MASTER PROJECT

Competition and Price Asymmetries in the Retail Gasoline Market: an Application to a Spanish Cartel Case

Master's Degree in Specialized Economic Analysis: Competition and Market Regulation Program

Authors:

Silvia Brumana (174059)
Roger Medina (102449)

Supervised by Miguel Espinosa

July 2020

Competition and Price Asymmetries in the Retail Gasoline Market: an Application to a Spanish Cartel Case

Silvia Brumana   Roger Medina

Master Project*
Barcelona Graduate School of Economics (BGSE)

July 2020

Abstract

Asymmetric pricing in the retail fuel market describes the situation in which retail prices are adjusted quickly in case of positive shocks in the wholesale costs, whereas the adjustment is much slower in case of negative ones. This asymmetry in the reaction of retail operators is captured by the well-known expression *rockets and feathers*. By relying on a reduced version of a database used in Moral and Gonzalez (2019) containing the series of diesel and Brent prices for the period 08/2014-06/2015, we address the following research question: how did the fuel prices of the retail gas stations in a subset of Spanish provinces react in this period to changes in the price of Brent? In order to answer this question, we construct a pricing equation that takes into account all the possible factors that can affect pricing with the aim of isolating the effect of negative shocks and the effect of positive shocks in the price of Brent on the price of gasoline of the various Spanish operators. Then, we check whether the difference between these two effects is statistically significant to see whether there are asymmetric reactions. Furthermore, since in the period covered by the database a cartel was active, we also investigate the relation between asymmetric pricing and competition.

Keywords: Asymmetric pricing, Competition, Cartel, Time Series.
JEL Classification: C22, C23, L11, L13

*We gratefully thank our supervisor, Miguel Espinosa, for his guidance and for his valuable comments that have definitely contributed to improve the quality of our master project. We also want to thank Jordi Perdiguero for giving us guidance about the most common models used in the asymmetric pricing literature and for providing useful references, and to Xulia González for solving doubts about her database. Finally, we also want to thank Javier Biscarri for clarifying important aspects about the econometrics and modelling of price asymmetries.

Contents

1 Introduction **1**

2 The Spanish cartel in the retail fuel market and asymmetric pricing **2**

2.1 The Spanish cartel 2

2.1.1 The decision of the CNMC 2

2.1.2 Relevant market definition, product market and geographical market 2

2.2 Asymmetric pricing 4

2.2.1 The econometrics behind asymmetric pricing analyses 4

2.2.2 Asymmetric pricing and competition/market power 6

3 Detecting pricing asymmetries in the Spanish retail fuel market **7**

3.1 Theoretical approach 7

3.2 Empirical analysis at the daily level 8

3.2.1 Data 8

3.2.2 Preliminary analysis at the daily level 9

3.2.3 The model with daily data 12

3.2.4 Analyses at the daily level and results 13

4 Conclusions and comments **15**

References **16**

Without data, you're just another person with an opinion

— W. Edwards Deming

What goes up must come down.

— Isaac Newton

1 Introduction

Asymmetric pricing is a phenomenon that occurs when retail prices react differently to shocks in costs. Consequently, output prices tend to respond faster to input increases than to decreases. This fact does not characterize exclusively the fuel market; however, because of the high volatility of oil prices it is more pronounced in this market than in others. In an important analysis done by Peltzman (2000) [17], the author found that the immediate response to a positive shock is at least twice the response to a negative shock, and the difference lasts between five and eight months. This effect is present in different markets and for different products.

However, the available evidence on price asymmetries is inconclusive. The divergent results in the literature are usually attributed to variations in the type of data used (frequency, periods, etc.), the specification of the econometric model, the estimation methodologies and the different parameters included in the models. Finally, country-specific particularities of each market are an important factor that could explain this divergence (Perdiguero, 2013). [18]

The well-functioning of the fuel market is of crucial importance in a modern economy, especially in the transportation sector, which plays a major role in the economy in terms of employment (in Spain, it directly employs more than 1 million people, which accounts for more than 5% of total employment) and economic growth (5% of GDP).[1] Despite the ambitious plans of European institutions to reduce the use of fuel in the medium and long term, it remains the major energy source in the transport sector. Moreover, almost 13% of the budget of a representative household is spend on transport goods and services.[2] All these reasons point out the importance of ensuring a competitive environment in this market in order to achieve efficiency and high social welfare levels.

Likewise, we chose to explore the topic of asymmetric pricing because we were intrigued by the fact that the common *rocket and feathers* phenomenon could be considered as a collusive device, as we explain in section 2.2.2. This represented a link between the asymmetric pricing literature and the one about competition policy.

Our paper tries to establish if a resolution by a National Competition Authority (NCA) can modify price setting in the oil market. Despite the important and exhaustive evidence of price asymmetries for the Spanish case, there is no paper that we are aware of that tries to establish if there is a link between pricing asymmetries and the conclusion of an investigation that ends by imposing a fine to the parties. Hence, we are interested in understanding better the peculiarities and how competition takes place in this market, and if collusion can be an explanatory factor of price asymmetries using as a *natural experiment* a fine imposed to members of a cartel.

This paper is organized as follows. In Section 2 we briefly describe the historical evolution of the Spanish retail gasoline market. We also explain the decision of the CNMC about a Spanish cartel and assess relevant characteristics of the market structure and its geographical scope. Additionally, we introduce some important elements behind the estimation and analysis of asymmetric pricing.

[1] Instituto Nacional de Estadística (INE). Encuesta de Población Activa (https://www.ine.es/jaxiT3/Tabla.htm?t=4128&L=0) y Contabilidad Nacional https://www.ine.es/jaxiT3/Tabla.htm?t=30678&L=0

[2] INE. Encuesta de Presupuestos Familiares https://www.ine.es/jaxiT3/Tabla.htm?t=24900

Section 3 gives details about the theoretical approaches to detect price asymmetries in the Spanish retail fuel market, describes the data and specifies the model that we estimate. Section 4 concludes and summarizes the paper. We also include an annex with all the tables and graphs that summarize our analysis and an appendix with some important concepts to understand times series.

2 The Spanish cartel in the retail fuel market and asymmetric pricing

2.1 The Spanish cartel

2.1.1 The decision of the CNMC

In 2015, the *Comisión Nacional de los Mercados y la Competencia* (CNMC), the Spanish competition authority, imposed a fine of more than 30 million euros to the five most important operators of the oil sector (Repsol, CEPSA, Shell, Galp and Meroil) for creating a cartel on retail fuel prices (more precisely, for reaching price agreements in gas stations and for exchanging sensitive information about fuel distribution market)[3]. The Spanish competition authority found that these firms infringed the Article 1 of the Spanish Competition Act (*Ley de Defensa de la Competencia*[3]) and Article 101 of the Treaty of Functioning of the European Union (TFEU)[4]. The investigation started on May 2013 and concluded in July 2014. The resolution with the amount of the fines that each firm had to face was made public on 20 February of 2015.

The lack of competition in this market has been persistent and is well documented by different reports of the CNC (the former CNMC) and the CNMC. In particular, the CNMC has found that retail prices are much more higher than prices in other countries. Due to historical reasons, Spain has had a long tradition of low levels of competition in the retail gasoline market. Between 1927 and 1992, there was a single state-monopoly undertaking that operated under a private concession: *Compañía Arrendataria del Monopolio del Petróleo S.A.* (CAMPSA). Under this state-granted monopoly, CAMPSA was the only operator that was allowed to export and import oil products and to provide service in the retail market (gas stations). It was not until the end of 1970s that Spain started to change its market structure and introduced small liberalizations, mostly driven by the entry to the European Union (EU) and the obligation to comply with EU competition rules. After the entry to the EU, CAMPSA started to loose its monopoly position due to foreign competition and the creation of new national firms (Contín-Pilart *et al*, 2009). [7].

2.1.2 Relevant market definition, product market and geographical market

The relevant market definition is a key element in the economic analysis of competition of a specific market. Competition authorities can use a different variety of tools in order to assess the relevant market in each particular case. There are different factors and parameters that can affect this definition. Typically, the most important element that defines a market is the **degree of substitutability**. Thus, the relevant market will be conformed by *the set of products that exercise some competitive constraint on each other* (Motta, 2004)[15].

[3] Ley 15/2007, de 3 de julio, de Defensa de la Competencia https://boe.es/buscar/act.php?id=BOE-A-2007-12946

[4] Consolidated version of the Treaty on the Functioning of the European Union https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012E%2FTXT

Operator	Market share
Repsol	40% - 45%
CEPSA	15% - 20%
BP	10% - 15%
Galp	5% - 10%
DISA	5% - 10%
Saras Energía	0% - 5%
Meroil	0% - 5%

Source: CNC (former CNMC)

Table 1: Market share in the wholesale segment of distribution to gas stations

Gasoline			Diesel oil		
Operator	**# stations**	**Volume**	**Operator**	**# stations**	**Volume**
Repsol	39%	45%	Repsol	38%	45%
CEPSA	17%	16%	CEPSA	16%	16%
BP	7%	12%	BP	7%	9%
Galp	7%	5%	Galp	7%	6%
DISA	5%	6%	DISA	5%	5%
Others	25%	16%	Others	27%	21%

Table 2: Market share of retail market for gasoline and diesel oil

In Table 1 we can see the market shares of each operator in the wholesale segment of distribution. Repsol is the most important undertaking in this market, followed by the former state-monopoly CEPSA. According to the CNC (the former Spanish Competition Authority), in Spain there are around 10,000 gas stations. Of those, 7,900 are integrated in oil operator networks, approximately 1,800 are owned by independent retailers and 300 are located in malls (CNMC, 2015). Table 2 shows disaggregated definitions of market shares.

The distribution network of a wholesale operator encompasses all supply facilities that it owns, as well as those facilities in which the wholesale operator has exclusive supply contracts with its owner. The contractual arrangements that typically exist in this market can be divided in wholesaler operators and retail distributors in the following basis:

1. COCO (Company Owned - Company Operated): Facilities that are owned and operated by the wholesale operator.
2. CODO (Company Owned - Dealer Operated): Facilities in which the wholesale operator keeps the ownership of the point of sale but gives the management to a third party with exclusivity of supply provision.
3. DOCO (Dealer Owned - Company Operated): Facilities owned by a dealer that gives the management of the point of sale to a wholesale operator.
4. DODO (Dealer Owned - Dealer Operated): Facilities owned by a dealer that are linked to the wholesale operator through a exclusivity contract of supply provision.

Differentiating between these contractual arrangements is useful to understand the intensity of competition. For instance, situations with COCO or DOCO imply a maximum degree of control by the wholesale operator and stability in supply provision (it is unfeasible that the gas station is provided by a wholesale competitor). In the CODO case it is not likely either that the gas station is provided by a wholesale competitor, but the dealer has a higher degree of independence when setting prices. Finally, in DODO situations the dealer has also a higher degree of independence when setting prices and the link to the wholesaler is weaker, because once the contract become due, it can switch to another operator. As reported by the previous National Energy Commission (CNE, integrated in the the CNMC in 2013), 83% of gas stations are integrated in wholesale networks through some of the aforementioned contractual arrangements (exclusivity in supply provision) . Of those, 20% are COCO arrangements, 38% are CODO, 8% are DOCO and the remaining 34% are DODO.

