Disentangling the importance of international border effects. Some evidence from Portugal–Spain based on diesel retailers

Jacint Balaguer[a] and Jordi Ripollés[a]
[a]*Department of Economics, Universitat Jaume I, Castelló, Spain.*

Corresponding author:

Jacint Balaguer
Department of Economics
Universitat Jaume I
12071-Castelló
Spain
E-mail: jacint.balaguer@uji.es

Acknowledgements:

The authors are grateful to the participants of 5th Meeting on International Economics (Villarreal, Spain) and especially to the Professor Lutz Kilian for their comments. However, any error or omission is our responsibility. Financial support from the Universitat Jaume I (UJI-B2017-17), European Union FEDER funds and the Spanish Ministry of Economy and Competitiveness (ECO2014-58975-P) are also gratefully acknowledged.

Highlights:

- Empirical studies on border effect typically disregard cross-country heterogeneity.
- We use a quasi-experiment to better identify international border effects.
- Results from a typical procedure and the quasi-experimental design are compared.
- It is shown that disregarding the heterogeneity may imply illusory border effects.
- Portugal–Spain border has a moderate impact on international fuel price differences.

1. Introduction

Welfare gains from geographical market integration are beyond discussion for economists and many policymakers. Thus, it is not surprising that efforts have been, and are still being, made to remove tariff and nontariff barriers to trade in important areas such as the NAFTA, the Mercosur or the European Union. Nevertheless, since the mid-nineties, a wide body of research has concluded that the elimination of trade barriers that was carried out was insufficient to reach a high degree of integration. In fact, the idea that the remaining political boundaries significantly hinder trade flows or fulfilment of the Law of One Price has been broadly supported by evidence in the literature.[1]

The transaction costs attributed to borders are rather unbelievable in numerous cases (e.g. McCallum, 1995; Engel and Rogers, 1996; Helliwell, 1996; Helliwell, 1997; Anderson and Smith, 1999; Head and Mayer, 2000; Nitsch, 2000), which led Obstfeld and Rogoff (2000) to consider the phenomenon as one of the major puzzles in International Macroeconomics. The early paper by Engel and Rogers (1996) constitutes a good example in this regard. These authors show that the US–Canada border affects consumer prices in cities like Seattle and Vancouver in the same way as an extra separation of locations of 163 million kilometres.[2] Since then, research economists have been quick to look for a convincing explanation for the empirical results. Several studies have made an interesting effort to obtain more accurate estimates on border effects by using disaggregated consumer data (Hillberry, 2002; Engel et al., 2003; Ceglowski, 2003) by adjusting for effects of non-tradable goods (Liu et al., 2010) and also by considering or improving the way in which other limiting factors of integration such as exchange rate variability are introduced into models (Parsley and Wei, 2001; Engel and Rogers, 2001; De Sousa and Lochard, 2005). This body of literature goes part of the way towards understanding the border effect puzzle. Thus, for example, De Sousa and Lochard (2005) found that currency barriers in countries in the CFA Franc Zone in

---

[1] Although a large part of the research in this area is focused on trade quantities and price dispersion, studies on the importance of borders are also carried out for other interesting economic variables such as amount of investments (e.g. Umber et al., 2014) and unemployment rates (e.g. Persyn and Torfs, 2016).

[2] The distance equivalent in Engel and Rogers is adequately recalculated in Parsley and Wei (2001) by considering the average distance between locations in different countries.

West and Central Africa decrease the effect attributed to borders by between 17% and 28%.

However, despite efforts to improve the estimates, authors often acknowledged that the border frictions obtained in their papers were still larger than can reasonably be expected. In fact, the resulting segmentation expressed in terms of an equivalent distance was commonly revealed to be rather inconsistent with the actual volume of trade across the countries analysed and, sometimes, even unbelievable. For instance, Parsley and Wei (2001) found that the impact of the US–Japan border on the cross-country volatility of relative prices is equivalent to adding about 69,000 trillion kilometres between the two countries, in spite of controlling for the exchange rate variability besides distance and unit-shipping cost. Moreover, the importance frequently attributed to the border per se is not entirely consistent with some evidence concerning the significant dependence across neighbouring countries of socio-political (e.g. Becker et al., 2009; Goel and Saunoris, 2014) and economic variables (e.g. Rietveld et al., 2001; Banfi et al., 2005; Connelly et al., 2009). The results for US–Mexico related to cigarettes in Connelly et al. (2009) are very illustrative of this phenomenon. Their findings show that the lower prices and other non-price benefits for smokers in Mexico, such as the availability of different brands, have a negative impact on cigarette sales in the US states located close to the border despite of the well-known difficulties involved in transporting this product across the border.

