UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Corso di Laurea Magistrale in International Economics and Commerce

Rockets and Feathers, an empirical analysis of fuel prices in Italy

Relatore: Chiar.mo
Prof. Riccardo Lucchetti

Tesi di Laurea di:
Ciarimboli Pietro

Anno Accademico 2018– 2019

Index

1. Introduction.....4
2. Overview of the literature.....8
 2.1 Worldwide contribution.....8
 2.2 Italian case.....20
3. Price analysis of crude oil and derivatives.....24
 3.1 Crude oil price benchmarks.....25
 3.2 Fuel prices in Italy.....28
  3.2.1 Petrol and Diesel price formation process.....28
  3.2.3 Petrol price series.....29
  3.2.3 Diesel price.....32
  3.2.3 Crude Oil price.....33
  3.2.4 Retail price components for Petrol and Diesel in Italy.....35
4. Econometrics.....37
 4.1 The ECM model.....37
  4.1.1 From ADL to ECM representation.....37
  4.1.2 The basic ECM representation.....40
  4.1.3 Bivariate ECM model.....41
  4.1.3 The NARDL model.....42
 4.2 Pass through and cointegration.....44
  4.2.1 Mark-up and Price pass-through.....44
  4.1.2 Lag selection.....47
  4.1.3 Cointegration and unit root analysis.....49
  4.1.4 Engle and Granger cointegration approach.....52
  4.1.5 Johansen test and VECM for petrol and diesel.....56
  4.1.6 Upward and downward price movement specification.....61
 4.3 Running the model.....63
  4.3.1 Asymmetric ECM from Engle and Grenger two step approach model 1.....63
  4.3.2 Asymmetric ECM for Engle and Grenger two step approach model 2.....68
  4.3.3 Asymmetric ECM from VECM.....72
  4.3.4 NARDL model for asymmetric behavior in petrol and diesel price adjustments.....75
5. Conclusions.....81
6.Bibliography and Sitography.....82

1. Introduction

The Rockets and Feathers (R&F) hypothesis has been a long discussed topic. First investigations of this hypothesis date back to the late 80's with the inquiries of the Monopolies and Mergers Commission in the UK, who prompted the work of Robert W. Bacon in 1991. His work can be seen as a milestone in the studies about price asymmetry in the fuel markets, i.e. the difference in the speed of reaction to price rises in crude oil on the pump prices of diesel and petrol.

So, what does the Rockets and Feathers hypothesis mean? The idea is quite simple, as the name already implies. It is about the effect that consumer price fuel has in relation to the price of crude oil, rocketing when the price of crude oil goes up and falling slowly when crude oil prices plunge.

Implications vary, although the most important one is that of competition. When there is an asymmetric price transmission it is possible to conjecture about monopolistic behavior.

Analyzing the structure of supply and demand of consumer fuels (specifically petrol and diesel, being the most relevant in terms of consumption) and crude oil, some assumptions can be made:

- The market for petrol or diesel is small compared to the one for crude oil, meaning that the downstream price is unable to drive the upstream one.

- The relation between the upstream price and downstream price in a competing environment should be stable. As no substitute to crude oil exists for petrol and diesel (to be precise, for this last one some do exist, for example biodiesels, but not enough to be a perfect substitute, at least in the short run).

That stated, one can understand the economic importance of this, which at first sight is a simple topic. If asymmetry is found one can assume that this market is not perfect, with the known economic policy implications, which are not discussed here.

However, asymmetry is not a straightforward concept. Many definitions must be added. Firstly, different types of asymmetry can exist or coexist:

- Asymmetry in time, taking into consideration how much time is taken to react to an upstream price variation no matter the size of the reaction.
- Short run asymmetry, taking into consideration the size of the reaction in relation to an upstream variation in the same period when the reaction takes place.
- Long run asymmetry, (that given a long run relationship must exist between the two variables) does measure the impact of the disequilibrium.

  This last point will be examined in more detail.

Firstly, it is important to choose the right lag when specifying the model. In other words, the right amount of past periods one has to consider in order to observe the downstream reaction to the upstream price variations. Different techniques will be exposed, from the correlogram for autocorrelation to the Breusch-Godfrey test for serial autocorrelation.

Short run asymmetry will be tested using a Nonlinear Auto Regressive Distributed Lag (NARDL) model with a software package function provided by Dhanasekaran Kuppusamy. While long-run asymmetry will be the focus of this work, using a bivariate Error Correction Model (ECM) and the NARDL model cited before.

To test the behavior of the time series a test for covariance stationarity of the data will be performed before running the model.

However, when defining the ECM model different specifications will be considered, resulting in different outcomes. Thus spotlighting the importance of model specifications for making inference. This is a widely discussed topic in the R&F literature, with many papers confronting different model specifications in same data set.

To provide a synthesis of the literature, the first chapter will be an overview of the most critical works according to Google Scholar citations.

The second chapter will be an examination of the data collected, their trends in time and an analysis of the price structure of petrol products for Italy.

Model specifications, data configuration and best approach to the model will be considered in the third chapter.

The fourth chapter will consist of a presentation of the different model outcomes and a discussion of the results. The final chapters will comprise a summary and the conclusions from the research.

2. Overview of the literature

2.1 Worldwide contribution

The so-called Rockets and Feathers (R&F) price asymmetry between crude oil price and retail petrol price is a long debated topic. In this section the most important studies on the subject are reviewed. It is common to start from the work of Robert W. Bacon published in 1991[1]. Bacon developed a quadratic quantity adjustment model to test the hypothesis put forward by the report from the Monopolies and Mergers commission, who claimed that speed of adjustment of UK retail petrol prices to cost changes is more rapid when costs rise than when they fall, and the adjustment path is more concentrated around the mean value of the upswing.

Bacon used fortnightly data of UK retail gasoline market prices and ex-refinery petroleum prices for a period from 1982 to 1989 published in the *Petroleum Times*, adjusted for the sterling/dollar exchange rate, as being oil quoted in dollars. The model defines a target or equilibrium level for gasoline price and then tries to assess if the adjustment path to equilibrium is equal in case of an upward or a downward price movement.

Bacon found that the upward adjustment process was slightly faster, and the period of adjustment more concentrated than was the case when cost fell, as suggested by the R&F hypothesis, finding evidence of asymmetry, even if small.

1 Robert W Bacon *"Rockets and feathers: the asymmetric speed of adjustment of UK retail gasoline prices to cost changes" (*1991)

In the same year a study from J. D. Karrenbrock of the Federal Reserve Bank of St. Louis[2] and based on monthly data ranging from January 1983 to December 1990 of retail and wholesale gasoline (data for US), attempted to test if the price transmission occurred at same speed between these two stages. Using Ordinary Least Squares (OLS) over wholesale prices to retail margins, the result is a full pass-through of price increases and decreases with the same speed, meaning no presence of asymmetry.

Rockets and feathers hypothesis was investigated accurately in the same period also by D. N. Manning in his 1991 study[3] using a two-step Engle and Granger methodology, first considering the level of cointegration of the data and then running an Equilibrium Correction Mechanism or Error Correction Model (ECM) to find presence of asymmetry. The dataset adopted comprised 3 time series, one for the spot price of Brent crude oil, one for the level of excise duty and one for the price of four-star quality motor 'spirit', over a period ranging from January 1973 to December 1988. Time series are tested for stationarity and cointegration, and later an ECM is run, finding evidence of an asymmetric price adjustment. In this case the long term equilibrium seems to take around 2 years to adjust and short run price asymmetries are fully absorbed within 4 months.

In 1992, a similar study was carried out in Germany by Gebhard Kirkgaessner and Knut Kuebler, comparing gasoline and light heating oil with Rotterdam oil prices, for

2 J. D. Karrenbrock *"The Behavior of Retail Gasoline Prices: Symmetric or Not?"* (1991)

3 D. N. Manning *"Petrol prices, oil price rises and oil price falls: some evidence for the UK since 1972"* (1991)

both wholesale and retail[4]. Data used are monthly Federal Republic of Germany consumer and product prices from 1972 to 1989 divided into two sub-periods, one referring to the seventies and one to the eighties. The model specified is again an ECM used in order to separate short run adjustment from long run equilibrium. The model found presence of asymmetry defined as a quicker reaction to increasing oil prices against a slower one to decreases. The authors reject the hypothesis of profit maximizing behavior, blaming political-economic conditions.

An interesting result is the one coming from a 1994 study of David Shin[5], who, after reviewing previous studies, found that data used aren't similar and can lead to different outcomes, specifically in the price transmission mechanism from crude oil to wholesale and from wholesale to retail. Using monthly *United States Department of Energy (DOE*) prices for two periods; 1986-1990 and 1990-1992 the author runs an OLS regression between weighted wholesale product prices and crude oil and between wholesale gasoline prices and crude oil, expressing a variable for price increases and one for price decreases. Asymmetry didn't emerged from the calculation. A result subsequently confirmed by a non linear partial adjustment model and other series of tests, concluding that no asymmetric adjustment results from the data. Moreover, in Shin (1994) if an asymmetry exists in oil price to wholesale it is in the opposite direction from the Rockets and Feathers assumption. Nonetheless this finding doesn't

4 Gebhard Kirgaessner, Knut Kuebler *"Symmetric or asymmetric price adjustments in the oil market"* (1992)
5 Shin David *"Do product prices respond symmetrically to changes in crude prices?"* (1994)

say anything about the wholesale to retail price transmission behavior, leaving it open for further studies.

Later on in 1997 a paper from Severin Borenstein, A. Colin Cameron and Richard Gilbert[6] investigated the price transmission mechanism, in order to find price asymmetries and possibly to draw conclusions related to economic theory.

By analyzing the price response at each level of distribution, the authors aim was to distinguish between the competing explanation for the price asymmetry. Borenstein, Cameron, Gilbert (BCG) used West Texas Intermediate (WTI) daily spot crude oil prices, generic gasoline prices (wholesale) which is taken from daily Gulf Coast prices. The retail prices are average weekly prices from a survey in 42 cities east of the Rocky Mountains. In this study they defined 3 main phases of price transmission; crude oil to wholesale, wholesale to retailer and crude oil to retailer. The time span was from January 1986 to December 1990 with weekly frequency, and the model used is a bivariate ECM for every level of distribution, in order to discriminate price transmission. Cumulative adjustment of retail price to spot crude oil price changes show that retail gasoline prices adjust more quickly to increases than to decreases in crude oil prices. Symmetry is rejected at a 1% level and the lag chosen is 10 weeks, assuming that all the asymmetry will be absorbed within this interval. On the crude oil to wholesale relationship no asymmetry was found while a strong one showed up in wholesale to retailer, leading the authors to draw three possible hypotheses:

6 S. Borenstein, A. C. Cameron, R. Gilbert *"Do gasoline prices respond asymmetrically to crude oil price changes?" (1997)*

1. Prices are sticky downward, because when input prices fall, the old output price offers a natural focal point for oligopolistic sellers.
2. Production lags and finite inventories of gasoline imply that negative shocks to the future optimal gasoline consumption path can be accommodated more quickly than positive shocks.
3. Volatile crude oil prices create a signal-extraction problem for consumers that lowers the expected pay-off from search, and makes retail outlets less competitive.

In 1998 another important study on the matter was published by Nathan S. Balke, Stephen P. A. Brown and Mine K. Yücel from the Federal Reserve Bank of Dallas[7].

Using weekly data for spot oil price, gasoline spot price for intermediate, retail price of gasoline at the pump (with and without taxes) and wholesale gasoline prices as an average of U.S. wholesale distributor's prices they settled on a level specification and an ECM, focusing on price transmission upstream and downstream, expanding Borenstein studies. Five series of prices are analyzed, ranging from January 1986 to August 1996. The 2 models differ only on their specification, as the ECM adopts the difference of the parameters compared to the levels model that didn't differentiate.

Results in this case are more inclined to an overall asymmetry, excluding symmetry in 9 out of 10 pairs. Magnitude of asymmetry is then tested for the two models, showing that the one implied by ECM is several times larger than that implied by the levels model, in particular during the first 8 weeks following a change in upstream price.

7 Balke et al. *"Crude oil and gasoline prices, an asymmetric relationship?" (1998)*

Asymmetry was found but the authors recognized that these findings are sensitive to model specification.

In the same period another work was carried out in United Kingdom (UK) by Berry Reilly and Robert Witt called “Petrol price asymmetries revisited”[8]. Using data for United Kingdom the focus is not only on the price asymmetry between crude oil and petrol prices but also on the exchange rate between sterling and the dollar and the impact of this. In this way it revises the work of Bacon (1991) and Manning (1991). The model used is an unrestricted dynamic ECM based on monthly data for gasoline net of VAT and excise duty retail prices for the UK, Brent crude oil prices from British Petroleum in cents per liter and exchange rate obtained as a monthly average of dollar against sterling prices. The period considered goes from January 1982 to June 1995.

Results show a short-run response of retail petrol prices to changes in input costs and the exchange rate, and the hypothesis of a symmetric response by petrol retailers to crude oil price rises and falls is rejected. Most of the increase in crude oil prices appear to be passed on within a given month with cost reductions taking somewhat longer to feed through to pump prices.

Short run exchange rate effects are found, operating with a longer lag than crude prices. However no evidence is found about the full transmission of oil prices and exchange rates in the long run. In addition, any shocks to the long-run steady state relationship are found to be eliminated quite rapidly, most of them within 3 months.

8 B. Reilly, R. Witt *“Petrol price asymmetries revisited”* (1998)

In 2003 a review of Borenstein et al. (1997) work is presented by Lance J. Bachmeier and James M. Griffin[9]. In this paper the authors test the assumption of Borensteins using both a standard two step ECM and a non standard ECM the same as used by Borenstein. Higher frequency data are adopted, with daily prices from February 1985 to November 1998 on a single phase of transmission, that is, the simple crude oil/regional wholesale gasoline relationship. This leads to quite different conclusions indeed. The null hypothesis of asymmetry is rejected in the response of regional wholesale gasoline prices to crude-oil price shocks. With daily regional gasoline prices adjusting almost instantaneously and symmetrically to crude-oil price changes. The authors blame the adoption of the standard approach to ECM and the frequency of data to be the source of the contrasting results.