According to the CNMC, the retail market for fuel distribution is characterized by an important local component. The demand is located in the vicinity of its center of activity. As a result, the substitutability between stations is geographically constrained.

However, it is also considered that many relevant parameters of the supply, such as the variety of products offered, the quality of the products, the level of the service (opening times, etc.), indirect taxes, are decided from a national perspective. Moreover, the main operators act at a national level following a national commercial strategy.

For all these reasons, national competition authority chose to consider a national definition in the geographical dimension of the market for the retail sale of fuels, admitting regional specificities and characteristics. This was previously recognised by the CNMC in 2009 when in a report analyzing the competition in the retail sector for this market concluded that *despite recognising the local component in competition between gas stations, it is considered that the geographical market to analyze has a national dimension.*[5]

Nonetheless, the CNMC didn't perform a Small but Significant Non-Transitory Increase in Price (SSNIP) test nor used alternative tools to do a good assessment of the relevant market. There is plenty of evidence that persistently shows that competition in this particular market has a strong local component, even if some of the commercial strategies are set at a national level. In particular, we know from the application of geographic information systems to the Spanish case that *the relevant geographic market is delineated by a 5 to 6 minutes travel time isochrone around each station* (Perdiguero & Ramon, 2018).[19] Hence, this local component should have had a larger weight in the decision of the NCA, since only close rivals seem to effectively compete with each other. In Annex 5 we have included maps with the spatial distribution of gas stations in all the provinces for which we did the analysis.

2.2 Asymmetric pricing

2.2.1 The econometrics behind asymmetric pricing analyses

In order to detect the presence of asymmetric pricing in the retail market of fuel, we first need to define the concept of asymmetry, and to do so, we need to understand what the variables affecting the retail price of fuel are.

The main explanatory variables for the retail fuel price (both gasoline and diesel fuel) are taxes, transportation costs and the price of oil on international markets. Among these, the one characterized by a higher degree of volatility – and thus by a higher frequency of shocks – is the oil price. In the literature, the type of oil price adopted in the analysis of asymmetric pricing is the spot price, that is, the price of oil on a given day. In Europe, the main price used for the oil price is the Brent.[6] If the series of this wholesale price is subject to variations, we expect that such variations

[5] Informe sobre la competencia en el sector de carburantes de automocíon. *Comisión Nacional de la Competencia* (CNC). Julio 2009. https://docplayer.es/11237819-Informe-sobre-la-competencia-en-el-sector-de-carburantes-de-automocion.html

[6] Some studies adopt the price of unleaded gasoline or the price of refined oil as a measure for the wholesale cost.

are mirrored in some way in the retail price of fuel. The relation existing between the time series of the spot price and the one of retail fuel prices is quite stable, that is, the two series generally tend to move together and not to drift apart in the long run; in other words, the long-term movements are somewhat constrained by the presence of the cointegration relation. When the spot price is subject to a shock, then the retail price of fuel reacts. This reaction may be quite immediate, although in the literature it was typically found that the retail fuel price reacts with some lags. Once the shock is over and once the spot price goes back to the pre-shock levels or reaches a new level of equilibrium, then the retail fuel price will adjust accordingly.

Hence, in the above described mechanism we can decompose the adjustment into two parts: the first is the short-run adjustment of the retail fuel price that occurs after the shock in the series of the spot price; the second is a long-term adjustment of the retail fuel price and can be conceptualized as a reversion of the fuel price to the relation with the Brent price. As stated in Engle and Granger (1987): *"long-run components of variables...obey equilibrium constraints while short-run components have a flexible dynamic specification"*.[6] By *asymmetric pricing* we refer to an asymmetry detected in the first component of the adjustment: in fact, it was observed that the retail fuel price reacts and adjusts faster to positive shocks in the spot price (that is, to a sudden increase in the latter) than to negative shocks; hence, the retail price does not react symmetrically to shocks, but its reaction depends on the sign of the shock. This asymmetry is well described by the expression *rockets and feathers*: in fact, the retail fuel price increases quite immediately after positive shocks, whereas it is stickier in case of negative ones.

In the literature we can find different types of asymmetries and different econometric models were built to represent the process generating these asymmetries. Suppose that x is the Brent spot price and y is the retail price of fuel; then, we can describe the main asymmetries as follows (Perdiguero, 2013):[18]

- Contemporaneous impact asymmetry: the contemporaneous effect of a positive change in x on y may be not symmetric to the effect of a negative change in x.
- Distributed lag effect asymmetry: the impact on y of a positive change in x may be not symmetric to the impact of a negative change depending on the lag we are considering.
- Cumulated impact asymmetry: the cumulated effect on y of positive changes in x may be not symmetric to the cumulated effect of negative changes in x.
- Reaction time asymmetry: the number of periods needed to y to react to a change in x may be different depending on the sign of the shock in x.
- Equilibrium adjustment path asymmetry: suppose we assume that there exists a target value for y, say y^*, to which y tends to. However, the actual value of y may be above or below the target; the speed of convergence of y to y^* depends on the value that y had at (t-1). The speed of convergence to y^* may be asymmetric depending on whether y is above or below y^*.[14]

The main methodologies adopted in the literature to analyse asymmetric pricing are:

- Error Correction Model (ECM): the idea behind the ECM is that there is an equilibrium relationship between a set of variables of interest. This equilibrium relationship could be thought of as "the long-run relationship between the variables" or "the relationship that would come out of an equilibrium model". In technical terms, when two variables are in a long run relation, they are said to be *cointegrated*. When the variables are not in equilibrium, they will tend to move towards that equilibrium; hence, the ECM models the changes in each

See Contin et al.(2008), Balaguer and Ripollés (2012) [1], Bettendorf et al. (2009) [13]

of the variables as being a function of past changes and as a function of the disequilibrium. If the current level of the variables is far from the equilibrium relationship, they will have a stronger "urge" to move towards equilibrium. By taking the example of fuel and Brent prices, the *error* that is corrected is thus the deviation of fuel prices from the long-run relation with the Brent prices.

- Auto-regressive Distributed Lag Model (ADL): in an ADL the variable y is modelled as a function of both its lagged values as well as the current and lagged values of a set of explanatory variables. In this case, the effect of the lagged variables are distributed over time rather than occurring all at one. It can be considered a special case of an ECM. For the estimates of the ADL to be consistent, the series must be stationary; otherwise, the model could treat the sources of non-stationarity as sources of asymmetry.

- Partial Adjustment Model (PAM): the PAM is suitable to model the equilibrium adjustment path asymmetry described above. As the ADL, it can be applied only in case of stationarity of the series. Since the ECM can be considered to be a model of partial adjustment, we can say that the PAM and ADL are incorporated in the ECM, and this explains why the ECM is also the most commonly used in the literature on asymmetric pricing.

- Regime Switching Model (in the deterministic or stochastic version) (RSM): this type of model lies on the idea that the relation between x and y is affected by a third variable, say z, and the different values of z give rise to different states (or regimes) of the world. For instance, the behavior of diesel prices responds differently to Brent prices depending on "some state of the economy". The deterministic version is applied if it is reasonable to assume that we know the state we are in at each moment in time; the stochastic one is suggested if the shift from one state to the other is random.

2.2.2 Asymmetric pricing and competition/market power

An important question for National Competition Authorities (NCA) is how to distinguish between legitimate and illegitimate conducts when they lack evidence of an explicit agreement that would be punished under the terms of article 101 of the Treaty of Functioning of the European Union (TFEU). Understanding the conditions and the context under which these price asymmetries can arise and their effects (the so called *rockets and feathers*) can help NCA to better assess collusion cases (OECD, 2013) [16].

The link between asymmetric pricing and competition is not clear. From a theoretical perspective, it might be feasible to think that a delay in price adjustments as a reaction of changes in wholesale prices could be attributed to some market inefficiency or distortion. In this sense, Hong and Lee (2020)[10] investigate how and why this link takes place. By using station-level panel data from the Korean retail gasoline market, they analyze how geographical separation can be used as a reliable source of market power. The fact that some gas stations are located in islands around the Korean peninsula is used as a natural mechanism of market power due to absence of competitive pressure. The main findings of their research are that there is a positive correlation between market power and price-response asymmetry. Those stations located in isolated islands do not adjust their prices as a response in cost shocks at the same speed as stations that face more competition (the ones that are located in the mainland). Moreover, the authors also find that, apart from adjusting their prices with more delay, those stations located in islands also keep their prices constant during some periods. The first fact is explained by the typical effect of *consumer search models*, that is, the fact that consumers have imperfect information about production costs and they form their expectations based on past cost realizations that are persistent across time. When costs are high, consumers expect them to remain at the same level in the next period, and this reduces stations' incentives to lower prices. On the contrary, with low costs realizations, consumers tend to intensify their search and as a result stations respond faster to cost shocks.

The authors hypothesize if there exists some degree of tacit collusion in this market, that is, if stations tacitly coordinate in order to maintain their prices artificially high. The idea is that past prices are used as a *focal point* at which gas stations coordinate (Borenstein *et al*, 1997) [24]. This means that sustainability of collusion depends on margin changes as a result of fluctuations in costs (wholesale prices). A positive cost shock is associated with a deviation from the collusive conduct due to the fact that keeping prices constant would reduce their margins. On the contrary, when there is a negative shock in costs then the collusive conduct is easier to sustain because keeping prices constant (past prices) increases the margins of all the undertakings. This allows the authors to construct a relation between sticky prices and market power. Other authors use the tacit collusion hypothesis in order to analyze the restructuration and liberalization of national markets and its main implications in price strategies (Perdiguero-García, 2010) [8]. In Eckert (2002) [5], the author found a particular asymmetric pricing device, and he showed how imperfect competition between different undertakings is the main explanation to this particular pattern.

Loy et. al (2018) [12] showed for the Australian case that the speed of transmission between the wholesale and retail prices is higher in more competitive environments, suggesting an inverse relation between sticky prices and delay in the adjustments through the degree of competition in a market. It is also important to find a good definitions of market power and the relevant market.

Another possible explanation of price asymmetries related to market power is that they could be the result of a punishment mechanism that typically exist in cartels. An undertaking might think that adjusting the price downwards could prompt other suppliers to reduce it more aggressively, engaging in a price war that could change the structure of market shares. However, it is difficult to observe this effect because comparing degrees of market power in different markets with strong local component is also difficult, and it is important to have disaggregated data to do a good assessment (if not, we may omit local characteristics). Apart from competition, there are other possible explanations to the price asymmetries phenomenon, but that are beyond the scope of this document. One alternative explanation are menu costs (fixed costs incurred when an undertaking changes its price, for instance, reprinting price tags) (Shuttleworth, 2001) [25]

3 Detecting pricing asymmetries in the Spanish retail fuel market

3.1 Theoretical approach

The exercise of detecting asymmetric pricing involves dealing with time series, and the properties of a series constrain our choice of a model. One of such crucial properties is stationarity. When a time series is generated by a stochastic process that is stationary, it means that the series has constant mean, variance and covariance, and we expect that after a shock the series reverts to its mean. Hence, if a variable is stationary, it means that a shock has a temporary effect on it.

