent can the global refining sector literally just dip into any part of the pool to get the crude oil it needs? Because of the different perspective, we focus on structural breaks in the long-run average size of quality differentials rather than modeling dynamics using cointegration models.[4]

Prior work in the literature has also discussed the occurrence and importance of structural breaks affecting differentials related to key benchmarks for light, sweet crude, such as West Texas Intermediate (WTI) and Brent. See, for example, Buyuksahin et al. (2013), Borenstein and Kellogg (2014), Scheitrum et al. (2018), and Agerton and Upton Jr. (2019). In contrast to those works, our main focus is on price differentials between crude oils of differing qualities. We also provide additional insight into the role of the downstream (refining) sector as an important market factor. The previous works have focused primarily on the implications of supply booms and logistical bottlenecks, i.e. issues related to the upstream and midstream oil sectors.

The rest of the paper is organized as follows. A brief introduction to crude quality and oil price differentials is contained in Section 2. In Section 3 we discuss our data and econometric methodology. Section 4 presents evidence regarding the presence of structural breaks and documents how they have affected the differentials. Section 5 discusses some potential explanations for our findings. We then conclude.

[4] Another line of work has shown the usefulness of threshold regression models when modeling the dynamic behavior of crude oil price differentials, for example Hammoudeh et al. (2008), Fattouh (2010) and Ghoshray and Trifonova (2014). Our work focuses on structural breaks in the means of price differentials, as opposed to modeling the dynamics of those differentials.

2 Crude oil properties and price differentials

While the previous literature has found that oil prices tend to move together over time, i.e. they are cointegrated, crudes usually do not sell for the same price due to differences in their physical characteristics. Two properties of particular importance are a crude oil's American Petroleum Institute gravity, hereafter API gravity, and sulfur content.[5] The industry has found it convenient to lump different crude oils into several major groups based on these properties. It is common to label oils as light, medium or heavy depending upon their API gravity and sweet or sour depending upon whether they have low or high sulfur content.

There is a hierarchy of quality in terms of density, with light at the top and heavy at the bottom, and in terms of sulfur content with sweet crudes preferred to sour. In terms of prices, light, sweet crudes usually sell at a premium to other grades, while heavy, sour crude oils usually sell at a discount to all other grades. In this section we discuss why these physical characteristics generate such price differentials and how refineries can influence these differentials. Although it is not the focus of our paper, we briefly discuss how transportation costs, the direction of trade and infrastructure issues can also influence price differentials as these factors play a role in some of the results we present later.

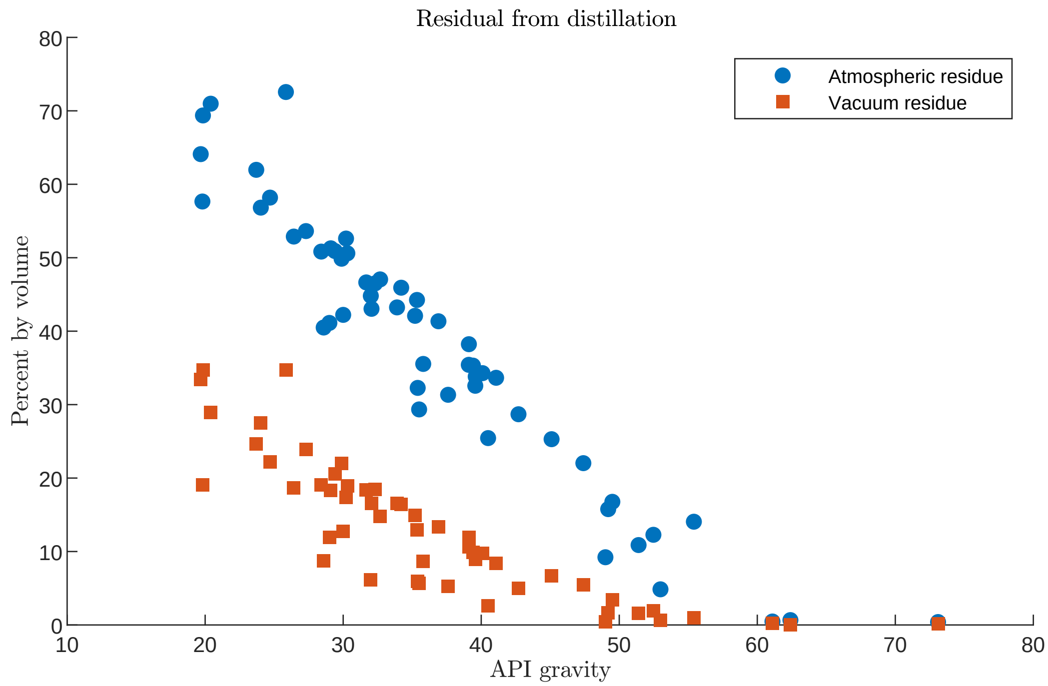
2.1 The refining process and API gravity

The first step of refining crude oil involves using an atmospheric distillation unit, also referred to as a crude distillation unit (CDU), to distill the crude into various "cuts" or fractions. All refiners, from the simplest to most complex, undertake this step. In general terms, it is helpful to imagine that every crude oil can be distilled into three fractions: light products (naphtha/gasoline), middle distillates (diesel/gas oil) and a residual, often referred to as atmospheric residue, which is literally the bottom of the barrel. These categories are determined by their boiling points and density, with light products possessing the lowest densities and boiling points and the atmospheric residue possessing the highest density and boiling point.

The API gravity of a crude is related to the proportions of the different cuts found in a specific crude oil. Light crudes, i.e. those with a high API gravity, tend to have greater proportions of gasoline and diesel than residual products, while medium and heavy crude oils usually contain greater amounts of residual products. The exact proportions for a specific crude oil are sometimes publicly available in the form of a chemical analysis known as a crude oil assay; we now use some of these analyses to discuss the relationship between API gravity and the residual content. As examples, the inherent yields of atmospheric residue for West

[5] API gravity for most crudes is a number between 10 and 70. The lower the value, the denser the oil.

Figure 2.1: Heavy crude oils typically contain greater volumes of residual



Notes: Figure plots the amount by volume of atmospheric residue and vacuum residue present as a function of API gravity for 54 crude oils. The data comes Exxon's library of crude oil assays.

Texas Intermediate (WTI) and Brent, two benchmarks for light, sweet crude, are 33.3 and 34.2 percent, respectively. Mars, a benchmark medium, sour crude in the U.S. Gulf Coast, contains about 47 percent residual while Maya, a heavy, sour crude produced by Mexico, has 61.2 percent residual.[6] The circles in Figure 2.1 show the relationship between API gravity and the amount of atmospheric residue present for 54 crude oils.

A major difference between simple and more complex refineries is in the latter's ability to transform the bottom of the barrel into other petroleum products. A simple refinery will have essentially no ability to do so. More complex refineries have additional machinery to convert the residual into higher valued petroleum products. Collectively, this capital is often referred to as secondary processing units, upgrading capacity or conversion capacity.

Moderately complex refineries will have a vacuum distillation unit (VDU), which further distills the residual from the CDU into vacuum gas oil (VGO) and vacuum residue, which is essentially residual fuel oil. They will also have crackers, equipment that processes the VGO into lighter products. The leftover, i.e. the residual fuel oil, is produced in greater concentrations in lower-quality crudes. The vacuum residue/residual fuel oil component of

[6] These are based on assays from the Oil&Gas Journal (08/15/1994), Exxon, BP and the Oil&Gas Journal (05/15/2000), respectively. Atmospheric residual here has a boiling point over 650 degrees Farenheit and includes both vacuum gas oil and residual fuel oil. For Mars, the boiling point listed is 696 degrees.

WTI is 9 percent, Brent 9.7 percent, Mars about 25 percent, and Maya 36.9 percent. The squares in Figure 2.1 show the relationship between API gravity and vacuum residue for 54 crude oils.

The most complex refineries, in addition to a VDU and crackers, also have a coker. This expensive piece of capital equipment allows the refiner to break down the residual fuel oil left over from the VDU and transform it into lighter products and petroleum coke. This equipment significantly reduces the amount of residual fuel oil produced during the refining process.

While the relative value of different refined products varies over time, residual fuel oil generally sells at a much lower price than gasoline or diesel. This inherently makes medium and heavy crude less valuable than light crude. Complex refineries take advantage of this price differential by using crackers and cokers to increase the production of higher-valued products, while at the same time reducing the production of residual fuel oil from lower grade crude oils.

2.2 The refining process and sulfur content

Sulfur is a pollutant and environmental regulations in many countries require that various petroleum products meet strict specifications for sulfur content. Due to these policies, sulfur is usually removed during refining. This requires investment in desulfurisation units or hydrotreaters. These regulations, and the costs associated with compliance, lead to sour crude oils, i.e. those with high sulfur content, selling at a discount to sweet crude oils.