In same year another important study was published by Marzio Galeotti, Alessandro Lanza and Matteo Manera[10]. It examines five European countries, namely Germany, France, UK, Italy and Spain using monthly data from 1985 to 2000 and allowing for an international comparison using same model. Two stages of price transmission are taken in account, crude to gasoline spot and spot to retail, and the exchange rate is added with prices coming from the International Monetary Fund (IMF), the International Energy Agency (IEA) and the Rotterdam Freight On Board (FOB) spot gasoline index. The model used is a two step asymmetric ECM and the results of the estimated

---

9 L. J. Bachmeier, J. M. Griffin *"New evidence on asymmetric gasoline price responses"* (2003)

10 Galeotti et al. *"Rockets and feathers revisited: an internationalcomparison on European gasoline markets"* (2003)

parameters generally point to widespread differences, both in adjustment speeds and short-run elasticities when input prices rose or fell. The usual test statistics fail to reject the null hypothesis of symmetry and a more powerful test was conducted, proving that “R&F appear to dominate the price adjustment mechanism of gasoline markets in many European countries”.

Stanislav Radchenko[11] from the University of North Carolina focused his attention on the volatility of prices and their impact, in order to find and test economic theory about the asymmetry. Weekly U.S. DOE deseasonalized prices for oil and gasoline, stretching from March 1991 to February 2003, where inspected for their volatility, using mostly a Vector Autoregressive model (VAR). Three measures of oil price volatility and 12 measures of gasoline price asymmetry were built comparing the impulse response functions of gasoline price asymmetry to a shock in oil price volatility. After dividing the sample in two for low and high oil price volatility the outcome shows that the decline in the asymmetry is attributed to a faster response of gasoline prices to oil price decreases, when oil price volatility increase leads to a possible oligopolistic behavior explanation. This makes a case for linking the analysis of the data to economic theory.

In line with the Borenstein et al. (1997) paper is the study of Li-Hsueh Chen, Miles Finney and Kon S. Lai published in 2005 which expands previous research in different stages of transmission to future markets, using a threshold ECM model.

---

11 S. Radchenko *“Oil price volatility and the asymmetric response of gasoline prices to oil price increases and decreases”* (2005)

Based on weekly data on WTI crude oil spot and future, gasoline spot and future New York Mercantile Exchange (NYMEX) and US retail gasoline (EIA weekly Motor Gasoline Price Survey) prices ranging from January 1991 through March 2003, it distinguished 3 stages split between spot and future prices. Results from the analysis show asymmetry at every stage in both long and short run and spot and future, coupled with the interesting finding that the asymmetry arises more on downstream stage (retail) than upstream (crude to refinery).

The study shows that differences arise from country to country, frequency and period of dataset, stages of price transmission and model used, with the latter claimed to be the source of the diverging outcomes. Following this study it was necessary to test the sensitivity of the empirical results against the model used. This is explored in a paper published by Margherita Grasso e and Matteo Manera in 2007[12]. They used a long dataset ranging from 1985 to 2003 with monthly frequency for 5 European countries (Italy, Spain, France, Germany and UK). The study tested the three main models in use, namely the asymmetric ECM, autoregressive threshold ECM and ECM with threshold cointegration. Three price transmission stages were taken in consideration, single stage (crude oil to gasoline retail), first stage (crude oil to gasoline wholesale) and second stage (gasoline wholesale to gasoline retail). Results show that in single stage analysis, the magnitude of ‘positive’ coefficients (the reaction to a positive impulse) is larger than the ‘negative’. And in the two-stage analysis the overall

12 M. Grasso, M. Manera *“Asymmetric error correction models for the oil–gasoline price relationship”* (2007)

magnitude of the coefficients appear to be larger in the first stage than in the second, with adjustment to equilibrium level faster in the first. All models are able to find the temporal delay in the reaction of retail prices to changes in wholesale gasoline and crude oil prices, as well as some evidence of asymmetric behavior. However, the types of stages and the number of countries that are characterized by asymmetric oil–gasoline price relations vary across models.

Also focused on Europe is the work of Aidan Meyler[13], who expanded the analysis to 12 countries (Belgium, Germany, Ireland, Greece, Spain, France, Italy, Luxembourg, the Netherlands, Austria, Portugal and Finland) and the Euro area as a whole, with weekly data for crude oil prices and consumer liquid (gasoline, diesel, heating) divided in four sub-periods of time (1994–2008, 1994–1999, 2000–2008, and 2004–2008). The paper directed attention to the pass-through mechanism, upstream and downstream. An ECM with cointegration was built, using raw data (not in logs) to achieve more robust results, following on from the analyses of causality, stationarity, cointegration and structural breaks. As for the price pass-through, results show a complete transmission of price at relative fast speed , with around 80% of price increase/decrease absorbed within 3 weeks. For asymmetry in response to price changes, no pervasive evidence is found.

13 A. Meyler *"The pass through of oil prices into euro area consumer liquid fuel prices in an environment of high and volatile oil prices"* (2009)

A different method is used by Afshin Honarvar[14], who in his 2009 paper uses a Crouching Error Correction Model (CECM) who defines cointegration as the sum of positive and negative errors (or white noises) when the differences in a time series are a random walk. So to test for hidden cointegration, the series are split in positive and negative movements and then accumulated for every time period *t*. Data used are monthly observations for gasoline prices and crude oil prices from US Department of Energy (EIA), covering the period Sep. 1981 – Dec. 2007.

Conclusions found that there is no evidence of a cointegrating relationship between positive movements of retail gasoline and crude oil prices, or, between the negative components of gasoline and crude oil prices. Asymmetry, according to the authors, can be due to consumer response to technological change instead of oligopolistic behaviors.

Ahmed Atil, Amine Lahiani and Duc Khuong Nguyen[15] in their 2013 study dropped the linearity assumption with a Nonlinear Autoregressive Distributed Lags model (NARDL) to examine the pass-through of oil prices to gasoline prices and natural gas. Employing US monthly spot closing prices for WTI crude oil, gasoline, and Henry Hub natural gas from January 1997 to September 2012, the author's model estimates, relative to gasoline, show a short run asymmetry and a non rejectable long run

---

14 A. Honarvar "*Asymmetry in retail gasoline and crude oil price movements in the United States: An application of hidden cointegration technique*" (2009)

15 A. Athil, A. Lahiani, D. K. "*Nguyen Asymmetric and nonlinear pass-through of crude oil prices to gasoline and natural gas prices*" (2013)

symmetry, paired with a lag response to crude oil prices, concluding that asymmetry exist in the short run.

This chapter ends with the 2015 study from Ladislav Kristoufek and Petra Lunackova[16] who re-investigated the oil-gasoline price asymmetry in the international market. Seven nations were studied; Belgium, France, Germany, Italy, The Netherlands, UK and US. They adoptied a redefined ECM with fractional integration, long-term memory and borderline (non) stationarity, thus examining ECM characteristics and proposing a different methodology for the study of the asymmetries.

After an exhausting overview of the literature a large dataset of weekly price observations from January 1996 to May 2014 was investigated, using two different price indexes, Brent for European countries and WTI for US, while gasoline prices are weekly averages for spot retailer prices. Using two new tests built by the authors to check for the asymmetry it appears that results of asymmetry aren't reliable, hence, no statistically significant signs of Rockets and Feathers hypothesis is found.

16 L. Kristoufek, P. Lunackova *"Rockets and feathers meet Joseph: Reinvestigating the oil–gasoline asymmetry on the international markets"* (2015)

2.2 Italian case

This section aim to give a closer look at the literature concerning Italy. The studies already cited have taken into consideration Europe and the main European countries including Italy, hence a brief recap of these is provided. Other papers are then scrutinized in order to provide a comprehensive insight.

From the already mentioned studies 4 were based on European data including Italy, namely the works of Galeotti et al. (2003), Grasso/Manera (2007), Mayler (2008) and Kristoufek/Lunackova (2015). The first two found evidence of asymmetry in some or all stages of price transmission for Italy and the second two didn't find them.

It would appear that a deeper investigation is needed.

Following the same year-of-publishing order, the first reviewed study is the 2009 work of Jorge Rodrigues[17] that is based on weekly price observation in all EU15 member states for a 4 year period starting in 2004. Data are WTI prices for crude oil and European Platt (an ex-refinery benchmark price) for gasoline and diesel. The model is the ECM with cointegration, taking into account the first price transmission as international and the second price transmission as national, using Platt as a common wholesale price for the single countries, EU15 as a whole and two subsets (Mediterranean countries and north western countries).

17 J. Rodrigues *"Asymmetries in the adjustment of motor diesel and gasoline pump prices in Europe"* (2009)

In Italy, asymmetry is found only in second stage (domestic) for diesel, while gasoline shows symmetric behavior.

The second study considered is the one of Giancarlo Fiorito[18], that covers 5 countries (Germany, Italy, France, Spain and UK) using Brent crude oil prices and European Commission gasoline and diesel prices. Frequency is weekly and time span is January 2000 to January 2010. The model is an ECM paired with an autoregressive model in order to search for maximum and minimum intervals. The outcome was a consistent asymmetry for almost all countries for gasoline and diesel. Italy showed asymmetry in both, with a larger one in diesel (common to all the countries considered, and possibly emerging from the larger consumption share of the latter against gasoline).

The first and only study found to be focused solely on Italy is "*Asymmetries in the adjustment of motor diesel and gasoline pump prices in Europe*" from Antonio Angelo Romano and Giuseppe Scandurra[19]. In this paper the authors investigated asymmetry for different stages of price transmission and also for different levels of volatility (like Radchenko did in 2005) of the time series, adopting weekly data for Brent crude oil price, Platt wholesale price and retailer price (provided by the Italian Ministry of Economic Development MISE). It starts with an investigation of oil and wholesale gasoline volatility using a Generalized Autoregressive Conditional Heteroskedasticity

18 G. Fiorito "*La dinamica dei prezzi dei carburanti rispetto alle quotazioni del petrolio, una analisi a livello europeo*" (2010)

19 A. A. Romano, G. Scandurra "*Price asymmetries and volatility in the Italian gasoline market*" (2012)

(GARCH) model and continues with a standard asymmetric ECM to compute an estimate of the gasoline price reaction to upward and downward movements. Final results show asymmetry in the second stage (Platt to retail gasoline prices) and an interesting issue is that a decline in asymmetry is attributed to a faster response of gasoline price to industrial price decrease when Platt's price volatility increases, meaning that asymmetric behavior increases with high price volatility.

The overview continues with the 2013 study by Fabrizio Venditti[20] from the Italian central bank (Banca d'Italia), who analyzed the response of weekly gasoline and diesel prices to oil prices in the US, the Euro area and the four largest Euro area countries (Germany, France, Italy and Spain). It used a weekly frequency for oil prices, as well as wholesale and retail gasoline and diesel prices. Data for the latter comes from Eurostat for European case and US department of Energy of United States, while oil prices are Brent for EU and WTI for US. range is January 1999 to September 2009.

The model specified by the author is an ECM with cointegration for two chain pass-through, direct from oil to retail and with wholesale in the middle. Results of the regression for all the countries taken shows limited evidence of asymmetries in the Euro area including Italy and some evidence for the US.

The last work cited was carried out in 2016 by Alberto Bagnai and Christian Alexander Mongeau Ospina[21]. It concentrated on the Eurozone, using a NARDL model to test for

---

20 F. Venditti *"From oil to consumer energy prices: How much asymmetry along the way?"* (2013)

21 A. Bagnai, C. A. Montegau Ospina *""Asymmetric asymmetries" in Eurozone markets gasoline pricing"* (2016)

R&F in both short run and long run. A low monthly frequency for crude oil prices, gasoline prices and exchange rate was used, coming from the European commission (EIA) and Pacific exchange rate service. The time span was January 1999 to December 2015 using around 200 observations. The results show no presence of asymmetry in short run for all the countries in the study.

What emerges from this overview is that differences arising when testing the R&F hypothesis can be attributed to the model specified, the number of stages of price transmission, frequency and span of the time series and the area or country analyzed (including the exchange rate parameter).

3. Price analysis of crude oil and derivatives

Drawing a model to test asymmetry requires finding a suitable dataset from which data can be collected. Once found, data must be made conformable to the model; difference in frequency, measurement unit, presence of gaps in the series etc., are worked out.

Crude oil price statistics are gathered first. Price is international as crude oil is globally traded.

The same doesn’t apply for the two consumer fuel prices of petrol and diesel, that are strongly related to the national market. For a study on Italy, Italian prices are collected.

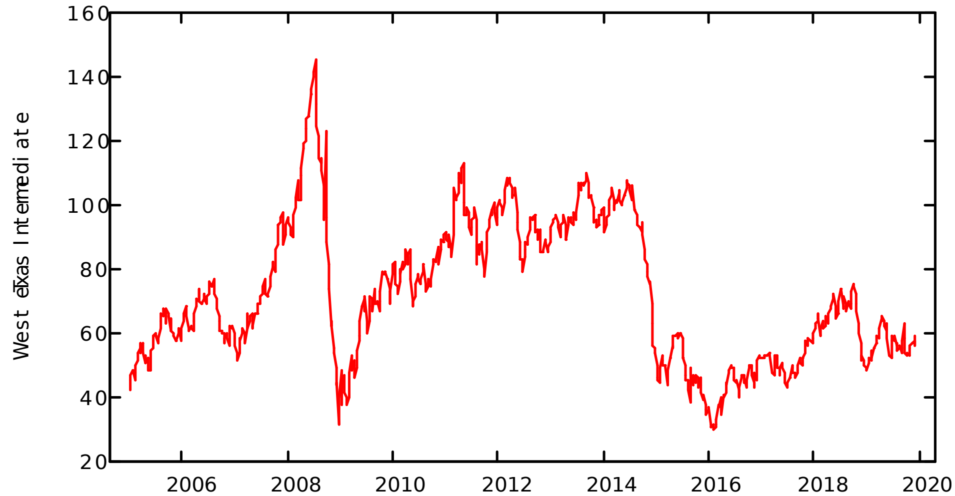
Given the sources and the type of data, a weekly frequency has been chosen to be a valid representation of reality. With the first observation recorded the 3rd of January 2005 and the last recorded on 9th December 2019, covering about 15 years, the sample size is quite consistent, involving 720 observations.

When using daily data, like the source of crude oil price, a one week price is the simple average of the 7 daily prices.

3.1 Crude oil price benchmarks

For crude oil price, data chosen is the WTI (West Texas Intermediate) weekly price. WTI is a crude stream produced in Texas and southern Oklahoma which serves as a reference or "marker" for pricing a number of other crude streams worldwide and which is traded in the US domestic spot market at Cushing, Oklahoma.