The two models that are mainly used to detect asymmetric pricing are the Error Correction model (ECM) and a special case of it, that is, the Auto-regressive Distributed Lag (ADL) model for stationary series. For the ECM to be applied, two requirements must be met:

1. The series of the variables of interest must be integrated of order one, that is, they can be non-stationary (or contain a unit root), but once the series are first-differenced, they must become stationary.[7]

[7]However, the paper of Engle and Granger (1987)[6] indicates that the framework applied to two variables integrated of the same order can be generalized to the case of *fractional cointegration*, that is, the situation in which the two variables have different orders of integration.

2. The two series must be cointegrated, that is, they must have a long-run equilibrium relation. In technical terms, a linear combination of the two variables is a stationary process and what makes their combination stationary is the *cointegrating vector*.

The presence of characteristics that make the series non-stationary, as for instance seasonality or structural breaks, may bias our estimates and lead us to the wrong conclusion that there is an asymmetry when actually such asymmetry reflects only the presence of features that make the series non-stationary. An example of this risk was described in a paper by Von Cramen-Taubadel and Meyer (2000)[4]: they generated through a Monte Carlo experiment two cointegrated time series, characterized by a structural break and a symmetric adjustment process. Then, they applied different models to the series in order to check whether each model could identify the presence of symmetry even if two potentially confounding factors were there. The experiment showed that many models frequently used to detect asymmetric pricing failed to recognize the presence of a symmetric process and reported the presence of (spurious) asymmetries. This implies that, in case we wish to use one of the two mentioned models, we need to carefully understand the behaviour of the series.

Another aspect that needs to be taken into account when designing an analysis for asymmetric pricing is the direction of causality. In fact, the aim of our project is to check the presence of asymmetric responses in the retail price of diesel in various Spanish provinces to shocks in the wholesale price of oil depending on the sign of the shock. This research question lies on the assumption that the direction of causality goes from the wholesale price of Brent to the retail price of fuel in Spain.

Furthermore, one issue that we may want to verify is that the panels are not cross-sectionally dependent. In other words, we must check whether the pricing of the different gas stations is independent. This assumption may well be violated because of the vicinity of the gas stations, resulting in the so-called "spatial effects" giving rise to correlation in pricing. Failing to take into consideration the potential existence of cross-sectional dependence may lead to inconsistent estimates.

Finally, the literature on asymmetric pricing stresses the importance of using data that are as disaggregated as possible on all dimensions in order to capture the heterogeneity in the responses to shocks. In fact, since asymmetries may take place in a very short time span, like days, even a low level of aggregation on time, as for instance weekly aggregation, can lead to inconsistencies. The same reasoning holds for the individual or geographic dimensions, because of the idiosyncratic component of responses to shocks that may be dependent on the geographical market considered.

3.2 Empirical analysis at the daily level

3.2.1 Data

Our data come from a paper published in the International Journal of Industrial Organization (González and Moral, 2019).[9] The researchers built a detailed database containing the daily series of retail diesel prices (pre-tax, pre-VAT and posted) for 8,080 fuel stations in 47 Spanish provinces (all except Balearic and Canary Islands) and the daily series of Brent prices covering the period 08/2014-06/2015 (that is, 302 days).[8] In addition to these variables, the database has also information about the brands and the location of each gas station, the brands of the three closest

[8]The use of diesel prices rather than gasoline ones is justified by the fact that diesel prices represents the majority of the fossil automotive fuel used in Spain (in González and Moral, (2019), this accounts for 85% of fuel consumption). As for the choice of Brent prices, the Brent is used as a price benchmark for crude oil, that is, the state in which oil is delivered to refineries. Since no intermediate prices between refineries and gas stations are available, we consider the Brent price series to be the closest available approximation to the wholesale cost of a gas station.

competitors (a competitor is considered to be close if it is located in a radius of 2 km from the gas station of interest), the distance from and the brand of the closest refinery. In the Annex, we reported descriptive statistics, like the number of gas stations per brand and per province (Table 4).

For the purpose of our project, we modified the database as follows: given the importance of having a balanced panel to be able to detect asymmetric pricing, we reduced the number of missing values as much as possible.[9] We adopted a conservative approach by choosing to keep only those gas stations (i) for which no brand mismatch was present in the original dataset and (ii) having at most 7 missing observations our of 302. In addition, since we ran our analysis for each province separately, we decreased the number of provinces by keeping only those having at least 200 gas stations. These changes led us to a final database consisting of 966,702 observations, corresponding to 3,201 gas stations observed for 302 days in 9 Spanish provinces.[10]

As for the remaining missing values in the dataset, we adopted the following solutions:

- For the series of the Brent price, we copied the price of the same day as reported for other gas stations since this price is the same for a given day regardless of the gas station considered;
- For the retail prices, we copied the price of the previous day; if this was not present, we copied the price of the previous day available; otherwise, we copied the price of the following day.

3.2.2 Preliminary analysis at the daily level

In order to choose the right model, we had to understand the properties of the time series of the retail prices and the price of Brent for each of the Spanish provinces considered. Our main objective was to check whether it was more adequate to adopt an Error Correction model (ECM) or a special case of it, that is, the Auto-regressive Distributed Lag model. As already mentioned, the ADL can be used in case of stationary variables, whereas the ECM can be used when the variables are cointegrated.

An alternative model that we could have used is the Regime-Switching model, in which the relation between x and y (in our case, between the retail fuel prices and the Brent prices) is affected by a third variable z. In our case, z could be the presence of the cartel fined by the CNMC.[11]. However, given the complexity of integrating a RSM into an ECM, we preferred to limit our choice to either an ECM or an ADL by carefully considering all the properties of the series.

Among the different measures of price available in our dataset (pre-tax, posted and pre-VAT), we chose to use the pre-tax price because it is not subject to changes in taxes, and thus it should be the one that best approximates the actual wholesale costs, like the Brent price.

Our preliminary analysis started by first verifying whether the series contained a unit root. In the literature, there are different types of tests that can be applied to panel data to check for stationarity and each of them makes different assumptions. Our idea, based on economic theory, is that our panels are heterogeneous, that is, the estimates of the constant and slope parameters vary across the gas stations.[12] Among the tests used to test stationarity in heterogeneous panels, there are some that are robust in the presence of cross-sectional dependence. The latter describes

[9]The reasons behind missing values of prices reported by the gas stations are varied, ranging from holidays and festivities to maintenance and closure days.

[10]The Spanish provinces that are left after we decreased the size of the dataset are the following: Alicante, Barcelona, Coruna, Granada, Madrid, Malaga, Murcia, Sevilla, Valencia.

[11]Papers that combine a regime switching model with an ECM in the analysis of asymmetric pricing are: Ihle and von Cramon-Taubadel (2008) [11] and Rezitis and Pachis (2015).[23]

[12]In section 3.2.4 we explain that we verified the hypothesis of parameter heterogeneity by running the Swamy's test for panel homogeneity.

the situation in which agents located in the same neighbourhood are subject to common shocks. As reported also in Balaguer and Ripollés (2015), this cross-sectional dependence may be detected in the behaviour of neighbouring gas stations. In order to check whether this was the case in our dataset, we thus first had to test for cross-sectional dependence in all the Spanish provinces considered and then choose the right unit root test. The test that we ran is the Pesaran (2015) test for weak cross-sectional dependence.[21] The idea behind this test is that some degree of cross-sectional dependence is very common in economics, and consequently those tests in which the null hypothesis is that the panels are cross-sectionally independent are too strict; consequently, relaxing the null hypothesis by stating that the panels are only weakly dependent means that the correlation between units converges to zero as the number of cross section increases. In all the provinces we got that the pricing behaviour of the gas stations showed this type of dependence (as shown in the results in Annex 2), and thus we needed to choose a unit root test that was robust in this sense. This requirement is satisfied by the Breitung (2005) Panel Unit Root test.[2][13] The results (in Annex 2) show that the retail prices do not present a unit root.[14] [15] The result of stationarity of the diesel prices is somewhat surprising because, as we can see from the graphs in Annex plotting the series of Brent and the series of retail fuel prices over the period considered, the process does not seem to show constant mean and variance. Our suggestion is that the increase of the series of the fuel prices after February 2015 may be interpreted as a mean reversion, and this is something that happens in stationary series. Consequently, we suggest that this result may need to be subject to further investigation. As for the Brent prices, they are the same regardless of the province and of the gas station. Consequently, we considered the Brent prices as a simple time series rather than a panel, and we thus applied the Phillips-Perron (1988) unit root test for time series, which is robust in presence of serial correlation.[22] Since the test showed that the series of Brent prices is non-stationary, we first-differenced the series and we ran the Phillips-Perron again. The null hypothesis of the presence of a unit root is rejected at 99 percent, implying that the series is integrated of order 1.[16]

Second, we checked the presence of a long-run equilibrium relation between the retail fuel prices and the prices of Brent by testing for the presence of cointegration. The test that is mainly used with heterogeneous panels characterized by cross-sectional dependence is the Westerlund (2007) bootstrap cointegration test.[26] The test computes four statistics: two of them are used to test the presence of cointegration for at least one panel in the dataset; the other two statistics test the presence of cointegration for the whole dataset. For the test statistics to be robust in the presence of cross-sectional dependence, a bootstrap method is applied: extracting bootstrap samples for a maximum of 800 iterations allows the estimation of the bias introduced by the presence of cross-sectional dependence in the data, and thus to generate a result of the cointegration test that is corrected accordingly (*robust*).[17] A summary of the results for both the unit root tests and the cointegration tests is provided in the following table:

[13] Given that we kept only those gas stations having at most 7 missing values, this means that for each gas station we copied at most 7 values out of 302. Consequently, we suggest that this modification did not affect the test on stationarity on the series of retail diesel prices.

[14] We carried out the Breitung test also on the series of the retail prices pre-VAT and the one of the posted prices. The results (not shown here, but verifiable from the do-files) is the same regardless of the series we consider.

[15] The Breitung (2005) unit root test robust to the presence of cross-sectional dependence can be carried out if T > N. In those provinces in which T < N, we applied the *demean* option which partially corrects for cross-sectional dependence in a similar way to the Pesaran (2003) unit root test in heterogeneous panels with cross-sectional dependence. In order to check whether our results about unit root were robust, we also divided the provinces in smaller datasets so that T > N and carried the Breitung (2005) unit root test in each of them. The results are robust and are showed in Annex 4.

[16] As we can see from the graphs in Annex 6 plotting the evolution of the retail diesel prices in the subset of Spanish provinces considered and the Brent prices, we can see that the two series mostly move together. We thus suggest that the result of non-stationarity for the Brent price may be due to the fact that at the beginning of 2015 the Brent series increased but started to fluctuate around a mean lower than the one it had at the beginning of our dataset. The lack of reversion to the previous mean is typically an indicator of a non-stationary series and so this may be justify the different result for the stationarity test for the diesel prices and the Brent prices.