Gorodnichenko and Tesar (2009) (henceforth, GT) have shed valuable light on the limitations of the empirical strategy commonly employed to identify the importance of border effects since the mid-nineties. They indicate that the typical empirical strategy used for this identification, consisting in the simple introduction of a dummy variable in regressions, would only be adequate if the distributions of the economic variable analysed (e.g. trade differences, price differences) were homogeneous across the regions involved. Otherwise, if there were cross-region heterogeneity, the measure of border frictions would be contaminated with factors beyond the border. In this latter case, the authors recommend the use of a structural model or a natural experiment.

Unfortunately, the problem originated by heterogeneity in distributions is often disregarded in the current empirical literature on the issue. Even though it is sometimes acknowledged, the empirical options frequently employed are unconnected with the idea underlying GT's paper. Some papers explicitly argue substitute solutions, such as considering a continuous variable for the degree of price stickiness instead of the typical dummy variable for border (Crucini et al., 2010);[3] the additional introduction of indicator variables for country-specific pairs (Aker et al., 2014); the application of quantile regressions (Borraz et al., 2016); or even the use of trade volumes rather than price differentials (Chen et al., 2016). It reasonable to think that ignoring the heterogeneity problem or adopting unconvincing alternatives to solve it can be the result of difficulties to implement GT's proposals. That is, structural models require very broad and detailed information on markets, and natural experiments can only be implemented in non-ordinary cases of establishment (elimination) of borders.

In this paper, we employ a quasi-experiment, as an alternative to a natural experiment or a structural model, to evaluate the importance of border effects between countries with the aim of contributing to the literature on international economics.[4] Specifically, the objective of this paper is twofold. On the one hand, we explore to what extent the empirical procedure commonly used in the literature could overstate the size of border effects on price dispersion. To do so, the results from a regression discontinuity design will be compared with those obtained from the standard procedure, considering in an original way both an illusory border and a real international border. On the other hand, we attempt to provide useful evidence on the importance of the Portugal–Spain border effect. This study framework is similar to the extensively studied case of the US and Canada (e.g. McCallum, 1995; Engel and Rogers, 1996; Yi, 2010; Feenstra, 2002; Anderson and Wincoop; 2003; Ishise and Matsuo, 2015) in the sense that both countries are contiguous and there is no outstanding geographical barrier between them that could be confused with the border influence.

---

[3] Because this paper only considers Japan, it is also indicated that GT's criticism cannot be applied to intra-country borders. However, in our paper we reexamine whether internal heterogeneity in distributions is possible.

[4] It is interesting to note that there is a another generation of literature concerned with knowing whether borders between regions within a same country have any economic relevance (e.g. Berkowitz and DeJong, 1999; Heliwell and Verdier, 2001; Gil-Pareja et al., 2005; Daumal and Zignago, 2010; Requena and Llano, 2010; Persyn and Torfs, 2016; Balaguer and Ripollés, 2017). In general, this generation of research papers has concluded that the sub-national borders are also of great economic significance. The paper by Balaguer and Ripollés (2017) constitutes a remarkable exception in this regard, as it provides some evidence that contrasts with that previously obtained for Spain (Gil-Pareja et al., 2005; Requena and Llano, 2010). Specifically, by employing an empirical approach similar to the one used in this paper (i.e. a quasi-experiment based on the typical geodesic distances), this more recent work suggests that the economic effects arising from sub-national borders are rather negligible.

The rest of this paper is organized as follows. Section 2 describes the data employed, their characteristics and their sources. Section 3 presents the framework to be studied. Section 4 offers a specification model with which to estimate the effect of the international border in line with the usual strategy derived from the seminal paper by Engel and Rogers (1996). Section 5 describes a regression discontinuity design to alternatively isolate the impact of the border. Section 6 provides the empirical evidence. Furthermore, robustness checks on the empirical results are performed in Section 7. Finally, concluding remarks will be given in Section 8.