2.3 Important factors affecting differentials

We next use the information just presented to develop some intuition about the factors likely to play an important role in the determination of quality differentials. We note that our goal is not to introduce a full model of a refinery or the refining sector. Instead, we want to provide guidance on what data might be useful when trying to explain our empirical findings. We discuss five factors: the amount of upgrading capacity; utilization of upgrading capacity; relative supplies of different types of crude oil; relative demand for types of petroleum products; government regulations about sulfur.

Amount of upgrading capacity: Upgrading capacity gives refiners the flexibility to adjust their outputs from a particular type of crude, rather than having to change the quality of their crude input. Medium and heavy crudes offer the most flexibility because they have more residue than light crudes, but the capacity must be available to transform the residue.

Utilization of upgrading capacity: There are physical limits on the utilization of refining capacity, whether they be crude distillation capacity or upgrading capacity. If utilization rates hit these upper limits, capacity is effectively constrained and a refinery, or more generally the refining sector, may be unable to arbitrage away price differentials even if profitable to do so.

Relative supplies of different crudes: The proportion of the world's crude supply attributable to specific types of crude oil has direct implications for the potential supply of various petroleum products. For example, if most of the world's crude is heavy, the potential supply of residual fuel oil would be much greater than if most of the world's crude was light.

Relative demand of different petroleum products: The mix of products desired by end users is also important. For example, if the world desires large amounts of gasoline or diesel relative to residual fuel oil, there would be implications for the relative values of those two products and, consequently, the relative value of different types of crudes. Obviously, this is intimately connected with supplies of different crude oils and the ability of the refining sector to upgrade low-quality crude oil.

Government regulations on sulfur emissions: Since it is costly to remove sulfur during the refining process, government regulations on sulfur emissions also figure in. These regulations target particular products rather than particular crude types, but should depress the prices of sour crude oils relative to sweet crudes. The impact on differentials will, of course, depend upon the amount of available desulfurization capacity and how much sour crude is produced.

2.4 Transportation costs and infrastructure issues

Finally, although not the main focus of this paper, transportation costs, the direction of trade, and infrastructure issues can also play a role in price differentials. If an area is a net importer of a particular type of oil, say light, sweet crude oil, then the price of light, sweet crude in that area will build in the transportation cost associated with importing the marginal barrel. For example, up until the late 2000s, the price of light, sweet crude oil in the U.S., given by either Louisiana Light Sweet (LLS) or WTI, carried a premium over similar grade crudes produced in Europe, such as Brent. The literature has previously pointed out that light, sweet crude differentials along the Gulf Coast (and in Cushing) have been subject to change as the shale boom has dramatically increased the supply of light, sweet crude in the area (see, for example, Buyuksahin et al. (2013) and Agerton and Upton Jr. (2019)).

Infrastructure issues, such as pipeline bottlenecks, can also affect price differentials between crude oils. This is true both for crudes of different types as well as for a single crude stream that is priced in more than one location. Over the past 10 years, these issues have become particularly prominent in the pricing of crude oil in Canada and the U.S. As production has grown in those two countries, large differentials have emerged at various points in time that have affected crude oil prices in Canada, the Bakken and the Permian Basin, to name a few locations.

3 Data and methodology

3.1 Prices

We work with a set of 14 crude oil prices. Table 3.1 lists the crude oils along with their API gravity and sulfur content. The crudes are divided by location, which refers to the geographic area where the crude is priced. For the U.S. crudes, these groupings are straight forward. Waterborne crude oils outside the U.S. are broken into two groups: a Europe/Atlantic Basin group and a Middle East/Asia group. We assign Dubai and Oman into the same group as Tapis because Dubai has long been an important benchmark for a large amount of oil sold into the Asian market (Energy Intelligence Research (2009)).

The table also categorizes crude oils into light, medium or heavy and sweet or sour. There are no formal definitions for these categories but we define a light crude oil as any oil with an API above 33 while heavy crudes have an API below 25, while a sweet crude is defined as any with a sulfur content below 0.50 percent. We note here that these categories are intended to help make the analysis more manageable by grouping together crude oils of roughly the same characteristics. In reality, there is a continuum of quality. With that being said, our series include light sweet crudes, such as Brent and Louisiana Light Sweet (LLS); medium, sour crudes, such as Dubai and Mars; and one heavy, sour crude, Maya. We have tried to include a broad set of crude oils that, while not necessarily on par with a benchmark crude, are relatively well known to ensure that the price data is of reasonable quality.

All price series come from Bloomberg except for Urals, which comes from the HAVER database. We consider a common sample that runs from January 1997 to December 2018. We start in 1997 as that is the first year we have data available for Mars. Our data is daily except for Urals, which comes as monthly averages.[7] The data appendix provides the exact series name for each crude stream. Data on API gravity and sulfur content come from Bloomberg for all of the crude streams except Brent and Urals, which comes from Platts (2018).

[7] We have access to daily data for Urals but only from 2002 to 2013. Our main conclusions for Urals-related differentials are unchanged whether we use daily or monthly data. Details can be found in the appendix and working paper version of this paper.

One point worth mentioning is the lack of a price series for Canadian heavy crude oil. Given our topic of interest, it would seem natural to include such a price. We do not, however, because Bloomberg data for the current benchmark price, Western Canadian Select (WCS), only starts in 2008. Given this, we have decided to exclude WCS prices from the analysis.

Finally, we also present some additional results using monthly averages, which allows us to include two additional light, sweet crude prices: Algerian Saharan and Bonny Light. Details on the data and results can be found in the appendix.

Table 3.1: Oil price series

Name	API gravity	Sulfur	API category	Sulfur category
Cushing, OK				
WTI Cushing (WTIC)	39.0	0.34	Light	Sweet
Midland, TX				
WTI Midland (WTIM)	39.0	0.34	Light	Sweet
West Texas Sour (WTS)	34.0	1.90	Light	Sour
U.S. Gulf Coast (USGC)				
Heavy Louisiana Sweet (HLS)	33.7	0.39	Light	Sweet
Louisiana Light Sweet (LLS)	35.7	0.44	Light	Sweet
Mars	28.9	2.05	Medium	Sour
Maya	21.1	3.38	Heavy	Sour
Europe/Atlantic Basin				
Brent	38.1	0.41	Light	Sweet
Saudi Heavy to Europe (SHE)	27.0	2.80	Medium	Sour
Urals	31.5	1.44	Medium	Sour
Middle East/Asia				
Dubai	31.0	1.70	Medium	Sour
Oman	33.0	1.10	Medium	Sour
Saudi Heavy to Asia (SHA)	27.0	2.80	Medium	Sour
Tapis	44.6	0.03	Light	Sweet

3.2 Differentials

We consider log-differentials of the price series, as in Gülen (1997), Gülen (1999), Hammoudeh et al. (2008) and Fattouh (2010). If we denote the level of two arbitrary oil prices as P_i and P_j, the log-differential between them in month t is given by

$$p_{ij,t} = \ln P_{i,t} - \ln P_{j,t}. \tag{3.1}$$

The use of log-differentials has the advantage of converting units to percent differences. An additional benefit is that the log-differential is equivalent to the log of a relative price. As such we do not need to worry about the effects of inflation on the differential over time.

We generally construct the differentials so that $P_{i,t}$ is the higher-quality crude. For the daily data, we construct pair-wise differentials on all days where there is an observation for both prices, and exclude any day where we are missing one or both prices. The number of observations, therefore, varies slightly from differential to differential but, in general, we have roughly 5500 data points per differential. For the analysis using monthly data, some of the price series are only available as monthly averages. To ensure comparability across series, we take a monthly average of the daily price data when it is available. Differentials are then calculated based on the monthly averages.

Even with the limited number of price series we work with, there are a large number of differentials that can be constructed. We have found it convenient to break the differentials into two groups. The first grouping contains differentials between various crude streams within the same area, as defined in Table 3.1. We hereafter refer to this group as the within-area differentials. The second group consists of differentials between crude oils that are priced in different areas. We hereafter refer to these as the across-area differentials.

In addition to being convenient, the breakdown into within-area and across-area differentials also has some intuitive appeal given our topic of interest. Over long periods of time, the within-area differentials, being priced closer to each other, should be less affected by transportation costs or infrastructure issues and better reflect the role of arbitrage across quality. The across-area differentials, on the other hand, should reflect not only differences in quality but also arbitrage across space. For example, the LLS-Dubai differential builds in not only the effect of quality but also the fact that LLS may be influenced more heavily by local conditions in the United States Gulf Coast (USGC) market, while Dubai may be influenced a bit more by conditions in Asia. On the other hand, we expect the LLS-Mars differential, in general, to be heavily influenced by arbitrage across quality in the USGC.

3.2.1 Within-area differentials

The within-area differentials are constructed starting with the crude oil that has the highest API gravity in the area and then working down. For example, in the USGC we construct differentials between LLS and the three other crudes. After LLS we calculate log-differentials between HLS and the two heavier crudes, Mars and Maya, and finally the differential between Mars and Maya. There are 16 differentials in this group.