[17] The command used in Stata is *xtwest* and was designed by Persyn and Westerlund. See [20].

Province	Presence of unit root (price_pretax)	Presence of unit root (pbrent)	Presence of unit root (diff_pbrent)	Presence of cointegration (Westerlund test)
Alicante	No	Yes	No	Yes
Barcelona	No	Yes	No	Yes
Coruna	No	Yes	No	Yes
Girona	No	Yes	No	Yes
Granada	No	Yes	No	Yes
Madrid	No	Yes	No	Yes
Malaga	No	Yes	No	Yes
Murcia	No	Yes	No	Yes
Sevilla	No	Yes	No	Yes
Valencia	No	Yes	No	Yes

Table 3: Summary of the results of the Breitung (2005) unit root test for panel data, the Phillips-Perron test for unit roots in time series and the Westerlund cointegration test.

Third, our theoretical framework relies on the idea that the retail prices of fuel are affected by the Brent price, and not vice versa. A test that is applied in these cases is the Granger causality test. The use of the term "causality" may be misleading because what the test actually does is checking whether a variable, say x, can be used to predict the values of another variable, say y. If this is the case, then x is said to "Grange-cause" y. However, the Granger causality test performs poorly in case in case the variables are non-stationary and cointegrated, which is our case. Hence, we decided to rely on economic theory and agree with what stated in Contin et al. (2008): *"Spain is a small country relative to the world market in trading crude oil and gasoline. This assumption ensures that causality, if present, is only in one direction, i.e. from the world market to the local market"*. Consequently, we assume that causality goes only in one direction, that is, from the Brent price to the Spanish retail fuel prices.

Given that the above results support the idea of constructing an Error Correction model to represent the process generating the retail fuel prices, and since the ECM contains both lagged values of the dependent variable (in our case, the retail fuel prices) as well as lagged values of the main regressor (that is, the Brent prices), we had to identify the optimal number of lags to include in our model. The usual procedures for lag selection are: (i) start from the maximum number of lags justified by economic theory, test the significance of the related coefficient; if this is not significant, decrease the size of the model, and repeat; (ii) use information criteria, like the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC); (iii) select the number of lags that ensure no autocorrelation is left in the residuals.[18] Given the above mentioned properties of our data and the estimation methodology we chose, we found it unfeasible to estimate the AIC; consequently, we adopted the following approach, which is a combination of the (i) and the (iii). First, we thought about the maximum number of lags to introduce that was justifiable according to economic theory and to the literature, that is, about 14 days.[19] Then, we checked for the presence of autocorrelation in the residuals by applying the Woolridge (2002) test for autocorrelation. If the residuals showed no autocorrelation, we removed the highest lag to decrease the size of the

[18] The use of the AIC and BIC is justified by the fact that both their formulas consist of two components: one component is the sum of residuals, that is, the unexplained part of the model, which decreases the more lags you add; the second component can be considered an adjustment component because it increase with the number of lags added. Consequently, one should stop when adding more lags becomes inefficient, that is, when increasing further the size of the model does not substantially improve its explanatory power.

[19] The same maximum number of lags was set also in a similar analysis in Balaguer and Ripollés (2015).

model and get more efficient estimates; afterwards, we applied the Woolridge (2002) test again and repeated the process from the start. The problem that emerged during the procedure was that in some provinces autocorrelation was still present also with 14 lags. Consequently, in some cases we had to add more than 14 lags (for a maximum of 21) to eliminate autocorrelation, thereby negatively affecting the efficiency of the model.[20] As for the number of lags of the Brent price, we considered the Root mean square error (RMSE), which can be considered as an indicator of the goodness of fit of the model used because it represents the average difference between the actual values and the fitted values. Hence, we added lags as long as the RMSE substantially decreased; once the decrease became negligible, we assumed that adding more lags would be more detrimental (in terms of efficiency of the results) than beneficial. This approach was applied to each province considered and the results of the autocorrelation tests performed on the residuals and the RMSE of the regressions for each province are in Annex 3.

3.2.3 The model with daily data

The preliminary analysis led to the following main results: (i) the retail fuel prices series show no unit root, whereas the series of Brent prices is integrated of order 1 (ii) the two series are cointegrated [21]. These two properties of our panel data indicate that the process that generated the series of retail fuel prices can be modelled with an Error Correction model, in which one term captures the short-run dynamics of price adjustment, whereas the other captures the push back towards the long-run equilibrium. As mentioned above, the ECM incorporates the ADL by including an auto-regressive component, and it exploits the existence of cointegration between the variables as a source of information. The type of asymmetries that we aim at capturing with our model are reaction time asymmetries, as represented by the *rocket and feather* metaphor.

We thus suggest an enriched version of the pricing equation in Balaguer and Ripollés (2015):

$$\Delta p_{it} = \theta_i(p_{i,t-1} - a_i - \gamma_i t - \sum_{q=0}^{6} \vartheta_{iq} Day_q - \phi_i w_{t-1}) + \sum_{m=1}^{M+} \beta_{im}^{+} \Delta p_{i,t-m}^{+} + \sum_{n=0}^{N} \delta_{in}^{+} \Delta w_{t-n}^{+}$$
$$+ \sum_{m=1}^{M-} \beta_{im}^{-} \Delta p_{i,t-m}^{-} + \sum_{n=0}^{N-} \delta_{in}^{-} \Delta w_{t-n}^{-} + \sum_{h=1}^{H} \lambda_h AvgPriceComp_{t-h} + \pi Fine + \mu Interaction + \varepsilon_{it} \tag{1}$$

where

- Δ is the first difference operator
- p is the retail price of fuel
- t is a linear time trend
- w is the price of Brent
- *Day* represent the daily dummies to control for seasonality in the demand
- *AvgPriceComp* is the lagged average of the prices of the three closest competitors in the area of 2 km radius
- *Fine* represents a dummy that is equal to 1 after February 2015 (the period in which the fine of the CNMC was made public)

[20] It is indeed common practice in dynamic analysis to include all the lags between 0 and the highest lag. Since in some cases we needed the 21st lag to eliminate autocorrelation from the residuals, we thus had to include also the previous 20 lags.

[21] Given that the two series are integrated of different orders, this means that they are *fractionally* cointegrated

- *Interaction* represents an interaction between *AvgPriceComp* and *MainFirms*, which is a dummy that is equal to 1 if the gas station belongs to either Repsol or Cepsa (the main Spanish operators)
- θ_i represents the speed of convergence to the long-run equilibrium
- ϕ_i identifies the cost pass-through of the wholesale price to the retail price of gas station i
- β^{+}_{im} and δ^{+}_{in} capture the short-run dynamics for positive variations in prices (as indicated by the +)
- β^{-}_{im} and δ^{-}_{in} capture the short-run dynamics for price reductions (as indicated by the -)

The term

$$(p_{i,t-1} - a_i - \gamma_i t - \sum_{q=0}^{6} \vartheta_{iq} Day_q - \phi_i w_{t-1})$$

represents the error correction term, whose coefficient captures the correction in current prices that helps push them back to the level implied by the long-term relation with Brent prices.

Our hypotheses are the following:

1. θ_i: negative because the correction should occur in the opposite direction to the deviation made by the retail fuel prices;
2. λ: significant because we consider the average price set by the closest competitors to matter;
3. μ: significant because we suggest that there is some variability in the pricing of Repsol and Cepsa that is not accounted for in the rest of the model and that may be due to their market power;
4. δ^{+}_{in}: positive and overall larger magnitude than δ^{-}_{in} given that we expect a larger response to positive shocks than to negative ones;
5. π: non-significant because of the results of the Breitung (2005) unit root test.

3.2.4 Analyses at the daily level and results

Before running the model, we test our assumption that the panel is heterogeneous, that is, the coefficients of the model vary across the gas stations by using the Swamy's test for parameter homogeneity, which can be applied also in the presence of cross-sectional dependence. The results of the test in Annex 2 support our assumption and are consistent with the idea that there is an idiosyncratic component in the pricing of gas stations.

We estimate the model by using the Common Correlated Effects Mean Group (CCEMG) estimator developed by Pesaran (2006) and adopted in the analysis of asymmetric pricing by Balaguer and Ripollés (2015). This methodology is adequate to address heterogeneous panels since it allows for heterogeneous slope coefficients across gas stations and takes into account the existence of cross-sectional (or spatial) dependence. The term *Mean Group* refers to the procedure implemented: first, a group specific regression is estimated; then, the estimated coefficients are averaged across groups. The outputs of the regressions are summarized in Annex 3 and from them we could establish whether the results meet our hypotheses.

First, we considered θ_i, which represents the speed of adjustment of the retail fuel price to the equilibrium relation with the Brent price series. As stated in Hypothesis (1), we expected θ_i to be negative and the results confirm our expectation because the coefficient is significant and negative in all the provinces considered, reflecting the fact that if the retail price temporarily deviates from the long-run equilibrium relation with the Brent price, then the adjustment must occur in the opposite direction to the one of the deviation. The average value of θ_i is about 0.56, which in percentage terms (56%) represents the daily correction. In other words, this tells us that in about two days the retail prices correct their deviations. However, we recognize that this value is likely to be overestimated because, for instance, in Balaguer and Ripollés (2015), the number of days needed for the correction to occur in Madrid and Barcelona was found to be about six days and when we discuss hypothesis (4) we suggest a potential justification for this.

Hypothesis 2 concerned λ_h, which represented the importance of the average price in the local market in the previous days and our expectation was that this coefficient should be significant. The results in Annex 3 show that this hypothesis is confirmed in 8 out of 9 provinces and is consistent with the idea that gas stations take into account the prices of their rivals in the neighbourhood and with the finding that the pricing of the gas stations is cross-sectionally dependent.

Hypothesis 3 investigated the interaction between the average price of the competitors in the previous days and the brand of the gas station, namely, if the gas station belonged to Repsol and Cepsa, which are the two largest operators in the Spanish oil market and the owners of eight out of nine refineries in Spain. The introduction of this interaction term was aimed at capturing some variability of the pricing of Repsol and Cepsa gas stations which is not accounted for in the rest of the model; consequently, we simply expected this coefficient to be significant. The results show that this is the case in 5 out of 9 provinces, whereas for the other provinces we suggest that the problem of imperfect collinearity (which is a likely downside of our model) may prevent the effect of this interaction to be adequately captured.