2. Data

In this study, we employ a large dataset for the automotive fuel sector. Specifically, we focus on diesel, which constitutes the most important petroleum-based fuel for road transportation in the whole of the EU, and on the two member countries involved in our analysis. Thus, following data for 2015 from FuelsEurope,[5] diesel represents 78.7% and 81.1% of the total automotive fuel consumption in Portugal and Spain, respectively. Besides the importance of this product, we also highlight two advantages in carrying out our analysis. First, there are no differences in the intrinsic characteristics of this consumer product between Portugal and Spain. Therefore, the only differences for consumers would be associated with the distance to sellers and brands. Second, as can be seen in Figure 1, there are a large number of sellers spread throughout both countries, which will facilitate our quasi-experimental design.

[Please insert Figure 1 about here]

According to the data collected on 23 November 2016 from the Portuguese *Directorate-General for Energy and Geology* and the Spanish *Ministry of Energy, Tourism and Digital Agenda,* altogether there are 11,513 petrol stations operating on the Iberian Peninsula. The data show that 20.43% of them are spread across Portuguese territory, while the remaining 79.57% of them are in Spain. For each of the stations, we have obtained detailed information on diesel prices (expressed in €/litre), brand affiliation

---

[5] See https://www.fuelseurope.eu/.

and geographical coordinates.[6] We have also collected the taxations to which the product is subjected from the *Associação Portuguesa de Empresas Petrolíferas* and *Asociación Española de Operadores de Productos Petrolíferos*.[7] Fuel taxes differ between Portugal and Spain and even from one Spanish autonomous community to another (as can be seen in Appendix A). Thus, by using the geographical coordinates, a tax burden has been attributed for each station according to its location.

3. Study framework and price data analysis

To carry out the empirical analysis we differentiate three sorts of borders, as shown in Figure 1. First, we simulate a border within Portugal, not supported by any real administrative separation. More specifically, by defining a line joining the coordinates (37.011, -7.875) and (42.108, -7.875), we divide the country into two virtual regions.[8] Second, we consider the existing borders between the contiguous Spanish autonomous communities (i.e. NUTS II). Third, we pay special attention to the international border between Portugal and Spain.

For analysis purposes, let us now build pairwise price comparisons based on the regions defined above. First, we build price differences within each virtual region, within each autonomous community in Spain, and within Portugal. We can then evaluate whether there is heterogeneity in the distribution of such price differences across the contiguous regions. To do so, we use the Kolmogorov-Smirnov test (Kolmogorov, 1933; Smirnov, 1939). As can be seen in Table 1, the null hypothesis of cross-region homogeneity in the distributions of within-region price differentials can be rejected at the 1% level of significance in all the cases considered. That is, we can reject the equality of distributions of the price differentials between the virtual regions (West region-East region), between the contiguous autonomous communities within Spain, and between Portugal and the contiguous Spanish autonomous communities (Portugal-Galicia,

---

[6] These data for Portugal and Spain were downloaded from http://www.precoscombustiveis.dgeg.pt/ and http://www.geoportalgasolineras.es/, respectively.

[7] Fuel tax information is available at http://www.apetro.pt/ and http://www.aop.es/ for Portugal and Spain, respectively.

[8] Portugal is one of the most centralized countries in Europe (Syrett, 1997; Magalhães, 2012) and, unlike Spain, its peninsula is not subdivided into regions with ample administrative autonomy that may interfere with simulation results.

Portugal-Castile-Leon, Portugal-Extremadura, Portugal-Andalusia). Therefore, in any of the three cases considered, we can expect that the method commonly applied to measure border frictions would be contaminated with factors beyond the border in accordance with GT's paper.