3.2.2 Across-area differentials

We follow the same procedure as before and construct across-area differentials beginning with the highest quality crude, with the following exception: the differentials between light crudes in the USGC and light crudes outside the U.S. Due to the direction of trade at the start the sample, i.e. the Gulf Coast was a net importer of light crude, LLS and HLS sold at a premium to many light crudes outside of the USGC. We put LLS and HLS in the numerator of those differentials. We have also excluded all but two of the across-area differentials involving WTI Midland, WTI Cushing and WTS. These differentials show extreme changes in behavior after 2010 due to shale boom and pipeline bottlenecks and, as many of these issues have been discussed elsewhere, for brevity's sake we do not include them in our analysis. This procedure leads to a total of 27 across-area differentials.

3.2.3 Summary statistics

To conserve space, tables containing the full set of summary statistics are in the appendix. To summarize, we find that the differentials are typically larger for those pairs of crude streams that are further apart in terms of API gravity and sulfur content, in line with the intuition developed in section 2 and previous works.[8] For example, the mean of the LLS-HLS differential was only 1.5 percent while it was almost 23 percent for the LLS-Maya differential. A few differentials, primarily light, sweet crude differentials, do not follow this pattern as trade costs, the direction of trade and infrastructure issues play a large role in their average values. The LLS and HLS differentials with Brent are both positive while the two WTI-LLS differentials have negative means, for example. We also find that the greater the differences in API gravity and sulfur content, the more volatile the differential tends to be.

3.3 Methodology

There are numerous econometric methods available to test for structural breaks in a time series. We use the sequential breakpoint test of Bai (1997), which allows one to determine both the number of breaks present and their timing. Here, we provide a brief sketch of the procedure. For details on the theory we refer the reader to Bai (1997).[9] Critical values come from Bai and Perron (2003), which also provides a discussion on more practical matters related to various structural break tests. We note here that we make use of the repartition technique introduced in Bai (1997), which makes the asymptotic distributions of the

[8] See, for example, Bacon and Tordo (2005) and Giulietti et al. (2015).

[9] Perron et al. (2006) provides a more general overview of structural breaks.

sequential test equivalent to those of the simultaneous breakpoint tests of Bai and Perron (1998).

For each differential, we consider a model of pure structural change where we estimate regression equations of the following form,

$$p_{ij,t} = c_{ij} + u_{ij,t}, \tag{3.2}$$

and test for breaks in the intercept term, c_{ij}. This specification has the advantage of allowing for fairly general properties of the residual, including serial correlation.[10]

Time is denoted by t and the sample runs from 1 to T. There are m possible breaks and $M = m + 1$ regimes. The test requires us to choose a maximum number of breaks to be considered. Visual inspection of the data usually pointed to no more than three breaks but we allow for a maximum of five, i.e. $0 \leq m \leq 5$. The breakpoint test also requires us to choose a trimming parameter, ϵ, which controls the minimum number of observations allowed for each regime. More specifically, if h is the minimum observations allowed, $h = \epsilon T$. The trimming parameter can be set as low as 0.05 but we set ϵ to 0.15. As discussed in Bai and Perron (2003), the higher value helps mitigate against potential size distortions that can occur when the data are serially correlated. For our time series, the minimum regime size is a little over 3 years.

The first step of the procedure is to estimate the regression equation for a price differential using the full sample of data. The test searches for breaks over all allowable sub-samples and the null of no breaks versus one break is then considered for the candidate break that maximizes the test statistic.[11] We use the robust version of the test statistic found in Bai and Perron (1998) where the estimate of the variance-covariance matrix is robust to heteroscedasticity and autocorrelation. The matrix is estimated using the Quadratic Spectral kernel and the automatic bandwidth method of Andrews (1991).[12] If the null can be rejected at the 1 percent level, we accept the candidate break. The sample is then split in two at the estimated breakpoint and the procedure is repeated individually for the two sub-samples. This process continues until the null hypothesis cannot be rejected for any of the subsamples

[10] We also considered regression equations that explicitly modeled auto-correlation by including lags of the dependent variable. In that case the test statistics are only valid when there is no serial correlation in the residuals. In many cases, particularly with monthly data, we found it difficult to a priori properly determine the lag length, which is not surprising given the nature of the breaks we are investigating. Given our interest in testing for breaks in the mean and our concern about potential misspecification, we decided to work with the more parsimonious model in equation (3.2).

[11] Technically speaking, the test and its asymptotic properties are defined in terms of the break fraction rather than the breakpoints. We follow Bai (1997) and base our discussion around the breakpoints.

[12] In preliminary analysis we also considered the Bartlett kernel. We found that this generally led to smaller standard errors for the estimates and, as a result, somewhat more breaks being accepted by the test.

or until we find 5 breaks. When a candidate break is accepted, the initial estimates for breakpoints and break fractions are denoted as k_s^0 and $\tau_s^0 = k_s^0/T$ for $s = 1, ..., m$.

We make use of a refinement of the sequential procedure, called repartitioning, that is introduced in Bai (1997). This process re-estimates the dates for the breakpoints, modifying the sub-samples to take into account the initial breakpoints identified by the sequential procedure. In the case of two breaks, the repartition process re-estimates the breaks using the subsamples $[1, k_2^0]$ and $[k_1^0, T]$. The final estimates for the break fractions and breakpoints, after the repartition process, are denoted as τ_s^* and k_s^*, respectively. Under the repartition technique, the asymptotic distributions for the sequential test are the same as those of the simultaneous breakpoint tests discussed in Bai and Perron (1998).

There is a well known issue with these types of structural breakpoint tests where the test can fail to reject the null of no breaks versus 1 but finds evidence for rejecting the null of 1 versus 2 breaks. This occurs particularly when the series experiences a second break where the mean shifts back to a level close to its initial value. Visual inspections show that several of our series experience potential breaks of these types. As a result, in the few cases where the null of no breaks is not rejected we also consider the UDmax test described in Bai and Perron (1998). This test reports the maximum test statistic up to m breaks, in this case 5. If the UDmax test provides statistical evidence for more than 1 break we then report the results for all of the cases up to and including the last break which is statistically significant at a 1 percent level. This occurs for only 2 cases.

Finally, the breakpoint test relies on an assumption that the oil price differentials are stationary. However, as discussed in Perron (1989) and many papers since then, stationarity tests can be biased by the structural breaks we are interested in. To account for this, we tested the differentials using a breakpoint unit root test. The test is an Augmented Dickey-Fuller test where the lag length is chosen using the SIC. All of the differentials are found to be stationary after allowing for a break in their mean. Additional details can be found in the appendix.

4 Results

4.1 Identifying the structural breaks

Our first goal is to document the presence of structural breaks in the crude oil differentials. To begin, we focus specifically on pairs of crude oils of different types, first for the within-area differentials and then for the across-area differentials. Results for same-type crudes, such as the light, sweet differentials, are introduced after.

We begin with the WTIM-WTS differential, which is a differential between a light, sweet crude and a light, sour crude. The test identifies two breaks that are significant at a 1 percent level. The dates and test statistics are listed at the top of the upper panel in Table 4.1. The first break is in December 2007 and the second in February 2013. These dates refer to the month that contains the last day of a given regime. The F-statistic for the first break is 156.51 vs. a critical value of 12.29. The second break has a test statistic of 14.14 vs. a critical value of 13.89. We list the breaks in the order the test finds them, which is related to the size of the test statistic that each break generates for the null of 0 or 1 break.

The middle portion of the upper panel in Table 4.1 shows the identified breakpoints for the USGC. Our main finding is that there is strong evidence for a break in the mean of all the series sometime between mid-2007 to mid-2008. This is similar to the timing of the first break in the WTIM-WTS differential. We also find evidence for the existence of a second break at the end of 2001 in HLS-Mars differential. A similar break is detected for the LLS-Mars differential but is not listed as it is only significant at a 5 percent level.

Finally, we run the breakpoint tests using the differentials in the Europe and Asia groups. As with the USGC differentials, we find evidence of a break affecting all of the differentials in or around 2008.

We next consider the across-area differentials for different crude types, with the results presented in the bottom panel of Table 4.1. The test finds that all of the differentials, with just two exceptions, experienced a break around 2008. The test identifies a few other breaks for differentials involving light, sweet crude in the USGC and two breaks involving Mars differentials after 2010.

As shown in Table 4.1, a very large number of breaks occurred between 2007 and early 2009. An immediate question of interest to us was whether this break affected oil price differentials generally speaking or if it was limited to differentials between different types of oil. To investigate this, we next tested for breaks in the differentials between crude oils of the same type, i.e. the light, sweet differentials and the medium, sour differentials. The results from those tests are shown in Table 4.2. The upper panel is for the within-area differentials while the bottom panel is for the across-area differentials.