Fig. 1: WTI price



Crude oil price time series are published by the US energy information administration[22] and available in DBnomics database. Figure 1 provides a visual pattern of the US dollar price of one barrel of WTI crude oil, where volatility can be easily seen.

22 https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RWTC&f=W

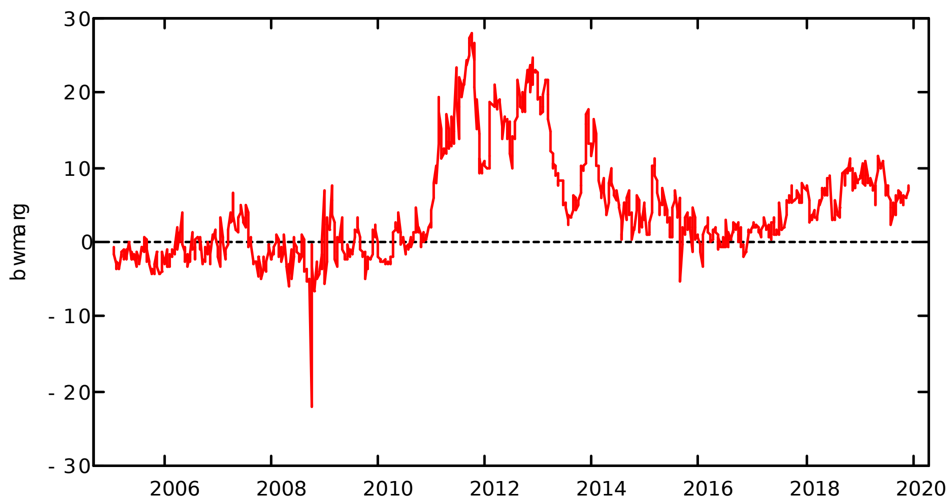
However, WTI is not the only benchmark for the price of crude oil, the second most important is Brent, which is the leading global price benchmark for the Atlantic basin crude oils. It is used to price two thirds of the world's internationally traded crude oil supplies.

Historically, price differences between Brent and other crude indexes have been based on physical differences in crude oil specifications and short-term variations in supply and demand. Prior to September 2010, a typical price difference per barrel of between ±3 USD/bbl compared to WTI and OPEC Basket existed. However, since autumn 2010 Brent has been priced much higher than WTI, reaching a difference of more than $11 a barrel by the end of February 2011 (WTI: 104 USD/bbl, LCO: 116 USD/bbl). In February 2011 the divergence reached $16 during a supply glut in the form of record stockpiles at Cushing, Oklahoma before peaking at above $23 in August 2012. It has since (September 2012) decreased significantly to around $18 after refinery maintenance settled down and supply issues eased slightly[23].

23 https://en.wikipedia.org/wiki/Brent_Crude

To get an insight, the spread between Brent and WTI in USD/bb is plotted in Fig.2.

Fig. 2: Brent WTI spread



For the purpose of this work WTI prices will be adopted, because of higher availability of data and no significant difference in outcome.

Source data frequency is daily, so, in order to match the weekly periodicity of petrol and diesel series, week price is the simple average of the daily prices of the relative week.

Gaps in the sequences are filled via Gretl using gap_filler function, included in package extra.gnf. Filling is done using interpolation instead of mean.

3.2 Fuel prices in Italy

Provided in this chapter is an overview about consumer fuel prices for the country in the analysis presented.

It will define how final price is formed, the steps it goes through and the single series patterns and statistics, including the application of duties and taxation.

3.2.1 Petrol and Diesel price formation process

Different phases encompass the process of final price formation from when the crude oil is bought, but they are not always the same. Nonetheless, these three price stages are deemed crucial:

1. Crude oil price before refinement and transport
2. Wholesale price of petroleum products
3. Retail price of petroleum products.

The first one is crude oil price as it is quoted and traded before being moved and transformed.

The second is the industrial price after refinement and transportation, where oil companies margins are made (if they do not control the whole production chain clearly).

The third is the retail price, that, together with VAT, excise duties and distributor margins constitute the price paid by the private user.

This work will focus on the first and the last price (excluding VAT and excise duties), which means searching for non competitive behaviors along the production chain that can be attributed to either wholesaler and retailer. The reason for that is the lower availability of data when considering wholesale and retail separately.

3.2.3 Petrol price series

Time series used here come from the weekly European Commission oil bulletin[24], which is a periodically published review of consumer prices for petroleum products and their components (wholesale price, taxes and excise duties) for every European country. The type of petrol is the Euro Super 95, which is a European standard 95 octane hydrocarbon mix, or simply the most available and consumed pump petrol in Europe.

Sources of the data are the 10 largest oil companies operating in Italy, plus a selection of traders and independent operators, alongside roughly 30 supermarkets/white pump filling stations.

Market share of each company is defined based on quantity sold in the previous year. Coverage is above 95% for each product.

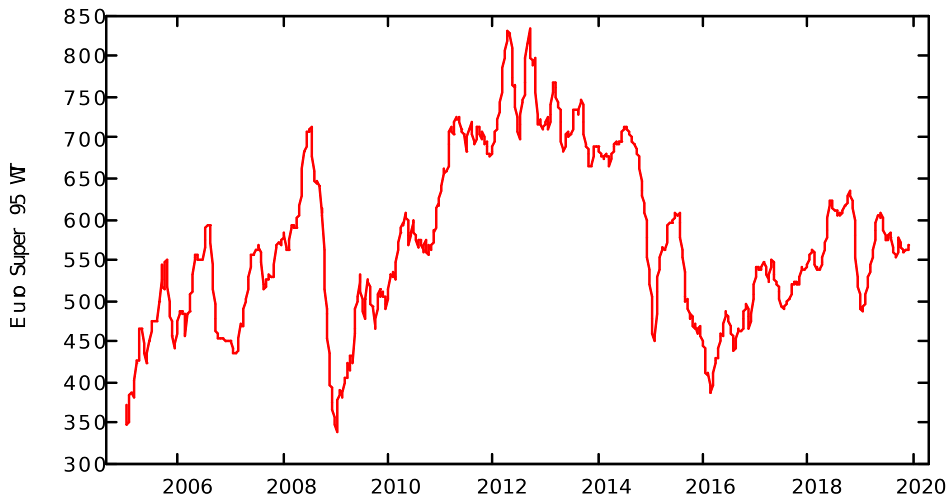
24 https://crude oil.europa.eu/energy/en/data-analysis/weekly-oil-bulletin

Companies report to the Italian Ministry of Economic Development one weighted average price based on sales volumes recorded for different models of self-service. The Ministry then carries out a monthly based survey on consumption of petroleum products in Italy.

Reported petrol price is expressed in Euros for 1000 liters and does not represent the pump price paid by the consumer, because it is not considering taxes and excise duties.

Sequence of petrol prices is plotted below:

Fig. 3 Petrol price Euro-Super 95 without taxes and excises in €/1000l.



At first glance some issues emerge; the almost vertical plunge of prices after the 2008 financial crises and the subsequent recovery, as well as the 2015 plunge reflecting a lack of demand in response to an oversupply triggered by 2 factors; a growth in production from the United States, due to technological development in hydraulic fracturing and horizontal drilling and an increased production from the Organization of Petroleum Producing Countries (OPEC) countries, notably Saudi Arabia which, after debating cutting production surprised the market by holding it, pushing prices further down[25].

Main statistics on petrol prices follows:

```
Summary Statistics, using the observations 2005-01-03 - 2019-12-09
for the variable Petrol (780 valid observations)
  Mean           Median           Minimum          Maximum
 573.21          561.35           338.20           833.43
  Std. Dev.      C.V.             Skewness         Ex. kurtosis
 104.37          0.18208          0.26802          -0.67592
  5% Perc.       95% Perc.        IQ range         Missing obs.
 424.25          742.91           181.64           0
```

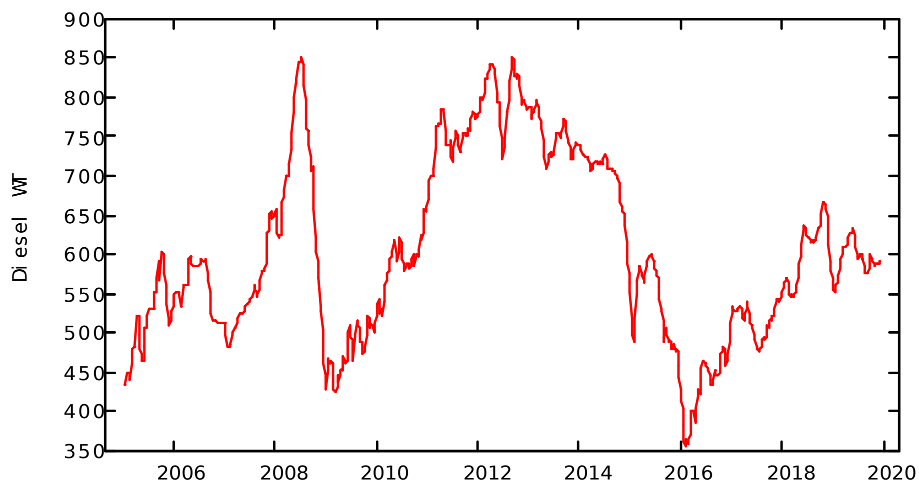
Price variations have been quite high during the sampled years, with a maximum almost 3 times the minimum reached, standard deviation and coefficient of variation reflects this variability.

25 D. Mead, P. Stieger, "*The 2014 plunge in import petroleum prices: What happened?*", U.S. Bureau of labor statistics (2015)

3.2.3 Diesel price

The following is a plot of diesel price out of taxes end excise duties for Italy in the same period, with gaps already filled. Also in this case the price is in Euros for 1000 liters. The graph shows a similar pace to the previous graph, being both diesel and petrol crude oil derivatives. 2015 crude oil price plunge seems to have reduced diesel prices more than petrol ones.

Fig. 4 Diesel price without tax and excise in €/1000l



Looking at the summary statistics, the coefficients are similar to the previous fuel, with a slightly bigger standard deviation and coefficient of variation, signaling even greater volatility.

Summary Statistics, using the observations 2005-01-03 - 2019-12-09
for the variable d (780 valid observations)

Mean	Median	Minimum	Maximum
604.45	585.74	354.85	851.86
Std. Dev.	C.V.	Skewness	Ex. kurtosis
115.40	0.19091	0.32842	-0.85804
5% Perc.	95% Perc.	IQ range	Missing obs.
444.95	799.49	198.69	0

3.2.3 Crude Oil price

Like the other series, a gap filling operation is needed for crude oil price, and implemented using the econometric software.

The spot price is in US dollars per barrel FOB (freight on board). Petrol and diesel prices expressed in Euros, with the exchange rate series needed to compare them.

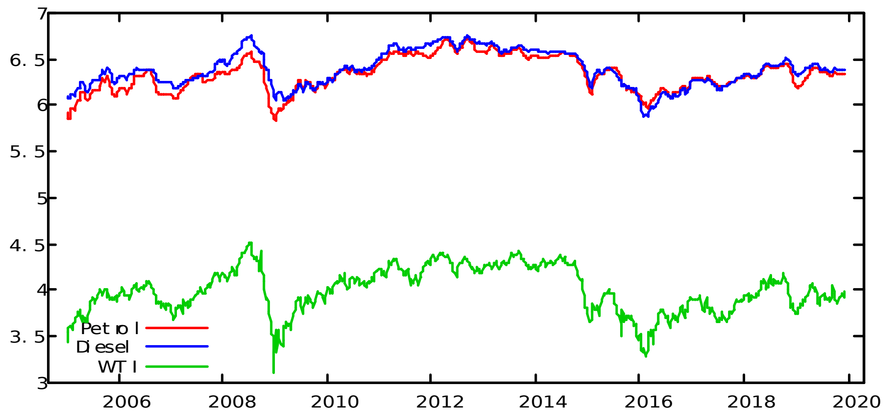
US Dollar/Euro exchange rate time series are available in FRED database from Federal reserve bank of St. Louis[26].

After dividing the crude oil series for the exchange rate expressed in Dollars per Euros the series are re-expressed in relative terms with their logarithms. As the focus of the work is to understand the price behavior in time, logarithms will always be used.

26 https://fred.stlouisfed.org/series/DEXUSEU/

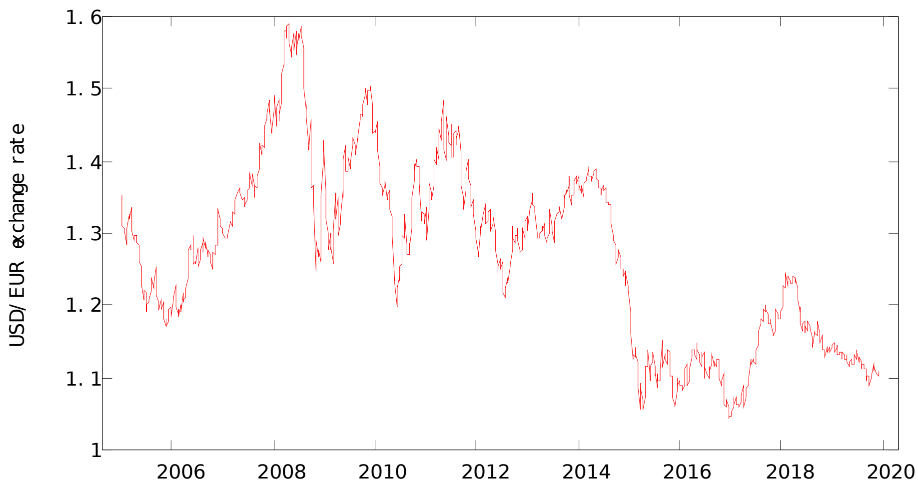
Figure 5 shows the joint behavior of the prices.

Fig. 5: Joint prices in log terms



The 3 series mirror each other quite closely. Data from WTI crude oil are in Euros, are already converted by the exchange rate and plotted below.