Hypothesis 4 is a crucial one because it directly concerns asymmetric pricing. In fact, we expected larger and more quick adjustments of the retail prices to positive shocks in the Brent price than to negative ones. Given the heterogeneous and large number of lags we included in the models for each province, we decided to report in Table 12 only the estimates for the first three lags for both positive and negative changes in the Brent price. The results that we got are mixed: first, it should be noted that these coefficients are rarely significant, probably because of the already mentioned issue of imperfect collinearity. In fact, we sometimes got a negative adjustment of retail prices to positive shocks of the Brent price, whereas we would expect adjustments of the same sign of the shock. As for the adjustment to negative shocks, they are mostly positive, which may still reflect the fact that prices tend to be stickier in presence of sudden decreases in the Brent price. On the one hand, we recognize that this heterogeneity in the signs may be a consequence of the presence of imperfect collinearity that prevents estimating precisely the coefficients of the lagged variables. On the other hand, we would like to highlight the evolution of Brent and retail diesel prices over the relevant period as plotted in Annex 6. From the graphs, we can indeed see that the relation seems quite complex: also by taking into consideration the fact that retail prices may adjust with some lag of delay to shocks in the Brent series, in some periods the retail fuel prices seem to follow an independent path: for instance, if we consider the period between August and December 2014, we can see that the retail prices happen to suddenly plunge also if the Brent is quite stable. There may be various local reasons that can explain the independent movements of the retail prices: since we are considering pre-tax prices, changes in the fuel taxes are to be excluded; however, one potential explanation could be that before February 2015 the decision of the CNMC on the cartel case in the retail fuel market had not been published yet. Consequently, it could be the case that there was some instability in the cartel pricing and this instability is reflected in the sometimes puzzling values taken by the retail fuel price series in the various provinces. Furthermore, the unusual relation between the diesel prices and the Brent ones may also be a potential explanation for the over-estimation of the θ_i parameter, capturing the convergence to the long-run equilibrium.

Finally, hypothesis 5 states that the coefficient of the dummy *Fine* should be non-significant given the result of stationarity of the retail fuel price. The coefficient of this dummy captures in general the variability in pricing after February 2015 that is not accounted for in the rest of the model. Against our expectations, the results show that this coefficient is significant in 6 out of 9 provinces, but it should be noted that the sign is heterogeneous across provinces. We cannot confidently state that this result is entirely due to the fine of the CNMC. Consequently, we suggest that this result may need to be further investigated: for instance, this coefficient may actually capture some variability in pricing that is due to the fact that before the decision of the CNMC became public, colluding firms may have incurred losses and thus firms may have started to recover these losses after February 2015; consequently, we may need to introduce in the model a control variable capturing the profitability of firms in the market. However, we suggest that, before adding further controls into our model, it would be better to try first to decrease its size by identifying a different method of lag selection that leads to a more efficient specification and this would potentially solve the problem of imperfect collinearity described above. After this, we would need to estimate the model again and check the significance of the coefficient of the dummy *Fine*: if this coefficient were still significant, then we would need to explore further its significance by introducing a control for the profitability of the firms.

4 Conclusions and comments

The aim of our project was to capture the presence of asymmetric pricing in nine Spanish provinces between 2014 and 2015 by taking into account the fact that during the period considered the CNMC fined five of the largest oil companies in Spain for collusion.

To do this, we estimated an Error Correction Model for each province by accounting for both panel heterogeneity and cross-sectional dependence to capture both the short-run and the long-run relation between the retail diesel prices and the Brent price. Our results show that there exists a correction mechanism that pushes the retail prices back to the long-run equilibrium relation with the Brent in case of shocks, even if the magnitude of the correction may be overestimated potentially because of the puzzling pattern of the retail fuel price in the period before the fine of the CNMC became public. As far as the short-term adjustments are concerned, the results are less clear because of the lack of significance of most of the coefficients; this is potentially due both to the presence of imperfect collinearity arising from issues with the lag selection in each province as well as to the already mentioned patterns of the diesel prices. Furthermore, as to what concerns the effect of the fine of the CNMC, the results show that in many provinces there is some variability in the pricing of the gas stations after February 2015 that is not accounted for in the rest of the model. We may not confidently state that this is entirely due to the fine of the CNMC: in fact, the related coefficient may capture some variability due to the attempts of firms of recovering the losses incurred before the fine was issued. Consequently, we suggested that we may need to first solve the issue of the lag selection and then, if the effect of the fine is still significant, we may eventually add some control for the profitability of the firms, even if this would imply expanding the database because of the absence of variables capturing the evolution of profits of the firms observed.

To conclude, although our analysis may still need some refinement especially in the lag selection procedure, we suggest that the relation between asymmetric pricing and competition is worth exploring. In addition, it would be also interesting to understand whether aggregating the data in the database we used at a weekly level, for instance, would lead to different estimates and insights or would simply lead to biased findings as suggested in the literature.

References

[1] Balaguer, J., and Ripollés, J. Testing for price response asymmetries in the Spanish fuel market. New evidence from daily data. *Energy Economics 34* (2012), 2066–2071.

[2] Breitung, J., and Das, S. Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica 59*, 4 (2005), 414–433.

[3] CNMC. RESOLUCIÓN (Expte. S/474/13 PRECIOS COMBUSTIBLES AUTOMOCIÓN).

[4] Cramon-Taubadel, S. V., and Meyer, J. Asymmetric price transmission: fact or artefact? *Working paper, Institut for Agricultural Economy, University Gottingen*.

[5] Eckert, A. Retail Price Cycles and Response Asymmetry. *The Canadian Journal of Economics 35*, 1 (2002), 52–77.

[6] Engle, R. F., and Granger, C. W. J. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica 55*, 2 (1987), 251–276.

[7] F.Correljé, I. C.-P. A., and Palacios, M. B. Competition, regulation, and pricing behaviour in the Spanish retail gasoline market. *Energy Policy 37* (2009), 219–228.

[8] García, J. P. Dynamic pricing in the spanish gasoline market: A tacit collusion equilibrium. *Energy Policy 38*, 4 (2010), 1931–1937.

[9] González, X., and Moral, M. J. Effects of antitrust prosecution on retail fuel prices. *International Journal of Industrial Organzation 67* (2019), 102537.

[10] Hong, W., and Lee, D. Asymmetric pricing dynamics with market power: investigating island data of the retail gasoline market. *Empirical Economics 58*, 5 (2020), 2181–2221.

[11] Ihle, R., and von Cramon-Taubadel, S. A Comparison of Threshold Cointegration and Markov-Switching Vector Error Correction Models in Price Transmission Analysis. *2008 Conference, April 21-22, 2008, St. Louis, Missouri 37603, NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*.

[12] Jens-Peter Loy, Carsten Steinhagen, C. W., and Koch, B. Price transmission and local market power: empirical evidence from the Austrian gasoline market. *Journal of Applied Economics 50*, 53 (2018), 5728–5746.

[13] L. Bettendorf, S. v. d. G., and Kuper, G. Do daily retail gasoline prices adjust asymmetrically? *Journal of Applied Statistics 36*, 4 (2009), 385–397.

[14] Manera, M., and Frey, G. Econometric Models of Asymmetric Price Transmission? *Fondazione Eni Enrico Mattei Research Paper Series 100*, 05 (2005).

[15] Motta, M. *Competition Policy: theory and practice.* Cambridge, 2004.

[16] OECD. Competition in Road Fuel. *Policy Roundtables* (2013).

[17] Peltzman, S. Prices rise faster than they fall. *Journal of Political Economy 108*, 3 (2000), 466–502.

[18] Perdiguero, J. Symmetric or asymmetric oil prices? A meta-analysis approach. *Energy Policy 57* (2013), 389–397.

[19] Perdiguero, J., and Ramon Borrell, J. Driving competition in local markets with near-perfect substitutes: an application on the Spanish retail gasoline market. *Empirical Economics 57* (2019), 345–364.

[20] Persyn, D., and Westerlund, J. Error Correction Based cointegration Tests for Panel Data. *Stata Journal 8*, 2 (2008), 232–241.

[21] Pesaran, M. H. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics 22*, 2 (2007), 265–312.

[22] Phillips, P. C. B., and Perron, P. Testing for a unit root in time series regression. *Biometrika 75*, 2 (1986), 335–346.

[23] Rezitis, A. N., and Pachis, D. N. Investigating the Price Transmission Mechanism of the Greek Fresh Tomato Market with a Markov Switching Vector Error Correction Model. *Journal of Agricultural & Food Industrial Organization 14*, 1 (2015).

[24] Severin Borenstein, A. Colin Cameron, R. G. Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes? *The Quarterly Journal of Economics 112*, 1 (1997), 205–339.

[25] Shuttleworth, G. Asymmetrical Price Response in Energy Supply: A Review of Ofgem's Analysis. NERA Economic Consulting.

[26] Westerlund, J. Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics 69*, 6 (2007), 709–748.

Annex 1: Descriptive statistics

Province	n^o of gids
Alicante	319
Barcelona	608
Coruna	220
Granada	203
Madrid	565
Malaga	230
Murcia	304
Sevilla	292
Valencia	460
TOTAL	3201

Table 4: Number of gids per province

Annex 2: tests on cross-sectional dependence, stationarity and cointegration on daily data

Pesaran (2015) test for weak cross-sectional dependence on retail fuel prices

H_0 : Panels are weakly cross-sectionally dependent

H_a : Panels are strongly cross-sectionally dependent

Province	CD	P-value
Alicante	3997.535	0.0000
Barcelona	7888.763	0.0000
Coruna	2720.510	0.0000
Girona	2474.131	0.0000
Granada	2559.761	0.0000
Madrid	7129.756	0.0000
Malaga	2879.981	0.0000
Murcia	3825.153	0.0000
Sevilla	3714.679	0.0000
Valencia	5837.670	0.0000

Table 5: Pesaran (2015) test for weak cross-sectional dependence

Breitung unit-root test for panel data

H_0 : The series of retail fuel prices contain a unit root

H_a : The series of retail fuel prices are stationary

Province	Statistic	P-value
Alicante	-1.1e+02	0.0000
Barcelona	-1.2e+02	0.0000
Coruna	-2.6436	0.0041
Granada	-4.3179	0.0000
Madrid	-1.6e+02	0.0000
Malaga	-3.2206	0.0006
Murcia	-99.1343	0.0000
Sevilla	-3.8124	0.0001
Valencia	-1.1e+02	0.0000

Table 6: Breitung unit root test for panel data (Note: for T>N: the *robust* option of the test was used; for T<N: the *demean* option of the test was used)

Phillips-Perron unit-root test for time series

H_0 : The series of Brent prices contains a unit root

H_a : The series of Brent prices is stationary

	Statistic	1% Critical value	5% Critical value	10% Critical value
Z(rho)	-2.388	-20.341	-14.000	-11.200
Z(t)	-1.495	-3.456	-2.878	-2.570

MacKinnon approximate p-value for Z(t) = 0.5360

Table 7: Phillips-Perron unit root test for Brent prices

H_0 : The first-differenced series of Brent prices contains a unit root

H_a : The first-differenced series of Brent prices is stationary

	Statistic	1% Critical value	5% Critical value	10% Critical value
Z(rho)	-305.567	-20.341	-14.000	-11.200
Z(t)	-17.125	-3.456	-2.878	-2.570

MacKinnon approximate p-value for Z(t) = 0.0000

Table 8: Phillips Perron unit root test for first-differenced Brent prices

Westerlund (2007) bootstrap panel cointegration tests

H_0 : No cointegration between retail fuel prices and the price series of Brent for all panels