Our main finding is that while the test identifies a number of breaks, evidence for a large set between 2007 and 2009 is non-existent. We find two breaks impacting the LLS-HLS differential after the start of the shale boom. We also find a set of breaks in the mid-2000s and during the shale boom that affect across-area, light, sweet differentials, and several breaks for medium, sour differentials that involve Mars crude.

Table 4.1: Breakpoint test results for crudes of different qualities

Part 1: Within-area differentials

Differential	Break 1	Break 2	Break 3	F-statistic 0 vs. 1	1 vs. 2	2 vs. 3
Midland, TX						
WTIM-WTS	12/2007	02/2013	-	157.83	14.36	-
U.S. Gulf Coast						
LLS-Mars	02/2008	-	-	62.98	-	-
LLS-Maya	05/2007	-	-	50.14	-	-
HLS-Mars	05/2008	12/2001	-	58.00	14.39	-
HLS-Maya	05/2007	-	-	50.44	-	-
Mars and Maya	04/2007	-	-	47.28	-	-
Europe/Atlantic Basin						
Brent-Urals$^{(m)}$	06/2008	-	-	31.96	-	-
Brent-SHE	02/2007	-	-	29.69	-	-
Middle East/Asia						
Tapis-Oman	05/2008	-	-	29.78	-	-
Tapis-Dubai	05/2008	-	-	39.15	-	-
Tapis-SHA	03/2009	-	-	25.27	-	-

Part 2: Across-area differentials

Differential	Break 1	Break 2	Break 3	F-statistic 0 vs. 1	1 vs. 2	2 vs. 3
Light-medium						
Tapis-Urals$^{(m)}$	05/2008	-	-	30.10	-	-
Tapis-Mars	02/2008	05/2011	-	32.51	20.00	-
Brent-Oman	05/2008	-	-	18.63	-	-
Brent-Dubai	05/2008	-	-	25.74	-	-
Brent-Mars	02/2008	08/2013	-	15.15	52.19	-
LLS-Oman	12/2008	-	-	100.62	-	-
LLS-Urals$^{(m)}$	05/2009	-	-	51.09	-	-
LLS-Dubai	12/2008	05/2005	-	116.83	14.39	-
HLS-Oman	11/2008	-	-	89.49	-	-
HLS-Urals$^{(m)}$	03/2007	04/2012	-	57.55	16.50	-
HLS-Dubai	11/2008	03/2005	-	105.34	17.24	-
Light-heavy						
Tapis-Maya	06/2007	-	-	47.47	-	-
Brent-Maya	07/2007	-	-	33.67	-	-
Medium-heavy						
Oman-Maya	05/2007	-	-	35.64	-	-
Dubai-Maya	03/2002	-	-	18.25	-	-
Urals-Maya$^{(m)}$	02/2002	-	-	14.53	-	-

Notes: Dates refer to the month of the last day of a given regime. The order of the breaks is determined by the test. The critical values are 12.29, 13.89, and 14.80 for tests of 0 or 1 break, 1 or 2 breaks, and 2 or 3 breaks, respectively. These reflect a significance level of 1 percent. A $^{(m)}$ refers to results based on monthly data.

Table 4.2: Breakpoint test results for crudes of similar type

Part 1: Within-area differentials

				F-statistic		
Differential	Break 1	Break 2	Break 3	0 vs. 1	1 vs. 2	2 vs. 3
U.S. Gulf Coast						
LLS-HLS	02/2011	07/2014	-	28.86	27.82	-
Europe/Atlantic Basin						
Urals-SHE$^{(m)}$	01/2001	-	-	13.55	-	-
Middle East/Asia						
Oman-Dubai	-	-	-	-	-	-
Oman-SHA	03/2009	-	-	13.61	-	-
Dubai-SHA	-	-	-	-	-	-

Part 2: Across-area differentials

				F-statistic		
Differential	Break 1	Break 2	Break 3	0 vs. 1	1 vs. 2	2 vs. 3
Light, sweet						
WTIC-LLS$^{\#}$	04/2010	02/2006	08/2013	11.65	84.81	12.77
WTIM-LLS	01/2011	11/2006	-	16.49	143.70	-
LLS-Tapis	01/2005	05/2011	03/2015	75.72	20.83	29.28
LLS-Brent	05/2011	01/2005	-	120.02	38.89	-
HLS-Tapis	05/2004	-	-	60.35	-	-
HLS-Brent	01/2005	08/2013	-	90.40	59.25	-
Medium, sour						
Oman-Mars$^{\#}$	01/2002	08/2013	-	9.39	36.70	-
Urals-Mars$^{(m)}$	07/2013	-	-	17.50	-	-
Dubai-Mars	08/2013	03/2002	11/2005	22.52	14.42	19.14
Oman-Urals$^{(m)}$	-	-	-	-	-	-
Urals-Dubai$^{(m)}$	-	-	-	-	-	-

Notes: Dates refer to the month of the last day of a given regime. The order of the breaks is determined by the test. The critical values are 12.29, 13.89, and 14.80 for tests of 0 or 1 break, 1 or 2 breaks, and 2 or 3 breaks, respectively. These reflect a significance level of 1 percent. A $^{\#}$ means the test failed to reject 0 vs. 1 break at 1 percent significance but did so for the null of 1 vs. 2 breaks. A $^{(m)}$ refers to results based on monthly data.

4.2 Grouping the structural breaks

Previous works in the literature have focused on breaks in light, sweet crude differentials. To help provide some context to our findings, which focus on quality differentials, we have found it useful to group the breaks in quality and light, sweet differentials based upon the affected differentials and the timing. Our analysis leads us to three major groupings, shown visually in Table 4.3. Any year with a break is marked with an X.

The set of breaks color-coded green occur from 2010 onwards and primarily affect light, sweet differentials that involve at least one U.S. crude oil. Given the timing and the differentials involved, it is natural to conclude that these breaks are a result of the oil production boom in the U.S., and to a lesser extent Canada. Most of these breaks, or ones of similar nature, have been documented previously in the literature. A few, such as the WTIM-WTS break in 2013, do not appear to have been noted before.

We have found another group of breaks, color-coded yellow, primarily affecting across-area, light, sweet differentials that involve either LLS or HLS prices. To the best of our knowledge, these breaks have not been documented before in the literature. Results based on monthly data provide further evidence for this group, as we find breaks in the LLS and HLS differentials involving Bonny Light and Algerian Saharan (both light, sweet crude oils in the Atlantic Basin) in late 2004 and early 2005.

Finally, the breaks involving the quality differentials are color-coded blue.

4.3 Quality differentials have shrunk

We next show in Table 4.4 how the means of the quality differentials have changed across regimes. The differentials are grouped in a similar manner to Tables 4.1 and 4.2. The final column shows how the means have changed from the initial to final regime. As a reminder, a change of -0.1 means a 10 percentage point decline.

Our main result is that the means of most quality differentials have shrunk in half, at least. We find many cases where the average values have declined by close to 10 percentage points or more, particulary for differentials connected with the USGC. We find only two notable exceptions to our main finding: the Tapis-Mars and the Brent-Mars differentials. Both experienced a sharp decline in their mean in early 2008, similar to other quality differentials, but then experienced a reversal after the beginning of the shale boom.

The table also shows how the means of light, sweet differentials have evolved. We find that the two WTI-LLS differentials have over time gone from being near 0 to negative. Likewise for differentials between USGC light, sweet crude and light crude oils outside the U.S. These findings are consistent with the major increase in the supply of light, sweet crude in the

Table 4.3: Grouping the breaks

		97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
Quality Differentials	WTIM-WTS											X						X					
	LLS-Mars												X										
	LLS-Maya											X											
	HLS-Mars					X							X										
	HLS-Maya											X											
	Mars-Maya											X											
	Brent-Urals												X										
	Brent-SHE											X											
	Tapis-Oman												X										
	Tapis-Dubai												X										
	Tapis-SHA													X									
	Tapis-Urals												X										
	Tapis-Mars												X			X							
	Brent-Oman												X										
	Brent-Dubai												X										
	Brent-Mars												X					X					
	LLS-Oman												X										
	LLS-Urals													X									
	LLS-Dubai									X			X										
	HLS-Oman												X										
	HLS-Urals											X					X						
	HLS-Dubai									X			X										
	Tapis-Maya											X											
	Brent-Maya											X											
	Oman-Maya											X											
	Dubai-Maya						X																
	Urals-Maya					X																	
Light Crude	LLS-HLS															X			X				
	WTIC-LLS										X				X			X					
	WTIM-LLS										X					X							
	LLS-Tapis									X						X				X			
	LLS-Brent									X						X							
	HLS-Tapis								X														
	HLS-Brent									X								X					