Fig. 6: USD per Euro



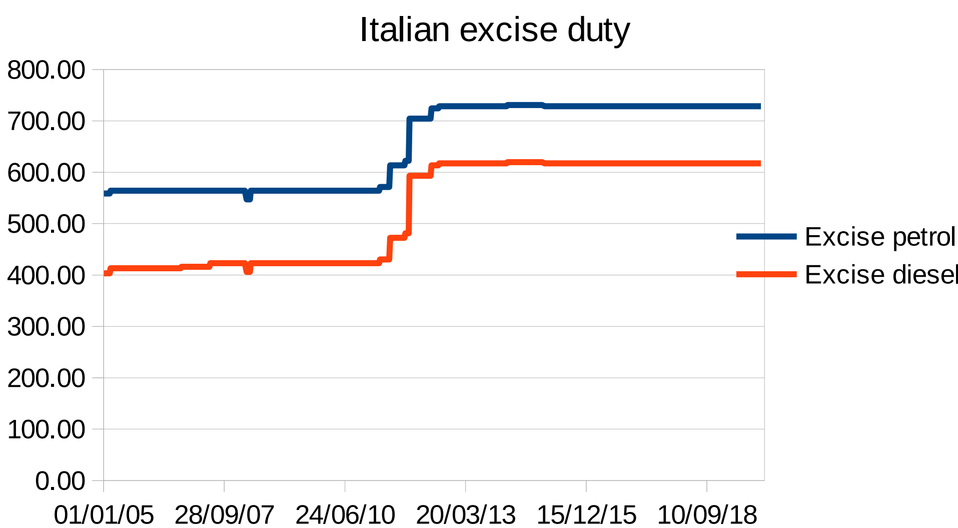
3.2.4 Retail price components for Petrol and Diesel in Italy

In Italy, retail price of fuels is burdened by excise duties and a quite high VAT rate compared to EU average.

Excise duties are fixed amounts, expressly 617,40 €/1000lit for diesel, accounting for around 42% of the final price and 728,40€/1000lit, around 46% of the retail price for petrol. The value of the duty is the sum of all the single entries for different purposes, ranging from financing the 1956 Suez crisis to the recent earthquakes in Emilia Romagna region.

Fig. 7 Shows the excise duties plotted against time:

Fig. 7 Excise duties

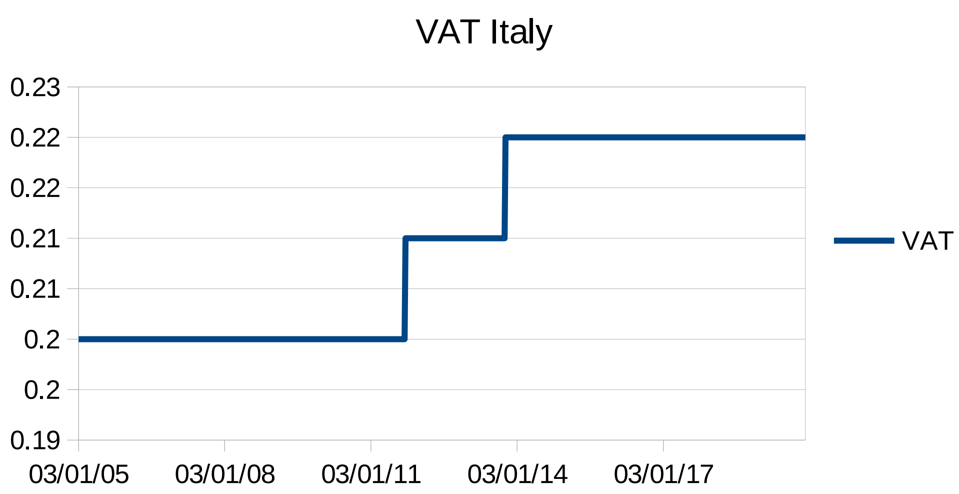


Government policies are quite clear looking at the evolution in time, with the introduction of 5 new excises in the year 2011 and 2012.

Subsequently, Value Added Tax applies, but with a peculiarity, that VAT does not only apply on retailer fuel price but also to the excise duty, meaning a higher overall tax rate and a double imposition.

VAT rate is a very debated topic in Italian politics, being settled at 22% against a mean of around 21% in EU countries.

Fig. 8 VAT in Italy



From 2005 VAT has been increased twice, first in September 2011 up to 21% and then in October 2013 to 22%.

Following the overview of the data and how they are composed, the next chapter will discuss the econometric model implementation, in order to investigate the presence of asymmetry in response to crude oil upswing and downswing.

4. Econometrics

4.1 The ECM model

To test for the presence of asymmetry in response to upward and downward movement in the price of oil a bivariate ECM model is build, but first a brief explanation of a basic ECM model derived from an ADL model.

4.1.1 From ADL to ECM representation[27]

In time series analysis, one way to express the relationship between two variables taking into account their history, is by using an Autoregressive Distributed Lag model (ADL or ARDL) that can be expressed as ADL(p,q):

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=0}^{q} \beta_i x_{t-i} + \epsilon_t$$

Where the dependent variable y in a certain period t depends from its past values until *p* times ago, and the present and past (until q) values of the x variable.

Given the above equation, the concept of Dynamic Multiplier can be introduced; a dynamic multiplier is the impact of a variable and its lags in a certain period on the dependent variable at time *t*.

For example, the first dynamic multiplier d_0 (called the impact multiplier) at time *t* is the effect x_t have on the dependent variable, namely β_0 .

27 ADL explanation comes from the lecture notes in “Basic Econometrics” from Riccardo Lucchetti

A multiplier can be defined as the partial derivative $d_o = \frac{\partial y_t}{\partial x_t}$, $d_1 = \frac{\partial y_t}{\partial x_{t-1}}$ and so on. To enable ease of interpretation d_h is the effect on y_t of something that happened h periods ago.

Worth studying is also the effect on y_t of a permanent change in x_t , that at time 0 would simply be d_0 and then overlapping the following dynamic multipliers. The new sequence of multipliers c_j , called the interim multipliers, is defined as:

$$c_j = d_0 + d_1 + d_2 + ... + d_j = \sum_{i=0}^{j} d_i$$

And the long-run effect of a permanent change in x_t is the limit of the interim multiplier $c = \lim_{j \to \infty} c_j$. Long run multiplier is of great importance, telling the proportion of y_t and x_t in steady state. In steady state neither variable changes until external shocks strike the system from outside.

Basically, the critical multipliers from an economic point of view are:

- the impact multiplier d_0
- the long run multiplier c

A good way to express these parameters immediately is by rewriting an ADL model in an Error Correction one. An example; having a ADL(1,1) in the form

$$y_t = \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1}$$

and taking the first difference using the Δ operator where $\Delta y_t = y_t - y_{t-1}$ gives

$$\Delta y_t = (\alpha - 1) y_{t-1} + \beta_0 \Delta x_t + (\beta_0 + \beta_1) x_{t-1}$$

which, after rearranging terms result in the ECM form

$$\Delta y_t = \beta_0 \Delta x_t + (\alpha - 1)[y_{t-1} - \frac{\beta_0 + \beta_1}{(1-\alpha)} x_{t-1}]$$

where β_0 is the impact multiplier and $c = \frac{\beta_0 + \beta_1}{(1-\alpha)}$ is the long run multiplier.

Long run relationship or ECM term is defined as $y_{t-1} - c x_{t-1}$ and $(1-\alpha)$ describes the fraction of disequilibrium that gets reabsorbed in one period.

4.1.2 The basic ECM representation

The basic ECM representation for the variables under scrutiny is the following

$$\Delta pd_t = \sum_{i=0}^{p} \beta_i \Delta wti_{t-i} + \sum_{i=1}^{q} \alpha_i \Delta pd_{t-i} + \theta z_{t-1} + \epsilon_t$$

Where the lower case letters represent the logarithms of the variables, with Δpd_t as the first difference of the log of price of oil derivative and Δwti_{t-i} the first difference of crude oil price logarithm, β_i and α_i the relative coefficients. p and q are the lags.

The first implication of the ECM model is that a long run equilibrium exists between the price of oil and the price of its derivative, either diesel or petrol. The implication is that the series move together, or, in a more specific term, they are said to cointegrate. The long run equilibrium is represented by the variable z_{t-1} that is $z_{t-1} = (pd_{t-1} - \gamma_0 - \gamma_1 wti_{t-1})$ with θ the relative coefficient.

The ECM model peculiarity is that it separates short run impact from long run equilibrium, in this case the long term multiplier is represented by the coefficient related to z_{t-1} and the short run by the ones related to Δwti_{t-i} , Δpd_{t-i} and their values lagged.

The focus is on the coefficient θ that tells how much of the disequilibrium relative to the long run gets re-absorbed in one period. In other words, if we expect that the system is not in the long run equilibrium, the coefficient θ measures the speed of adjustment needed to go back to equilibrium in one period, the closer this is to 1 the faster the adjustment will be.

4.1.3 Bivariate ECM model

In order to check for the presence of asymmetry in response to a price increase or decrease, that is, the parameter θ when the price moves up or down, the model must be re-expressed in the function of the upward and downward movements. This results in the following:

$$\Delta pd_t = \sum_{i=0}^{p} \beta_i^+ \Delta wti_{t-i} + \sum_{i=1}^{q} \alpha_i^+ \Delta pd_{t-i} + \theta^+ z_{t-1} + \sum_{i=0}^{p} \beta_i^- \Delta wti_{t-i} + \sum_{i=1} \alpha_i^- \Delta pd_{t-i} + \theta^- z_{t-1}$$

The coefficients are the same as before but are now divided by their upward and downward movement; the subscript $^+$ and $^-$ define positive and negative variation, meaning that β_i^+ applies when $\Delta wti_{t-i} \geq 0$ and similarly for β_i^- when $\Delta wti_{t-i} < 0$. Obviously the same happens for α_i in presence of the positive and negative movements of Δwti_{t-i} . In this case, if the behavior of price is symmetrical, θ^+ and θ^- must not be statistically different. The ECM will be specified in 2 different models depending on how positive and negative variations are defined. The two ways are:

- Positive and negative movement of the upstream price, in this case, the crude oil price.
- Positive and negative disequilibrium, which means if the price is above or below the long run coefficient.

4.1.3 The NARDL model

Being the ECM model shown before a simple reparametrization of an ADL (autoregressive distributed lags) model, split into positive and negative variations against the long run relationship. An alternative way is to set a NARDL (non-linear autoregressive distributed lags) model, where the long run relationship is split into upward and downward movements. In this case the long run relationship that was previously stated as $pd_{t-1}=\gamma_0+\gamma_1 wti_{t-1}$ becomes

$$pd_{t-1}=\gamma_0+\gamma_1 wti^{+}_{t-1}+\gamma_2 wti^{-}_{t-1}$$

Setting the NARDL model as

$$y_t=\beta_0+\sum_{j=1}^{p}\beta_j y_{t-j}+\sum_{j=0}^{q}(\delta_j^{+} x_j^{+}+\delta_j^{-} x_j^{-})+u_t$$

This, after differentiation takes the form of a conditional ECM

$$\Delta y_t=\alpha_0+\rho y_{t-1}+\pi^{+} x^{+}_{t-j}+\pi^{-} x^{-}_{t-1}+\sum_{j=1}^{p}\theta_j \Delta y_{t-j}+\sum_{j=0}^{q}(\delta^{+}\Delta x^{+}_{t-j}\,\delta^{-}\Delta x^{-}_{t-j})+u_t$$

where $\rho=(\sum_{j=1}^{p}\beta_j-1)$ is the error correction speed, and $\pi^{+}=\frac{(\sum_{j=0}^{q}\delta_j^{+})}{-\rho}$ is the long run coefficient related to the positive variation of the upstream variable,

$\pi^{-}=\frac{(\sum_{j=0}^{q}\delta_j^{-})}{-\rho}$ is the one related to the negative variation.

Inference can be made on these coefficients (there is also the constant term after differentiation $\alpha_0=\frac{\beta_0}{-\rho}$) as well as in the short run ones. To test for cointegration the Bound test is performed (see cointegration chapter).

After defining the equation and finding that cointegration does exist, restrictions are applied to the long run coefficients to check for their similarity. If their differences do not statistically diverge from zero, no asymmetry is present in the long run. The same tests are conducted to short run relationship in order to check if the dependent variables respond symmetrically.

In the next chapter a more deep analysis is undertaken, first, looking at price pass-through and mark-up, how they are related and how they behave. Then, giving these characteristics, the study goes on to define different models that can catch presence of asymmetry in price variation.

4.2 Pass through and cointegration

As first step, the focus is a more in deep analysis on the price structure, his relationship with price of crude oil, the value of mark-up and pass-through between prices. This is made initially using OLS to get a first indication. However, an OLS regression, without a previous cointegration analysis, can be spurious and providing misleading statistical inference. Cointegration will be treated later. This chapter starts with defining and investigating key elements needed in the study, starting with mark-up.

4.2.1 Mark-up and Price pass-through

The mark-up is defined as the ratio between price and cost of a good related to the cost, in other words it explains the relationship between the margin and the cost.

In this case the final price is intended as the price that is paid by the retailer free of taxes and excises.

In formulas; if mark-up in a certain period is M_t , with P_t being the final price and C_t the cost, $M_t = (\frac{P_t}{C_t})$. After rearranging data in logs is can be re-expressed as $m_t = (p_t - c_t)$ where the lower case represent the natural logarithm of the variables.

If the cost of a good is, let's say 60, with a final price of 100, the markup is then (100/60)=1,6666, that is, a markup of around 66%.

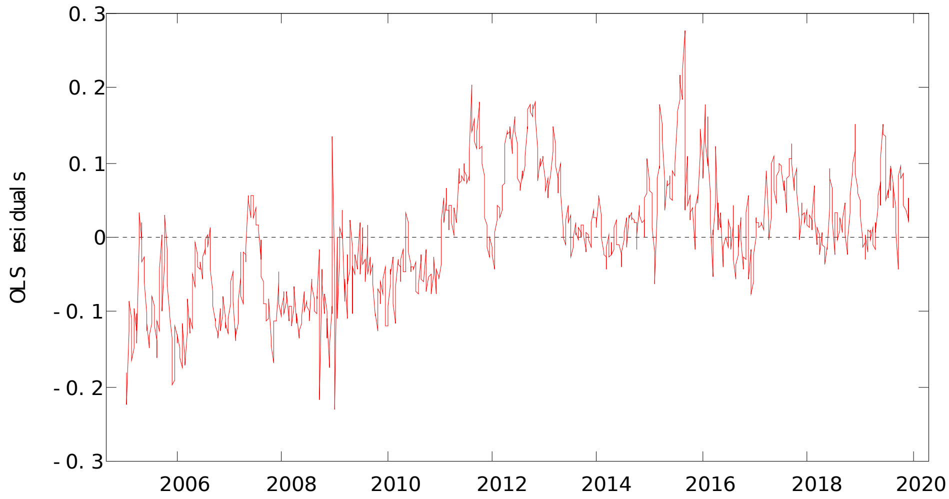
When there is a relationship between the two variables, which is something that is expected (and can be proved by cointegration analysis), as one variable is a derivative of the other one, it should be possible to define a linear combination between the two. Given the two time series, a simple OLS can provide a first quick look at the behavior of the data.