[Please insert Table 1 about here]

4. Standard approach

4.1. A typical specification

In line with the empirical strategy adopted in an influential generation of papers (e.g. Engel and Rogers, 1996; Parsley and Wei, 2001; Engel et al., 2003), let us specify the following baseline regression model:

$$ln\left(\frac{p_i}{p_j}\right) = \beta \;\; Border_{ij}^{(K)} + \; f(TC_{ij}) + \; Z_{ij}^{'}\,\theta + u_{ij},\; p_i > p_j \qquad (1)$$

where price dispersion is measured as the log ratio of prices fixed by retailers located at i and j, ordered such that $p_i > p_j$. $Border_{ij}^{K}$ represents a dummy variable which is equal to one if retailers are separated by a particular sort of border *(K),* and zero otherwise. The function $f(TC_{ij})$ would capture the effect of the transportation cost (TC) of engaging in arbitrage activity between locations i and j. The vector $Z_{ij}^{'}$ controls for other potential determinants of price dispersion such as differences in brand affiliation and local taxes. Lastly, u_{ij} is an error term that is assumed to be independent and distributed normally.

The Engel and Rogers-type coefficient (β) has been commonly interpreted as the border effect. However, as has been commented, this interpretation could lead to erroneous conclusions. In fact, the estimation on the coefficient could be determined, in part or even completely, by possible differences in the distribution of price discrepancies within the regions included in the analysis.

4.2. Results from the simulated border

Let us take into account the simulated border previously defined to illustrate the feasible erroneous interpretation of coefficient β in the Eq. (1). With regard to the specification for $f(TC_{ij})$, two issues have to be considered. First, we proxy the transportation costs (TC) by using the driving time between each pair of petrol stations. This has been calculated using the Stata program *osrmtime* developed by Huber and Rust (2016). It determines the driving time corresponding to the shortest route by car between any two pairs of coordinates by means of the Open Source Routing Machine software based on *OpenStreetMap.*[9] The algorithm takes into account the speed limits and bends in the roads, considering normal traffic conditions without disruptions. This strategy is expected to avoid an important restriction associated to the conventional use of straight-line distance between two points in Euclidean space. In fact, since road networks are generally complex structures, it is possible that some geographically closer service stations (i.e. within a few kilometres) may not be good substitutes for drivers. The case of neighbouring petrol stations located on opposite sides of divided roadways is a very illustrative example.

Second, it is reasonable to expect that as transport costs increase with the separation between sellers, arbitrage by consumers will tend to be discouraged, thus leading to increasing differences in prices. Therefore, regarding the functional form for transportation costs, $f(TC_{ij})$, many researchers have employed a logarithmic function to capture this phenomenon (e.g. Berkowitz and DeJong, 1999; Borraz et al., 2016). In our case, using our large dataset, we have alternatively opted for considering a step function varying with transportation costs at discrete intervals. We expect this decision to provide our specification with a more realistic approximation of the effect of transportation cost on price differences. Specifically, we built a set of dummies denoted by $D\_TC_{ij}[a,b)$ that take a value one if petrol stations i and j are separated within the interval [a,b), and zero otherwise. Then, it is expected that as the interval [a,b) represents a greater separation between sellers, the effect of the associated dummy variables will tend to be greater until a point where arbitrage becomes practically discouraging.

---

[9] We employ maps updated to 2016, which are available at http://download.geofabrik.de/.

Lastly, vector $Z_{ij}^{'}$ further includes a dummy variable ($Brand_{ij}$) that is equal to one if petrol stations i and j belong to different brand categories, and zero otherwise. To introduce this variable, we distinguished between eight brand categories: Repsol, Cepsa, Galp, BP, Shell, Petronor, Campsa, and others with a market share equal to or lower than 1.5%.

The results are displayed in Table 2, which contains the estimates calculated by using OLS, where White heteroskedasticity-robust standard errors are applied. Let us first focus on the coefficients of our step function related to transportation costs.[10] They suggest that, within 30 minutes' travelling time by car, the closer petrol stations are to each other, the more similar prices are. For longer driving times, price dispersion remains quite constant. In fact, we cannot reject the null hypothesis of equality between $D\_TC_{ij}[30, 35)$ and $D\_TC_{ij}[35, maximum)$ at the standard levels (p-value of 0.160). Indeed, if we estimate an auxiliary regression including some additional staggered dummies, we can observe that increases in travel time after about 35 minutes no longer cause significant changes in price dispersion. This fact can be seen from Figure 2. Moreover, the estimated coefficient associated to $Brand_{ij}$ also seems reasonable. It indicates that price dispersion is significantly higher if petrol stations belong to different brand categories.