Table 4.4: Regression constant across regimes

<table>
<tr><th>Differential</th><th>97</th><th>98</th><th>99</th><th>00</th><th>01</th><th>02</th><th>03</th><th>04</th><th>05</th><th>06</th><th>07</th><th>08</th><th>09</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th><th>18</th><th>Change</th></tr>
<tr><td>WTIM-WTS</td><td colspan="11">0.079</td><td colspan="5">0.024</td><td colspan="6">0.003*</td><td>-0.076</td></tr>
<tr><td>LLS-Mars</td><td colspan="11">0.152</td><td colspan="11">0.065</td><td>-0.087</td></tr>
<tr><td>HLS-Mars</td><td colspan="5">0.161</td><td colspan="6">0.108</td><td colspan="11">0.055</td><td>-0.106</td></tr>
<tr><td>LLS-Maya</td><td colspan="10">0.312</td><td colspan="12">0.151</td><td>-0.161</td></tr>
<tr><td>HLS-Maya</td><td colspan="10">0.292</td><td colspan="12">0.142</td><td>-0.150</td></tr>
<tr><td>Mars-Maya</td><td colspan="10">0.158</td><td colspan="12">0.083</td><td>-0.075</td></tr>
<tr><td>Brent-Urals$^{(m)}$</td><td colspan="11">0.061</td><td colspan="11">0.018</td><td>-0.043</td></tr>
<tr><td>Brent-SHE</td><td colspan="10">0.198</td><td colspan="12">0.087</td><td>-0.111</td></tr>
<tr><td>Tapis-Oman</td><td colspan="11">0.116</td><td colspan="11">0.069</td><td>-0.063</td></tr>
<tr><td>Tapis-Dubai</td><td colspan="11">0.130</td><td colspan="11">0.074</td><td>-0.056</td></tr>
<tr><td>Tapis-SHA</td><td colspan="12">0.195</td><td colspan="10">0.109</td><td>-0.087</td></tr>
<tr><td>Tapis-Urals$^{(m)}$</td><td colspan="11">0.122</td><td colspan="11">0.062</td><td>-0.060</td></tr>
<tr><td>Tapis-Mars</td><td colspan="11">0.149</td><td colspan="3">0.080</td><td colspan="8">0.110</td><td>-0.039</td></tr>
<tr><td>Brent-Oman</td><td colspan="11">0.055</td><td colspan="11">0.024</td><td>-0.031</td></tr>
<tr><td>Brent-Dubai</td><td colspan="11">0.069</td><td colspan="11">0.030</td><td>-0.039</td></tr>
<tr><td>Brent-Mars</td><td colspan="11">0.088</td><td colspan="6">0.034</td><td colspan="5">0.080</td><td>-0.008</td></tr>
<tr><td>LLS-Oman</td><td colspan="12">0.117</td><td colspan="10">0.030</td><td>-0.086</td></tr>
<tr><td>LLS-Urals$^{(m)}$</td><td colspan="12">0.121</td><td colspan="10">0.020</td><td>-0.101</td></tr>
<tr><td>LLS-Dubai</td><td colspan="8">0.143</td><td colspan="4">0.101</td><td colspan="10">0.035</td><td>-0.107</td></tr>
<tr><td>HLS-Oman</td><td colspan="12">0.098</td><td colspan="10">0.021</td><td>-0.077</td></tr>
<tr><td>HLS-Urals$^{(m)}$</td><td colspan="10">0.108</td><td colspan="5">0.047</td><td colspan="7">0.003*</td><td>-0.105</td></tr>
<tr><td>HLS-Dubai</td><td colspan="8">0.124</td><td colspan="4">0.083</td><td colspan="10">0.026</td><td>-0.098</td></tr>
<tr><td>Tapis-Maya</td><td colspan="10">0.308</td><td colspan="12">0.186</td><td>-0.121</td></tr>
<tr><td>Brent-Maya</td><td colspan="11">0.245</td><td colspan="11">0.140</td><td>-0.105</td></tr>
<tr><td>Oman-Maya</td><td colspan="10">0.192</td><td colspan="12">0.113</td><td>-0.078</td></tr>
<tr><td>Dubai-Maya</td><td colspan="6">0.210</td><td colspan="16">0.119</td><td>-0.091</td></tr>
<tr><td>Urals-Maya$^{(m)}$</td><td colspan="5">0.219</td><td colspan="17">0.131</td><td>-0.088</td></tr>
<tr><td>WTIC-LLS</td><td colspan="9">0.002*</td><td colspan="4">-0.035</td><td colspan="3">-0.138</td><td colspan="6">-0.055</td><td>-0.057</td></tr>
<tr><td>WTIM-LLS</td><td colspan="10">-0.008</td><td colspan="4">-0.043</td><td colspan="8">-0.125</td><td>-0.117</td></tr>
<tr><td>LLS-HLS</td><td colspan="14">0.019</td><td colspan="4">-0.002*</td><td colspan="4">0.013</td><td>-0.007</td></tr>
<tr><td>LLS-Tapis</td><td colspan="8">0.015</td><td colspan="6">-0.019</td><td colspan="4">-0.067</td><td colspan="4">-0.027</td><td>-0.042</td></tr>
<tr><td>LLS-Brent</td><td colspan="8">0.075</td><td colspan="6">0.039</td><td colspan="8">-0.005*</td><td>-0.080</td></tr>
<tr><td>HLS-Tapis</td><td colspan="7">-0.002</td><td colspan="15">-0.045</td><td>-0.043</td></tr>
<tr><td>HLS-Brent</td><td colspan="8">0.056</td><td colspan="9">0.016</td><td colspan="5">-0.019</td><td>-0.075</td></tr>
<tr><td>Differential</td><td>97</td><td>98</td><td>99</td><td>00</td><td>01</td><td>02</td><td>03</td><td>04</td><td>05</td><td>06</td><td>07</td><td>08</td><td>09</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>Change</td></tr>
</table>

Notes: Change is the difference between the final regime and the first regime for each regression equation. A * means the coefficient is not statistically different from 0 at a 5 percent confidence level. In the table, breaks that occur from July to December in a particular year are assigned to the following year. A $^{(m)}$ refers to results based on monthly data.

U.S., and reconfirm previous results in the literature regarding structural breaks affecting light, sweet crude differentials.

4.4 Less volatile quality differentials

Figure 1.1 is suggestive that there may have been changes in the volatility of quality differentials, particularly for those experiencing a break around 2008. We investigate this a little deeper by comparing the means and standard deviations of the quality differentials before and after 2008. While the actual breakpoint for many series varies, we decided to work with a "pre-break" period that runs until the end of 2008 as this simplifies the exposition.

The statistics are shown in Table 4.5. We include any differential that experienced a permanent drop in its mean since 2008. Overall, we find a marked reduction in both the average level of the differentials, as well as their volatilities. In most cases, the mean in the post-break sample is less than half the size of the pre-break mean. Post-break volatilities are about 1/2 to 3/4 the size of the pre-break volatilities.

4.5 Breaks in residual fuel oil differentials

In Section 2 we discussed the connection between a crude oil's API gravity and its inherent yield of residual products that come from the distillation process. This should create a relationship between quality-related oil price differentials and the value of residual fuel oil relative to other petroleum products. Given this, we investigated whether differentials related to residual fuel oil have experienced breaks similar to those affecting quality-related oil price differentials.

We use daily data to calculate differentials between the spot price of high-sulfur residual fuel oil and the following spot prices, all for delivery in the Gulf Coast: heating oil, gasoline, LLS and Mars. A plot of the differentials, found in the appendix, shows a remarkable similarity between them and the differentials plotted in Figure 1.1. This similarity exists if one uses product prices for New York Harbor and replaces LLS and Mars prices with Brent and Dubai.

More formally, we run breakpoint tests on the Gulf Coast fuel oil differentials and found that all of them experienced a break in their mean around the same time that many quality-related oil price differentials did. The gasoline-residual fuel oil differential has a break in September 2007, while the other differentials have a break in January 2009. The decline in the means is on the same order as was documented for the oil price differentials.