Defining PD_t as the price of the derivative, either petrol or diesel, and $POIL_t$ the price of crude oil, a linear relationship can be defined as $pd_t = \gamma_0 + \gamma_1 poil_t + \epsilon_t$ with γ_0 the constant term, γ_1 the coefficient related to price of oil and ϵ_t the error term of the regression.

Rearranging terms it is possible to look at ϵ_t as a proxy of the mark-up, in other words, given the above linear relationship, m_t can be expressed as $m_t = pd_t - \gamma_0 - \gamma_1 poil_t = \epsilon_t$.

After running an OLS for petrol/crude oil relationship, in the form of $p_t = \gamma_0 + \gamma_1 wti_t + \epsilon_t$, with p_t price of petrol and wti_t price of WTI crude oil (as above, lower case = logarithm of the variable), plotting the residuals, outcome is the following:

Fig 9: OLS residuals



What comes clear at first sight is the presence of a linear trend, something really interesting that would require more investigation, and probably a study by his own. For the moment, and for the sake of this study, a second OLS is run to get rid of the trend, in the form of $p_t = \gamma_0 + \gamma_1 wti_t + \gamma_3 t + \epsilon_t$ with t being the time variable.

As expected, the residuals of the second equation doesn't show sign of a trend, which is captured by the time variable. However, this is good just enough to get a grip on the subject. To have better estimates and something where inference can be done, the concept of cointegration must be introduced. But before another notion is placed, that is the one of price pass-through.

Price pass-through is the amount of price that is passed downstream from a variable to another in the production process. In this case, the amount of crude oil price that is transferred to either petrol or diesel.

In the previous specification of a linear relationship between the 2 variables the pass-through would be identified with the coefficient γ_1, which can be interpreted as the amount of crude oil price that is found in the price of petrol.

4.1.2 Lag selection

Before delving deeper into the analysis, it is important to check the number of lags that will be needed in order to explicate the model. A quick way of checking the number of lags is using a correlogram.

The correlogram basically checks the autocorrelation in the series in a defined number of lags and sets a confidence interval for the significance of the coefficients.

Two different methods are defaulted by Gretl;

- ACF (Auto-Correlation Function) that is the standard version as expressed above and
- PACF (Partial Auto-Correlation Function) which gives the partial correlation of a stationary time series with its own lagged values, regressing the values of the time series at all shorter lags.

Running a correlogram for the variation of the log of petrol Δb and diesel Δd result in the following graphs.

Fig. 10: Petrol price variation correlogram

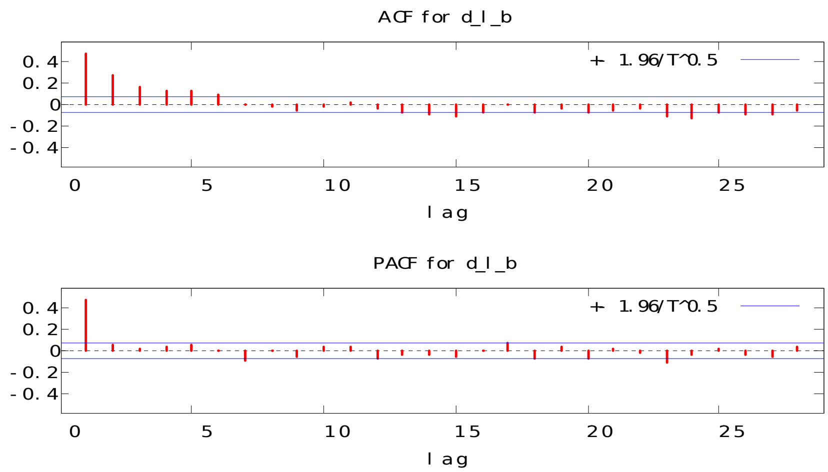
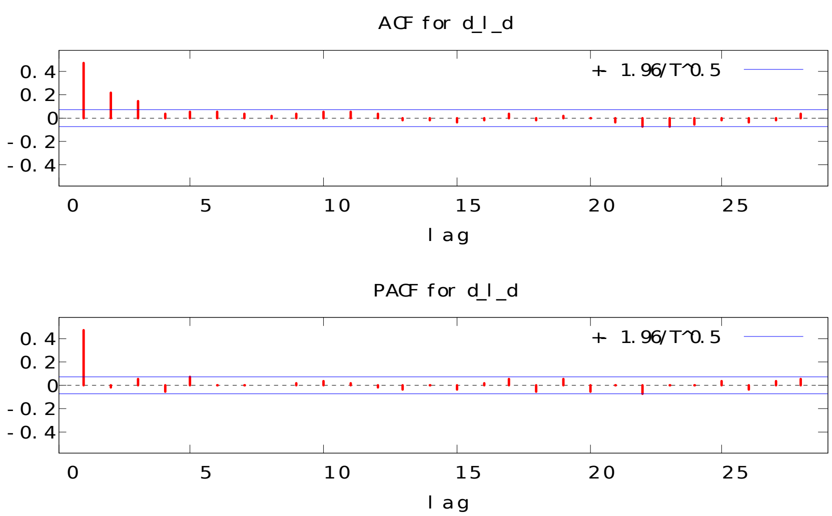


Fig. 11: Correlogram for diesel



Looking at PACF for the log variation of diesel and petrol partial autocorrelation is found one lag ahead.

In order to find the best lag in a VAR analysis which is needed for Johansen cointegration testing between variables, the `--lagselect` option is used when specifing the VAR. It provides the maximum lag order given the information criteria AIC, BIC and HQC, and will be shown in the cointegration analysis later.

Finally, the Breusch-Godfrey test can be performed to check for presence of serial autocorrelation in the residuals of a regression, that can result from a mispecification of the lags.

4.1.3 Cointegration and unit root analysis

Two variables are said to cointegrate when they are both integrated of order one and there is a linear relationship that connects them. The order of integration is, very briefly, the number of times a variable must be differentiated in order to be stationary, this characteristic is translated in the unit root.

A unit root is a characteristic that a stochastic process can have that makes statistical inference harder. A linear stochastic process has a unit root, if 1 is a root of the process's characteristic equation[28]. If series have a unit root they are also said to be integrated of order one *I(1)* and difference stationary when their first difference is not integrated, expressly *I(0)*, that is, a process y_t is *I(1)* if the first difference Δy_t is a white noise. In case that a process is in fact *I(1)* inference can be made on his first difference. This is the reason why the ECM model is specified in differences.

Having *I(1)* processes, some speculations about them can be done; while it is known that a linear combination of series who are stationary will result in a stationary process, it is not always certain when combining different *I(1)* variables. The result of this combination can be an *I(0)* process and can be not. When it is *I(0)*, it is possible to speak of cointegration.

28 https://en.wikipedia.org/wiki/Unit_root

If cointegration exist, the series are linked together by a linear combination, so they are free to go wherever they want, but they will be always tied together in a way.

Formally, having an *I(1)* process y_t there is cointegration when at least a vector β do exist from which the following relation $z_t = \beta^{'} y_t$ result in an I(0) process.

Different methods can be adopted for finding a cointegration relationship, the first is the classical Engle and Granger[29] two-step approach; which is based on unit root analysis first and an OLS after to define the cointegration relationship, with tests on the residuals. The second is the Johansen[30] test, which can be applied in multivariate field and whose target is to find the rank of the matrix $\Pi = \alpha\beta'$ (having more variables), in order to tell if there is presence of cointegration and the number of cointegrating vectors. It does it in two ways; these are the "λ-max" test, for hypotheses on individual eigenvalues, and the "trace" test, for joint hypotheses[31]. Cointegration relationship is later described via a Vector Error Correction Model (VECM).

29 RF Engle, CWJ Granger, "*Co-integration and error correction: representation, estimation and testing*" (1987)
30 Johansen Søren, *"Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models"*. *Econometrica* (1991)
31 Gretl User's Guide (2019)

The third is the bound test to apply in NARDL and ARDL models. It is computed in 5 steps according to Peasaran et al. (2001)[32]:

1. Identification to a tentative model, that is trying to specify a model using the information criteria lowest value.
2. Estimating a conditional ECM equation using the OLS technique
3. Diagnosis testing (Breusch-Godfrey LM test)
4. Using Wald test (F-test) to test the null and alternative hypothesis
   - H0 ~ long run coefficients = 0 (No long run levels relationship)
   - H1 ~ long run coefficients ≠ 0 (Long run levels relation exist)
5. Compare computed statistics to critical values in order to reject or not null hypothesis

Bound test is computed in the NARDL paragraph alongside the model.

In order to find the best estimates of a cointegration relationship needed to run a model to test the asymmetric price behavior tests will be performed starting from an Engle and Granger approach.

32 Peasaran et al (2001) *"Bounds testing approaches to the analysis of level relationships"* Journal of applied econometrics

4.1.4 Engle and Granger cointegration approach

It starts testing for a unit root in the price series studied. To check for the presence of integration, many statistical tests can be performed. Here 3 of the most common are used;

- ADF Augmented Dickey – Fueller test
- KPSS test, named after Kwiatkowski, Phillips, Schmidt and Shin
- PP test or Phillips Perron test

Given the price of crude oil, petrol and diesel, results of the tests are plotted below, indicating the test null hypothesis and the related p-value for both the 2 oil derivatives.

Lag order is set to be 13 for ADF, testing the series unit root in presence of a constant and a constant with a trend. PPtest is conducted including a constant and trend with Barthlett window size zero. Kpss with a lag 5.

	ADF test $H_0 = y_t \sim I(1)$		PP test $H_0 = y_t \sim I(1)$	KPSS test $H_0 = y_t \sim I(0)$
	const	const+trend	const+trend	const
Petrol	0.009684	0.05924	0.0790	< .01
Diesel	0.09618	0.2876	0.2651	< .01
WTI crude oil	0.02252	0.0827	0.0220	< .01

Results are pretty concordant, excluding PPtest for crude oil and ADF without trend for crude oil and petrol, whose null hypothesis is rejected at 95% confidence level, all the 3 series are non-stationary and integrated of order one when including a trend.

Second step is to run OLS regressions for petrol and diesel against WTI crude oil, save residuals and test for stationarity. But first the right number of lags must be specified, in this case is done using the command `--lagselect` in Gretl script, which returns the maximum number of lags possible given the information criteria. Result from a VAR lag selection between log of petrol and log of WTI crude oil follows.

```
VAR system, maximum lag order 13

The asterisks below indicate the best (that is, minimized) values
of the respective information criteria, AIC = Akaike criterion,
BIC = Schwarz Bayesian criterion and HQC = Hannan-Quinn criterion.

lags        loglik    p(LR)       AIC          BIC          HQC

   1    3210.64891              -8.356321    -8.320004    -8.342342
   2    3395.21273  0.00000   -8.827152    -8.766624    -8.803854
   3    3428.82071  0.00000   -8.904356    -8.819617*   -8.871739*
   4    3432.81841  0.09175   -8.904350    -8.795400    -8.862414
   5    3439.20567  0.01243   -8.910575    -8.777414    -8.859319
   6    3439.77301  0.88873   -8.901625    -8.744252    -8.841049
   7    3442.78470  0.19741   -8.899047    -8.717464    -8.829153
   8    3451.85283  0.00116   -8.912263    -8.706468    -8.833049
   9    3455.04951  0.17163   -8.910168    -8.680162    -8.821635
  10    3461.38537  0.01300   -8.916259    -8.662042    -8.818407
  11    3463.06668  0.49907   -8.910213    -8.631785    -8.803041
  12    3468.21831  0.03562   -8.913216    -8.610577    -8.796725
  13    3473.62270  0.02880   -8.916878*   -8.590028    -8.791068
```

As comes clear from the table, 2 out of 3 information criteria choose a lag of 3 as the maximum lags possible.

Given this outcome a first OLS regression can be run, for which results are plotted above, *l_b* is the log price of petrol while *l_wteu* is the price of WTI crude oil expressed in Euros, the lag chosen is 3 .

```
Cointegrating regression -
OLS, using observations 2005-01-03:2019-12-09 (T = 780)
Dependent variable: l_b

             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const       3.67106      0.0473047     77.60    0.0000    ***
  l_wteu      0.666527     0.0118151     56.41    3.73e-277 ***

Mean dependent var   6.334611   S.D. dependent var   0.183284
Sum squared resid    5.140677   S.E. of regression   0.081287
R-squared            0.803558   Adjusted R-squared   0.803305
Log-likelihood       851.8505   Akaike criterion    -1699.701
Schwarz criterion   -1690.382   Hannan-Quinn        -1696.117
rho                  0.900550   Durbin-Watson        0.198514
```

With ADF test on regression residuals strongly rejecting the null hypothesis of a unit root. That is, residuals look stationary and cointegration between the 2 variables seems to exist. In this case the amount of crude oil price that is passed to petrol is around 66%. However when plotted, the residuals show a trend. This can be overcome as did in the mark-up section adding a time variable.

For diesel the OLS regression is the following, with *l_d* being the log of diesel price and *l_wteu* the same as before, the lag chosen comes from a second VAR lag selection that resulted in choosing the best lag as 3.

```
Cointegrating regression -

OLS, using observations 2005-01-03:2019-12-09 (T = 780)
Dependent variable: l_d

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const       3.52608      0.0422917     83.38    0.0000   ***
  l_wteu      0.715712     0.0105630     67.76    0.0000   ***

Mean dependent var   6.386187   S.D. dependent var   0.190786
Sum squared resid    4.108874   S.E. of regression   0.072673
R-squared            0.855092   Adjusted R-squared   0.854906
Log-likelihood       939.2245   Akaike criterion    -1874.449
Schwarz criterion   -1865.130   Hannan-Quinn        -1870.865
rho                  0.866966   Durbin-Watson        0.263145
```

And also here the ADF test rejects strongly the null of no stationarity, leading to the same conclusions as for petrol. However in this case pass-through seems to be higher up to 71%, and residuals doesn’t show a trend visually.