H_a : At least one panel is cointegrated

Province	Gt statistic (Value)	Gt statistic (Z-value)	Gt statistic (P-value)	Gt statistic (Robust p-value)	Ga statistic (Value)	Ga statistic (Z-value)	Ga statistic (P-value)	Ga statistic (Robust p-value)
Alicante	-7.005	-101.951	0.0000	0.0000	-67.040	-144.531	0.0000	0.0000
Barcelona	-7.024	-141.340	0.0000	0.0000	-66.659	-198.155	0.0000	0.0000
Coruna	-4.755	-43.608	0.0000	0.0000	-37.882	-56.429	0.0000	0.0000
Granada	-7.022	-81.623	0.0000	0.0000	-67.097	-115.417	0.0000	0.0000
Madrid	-7.470	-149.288	0.0000	0.0000	-73.644	-215.436	0.0000	0.0000
Malaga	-7.359	-93.173	0.0000	0.0000	-71.000	-131.556	0.0000	0.0000
Murcia	-7.327	-106.429	0.0000	0.0000	-71.783	-153.254	0.0000	0.0000
Sevilla	-7.020	-97.864	0.0000	0.0000	-68.013	-140.726	0.0000	0.0000
Valencia	-7.456	-96.238	0.0000	0.0000	-67.148	-157.143	0.0000	0.0000

Table 9: Westerlund error-correction-based panel cointegration tests

H_0 : No cointegration between retail fuel prices and the price series of Brent for all panels

H_a : All the panels are cointegrated

Province	Pt statistic (Value)	Pt statistic (Z-value)	Pt statistic (P-value)	Pt statistic (Robust p-value)	Pa statistic (Value)	Pa statistic (Z-value)	Pa statistic (P-value)	Pa statistic (Robust p-value)
Alicante	-124.168	-98.834	0.0000	0.0000	-65.712	-165.395	0.0000	0.0000
Barcelona	-172.332	-137.488	0.0000	0.0000	-65.346	-226.865	0.0000	0.0000
Coruna	-70.295	-44.550	0.0000	0.0000	-36.301	-66.206	0.0000	0.0000
Granada	-98.324	-78.011	0.0000	0.0000	-64.172	-128.360	0.0000	0.0000
Madrid	-175.982	-143.806	0.0000	0.0000	-72.192	-245.236	0.0000	0.0000
Malaga	-109.340	-88.389	0.0000	0.0000	-67.634	-145.193	0.0000	0.0000
Murcia	-125.708	-101.622	0.0000	0.0000	-69.193	-171.358	0.0000	0.0000
Sevilla	-118.648	-94.389	0.0000	0.0000	-65.963	-158.941	0.0000	0.0000
Valencia	-128.765	-91.326	0.0000	0.0000	-66.413	-183.488	0.0000	0.0000

Table 10: Westerlund error-correction-based panel cointegration tests

Swamy's test for panel homogeneity

H_0 : Parameters are homogeneous across panel units

H_a : Parameters are heterogeneous across panel units

Province	Chi2 statistic	Prob > Chi2
Alicante	60349.85	0.0000
Barcelona	1.3e+05	0.0000
Coruna	21862.10	0.0000
Granada	41234.90	0.0000
Madrid	1.2e+05	0.0000
Malaga	41141.79	0.0000
Murcia	67398.90	0.0000
Sevilla	55123.90	0.0000
Valencia	92028.24	0.0000

Table 11: Swamy's test for panel homogeneity

Annex 3: regression results and Woolridge (2002) test

	Alicante	Barcelona	Coruna	Granada	Madrid	Malaga	Murcia	Sevilla	Valencia
θ_i	-0.579***	-0.561***	-0.443***	-0.503***	-0.640***	-0.541***	-0.670***	-0.567***	-0.515***
	(0.0158)	(0.0130)	(0.0124)	(0.0198)	(0.0130)	(0.0182)	(0.0279)	(0.0182)	(0.0131)
ϕ_i	-0.000592	0.00269	-0.00179	-0.0102*	-0.00101	-0.00723	-0.00519	-0.00705	-0.00290
	(0.00380)	(0.00335)	(0.00250)	(0.00476)	(0.00275)	(0.00381)	(0.00504)	(0.00419)	(0.00427)
λ_{t-1}	0.0929***	0.0354***	0.0949***	0.0148	0.0602***	0.0705***	0.0401**	0.0379**	0.0501***
	(0.0127)	(0.00896)	(0.0144)	(0.0114)	(0.00767)	(0.0147)	(0.0124)	(0.0116)	(0.00972)
λ_{t-2}	-0.0108	0.0378***	0.0124	0.0239	0.00196	0.00717	0.0336*	0.0334*	-0.0144
	(0.0150)	(0.00857)	(0.0141)	(0.0140)	(0.0104)	(0.0189)	(0.0157)	(0.0135)	(0.0117)
π	-0.119***	0.161***	0.0898***	0.0104	-0.126***	0.0243	0.106***	-0.0970**	0.0910***
	(0.0249)	(0.0178)	(0.0157)	(0.0344)	(0.0194)	(0.0311)	(0.0315)	(0.0314)	(0.0245)
μ	0.0126**	0.00269	0.00709	0.00418	0.0535***	0.0411***	0.0126	0.0191*	0.0241***
	(0.00458)	(0.00225)	(0.00408)	(0.00723)	(0.00675)	(0.0101)	(0.00783)	(0.00763)	(0.00548)
$\delta^+_{i,0}$	-0.00452	0.00231	0.00369	-0.00431	-0.00342	-0.00325	-0.000855	-0.00153	0.00315
	(0.00569)	(0.00438)	(0.00608)	(0.00682)	(0.00369)	(0.00522)	(0.00544)	(0.00552)	(0.00481)
$\delta^+_{i,1}$	-0.00415	-0.00626	0.00106	0.0118	0.00307	0.00819	0.00626	0.00297	0.00721
	(0.00674)	(0.00537)	(0.00669)	(0.00838)	(0.00452)	(0.00640)	(0.00744)	(0.00768)	(0.00625)
$\delta^+_{i,2}$	-0.00560	-0.00618	0.00599	-0.00464	0.00340	0.0183	0.00275	0.00216	0.00709
	(0.00866)	(0.00737)	(0.00936)	(0.0126)	(0.00709)	(0.00941)	(0.0101)	(0.0103)	(0.00820)
$\delta^+_{i,3}$	-0.00257	-0.00140	0.0133	0.00819	-0.00166	0.00581	0.00631	0.00859	0.00698
	(0.00688)	(0.00538)	(0.00737)	(0.00995)	(0.00621)	(0.00790)	(0.00861)	(0.00745)	(0.00773)
$\delta^-_{i,0}$	0.00519	-0.00385	0.00364	0.00222	-0.00507	0.00255	-0.00376	0.000985	-0.00156
	(0.00480)	(0.00434)	(0.00491)	(0.00797)	(0.00360)	(0.00536)	(0.00569)	(0.00565)	(0.00425)
$\delta^-_{i,1}$	0.00453	-0.00230	0.00483	0.0111	-0.00120	0.00469	0.00879	0.00775	0.000910
	(0.00613)	(0.00644)	(0.00675)	(0.0107)	(0.00553)	(0.00763)	(0.00862)	(0.00706)	(0.00725)
$\delta^-_{i,2}$	0.00419	-0.00424	-0.00192	0.0120	0.00488	0.00441	0.0119	-0.000352	0.00702
	(0.00790)	(0.00686)	(0.0102)	(0.00975)	(0.00691)	(0.00649)	(0.00864)	(0.00799)	(0.00744)
$\delta^-_{i,3}$	-0.000751	-0.0101	0.00252	0.0138	-0.00173	0.00653	0.0202*	0.00200	0.00752
	(0.00795)	(0.00610)	(0.00929)	(0.00958)	(0.00504)	(0.00801)	(0.00790)	(0.00845)	(0.00618)
_cons	-0.291	-0.00265	0.00271	-1.033**	-0.0978	-0.330*	-0.549	-0.370	-0.276
	(0.174)	(0.174)	(0.217)	(0.320)	(0.152)	(0.159)	(0.305)	(0.229)	(0.207)

Note: To limit the size of the table, we omitted the coefficients of the lagged retail fuel prices, Brent prices and the daily dummies.
Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Regression results. (Common Correlated Effect Mean Group estimator, Pesaran (2006))

Summary of the number of lags used in the model of each province

Province	p_{t-m}	w_{t-n}
Alicante	14	10
Barcelona	21	5
Coruna	12	8
Granada	18	5
Madrid	19	6
Malaga	13	7
Murcia	20	9
Sevilla	14	11
Valencia	17	10

Woolridge (2002) test for autocorrelation in panel data per province

H_0 : No first-order autocorrelation in the residuals

H_a : Presence of first-order autocorrelation in the residuals

Province	Statistic	Prob > F
Alicante	2.476	0.1166
Barcelona	0.040	0.8416
Coruna	2.356	0.1263
Granada	1.074	0.3013
Madrid	1.316	0.2517
Malaga	2.652	0.1048
Murcia	2.760	0.0977
Sevilla	0.086	0.7701
Valencia	0.838	0.3604

Table 13: Wooldridge (2002) test for autocorrelation in panel data

Root mean square error of the model

Province	RMSE
Alicante	0.6523
Barcelona	0.6742
Coruna	0.6172
Granada	0.7025
Madrid	0.5919
Malaga	0.6520
Murcia	0.6048
Sevilla	0.6931
Valencia	0.6740

Table 14: Root mean square error of the model per province

Annex 4: Robustness check: Breitung (2005) unit root test

In the following table we present the results of the Breitung unit root test for panel data performed on subsets of the province datasets for which T < N.
The subsets where created according to the following criteria:

- If N/2 < T, then the dataset was split into two subsets accordingly;
- If N/2 > T, then we created two subsets of 200 observations each, and then a third one with the remaining observations.