Table 4.5: Summary statistics pre and post-break

Part 1: Within-area differentials

Differential	Pre-break Mean	Pre-break Standard deviation	Post-break Mean	Post-break Standard deviation	Ratio of mean (post/pre)	Ratio of std. dev. (post/pre)
Midland, TX						
WTIM-WTS	0.076	0.031	0.010	0.018	0.13	0.58
U.S. Gulf Coast						
LLS-Mars	0.147	0.052	0.063	0.029	0.45	0.56
LLS-Maya	0.299	0.089	0.141	0.055	0.47	0.62
HLS-Mars	0.128	0.053	0.054	0.023	0.42	0.43
HLS-Maya	0.279	0.086	0.133	0.051	0.48	0.59
Mars-Maya	0.151	0.062	0.079	0.040	0.52	0.65
Europe/Atlantic Basin						
Brent-Urals$^{(m)}$	0.058	0.036	0.019	0.017	0.33	0.47
Middle East/Asia						
Tapis-Oman	0.114	0.058	0.069	0.033	0.61	0.57
Tapis-Dubai	0.128	0.058	0.074	0.032	0.58	0.55

Part 2: Across-area differentials

Differential	Pre-break Mean	Pre-break Standard deviation	Post-break Mean	Post-break Standard deviation	Ratio of mean (post/pre)	Ratio of std. dev. (post/pre)
Light-medium						
Brent-Oman	0.053	0.048	0.025	0.031	0.47	0.65
Brent-Dubai	0.067	0.051	0.030	0.030	0.45	0.59
LLS-Oman	0.116	0.055	0.030	0.043	0.26	0.78
LLS-Urals$^{(m)}$	0.121	0.049	0.025	0.037	0.21	0.75
LLS-Dubai	0.130	0.056	0.035	0.043	0.27	0.77
HLS-Oman	0.096	0.055	0.022	0.042	0.22	0.78
HLS-Urals$^{(m)}$	0.101	0.046	0.017	0.033	0.17	0.71
HLS-Dubai	0.110	0.057	0.027	0.042	0.25	0.74
Light-heavy						
Tapis-Maya	0.296	0.095	0.181	0.055	0.61	0.59
Brent-Maya	0.235	0.081	0.136	0.055	0.58	0.68
Medium-heavy						
Oman-Maya	0.182	0.078	0.112	0.050	0.62	0.64

Notes: The pre-break sample runs from January 1997 to December 2008. The post-break sample runs from January 2009 to December 2018. A $^{(m)}$ means the statistic is based on monthly data.

4.6 Additional results

We also repeated our analysis using monthly price data for a slightly larger set of crude oils. We find additional evidence of breaks affecting quality differentials between 2007-2009, as well as additional breaks in late 2004 and early 2005 affecting differentials between light, sweet crude oils in the U.S. Gulf Coast and light crude outside the U.S. Details on the data and results can be found in the appendix.

As a further robustness check, we also used the cointegration breakpoint test of Gregory and Hansen (1996). This tests for a break in the intercept term of the long-run equilibrium equation, given by $\ln P_{i,t} = c + \beta \ln P_{j,t} + u_t$. Using daily data, the test also finds a large number of breaks in the quality differentials between 2007 and 2009.[13]

5 Discussion

In this section, our goal is to provide explanations for what we believe are the three most important and interesting findings in the results section: (1) the means of most quality differentials have shrunk over time; (2) residual fuel oil differentials have shrunk in a similar manner; (3) there is a large cluster of breaks around 2008.

Based on our discussion in Section 2.3, we focus on five specific factors to explain findings (1) - (3): environmental regulations; trends in the use of residual fuel oil; global upgrading capacity; the shale boom; and utilization of refining capacity. For each factor, we introduce relevant data and discuss to what extent the factor seems to be able to explain our findings. After presenting all the data, we also discuss briefly to what extent the Great Recession may have played in the timing of the breaks, which seems natural to investigate given the cluster occurs around the start of the Recession.

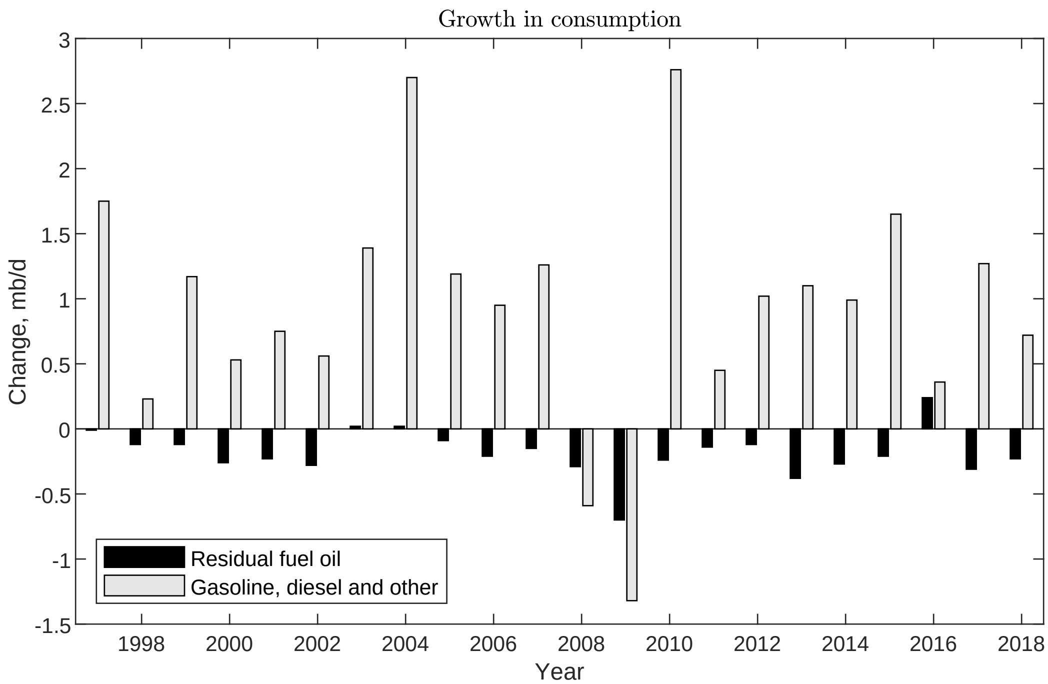
Our analysis of the data leads us to conclude that key findings (1) and (2) can be explained by long-run changes in the oil market due to the shale boom and significant growth in global upgrading capacity. Environmental regulations and trends in consumer demand should, on the contrary, be contributing to wider differentials over time. In regards to key finding (3), we find a connection with the Great Recession due to its impact on consumer demand and utilization rates for refining capacity.

5.1 Factors

Environmental regulations tightened: A weakening of environmental regulations regarding sulfur emissions, particularly for residual fuel oil, would be a straight forward ex-

[13]Full results and details are available upon request.

Figure 5.1: Residual fuel oil consumption declining over time



Notes: Units are annual change in millions of barrels per day. Other petroleum products includes naphtha, gasoline, jet fuel, and middle distillates but excludes natural gas liquids, such as ethane.

planation for key findings (1) and (2). However, in general, standards regarding sulfur have actually been tightening over time, which should be contributing to wider differentials. These standards have been applied in several large consuming countries, including the U.S., the EU and China, to a variety of petroleum products, including gasoline, diesel and residual fuel oil. Of particular note, requirements for residual fuel oil have been tightened several times since 2008 and very stiff regulations will go into effect in 2020.

Use of residual fuel oil declining over time: An increase in the demand for residual fuel oil relative to other petroleum products since 2008 is another explanation that could potentially explain key findings (1) and (2). However, consumption data does not support this argument. We show this visually in Figure 5.1, where we plot annual changes in the consumption of residual fuel oil and all other petroleum products (excluding NGLs) from 1997 to 2018. This data comes from various Annual Statistical Supplements from the International Energy Agency. Use of residual fuel oil has declined almost every year since 1997 while use of other petroleum products has been increasing at a relatively rapid pace, with the exception of the Great Recession. We will return to this point when we discuss the timing of the breaks.

Refining sector becoming more complex: Cokers and crackers increase the amount of gasoline and diesel that can be produced from a given barrel of medium and heavy crude oil while reducing the supply of residual fuel oil. Given this, an increasingly complex refining sector is a natural candidate for explaining both the breaks in quality-related oil price differentials and the residual fuel oil price differentials.

It has long been noted that the refining sector in the U.S. has become increasingly complex over time. Publicly available data for the rest of the world is limited but it also shows an increasing ability of the refining sector worldwide to convert medium and heavy crude oils into high-valued petroleum products. Our main discussion is based on data from Eni's World Oil Review and World Oil and Gas Review publications, and data from the International Energy Agency.

Table 5.1 shows the data from Eni on global refining capacity, as well as two measures of how complex the refining sector is overall.[14] The second column shows data on primary capacity, which is crude distillation capacity and condensate splitters. The third column shows data on conversion capacity, which measures how much cracking and coking capacity is available.[15] The fourth column shows the first measure of complexity, which is simply the ratio of conversion capacity to primary capacity. The final column is the Nelson Complexity Index (NCI). This is a commonly used measure of refinery complexity where higher values reflect greater complexity, either at a particular refinery or for a particular area.[16] Unlike the conversion capacity data, the NCI reflects not only the amount of upgrading capacity available but also the amount of desulfurisation capacity available.