As said before, a time trend can be captured by a time variable in the OLS regression, including it and running again gives the outcome above:

	Const	*l_wteu*	*time*
l_b (petrol)	3.52188	0.682605	0.000217473
l_d (diesel)	3.47754	0.720944	7.07676e-05

As expected, time variable in petrol looks like capturing more information than for diesel, pass-through is slightly increased for both variables.

4.1.5 Johansen test and VECM for petrol and diesel

Petrol

A Johansen test is run for both petrol and diesel. As said previously, the object of this test is to find the rank of the cointegration matrix, or the number of cointegrating vectors. After the test results, a VECM is specified, from which an error correction term is derived.

First, the test is conducted on *l_b*, the petrol price series, and the specification for the deterministic term is case 4, with a unrestricted constant and a restricted trend, which seem to be the best specification for the data under investigation.

```
Johansen test:
Number of equations = 2
Lag order = 3
Estimation period: 2005-01-24 - 2019-12-09 (T = 777)
Case 4: Restricted trend, unrestricted constant

Log-likelihood = 5662.74 (including constant term: 3457.71)

Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0   0.036964     40.195 [0.0003]     29.265 [0.0008]
   1   0.013968     10.930 [0.0911]     10.930 [0.0908]

Corrected for sample size (df = 769)
Rank Trace test p-value
   0     40.195 [0.0003]
   1     10.930 [0.0908]

eigenvalue    0.036964     0.013968
```

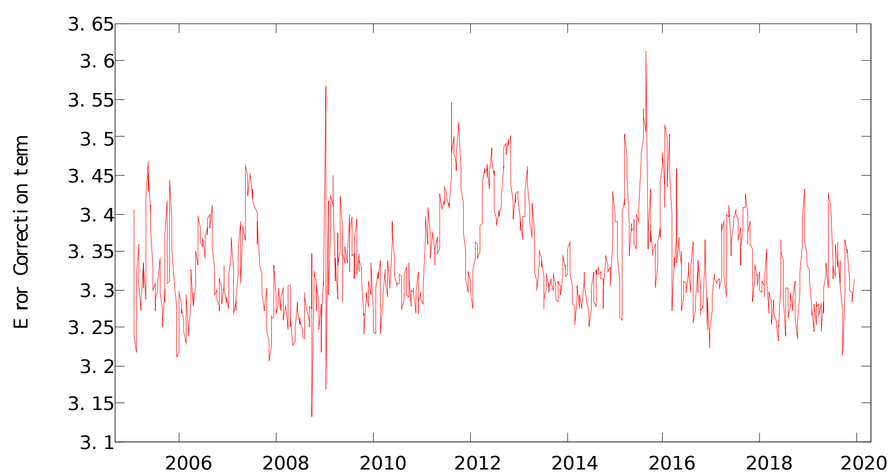
Trace test and Lmax test confirm that cointegration does exist as the rank of Π is not zero.

With a rank of 1 a VECM model is run to have a view of the cointegration matrix. Outcome is:

```
VECM system, lag order 3
Maximum likelihood estimates, observations 2005-01-24-2019-12-09 (T = 777)
Cointegration rank = 1
Case 4: Restricted trend, unrestricted constant

beta (cointegrating vectors, standard errors in parentheses)

l_b           1.0000
            (0.0000)
l_wteu     -0.72993
          (0.045011)
trend   -0.00020402
        (4.8854e-05)

alpha (adjustment vectors)

l_b       -0.034896
l_wteu     0.087276

Log-likelihood = 3452.2456
Determinant of covariance matrix = 4.7410208e-07
AIC = -8.8501
BIC = -8.7662
HQC = -8.8178
```

The results show a pass-through of around 73% which is higher than the OLS estimate, and a time trend that is pretty the same. The error correction term is shown below.

Fig. 12: Petrol VECM error correction term



Diesel

The test and VECM are then run for *l_d*, the diesel variable, with same specifications as above:

```
Johansen test:
Number of equations = 2
Lag order = 3
Estimation period: 2005-01-24 - 2019-12-09 (T = 777)
Case 4: Restricted trend, unrestricted constant

Log-likelihood = 5774.02 (including constant term: 3568.99)

Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0   0.042848     41.662 [0.0001]     34.028 [0.0001]
   1  0.0097779     7.6349 [0.2916]     7.6349 [0.2919]

Corrected for sample size (df = 769)
Rank Trace test p-value
   0     41.662 [0.0001]
   1     7.6349 [0.2916]

eigenvalue    0.042848    0.0097779
```

With cointegration confirmed and subsequent VECM as:

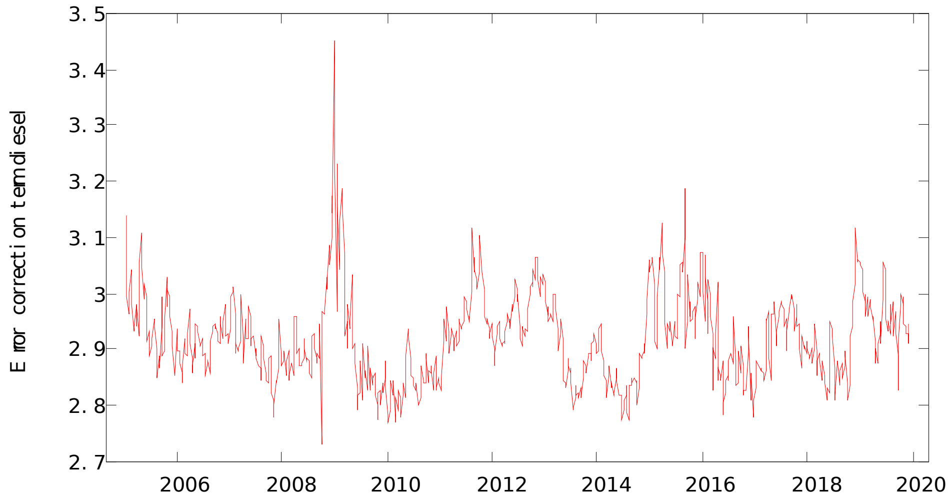
```
VECM system, lag order 3
Maximum likelihood estimates, observations 2005-01-24-2019-12-09 (T =
777)
Cointegration rank = 1
Case 4: Restricted trend, unrestricted constant

beta (cointegrating vectors, standard errors in parentheses)

l_d          1.0000
            (0.0000)
l_wteu     -0.85971
          (0.047341)
trend   -7.9219e-05
        (5.0987e-05)

alpha (adjustment vectors)

l_d       -0.030162
l_wteu     0.073304
```

```
Log-likelihood = 3565.1704
Determinant of covariance matrix = 3.545162e-07
AIC = -9.1407
BIC = -9.0568
HQC = -9.1085
```

Also in this case the pass-through is higher, and the time trend is similar to the one of the OLS.

Error correction term is plotted below:

Fig. 13: Diesel VECM error correction term



Now, having good estimates of the long term relationship who exist between the variables, it is possible to run the bivariate ECM model. It will be done using the error correction term of the VECM and the OLS residuals, even though, using maximum likelihood instead of linear regression, results from the Johansen test are more reliable.

4.1.6 Upward and downward price movement specification

Before running the model, another specification must be done, as said previously, the bivariate ECM model defines upward and downward price movements. This can be modeled in 2 ways:

- Model 1 - Recognizing as upward movements the positive variation of the upstream price, that is $\Delta wti_t > 0$ and specify the model under this assumption, so the coefficients of the long term relationship assumes positive and negative values due to the variation of crude oil price.
- Model 2 - Recognizing also as upward movements the negative disequilibrium from the long term relationship. If the long run term is defined as $pd_{t-1} = \gamma_0 + \gamma_1 wti_{t-1} + \epsilon_{t-1}$, rearranged gives $\epsilon_{t-1} = pd_{t-1} - \gamma_0 - \gamma_1 wti_{t-1}$. This imply that if the price of crude oil have a sudden shock upward, the price of the derivative (either petrol or diesel) will be lower than the equilibrium, so that $pd_{t-1} < \gamma_0 + \gamma_1 wti_{t-1}$ and consequently $\epsilon_{t-1} < 0$. This would require an upward movement to go back to equilibrium by the derivative price. The opposite is expected in the case of a fall in crude oil price. This is the specification used by Chen, Finley and Lai in their 2005 threshold ECM analysis. In this model, the long run coefficient applies when the relationship is above or below the equilibrium. In this case

the change in regime, positive or negative, is not given by the price variation in one period but on whether negative or positive is the disequilibrium. Using the word of Chen, Lai and Finley "Consider two alternative situations: *(1)* The crude oil price jumped 5 dollars in the first period and then fell 10 cents in the second period, and (2) the crude oil price jumped 4 dollars in the first period, followed by another 90 cents increase during the second period. Both led to a net increase of 4.9 dollars for the crude oil price. Because of lagged responses to the higher crude oil price, the retail price would rise over time in both situations. Under the conventional approach, the first scenario would trigger a regime witch but the second scenario would not."[33]

In this work both configurations will be considered using the OLS residuals, and results will be compared and discussed.

33 Chen, Finley, Lai, "*A threshold cointegration analysis of asymmetric price transmission from crude oil to gasoline prices*" (2005)

4.3 Running the model

Once all the specifications are made, it is possible to run the model.

First thing, having the logs of the 3 time series; petrol, diesel and crude oil, they are differentiated one time to be stationary.

4.3.1 Asymmetric ECM from Engle and Grenger two step approach model 1

As written before, using the residuals from the OLS cointegration relationship seen before, two models are defined, one on price variation and one on disequilibrium:

Model 1 – Long run response to upstream price variations:

Once the series are differentiated they are split in upward movements and downward movements of the upstream variable, in this case, the price of crude oil:

- $\Delta wti_t^{+} = max(\Delta wti_t ; 0)$
- $\Delta wti_t^{-} = min(\Delta wti_t ; 0)$

Therefore Δwti_t^{+} represent the upward movement in one period, while Δwti_t^{-} the downward one.

Number of lags chosen is two so the model can be expressed for petrol as:

$$\Delta p = \beta_0^{+} \Delta wti_t + \beta_1^{+} \Delta wti_{t-1} + \beta_2^{+} \Delta wti_{t-2} + \alpha_0^{+} \Delta p_{t-1} + \alpha_1^{+} \Delta p_{t-2} + \theta^{+} z_{t-1}$$

$$+ \beta_0^{-} \Delta wti_t + \beta_1^{-} \Delta wti_{t-1} + \beta_2^{-} \Delta wti_{t-2} + \alpha_0^{-} \Delta p_{t-1} + \alpha_1^{-} \Delta p_{t-2} + \theta^{-} z_{t-1}$$

Where Δp_t is the price variation of petrol, Δwti_t the price variation of WTI crude oil and z_{t-1} the long run equilibrium between price of petrol and price of WTI crude oil, defined as $z_{t-1}=(p_{t-1}-\gamma_0-\gamma_1 wti_{t-1})$, θ^{+} defines the speed of adjustment to equilibrium in case of upward movements of crude oil price Δwti_t^{+} and θ^{-} the respective one for downward movements Δwti_t^{-} . The same apply for the log variation of diesel Δd_t .

Petrol robust OLS regression model 1

```
OLS, using observations 2005-01-24:2019-12-09 (T = 777)
Dependent variable: ld_b
HAC standard errors, bandwidth 6 (Bartlett kernel)

                coefficient    std. error   t-ratio   p-value
  -----------------------------------------------------------
  const_p        0.00246258    0.00111466    2.209    0.0275   **
  ld_b_1_p       0.189383      0.0963709     1.965    0.0498   **
  ld_b_2_p       0.0333439     0.0662416     0.5034   0.6149
  ld_wteu_p      0.0132980     0.0301636     0.4409   0.6594
  ld_wteu_1_p    0.128357      0.0336097     3.819    0.0001   ***
  ld_wteu_2_p    0.0789654     0.0287370     2.748    0.0061   ***
  z_1_p         −0.0451364     0.0135177    −3.339    0.0009   ***
  const_m       −0.000616840   0.00117834   −0.5235   0.6008
  ld_b_1_m       0.353919      0.0501944     7.051    3.98e-12 ***
  ld_b_2_m       0.0977637     0.0558747     1.750    0.0806   *
  ld_wteu_m      0.0418767     0.0266920     1.569    0.1171
  ld_wteu_1_m    0.169297      0.0166101    10.19     5.86e-23 ***
  ld_wteu_2_m    0.0880898     0.0235126     3.747    0.0002   ***
  z_1_m         −0.0431145     0.0130486    −3.304    0.0010   ***

Mean dependent var   0.000628   S.D. dependent var   0.019699
Sum squared resid    0.154916   S.E. of regression   0.014249
R-squared            0.485558   Adjusted R-squared   0.476793
F(13, 763)           47.35783   P-value(F)           6.65e-89
Log-likelihood       2207.625   Akaike criterion    −4387.251
Schwarz criterion   −4322.075   Hannan-Quinn        −4362.179
rho                 −0.018994   Durbin-Watson        2.036557

P-value was highest for variable 29 (ld_wteu_p)
```

The coefficients under investigation are `z_1_p` and `z_1_m` (respectively θ^{+} and θ^{-}) who represent the response of petrol to a shock in crude oil price against the long run equilibrium.

Then Breusch-Godfrey test is performed to check the presence of serial autocorrelation in the residuals, outcome is the following:

```
Test statistic: LMF = 1.262528,
with p-value = P(F(52,711) > 1.26253) = 0.107

Alternative statistic: TR^2 = 65.680941,
with p-value = P(Chi-square(52) > 65.6809) = 0.0962

Ljung-Box Q' = 64.6802,
with p-value = P(Chi-square(52) > 64.6802) = 0.111
```

The null hypothesis of no autocorrelation can't be rejected, meaning that residuals of the regression aren't statistically correlated.

A later test is conducted to check the hypotheses of the 2 coefficients being identical.

Via the *restrict* command in Gretl, imposing the null hypothesis of $H_0 \sim \theta^{+} - \theta^{-} = 0$.

The software automatically runs an asymptotic F-test for the null.

```
Restriction:

 b[z_1_p] - b[z_1_m] = 0

Test statistic: Robust F(1, 763) = 0.011623, with p-value = 0.914175
```

As the p-value of the test doubtlessly state, it is not possible to reject the null, meaning that as far as this model can go, sign of asymmetric price behavior cannot be found in the reaction of retail petrol price to a positive or negative crude oil shock.