[Please insert Table 2 about here]

[Please insert Figure 2 about here]

Finally, we focus on the coefficient associated to $Border^{(simulated)}_{ij}$, which we are mainly interested in. It captures increase in price dispersion when the simulated border separates petrol stations. Because it is statistically significant, it could lead us to wrongly conclude that there is a relevant effect derived from a presumed border. Specifically, under the common interpretation, we would think that price differences would increase by 0.034% due to the presence of a border.

5. An alternative estimation strategy

---

[10] As we expected, estimating a model with a step function to approximate the effect of transportation costs yields a better fit than one with a logarithmic function. To be more precise, the adjusted R^2 for the estimated Eq. (1) is 0.595 when a step function is considered, while it is 0.008 when a continuous logarithmic function is used.

5.1. Quasi-experimental design

From the results derived above, we can extract that, in the presence of a significant heterogeneity between imaginary regions, the standard methodology can oversize the estimated coefficient β to the point of obtaining a significant border effect when, in fact, it does not exist. So, we need another procedure to estimate the border effects more accurately. The paper by Gorodnichenko and Tesar (2009) concluded that, in this case, it is possible to disentangle the impact of border by employing a (natural) experiment. However, this sort of experimentation can only be implemented in extraordinary situations where formation (elimination) of borders takes place. Therefore, taking into account that borders raise price discontinuities (Deardorff, 2014), here we alternatively design a quasi-experiment to isolate their effect.

We assume that retailers in each region can be easily separated into two different groups, giving rise to the following regression discontinuity (RD) specification based on Eq. (1):

$$ln\left(\frac{p_i}{p_j}\right) = \beta_0 Border_{ij}^{(K)}\Big|_{TC_{ij} \leq \delta} + (\beta_0 + \beta_1) Border_{ij}^{(K)}\Big|_{TC_{ij} > \delta} + f(TC_{ij}) + Z_{ij}^{'}\theta + u_{ij}, \quad p_i > p_j \qquad (2)$$

where δ represents a threshold value referring to the transportation cost between each pair of retailers. This threshold value is set as being small enough to ensure that both retailers face the same local market characteristics. Therefore, $Border_{ij}^{(K)}\big|_{TC_{ij} \leq \delta}$ is a dummy variable that only equals one if there is a border (*K*) between the retailers' locations and, in addition, transportation costs between locations is equal to or lower than the threshold value. That is, this variable would capture the effect on price dispersion of an experimental group of neighbouring pairs of retailers belonging to different regions. Moreover, $Border_{ij}^{(K)}\big|_{TC_{ij} > \delta}$ is a dummy variable that only equals one if there is a particular border (*K*) between retailers' locations and the transportation costs between them are larger than the threshold. This variable identifies a control group composed of retailers belonging to different regions, which can be affected by the border as well by the effect derived from heterogeneity of local market characteristics that may be occurring across the territory. Price dispersion of this control group may be determined by an effect from the border (β_0) as well as a residual effect derived from the existence of heterogeneity (β_1). Obviously, we expect the coefficient associated to this control group $(\beta_0 + \beta_1)$ to be similar to the Engel and Rogers-type coefficient (β) in Eq. (1).

5.2. Testing the specification design

Let us test the validity of our quasi-experimental design as a means to perform an appropriate cleansing of the potential heterogeneity contamination from the previously estimated coefficient of the simulated border. To do so, ideally we should choose a threshold value (δ) tending to zero. The reason for this lies in the necessity to establish an appropriate benchmark that ensures the existence of identical local conditions for petrol stations located on both sides of the border (e.g. consumers, competition and production costs). However, this “ideal” context would imply an insufficient number of observations (price comparisons) to carry out a reliable empirical analysis. Hence, we expect that choosing a threshold driving time of 16 minutes, which implies 401 observations, does not represent a relevant limitation, as it allows similar local conditions to be achieved for petrol stations included in the experimental group.

[Please insert Table 3 about here]

The regression results from Eq. (2) are presented in Table 3. The coefficients related to transportation costs and brands are similar to those obtained from Eq. (1) in the section above. Interestingly, the new coefficient associated to the effect of the border variable in the experimental group is not at statistically conventional levels. That is, as is reasonable, we could conclude that the simulated border is an irrelevant barrier to consumers. Thus, according to the results from the control group, the cross-border heterogeneity in distributions constitutes an important source of the observed price dispersion between the virtual regions.