Table 5.1: Global refineries increasingly complex

Year	Primary capacity (mb/d)	Conversion capacity (mb/d)	Conversion capacity ratio (percent)	Complexity Ratio Nelson Complexity
2000	83.2	31.6	38	7.9
2005	87.3	37.5	43	8.2
2010	92.4	43.4	47	8.7
2015	96.5	50.2	52	9.1
2016	98.1	52.0	53	9.3
2017	98.7	53.3	54	9.3

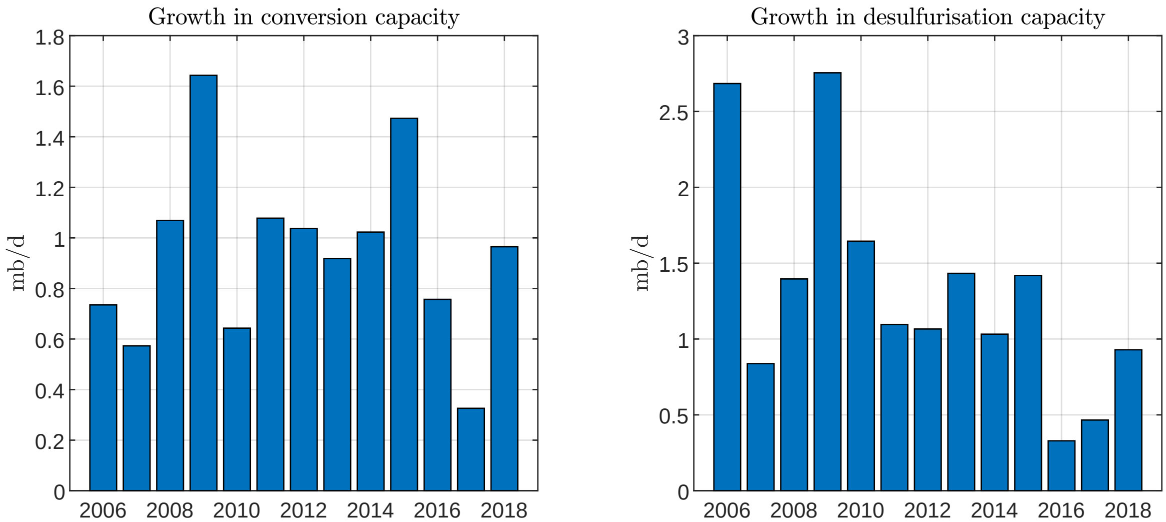
Notes: The conversion capacity ratio is conversion capacity divided by primary capacity. Sources: Eni World Oil Review 2018, Eni World Oil Review 2017, Eni World Oil and Gas Review 2016.

[14] A more complete time series can be put together using older versions of Eni publications. However, the data has been revised several times, most recently in 2015. As a result, while the numbers in Table 5.1 are comparable to each other, the longer time series one can put together using older reports are not comparable, strictly speaking. The appendix contains the full series.

[15] Conversion capacity is fluid catalytic cracking equivalent. Details on the calculation can be found in Eni's World Oil Review 2018.

[16] Johnston (1996) provides a good introduction to the index and how it is calculated.

Figure 5.2: IEA data shows substantial additions to upgrading capacity



Notes: Units are millions of barrels per day. The data come from various International Energy Agency Medium-Term Oil Market Reports and Market Report Series which are publicly available.

Since the year 2000, primary capacity has been growing at about 1 percent per year, on average, or 0.9 million barrels per day (mb/d).[17] Conversion capacity has been growing at a more rapid pace, about 4 percent a year, on average, or 1.3 mb/d. This has led to an increase in the conversion capacity ratio and contributed to higher values of the NCI. The ratio of conversion capacity to primary capacity rose from 38 percent in 2000 to 54 percent in 2017 while the NCI rose from 7.9 in 2000 to 9.3 in 2017. Some of the largest increases in complexity have occurred in Asia, where the conversion capacity ratio has risen from 36 percent in 2000 to 66 percent in 2017, with the NCI rising from 7.0 to 9.7.

International Energy Agency reports also provide data on additions to global conversion capacity and desulfurisation capacity from 2006 to 2018. We plot both of those series in Figure 5.2. This data reinforces the findings of the Eni data, as it shows significant additions to conversion capacity. We also note here that there were some fairly large capacity additions in 2008 and 2009, a point we will come back to when we discuss the timing of the breaks.

Another source is the Oil&Gas Journal Worldwide Refinery Survey, which provides both primary capacity and conversion capacity numbers. Ideally, we would prefer to use this data as it is available at an annual frequency and begins earlier than other data series we have available. We do not, however, because participation in the survey is voluntary and it

[17] Annual primary capacity data is also available from the British Petroleum (BP) Statistical Review of World Energy. The differences between the BP and Eni series are relatively modest, usually less than one percent, so none of our conclusions are sensitive to using one series or the other.

appears that the survey is not accurately measuring capacity in some important developing countries, particularly China.[18]

Unexpected growth in light oil production: At the same time the global refining sector has increased its ability to transform low-quality crude oil into gasoline and diesel, the production of light, sweet crude has unexpectedly increased since the late 2000s due in large part to the U.S. shale boom. The first column in Table 5.2 shows that U.S. shale production grew by more than 5.5 million barrels per day from 2010 to 2018, averaging 6.5 mb/d in 2018. The next columns show data from Eni on global production of ultra light (API 50 and above) and light (API 35 - 50) crude oil. Since 2010, world production of ultra light is up almost 1 million barrels per day, while light is up about 2 million barrels per day. These increases are important because, as shown in Figure 2.1, lighter oil naturally produces less residual fuel oil than medium and heavy crude oils, particularly for very light crude oils. All else equal, this increase in supply, therefore, would reduce the spread between light crude and other types of crude, while also contributing to smaller spreads between residual fuel oil and other, higher-valued petroleum products.

Table 5.2: Light crude production up since 2005

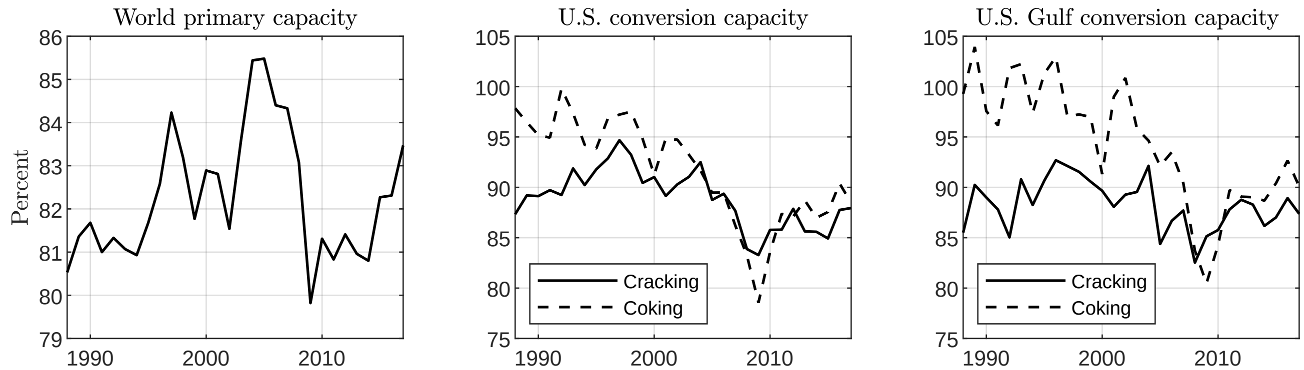
Year	U.S. light tight oil	Ultra light	Light	Other
2000	0.40	1.46	20.71	46.49
2005	0.41	1.94	19.81	53.10
2010	0.83	2.46	20.43	52.11
2015	4.77	3.43	21.98	56.35
2016	4.43	3.42	21.56	56.73
2017	4.96	3.40	22.48	55.82
2018	6.48	N/A	N/A	N/A

Notes: Units are millions of barrels per day. Eni defines ultra light as crude oil with API gravity of 50 or above while light crude oil has an API gravity from 35 up to but not including 50. Sources: Eni World Oil Review 2018, Eni World Oil Review 2017, Eni World Oil and Gas Review 2016, U.S. Energy Information Administration.

Utilization rates lower since Great Recession: Utilization rates for primary and upgrading capacity can be varied over time. We construct a utilization rate for world primary capacity using data from the BP Statistical Review of World Energy on crude throughput and primary capacity. Unfortunately, no publicly available data exist that allow us to construct

[18]The discrepancies appear to be large enough to be important for our discussion. The survey reported that at the start of 2011 that China's crude distillation capacity was little under 7 mb/d. Wu (2011), however, brings additional data to bear and reports distillation capacity over 10 mb/d. Likewise, the survey reports distillation capacity of 7 mb/d and coking capacity of 156,000 b/d at the start of 2013 but the International Energy Agency's Medium-Term Oil Market Report 2013 shows distillation capacity of 13.4 mb/d and coking capacity of 1.8 mb/d (see table on page 98).