Diesel robust OLS regression model 1

For diesel the lag chosen is 3, because for a lag of 1 and 2 the Breusch-Godfrey test rejects the null of no autocorrelation, regression is the following.

```
OLS, using observations 2005-01-31:2019-12-09 (T = 776)
Dependent variable: ld_d
HAC standard errors, bandwidth 6 (Bartlett kernel)

                coefficient   std. error    t-ratio   p-value
  ------------------------------------------------------------
  const_p        0.00221076   0.000890121    2.484    0.0132   **
  ld_d_1_p       0.235215     0.0657504      3.577    0.0004   ***
  ld_d_2_p      −0.110416     0.0571735     −1.931    0.0538   *
  ld_d_3_p       0.0525274    0.0389914      1.347    0.1783
  ld_wteu_p      0.0139681    0.0230000      0.6073   0.5438
  ld_wteu_1_p    0.115602     0.0244394      4.730    2.68e-06 ***
  ld_wteu_2_p    0.0881190    0.0203793      4.324    1.74e-05 ***
  ld_wteu_3_p    0.0164825    0.0167984      0.9812   0.3268
  z_1_p         −0.0348334    0.00901652    −3.863    0.0001   ***
  const_m       −0.00103746   0.00103590    −1.002    0.3169
  ld_d_1_m       0.241883     0.0543359      4.452    9.80e-06 ***
  ld_d_2_m      −0.0381773    0.0750051     −0.5090   0.6109
  ld_d_3_m       0.0663865    0.0772465      0.8594   0.3904
  ld_wteu_m      0.0353135    0.0272641      1.295    0.1956
  ld_wteu_1_m    0.178987     0.0160039     11.18     5.54e-27 ***
  ld_wteu_2_m    0.120367     0.0204728      5.879    6.17e-09 ***
  ld_wteu_3_m    0.0578572    0.0202131      2.862    0.0043   ***
  z_1_m         −0.0248565    0.0101191     −2.456    0.0143   **

Mean dependent var   0.000402   S.D. dependent var   0.017604
Sum squared resid    0.112440   S.E. of regression   0.012179
R-squared            0.531845   Adjusted R-squared   0.521345
F(17, 758)           47.47183   P-value(F)           5.6e-107
Log-likelihood       2328.624   Akaike criterion    −4621.248
Schwarz criterion   −4537.473   Hannan-Quinn        −4589.019
rho                  0.001317   Durbin-Watson        1.993634

P-value was highest for variable 34 (ld_d_2_m)
```

Breusch-Godfrey test results:

```
Test statistic: LMF = 1.232300,
with p-value = P(F(52,706) > 1.2323) = 0.132

Alternative statistic: TR^2 = 64.572246,
with p-value = P(Chi-square(52) > 64.5722) = 0.113

Ljung-Box Q' = 63.7252,
with p-value = P(Chi-square(52) > 63.7252) = 0.128
```

Also in this case, taking the 2 coefficients linked to the long run equilibrium for upward and downward movements and testing them gives the following outcome:

```
Restriction:
 b[z_1_p] - b[z_1_m] = 0
Test statistic: Robust F(1, 758) = 0.525894, with p-value = 0.468562
```

As well as with petrol, an asymmetric behavior in diesel price does not seem to be statistically significant. A more complete description will be given later.

4.3.2 Asymmetric ECM for Engle and Grenger two step approach model 2

Model 2 – long run response to disequilibrium

Here upward and downward movements are specified as situations when the price of the downstream variable is over or under the equilibrium. That is;

- $z_{t-1}^{+}=min(z_{t-1};0)$
- $z_{t-1}^{-}=max(z_{t-1};0)$

Because, as said previously, a crude oil price increase imply a situation where the equation is below the equilibrium, meaning that price of petrol or diesel needs to catch back up the price of crude, while a crude oil price decrease means equation above equilibrium.

Stating $z_{t-1}=pd_{t-1}-\gamma_0-\gamma_1 wti_{t-1}$ after a positive shock in crude oil price $\Delta wti_{t-1}>0$ it is expected that $pd_{t-1}<\gamma_0+\gamma_1 wti_{t-1}$ so $z_{t-1}^{+}<0$.

On this assumption the model is shaped over the long run disequilibrium with positive and negative Δwti_t applying as before, but positive and negative θ applying to positive and negative z_{t-1} .

Petrol robust OLS cointegration relationship Model 2

Running the model under this assumptions gives the following outcome. Number of lags chosen is 2.

```
OLS, using observations 2005-01-24:2019-12-09 (T = 777)
Dependent variable: ld_b
HAC standard errors, bandwidth 6 (Bartlett kernel)

                coefficient    std. error   t-ratio   p-value
  ---------------------------------------------------------------
  const_p       −0.000796361   0.00132177   −0.6025   0.5470
  ld_b_1_p       0.204874      0.0943273     2.172    0.0302   **
  ld_b_2_p       0.0291585     0.0668596     0.4361   0.6629
  ld_wteu_p      0.00892426    0.0315606     0.2828   0.7774
  ld_wteu_1_p    0.123879      0.0315116     3.931    9.22e-05 ***
  ld_wteu_2_p    0.0685337     0.0287953     2.380    0.0176   **
  const_m       −0.00376526    0.00137836   −2.732    0.0064   ***
  ld_b_1_m       0.355545      0.0483460     7.354    4.96e-13 ***
  ld_b_2_m       0.106299      0.0578155     1.839    0.0664   *
  ld_wteu_m      0.0529649     0.0262410     2.018    0.0439   **
  ld_wteu_1_m    0.154931      0.0144557    10.72     4.61e-25 ***
  ld_wteu_2_m    0.0830032     0.0229110     3.623    0.0003   ***
  zplus_1       −0.130151      0.0223197    −5.831    8.12e-09 ***
  zminus_1       0.00446217    0.0131019     0.3406   0.7335

Mean dependent var   0.000628   S.D. dependent var   0.019699
Sum squared resid    0.150520   S.E. of regression   0.014045
R-squared            0.500158   Adjusted R-squared   0.491642
F(13, 763)           51.81065   P-value(F)           1.33e-95
Log-likelihood       2218.811   Akaike criterion    −4409.622
Schwarz criterion   −4344.445   Hannan-Quinn        −4384.549
rho                 −0.006540   Durbin-Watson        2.010445

P-value was highest for variable 34 (ld_wteu_p)
```

Here the coefficient related to z^{+}_{t-1} , *zplus_1* in the regression and z^{-}_{t-1} *zminus_1,* takes different values depending on the upper or lower distance to the equilibrium. In fact, in case of a positive crude oil price shock (*zplus*), prices are below the equilibrium level, and need to readjust. From what it looks like in the regression, the θ^{+} coefficient related to an increase in price of WTI crude oil is quite larger than the coefficient θ^{-} related to a decrease in crude. Breusch-Godfrey test can't reject the null and a consequent F-test confirms what is seen:

```
Restriction:
 b[zplus_1] - b[zminus_1] = 0

Test statistic: Robust F(1, 763) = 20.1619, with p-value = 8.21515e-06
```

This could be seen as a clear sign of asymmetry depending on where the prices lay against the equilibrium. And supporting evidence of R&F hypothesis.

Diesel robust OLS regression Model 2

Having diesel as downstream variable the Model 2 regression is the following (also here 2 lags are chosen):

```
OLS, using observations 2005-01-24:2019-12-09 (T = 777)
Dependent variable: ld_d
HAC standard errors, bandwidth 6 (Bartlett kernel)

                coefficient   std. error    t-ratio   p-value
  ------------------------------------------------------------
  const_p        0.00116536   0.000927653    1.256    0.2094
  ld_d_1_p       0.243983     0.0621195      3.928    9.36e-05 ***
  ld_d_2_p      −0.0628027    0.0463242     −1.356    0.1756
  ld_wteu_p      0.00354605   0.0228655      0.1551   0.8768
  ld_wteu_1_p    0.114358     0.0235482      4.856    1.45e-06 ***
  ld_wteu_2_p    0.0827913    0.0204051      4.057    5.47e-05 ***
  const_m       −0.00207400   0.00118228    −1.754    0.0798   *
  ld_d_1_m       0.310209     0.0507017      6.118    1.51e-09 ***
  ld_d_2_m       0.0337128    0.0530533      0.6355   0.5253
  ld_wteu_m      0.0432695    0.0248970      1.738    0.0826   *
  ld_wteu_1_m    0.162091     0.0143509     11.29     1.83e-27 ***
  ld_wteu_2_m    0.0966539    0.0191731      5.041    5.79e-07 ***
  zplus_1       −0.0632571    0.0123027     −5.142    3.46e-07 ***
  zminus_1      −0.0153414    0.0108870     −1.409    0.1592

Mean dependent var   0.000401   S.D. dependent var   0.017593
Sum squared resid    0.114366   S.E. of regression   0.012243
R-squared            0.523828   Adjusted R-squared   0.515715
F(13, 763)           63.51336   P-value(F)           4.9e-112
Log-likelihood       2325.528   Akaike criterion    −4623.056
Schwarz criterion   −4557.880   Hannan-Quinn        −4597.984
rho                 −0.029630   Durbin-Watson        2.057158

P-value was highest for variable 35 (ld_wteu_p)
```

The 2 coefficients related to the long run equilibrium look similar also here, B-G test can't reject the null and F-test rejects the symmetry.

```
Restriction:
 b[zminus_1] - b[zplus_1] = 0

Test statistic: Robust F(1, 763) = 6.51402, with p-value = 0.0108965
```

Also here, the null hypothesis of the 2 variables being the same is rejected at 95% confidence level, showing a different reaction of diesel price for positive and negative deviations from equilibrium.

4.3.3 Asymmetric ECM from VECM

Here the long run relationship used to run the model is not the residual of the OLS but the error correction term resulting from the VECM built before. Equation is:

$$\Delta\, pd_t = \sum_{i=0}^{p} \beta_i^{+}\, \Delta\, wti_{t-i} + \sum_{i=1}^{q} \alpha_i^{+}\, \Delta\, pd_{t-i} + \theta^{+}\, EC + \sum_{i=0}^{p} \beta_i^{-}\, \Delta\, wti_{t-i} + \sum_{i=1} \alpha_i^{-}\, \Delta\, pd_{t-i} + \theta^{-}\, EC$$

where EC is the error correction term resulting from the VECM cointegration relationship.

Petrol ECM with EC

Regression for petrol, using a lag of 2 follows:

```
OLS, using observations 2005-01-24:2019-12-09 (T = 777)
Dependent variable: ld_b
HAC standard errors, bandwidth 6 (Bartlett kernel)

                coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const_p        0.147820     0.0438062     3.374    0.0008   ***
  ld_b_1_p       0.145377     0.0989447     1.469    0.1422
  ld_b_2_p       0.0311279    0.0665888     0.4675   0.6403
  ld_wteu_p      0.0193268    0.0315198     0.6132   0.5400
  ld_wteu_1_p    0.159948     0.0332349     4.813    1.80e-06 ***
  ld_wteu_2_p    0.0794926    0.0288974     2.751    0.0061   ***
  z_1_p         −0.0435992    0.0132213    −3.298    0.0010   ***
  const_m        0.131306     0.0432652     3.035    0.0025   ***
  ld_b_1_m       0.313601     0.0534586     5.866    6.64e-09 ***
  ld_b_2_m       0.0959073    0.0555256     1.727    0.0845   *
  ld_wteu_m      0.0391691    0.0268648     1.458    0.1452
  ld_wteu_1_m    0.199571     0.0171286    11.65     5.46e-29 ***
  ld_wteu_2_m    0.0885733    0.0237478     3.730    0.0002   ***
  z_1_m         −0.0395512    0.0129338    −3.058    0.0023   ***

Mean dependent var   0.000628   S.D. dependent var   0.019699
Sum squared resid    0.155439   S.E. of regression   0.014273
R-squared            0.483823   Adjusted R-squared   0.475029
F(13, 763)           46.53751   P-value(F)           1.22e-87
Log-likelihood       2206.318   Akaike criterion    −4384.635
Schwarz criterion   −4319.459   Hannan-Quinn        −4359.563
rho                 −0.019519   Durbin-Watson        2.037734

P-value was highest for variable 30 (ld_b_2_p)
```

Breusch-Godfrey test does not reject the null for residuals autocorrelation and plot is omitted for the sake of brevity. The test for the long run coefficient in plotted above.

```
Restriction:
 b[z_1_p] - b[z_1_m] = 0
Test statistic: Robust F(1, 763) = 0.0476041, with p-value = 0.827344
```

Ensuing a strong inability of rejecting the null hypothesis. With result similar to the model 1 regression for petrol resulting from the Engle and Granger approach.

Diesel ECM with EC

Lag is 3 periods.

```
OLS, using observations 2005-01-31:2019-12-09 (T = 776)
Dependent variable: ld_d
HAC standard errors, bandwidth 6 (Bartlett kernel)

                coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const_p        0.123668     0.0353951     3.494    0.0005   ***
  ld_d_1_p       0.202884     0.0655786     3.094    0.0020   ***
  ld_d_2_p      −0.106634     0.0548843    −1.943    0.0524   *
  ld_d_3_p       0.0511501    0.0372180     1.374    0.1697
  ld_wteu_p      0.0132786    0.0214378     0.6194   0.5358
  ld_wteu_1_p    0.142904     0.0231241     6.180    1.05e-09 ***
  ld_wteu_2_p    0.0906206    0.0206687     4.384    1.33e-05 ***
  ld_wteu_3_p    0.0147391    0.0159812     0.9223   0.3567
  z_1_p         −0.0363653    0.0106228    −3.423    0.0007   ***
  const_m        0.0835099    0.0357073     2.339    0.0196   **
  ld_d_1_m       0.217755     0.0532456     4.090    4.78e-05 ***
  ld_d_2_m      −0.0401403    0.0752784    −0.5332   0.5940
  ld_d_3_m       0.0779522    0.0768231     1.015    0.3106
  ld_wteu_m      0.0410197    0.0252031     1.628    0.1040
  ld_wteu_1_m    0.200449     0.0150083    13.36     1.08e-36 ***
  ld_wteu_2_m    0.121587     0.0208262     5.838    7.83e-09 ***
  ld_wteu_3_m    0.0613416    0.0199102     3.081    0.0021   ***
  z_1_m         −0.0252749    0.0106985    −2.362    0.0184   **

Mean dependent var   0.000402   S.D. dependent var   0.017604
Sum squared resid    0.112525   S.E. of regression   0.012184
R-squared            0.531492   Adjusted R-squared   0.520985
F(17, 758)           46.55134   P-value(F)           2.4e-105
Log-likelihood       2328.332   Akaike criterion    −4620.663
Schwarz criterion   −4536.889   Hannan-Quinn        −4588.435
rho                  0.001007   Durbin-Watson        1.995238
P-value was highest for variable 34 (ld_d_2_m)
```

B-G and restriction non rejected.

```
Restriction:

 b[z_1_p] - b[z_1_m] = 0

Test statistic: Robust F(1, 758) = 0.62484, with p-value = 0.429502
```

In this case as well, results are similar to the one from the Engle and Granger approach.