Province	Statistic	P-value
Alicante (1)	-3.4181	0.0000
Alicante (2)	-3.6737	0.0001
Barcelona (1)	-4.1663	0.0000
Barcelona (2)	-4.2406	0.0000
Barcelona (3)	-4.4697	0.0000
Coruna	-2.6436	0.0041
Granada	-4.3179	0.0000
Madrid (1)	-3.7355	0.0001
Madrid (2)	-3.9824	0.0000
Malaga	-3.2206	0.0006
Murcia (1)	-3.9385	0.0000
Murcia (2)	-3.8750	0.0001
Sevilla	-3.8124	0.0001
Valencia (1)	-3.9305	0.0000
Valencia (2)	-4.3424	0.0000

Table 15: Breitung unit root test for panel data (Option: *robust*)

Annex 5: Spatial distribution of gas stations per province

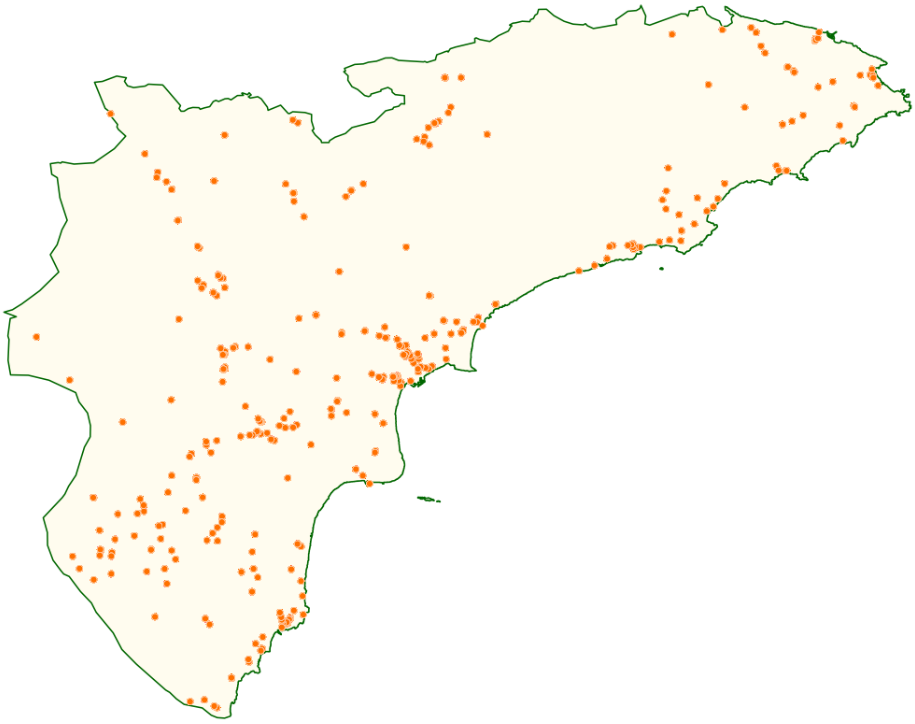

Figure 1: Alicante

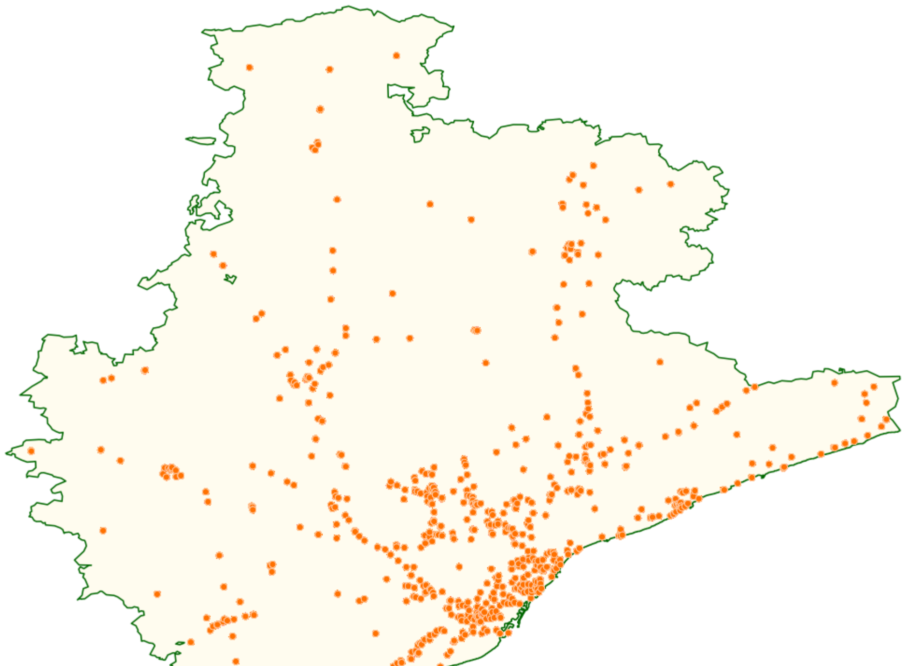

Figure 2: Barcelona

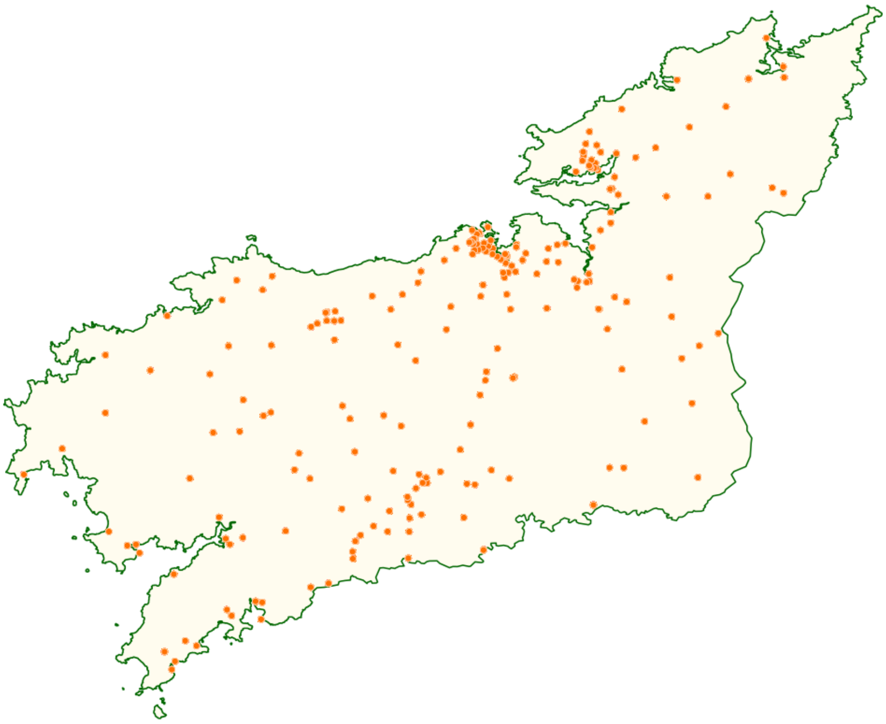

Figure 3: Coruna

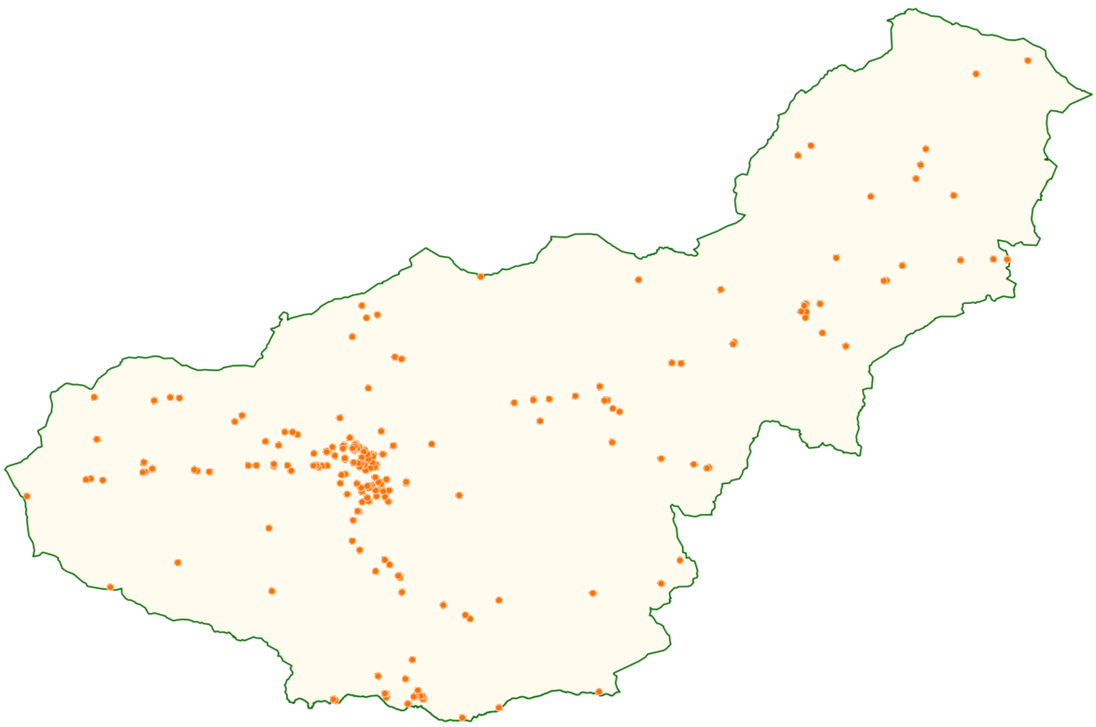

Figure 4: Granada

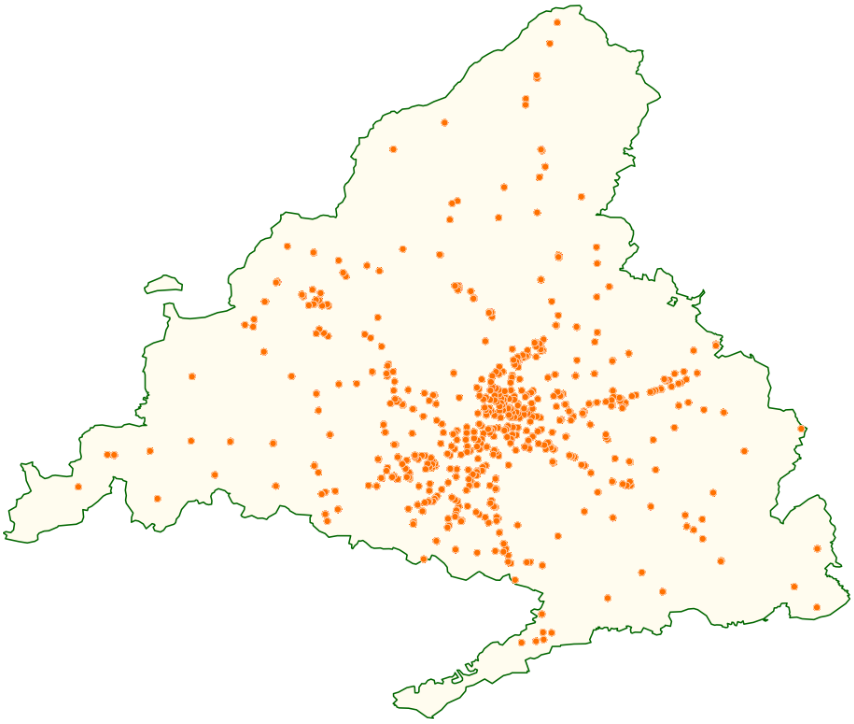

Figure 5: Madrid

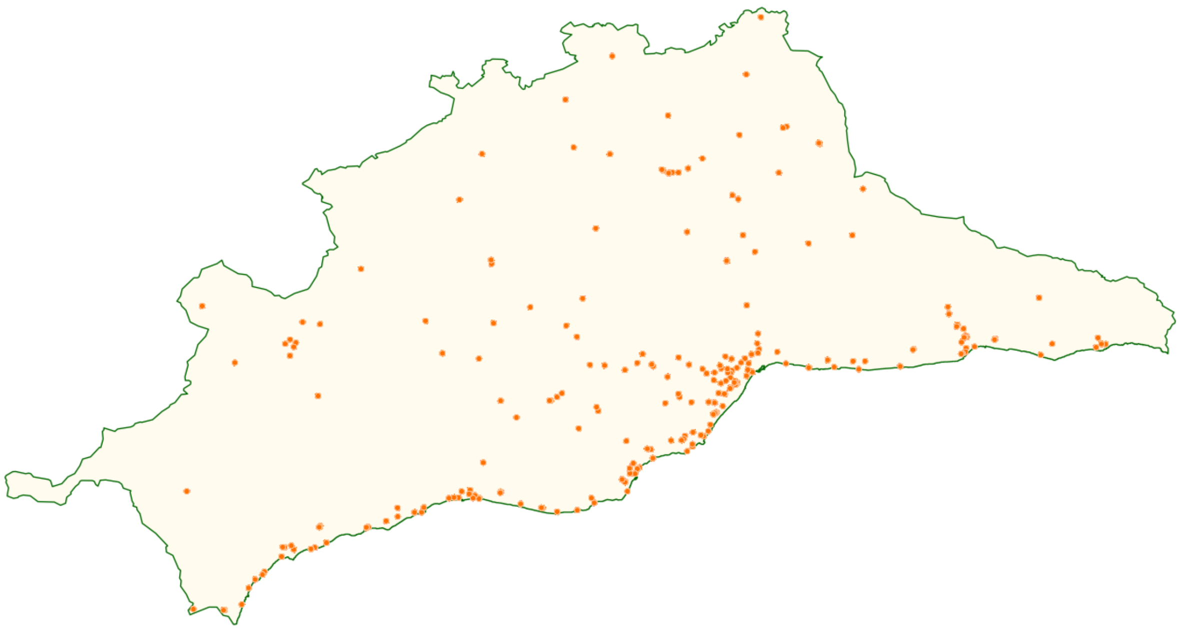

Figure 6: Malaga

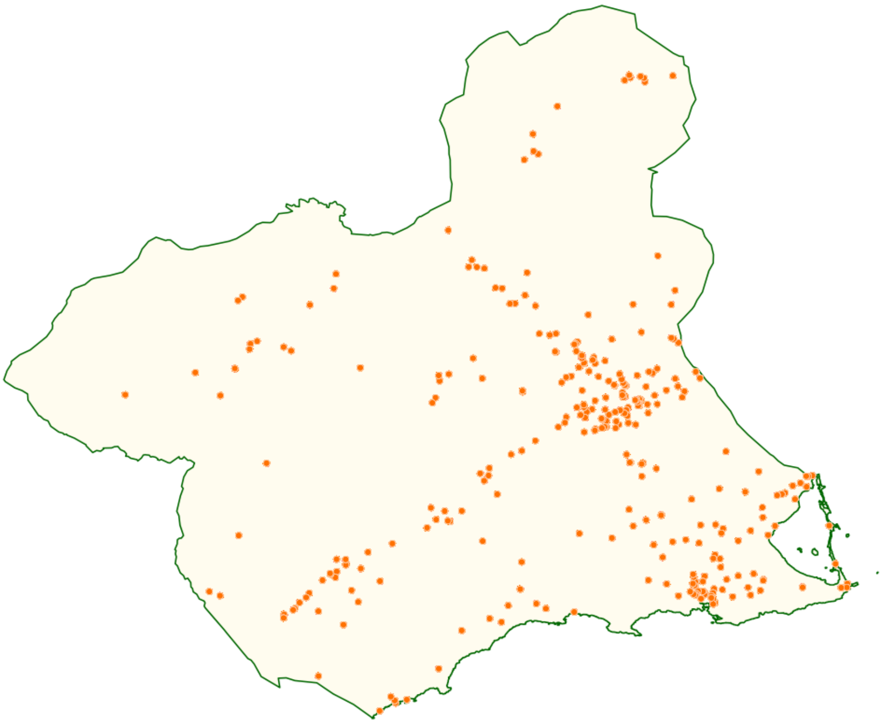
Figure 7: Murcia

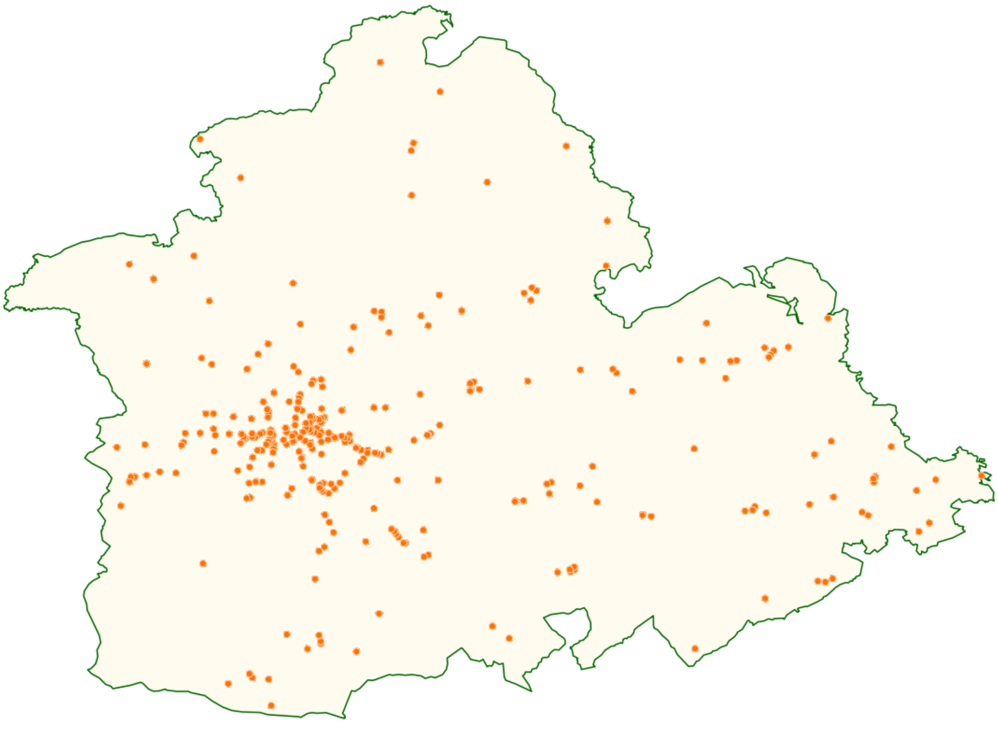
Figure 8: Sevilla

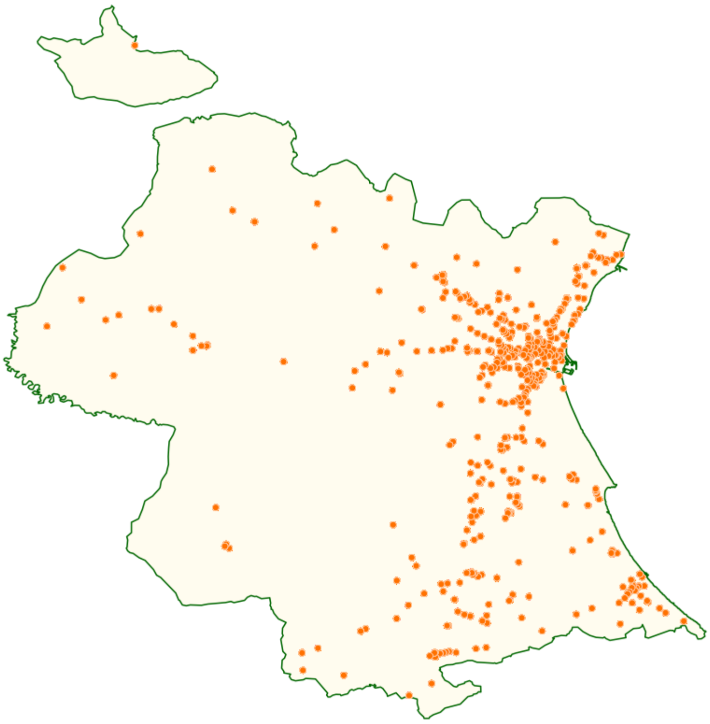

Figure 9: Valencia

Annex 6: Evolution of diesel and Brent prices per province

Black lines are the evolution of diesel prices before taxes (euros cents per litre). Blue lines are Brent prices (euros cents per barrel). Primary axises are Diesesl Prices before taxes (euros cents per litre) and secondary axis are Brent Prices (euros cents per barrel).

Figure 10: Prices

Appendix: A primer on time series

A time series variable is observed at a regular frequency. This can be once per year, once per month, every day, and sometimes, like in some areas of finance, even every second. In this section we provide some insights to understand basic concepts of time series.

Stationarity: Let y_t be a time series variable, where $t = 1, ..., n$ is time index. We say that y_t is stationary if its mean, variance, and covariances with past observations are constant over time

- Mean $E(y_t) = \mu$ is fixed (same for all t)
- Autocovariance $E((y_t - \mu)(y_{t-k} - \mu) = \gamma_k$. If $\gamma_k = 0$ for all $k = 1, 2...$ then the past does not have predictable value for the future. This time series is called **white noise**