Figure 5.3: Utilization rates for refining capacity



Notes: World primary capacity utilization is calculated as crude throughput divided by world primary capacity, which is a nameplate capacity figure. Cracking and coking utilization is calculated as fresh feed input divided by the Energy Information Administration's measure of calendar day capacity.

a world utilization rate for either cracking or coking capacity. However, data is available from the U.S. Energy Information Administration for the U.S. refining sector from 1987 to 2017. As the U.S. refining sector is the largest in the world, composed of profit-maximizing firms, and fully integrated with the global fuel market, we believe it should be reflective to some extent of conditions elsewhere. We use data for the U.S. as a whole and the U.S. Gulf Coast, which is home to a substantial portion of U.S. conversion capacity and has seen the largest increase in coking capacity over the sample period.[19]

Figure 5.3 plots the utilization rates. The BP data reflect the boom in demand that occurred before the Great Recession, as well as the effects of the Recession itself. Utilization rates for both coking and cracking capacity exhibit what appear to be structural breaks right around the time of the Great Recession. Before, the utilization rates were relatively high and, for coking capacity, the data is suggestive of some potential capacity constraints early in the sample. Utilization rates for coking capacity in the U.S. Gulf Coast were even higher, with rates exceeding 100 percent several years in the 1990s and in 2002.[20] During the recession, however, utilization rates for conversion capacity declined sharply and since then they have remained at levels below those seen from 1987 to the mid-2000s.

The slump in the rates around 2008 can be explained by the sharp drop in demand during

[19]We use data on fresh feed input to cokers along with annual data on coking capacity to construct the utilization rate for coking capacity. Likewise, we use equivalent data for catalytic crackers and hydrocrackers to construct a utilization rate for cracking capacity. The appendix provides details on the calculations and a full time series.

[20]The EIA capacity data is designed to take into account downtime at units and, as a result, allows for the possibility of utilization rates above 100 percent. However, episodes such as that suggest the machinery is being pushed to its physical limits.

the Recession, shown in Figure 5.1, which coincided with some extremely large capacity additions, as shown in Figure 5.2. The fact that rates have remained relatively low since then suggests global capacity additions since the Recession have been sufficient to process new supplies of medium and heavy crude and meet the growing demand for gasoline and diesel, without significantly increasing the production of residual fuel oil.

This discussion highlights the role of the Great Recession in influencing the timing of the breaks. The Recession, by significantly reducing the demand for petroleum products, particularly for products besides residual fuel oil, in essence allowed capacity additions to catch up with demand.

6 Conclusion

Crude oil can vary significantly in some key physical properties, making them imperfect substitutes for each other and leading to the existence of price differentials among crude oils. In a certain sense, these differentials reflect the limits to arbitrage that exist across crude oil quality. In this paper, we documented that a large number of differentials between crude oils of different types have experienced structural breaks where their means have become smaller over time. In particular, we show that many quality-related differentials experienced a major break in and around the time of the Great Recession.

Our analysis leads us to conclude that these quality differentials have narrowed over time due to several fundamental, long-lasting changes in the oil market. One is the fact that the global refining sector has become increasingly complex over time, as capacity additions have increased the ability of the sector to transform lower-grade crude oil into high-valued petroleum products. The other is the shale boom, which has unexpectedly increased the production of light crude oil, reducing, on the margin, the need for such complex refineries. This narrowing of the differentials has occurred despite the fact that increasingly stringent environmental regulations and trends in consumer demand should be pushing them apart.

We believe a number of possible avenues exist for future research. For one, our paper has focused on changes in the long-run means of crude oil price differentials. More sophisticated time-series analysis could try to disentangle the structural factors behind the short-run dynamics of those differentials. One could also consider setting up a theoretical model of the global refining sector to explore how theory suggests that changes in that sector should affect oil price differentials.

References

Adelman, M. A. (1984): "International Oil Agreements," *The Energy Journal*, 5, 1–9, https://doi.org/10.5547/issn0195-6574-ej-vol5-no3-1.

Agerton, M. and G. B. Upton Jr. (2019): "Decomposing Crude Price Differentials: Domestic Shipping Constraints or the Crude Oil Export Ban?" *The Energy Journal*, 40, https://doi.org/10.2139/ssrn.2942989.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858, https://doi.org/10.2307/2938229.

Bachmeier, L. J. and J. M. Griffin (2006): "Testing for market integration crude oil, coal, and natural gas," *The Energy Journal*, 55–71.

Bacon, R. and S. Tordo (2005): "Crude oil price differentials and differences in oil qualities: A statistical analysis," ESMAP Technical Paper.

Bai, J. (1997): "Estimating Multiple Breaks One at a Time," *Econometric Theory*, 13, 315–352, https://doi.org/10.1017/s0266466600005831.

Bai, J. and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78, https://doi.org/10.2307/2998540.

——— (2003): "Critical values for multiple structural change tests," *The Econometrics Journal*, 6, 72–78, https://doi.org/10.1111/1368-423x.00102.

Borenstein, S. and R. Kellogg (2014): "The Incidence of an Oil Glut: Who Benefits from Cheap Crude Oil in the Midwest?" *The Energy Journal*, 35, 15–33, https://doi.org/10.5547/01956574.35.1.2.

Buyuksahin, B., T. K. Lee, J. T. Moser, and M. A. Robe (2013): "Physical Markets, Paper Market and the WTI-Brent Spread," *The Energy Journal*, 34, 129–151, https://doi.org/10.5547/01956574.34.3.7.

DiPaola, A. (2019): "Arcane pollution rule may cost Saudis and neighbors billions," *Bloomberg*.

Energy Intelligence Research (2009): *The International Crude Oil Market Handbook 2009*, Energy Intelligence Research.

Evans, B. and J. Mowler (2002): "Switch to sour crude haunts US refiners," *Platts Oilgram News*, 80.

Fattouh, B. (2010): "The dynamics of crude oil price differentials," *Energy Economics*, 32, 334–342, https://doi.org/10.1016/j.eneco.2009.06.007.

Ghoshray, A. and T. Trifonova (2014): "Dynamic Adjustment of Crude Oil Price Spreads," *The Energy Journal*, 35, 119–136, https://doi.org/10.5547/01956574.35.1.7.

Giulietti, M., A. M. Iregui, and J. Otero (2015): "A pair-wise analysis of the law of one price: Evidence from the crude oil market," *Economics Letters*, 129, 39–41, https://doi.org/10.1016/j.econlet.2015.02.002.

Gregory, A. W. and B. E. Hansen (1996): "Residual-based tests for cointegration in models with regime shifts," *Journal of Econometrics*, 70, 99–126, https://doi.org/10.1016/0304-4076(69)41685-7.

Gülen, G. (1997): "Regionalization in the World Crude Oil Market," *The Energy Journal*, 18, 109–126, https://doi.org/10.5547/issn0195-6574-ej-vol18-no2-6.

——— (1999): "Regionalization in the World Crude Oil Market: Further Evidence," *The Energy Journal*, 20, 125–139, https://doi.org/10.5547/issn0195-6574-ej-vol20-no1-7.

Hammoudeh, S. M., B. T. Ewing, and M. A. Thompson (2008): "Threshold Cointegration Analysis of Crude Oil Benchmarks," *The Energy Journal*, 29, 79–95, https://doi.org/10.5547/issn0195-6574-ej-vol29-no4-4.

Johnston, D. (1996): "Refining Report Complexity index indicates refinery capability, value," *Oil&Gas Journal*.

Kemp, J. (2009): "Saudi move is bid to realign oil market," *Reuters*.

Khrennikova, D. and E. Mazneva (2018): "Russian oil set to lose billions in ship-fuel overhaul," *Bloomberg*.

Perron, P. (1989): "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica*, 1361–1401, https://doi.org/10.2307/1913712.

Perron, P. et al. (2006): "Dealing with structural breaks," *Palgrave Handbook of Econometrics*, 1, 278–352.

Piotrowksi, M. (2009): "Tight differentials for heavy/sour crudes hurt complex refiners," *The Oil Daily*.

PIW (2005): "US refining trends ride on shifting spreads," *Petroleum Intelligence Weekly*.

Platts (2018): *Methodology and Specifications Guide: Crude Oil*, Platts.

Ripple, R. and P. R. Wilamoski (1995): "Is the world oil market one great pool?: revisited," *OPEC Review*, 19, 283–292.

Sauer, D. G. (1994): "Measuring Economic Markets for Imported Crude Oil," *The Energy Journal*, 15, 107–124, https://doi.org/10.5547/issn0195-6574-ej-vol15-no2-6.

Scheitrum, D. P., C. A. Carter, and C. Revoredo-Giha (2018): "WTI and Brent Futures Pricing Structure," *Energy Economics*, 72, 462–469, https://doi.org/10.1016/j.eneco.2018.04.039.

Tuttle, R. (2019): "Ship-fuel rule to keep refiners running hard as 2020 approaches," *Bloomberg*.

Weiner, R. J. (1991): "Is the World Oil Market 'One Great Pool'?" *The Energy Journal*, 12, 95–107, https://doi.org/10.5547/issn0195-6574-ej-vol12-no3-7.

Wu, K. (2011): "Capacity, complexity expansions characterize China's refining industry past, present, future," *Oil&Gas Journal*.