4.3.4 NARDL model for asymmetric behavior in petrol and diesel price adjustments

Petrol

The output of the NARDL model comes from the "Addon for Function Package of GRETL on Asymmetric Cointegration and Dynamic Multiplier in a NARDL Framework" provided by Kuppusamy Dhanasekaran.

After an ADF test to check for difference stationarity of the data with same results as the previous, a VAR system to select best lags is done, choosing 3 as the best lag length. The OLS in levels is then provided, where Y is the price of petrol with the 3 lags, Xp and Xn respectively the positive and negative price variations of crude oil price, and their relative lags. The related OLS model is plotted above:

```
OLS, using observations 2005-01-31:2019-12-09 (T = 776)
Dependent variable: Y

             coefficient    std. error    t-ratio     p-value
  ---------------------------------------------------------------
  const       0.256959      0.0526843     4.877      1.31e-06  ***
  Y_1         1.23819       0.0353431    35.03       4.29e-161 ***
  Y_2        −0.222979      0.0554942    −4.018      6.45e-05  ***
  Y_3        −0.0584309     0.0304994    −1.916      0.0558    *
  Xp          0.0262337     0.0211312     1.241      0.2148
  Xp_1        0.193452      0.0316952     6.104      1.65e-09  ***
  Xp_2       −0.139665      0.0324762    −4.301      1.92e-05  ***
  Xp_3       −0.0532005     0.0212878    −2.499      0.0127    **
  Xn          0.0739895     0.0190165     3.891      0.0001    ***
  Xn_1        0.0724256     0.0295230     2.453      0.0144    **
  Xn_2       −0.000261883   0.0295033    −0.008876   0.9929
  Xn_3       −0.119776      0.0207187    −5.781      1.08e-08  ***
```

```
Mean dependent var   6.337008   S.D. dependent var   0.180668
Sum squared resid    0.157575   S.E. of regression   0.014361
R-squared            0.993771   Adjusted R-squared   0.993681
F(11, 764)           11080.63   P-value(F)           0.000000
Log-likelihood       2197.683   Akaike criterion    −4371.365
Schwarz criterion   −4315.515   Hannan-Quinn        −4349.880
rho                 −0.005117   Durbin's h          −0.813844

Excluding the constant, p-value was highest for variable 46 (Xn_2)
```

From which the above coefficients can be developed:

```
The error-correction speed is Rho = -0.0432

The long-run coefficient of the LRconstant is LRCon = 5.9457

The long-run coefficient of the XP is LRbeta = 0.6206

The long-run coefficient of the XN is LRbeta = 0.6103
```

Which means that the long run relationship is stated as

$$p = 5.9457 + 0.6206\, wti^{+} + 0.6103\, wti^{-}$$

A subsequent Breusch-Godfrey test can’t reject the null, so no serial autocorrelation is found.

The Wald test/Bound test for cointegration results in the following outcome for constant and positive and negative LR coefficients.

```
Long Run Intercept/Constant
------------------------------------------------------------
5.9457496

Wald test of a (non)linear restriction:

b[1]/b[2] = 0

Chi(1) = 12129, with p-value = 0
```

```
Long Run X_Positive Coefficient
------------------------------------------------------------
0.62057749

Wald test of a (non)linear restriction:

b[3]/b[2] = 0

Chi(1) = 114.31, with p-value = 1.11448e-26

Long Run X_Negative Coefficient
------------------------------------------------------------
0.61033662

Wald test of a (non)linear restriction:

b[4]/b[2] = 0

Chi(1) = 112.6, with p-value = 2.64028e-26
```

All 3 strongly rejecting the null of the variable being zero, meaning that cointegration does exist between the variables.

Finally a Wald test for Long Run asymmetry and Short Run asymmetry is performed.

```
Wald  test for  Long-Run Asymmetry
------------------------------------------------------------

Wald test of a (non)linear restriction:

b[3]/b[2] - b[4]/b[2] = 0

Chi(1) = 10.494, with p-value = 0.0011976


Wald  test for  Short-Run-Asymmetry
------------------------------------------------------------
Restriction:
 b[d_Xp] + b[d_Xp_1] + b[d_Xp_2] - b[d_Xn] - b[d_Xn_1] - b[d_Xn_2] = 0

Test statistic: F(1, 764) = 0.879295, with p-value = 0.348691
```

The null of no long run asymmetry is rejected and the short run one is not rejected, meaning that signs of asymmetry can be found in long run but aren't present in short run.

Diesel

NARDL model for diesel follows the same process. After an ADF test for stationarity and a VAR model to choose best lag, the OLS result is the following:

```
OLS, using observations 2005-02-07:2019-12-09 (T = 775)
Dependent variable: Y

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       0.207272     0.0450344     4.603    4.89e-06  ***
  Y_1         1.20744      0.0360805    33.47     1.29e-151 ***
  Y_2        −0.323054     0.0561039    −5.758    1.23e-08  ***
  Y_3         0.153333     0.0548004     2.798    0.0053    ***
  Y_4        −0.0729069    0.0295180    −2.470    0.0137    **
  Xp          0.0458830    0.0181119     2.533    0.0115    **
  Xp_1        0.175808     0.0271344     6.479    1.66e-10  ***
  Xp_2       −0.118168     0.0278227    −4.247    2.43e-05  ***
  Xp_3       −0.0323811    0.0280490    −1.154    0.2487
  Xp_4       −0.0425421    0.0183638    −2.317    0.0208    **
  Xn          0.0642256    0.0164268     3.910    0.0001    ***
  Xn_1        0.0665122    0.0254276     2.616    0.0091    ***
  Xn_2        0.0267403    0.0252843     1.058    0.2906
  Xn_3       −0.108754     0.0253140    −4.296    1.96e-05  ***
  Xn_4       −0.0202700    0.0185806    −1.091    0.2757

Mean dependent var   6.388132   S.D. dependent var   0.189847
Sum squared resid    0.113846   S.E. of regression   0.012239
R-squared            0.995919   Adjusted R-squared   0.995844
F(14, 760)           13247.69   P-value(F)           0.000000
Log-likelihood       2320.310   Akaike criterion    −4610.620
Schwarz criterion   −4540.827   Hannan-Quinn        −4583.769
rho                  0.010058   Durbin-Watson        1.979531

Excluding the constant, p-value was highest for variable 46 (Xn_2)
```

and the ensuing coefficients are:

```
The error-correction speed is Rho = -0.0352

The long-run coefficient of the LRconstant is LRCon = 5.8904

The long-run coefficient of the XP is LRbeta = 0.8128

The long-run coefficient of the XN is LRbeta = 0.8086
```

Long run relationship for diesel is $d = 5.8904 + 0.8128\, wti^{+} + 0.8086\, wti^{-}$

With a chosen lag of 4 the Breusch-Godfrey test can't reject the null of no serial autocorrelation and the Bound test confirms that the series cointegrate. Tests for asymmetry are reported below:

```
------------------------------------------------------------
11.1.Wald  test for  Long-Run Asymmetry
------------------------------------------------------------

Wald test of a (non)linear restriction:

b[3]/b[2] - b[4]/b[2] = 0

Chi(1) = 1.60374, with p-value = 0.205374

------------------------------------------------------------
11.2. Wald  test for  Short-Run-Asymmetry
------------------------------------------------------------
Restriction:
 b[d_Xp] + b[d_Xp_1] + b[d_Xp_2] + b[d_Xp_3] - b[d_Xn] - b[d_Xn_1] -
b[d_Xn_2] - b[d_Xn_3] = 0

Test statistic: F(1, 760) = 0.892437, with p-value = 0.345118
```

In contrast with the results for petrol, no signs of asymmetry are found in either long run and short run.

Results from the NARDL model provide different conclusions about the R&F hypothesis, finding long run asymmetry in petrol price response. Short run asymmetry is rejected in both series and diesel does not show sign of asymmetry at all. This results are similar to the Engle and Granger ECM methodology using model 2 specification, where long run asymmetry couldn’t be rejected for the two series at 95% confidence level. However, 99% confidence level would have rejected it for diesel.

5. Conclusions

Rockets and Feathers hypothesis is sensitive to model specification. The estimated models show strong differences in outcome, because price asymmetry is explored using different tools and shapes.

Stating if the response of petroleum products prices to crude oil prices variation is symmetrical in Italy is not simple. However, some conclusions can be drawn; if cointegration is statistically confirmed and a long run equilibrium exist, departures from this equilibrium have different impact when they are positive or negative.

In model 2 in the Engle and Granger approach, this departures from equilibrium are asymmetrical for both fuels, meaning that downstream prices will tend to go back to equilibrium at different pace. Specifically, when the long run relationship is below equilibrium (resulting from an upstream price increase) the downstream variable tend to go back to equilibrium faster, a typical R&F behavior.

Similar results appeared from the NARDL model, which investigated short run and long run price asymmetries given a long run relationship split in upward and downward crude oil price changes. While finding no evidence of asymmetry for short run, a strong one emerged for petrol prices in the long run. Even here asymmetry is consistent with the R&F hypothesis.

Model 1, using both Engle and Granger methodology and Johansen VECM couldn't find presence of asymmetry. As previously said, disequilibrium in this case is specified as the upstream price variation in one period, implying an immediate departure from equilibrium following a crude oil price change.

Nonetheless, having used a single stage of price transmission instead of distinguishing between wholesaler and retailer, it is not possible to assess exactly where in the price transmission process the asymmetric price behavior takes place.

In conclusion, signs of R&F hypothesis appear mostly for petrol price when it diverge from the long run equilibrium, suggesting a possible market imperfection.

6.Bibliography and Sitography

1. Bacon, Robert W., “Rockets and Feathers: The Asymmetric Speed of Adjustment of U.K. Retail Gasoline Prices to Cost Changes,” Energy Economics 13 (July 1991).
2. J. D. Karrenbrock, “The Behavior of Retail Gasoline Prices: Symmetric or Not?”, Federal reserve of Saint Louis review, (1991).
3. D. N. Manning, “Petrol prices, oil price rises and oil price falls: some evidence for the UK since 1972”, Applied Economics, (1991).
4. Gebhard Kirgaessner, Knut Kuebler, “Symmetric or asymmetric price adjustments in the oil market”, Energy Economics, (1992).
5. Shin David “Do product prices respond symmetrically to changes in crude prices?”, OPEC review, (1994).
6. Borenstein, Severin, Colin A. Cameron, and Richard Gilbert, “Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?” Quarterly Journal of Economics 112 (February 1997).
7. Balke et al. “Crude oil and gasoline prices, an asymmetric relationship?”, Economic review – Federal Reserve of Dallas, (1998).
8. B. Reilly, R. Witt,“Petrol price asymmetries revisited”, Energy Economics, (1998).
9. L. J. Bachmeier, J. M. Griffin “New evidence on asymmetric gasoline price responses”, Review of Economics and Statistics, (2003).
10. Galeotti et al. “Rockets and feathers revisited: an internationalcomparison on European gasoline markets”, Energy Economics, (2003).
11. S. Radchenko “Oil price volatility and the asymmetric response of gasoline prices to oil price increases and decreases”, Energy Economics, (2005).
12. M. Grasso, M. Manera “Asymmetric error correction models for the oil–gasoline price relationship”, Energy policy, (2007).

13. A. Meyler “The pass through of oil prices into euro area consumer liquid fuel prices in an environment of high and volatile oil prices”, Energy Economics, (2009).
14. A. Honarvar “Asymmetry in retail gasoline and crude oil price movements in the United States: An application of hidden cointegration technique”, Energy Economics, (2009).
15. A. Athil, A. Lahiani, D. K. “Nguyen Asymmetric and nonlinear pass-through of crude oil prices to gasoline and natural gas prices”, Energy Policy, (2013).
16. L. Kristoufek, P. Lunackova “Rockets and feathers meet Joseph: Reinvestigating the oil–gasoline asymmetry on the international markets”, Energy Economics (2015).
17. J. Rodrigues “Asymmetries in the adjustment of motor diesel and gasoline pump prices in Europe”, Concorrencia.pt, (2009).
18. G. Fiorito “La dinamica dei prezzi dei carburanti rispetto alle quotazioni del petrolio, una analisi a livello europeo”, conference paper, (2010).
19. A. A. Romano, G. Scandurra “Price asymmetries and volatility in the Italian gasoline market”, OPEC Energy review, (2012).
20. F. Venditti “From oil to consumer energy prices: How much asymmetry along the way?”, Energy Economics, (2013).
21. A. Bagnai, C. A. Montegau Ospina ““Asymmetric asymmetries” in Eurozone markets gasoline pricing”, The Journal of economic asymmetries, (2016).
22. D. Mead, P. Stieger, “The 2014 plunge in import petroleum prices: What happened?”, U.S. Bureau of labor statistics (2015), 4.
23. Engle, Robert F., and Clive W. J. Granger, “Co-integration and Error Correction: Representation, Estimation, and Testing,” Econometrica 55 (March 1987), 251–276.
24. Johansen Søren, "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", Econometrica, (1991).
25. Gretl User’s Guide (2019)

26. Peasaran et al., “Bounds testing approaches to the analysis of level relationships”, Journal of applied econometrics, (2001).

Sitography

1. https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RWTC&f=W
2. https://en.wikipedia.org/wiki/Brent_Crude
3. https://crude oil.europa.eu/energy/en/data-analysis/weekly-oil-bulletin
4. https://fred.stlouisfed.org/series/DEXUSEU/
5. https://en.wikipedia.org/wiki/Unit_root