There are several ways to transform a series that is non-stationarity into a stationary one. The first option involves running a regression in which the dependent variable is the non-stationary variable and the regressor is the source of non-stationarity, as for instance a time trend. The residuals of the regression are thus the original variable after having been detrended. However, this strategy may not eliminate the other sources of non-stationarity, that is, the presence of seasonality. Consequently, another stronger strategy may be adopted, which consists in taking the first differences of the series, that is, subtracting from the current values of the series the values of the previous lag. A strategy that leads to the same result of first differencing involves taking the logarithm of the series. In order to check whether the series became stationary, we have to run a unit root test on the new series (either the residuals of the regression or the first-differenced variable or the series in logarithms). If the first-differenced variable is stationary, then we say that the process is **integrated of order 1**, or **I(1)**. If the first-differenced variable is still non-stationary, then we have to take the second differences, and if the second-differenced variable is stationary, then it means that the series is integrated of order 2, or I(2), and so on.

Autoregressive model: A model in which the current observation of y in period t is explained by the previous observation of y in period $t-1$. In other words, the variable keeps memory of itself. A particular example of this model is the Autoregressive Model of Order 1, AR(1): $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$. This simple model provides a nice way to illustrate the relevance of stationarity. If the slope parameter β lies between -1 and 1 the effects of past shocks ε_t die out. Formally, if $-1 < \beta < 1$ then

$$\begin{aligned} y_t &= \alpha + \beta y_{t-1} + \varepsilon_t = \alpha + \beta(\alpha + \beta y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \alpha(1+\beta) + \varepsilon_t + \beta\varepsilon_{t-1} + \beta^2 y_{t-2} = ... \\ &= \alpha \sum_{j=0}^{t-2} \beta^j + \sum_{j=0}^{t-2} \beta^j \varepsilon_{t-j} + \beta^{t-1} y_1 \end{aligned}$$

For $t \to \infty$ we get $\beta^{t-1} y_1 \to 0$ and $y_t = \frac{\alpha}{1-\beta} + \sum_{j=0}^{\infty} \beta^j \varepsilon_{t-j}$

Cointegration: If two time series share the same stochastic trend we say that they are cointegrated. We can also say that two time series are cointegrated if both series are non-stationary, but a linear combination is stationary.