6. Measuring the impact of the Portugal–Spain border

The main aim of this section is to measure the effect of the international border between Portugal and Spain. Moreover, since it could be interesting to compare its impact with those corresponding to intra-national borders belonging to the autonomous communities, our analysis also comprises the intra-national borders within Spain. Dummy variables are included in the specification to capture the effect of brand differences. On this occasion, we allow the effect of brand differences to vary according to whether stations are within Portugal ($Brand_{ij}\,[Portugal]$), within Spain ($Brand_{ij}\,[Spain]$) or belong to different countries ($Brand_{ij}\,[Portugal - Spain]$). In this way we will allow for the possibility of a company of the same brand having a different position and pricing strategy in each country. Thus, for example, Repsol is the leading company in Spain but does not have this advantage in Portugal and one can therefore expect that its pricing behaviour may vary. To introduce these variables, we distinguish between the main brand categories within each country.[11] Finally, we also control for tax differences (Tax_{ij}).

In Table 4 we present the empirical results, where the first column reports the estimates using the typical approach (Eq. 1) and the remaining columns contain the estimates from the regression discontinuity design (Eq. 2). Following the same reasoning as in the section above, we also use a threshold value of 16 minutes' driving time. This implies 400 observations for the experimental group close to the international border, and 3,000 observations belonging to the borders with the contiguous autonomous communities.

[Please insert Table 4 about here]

We find that the conclusions associated to transportation cost, brands and tax differences are independent from Eq. (1) or Eq. (2). Specifically, prices become more similar as petrol stations are closer but within 20 minutes' driving time. A longer driving time has no further effect on the observed price differences. Indeed, we cannot reject the null hypothesis of equality between $D\_TC_{ij}[20, 25)$ and $D\_TC_{ij}\,[25, maximum)$. Figure 3 illustrates how the introduction of some additional staggered dummies in an auxiliary regression would be statistically equivalent after 25 minutes'

---

[11] Specifically, Galp (29.45%), Repsol (18.72%), BP (13.58%), Cepsa (8.03%) and other minor brands are considered in Portugal, whereas we introduce Repsol (28%), Cepsa (13.86%), Galp (5.69%), Shell (3.48%), Petronor (2.25%), Campsa (2.19%), BP (1.90%) and other minor brands in Spain.

driving time between petrol stations. Moreover, the estimated coefficients associated to brands and tax differences are positive, which seems quite reasonable.

In accordance with the results of Eq. (1), the effects of the sub-national and the international borders are both positive and statistically significant at standard levels. Results would indicate that borders among the Spanish autonomous communities would imply that dispersion rises by 0.123%. The international border increases price dispersion to a much greater extent. Specifically, their estimated effect is 4.074%, which would be equivalent to more than 20 minutes' travel time between sellers.

To evaluate whether the border effects discussed above are oversized, we now focus our attention on the results provided by Eq. (2). It is interesting to note that we cannot obtain significant effects for borders belonging to the autonomous communities, unlike the results from Eq. (1). The impact of the international border is once again positive and statistically significant at standard levels, although its magnitude is clearly lower than that obtained from Eq. (1). More particularly, crossing the Portugal–Spain border adds 3.689% to the price dispersion. This is equivalent to a maximum of five minutes' separation between petrol stations. In fact, the effect of the international border is statistically equivalent to the estimated coefficient linked with the step dummy $D\_TC_{ij}[0,5)$ (with a p-value of 0.858).

7. Robustness check

In order to test the robustness of the results concerning the regression discontinuity design, we replicated the analysis by considering different threshold values of driving time. We increased the threshold time in a reasonable way with the idea of maintaining as far as possible similar local conditions for petrol stations included in the experimental group. However, the advantage of a moderate increase is that it considerably enlarges the number of observations within the experimental group. Indeed, in the case of the international border, they increase by 50% on enlarging the threshold time from 16 to 18 minutes. The results obtained from Eq. (2) for some different threshold values are reported in Table 5. As can be seen, our findings are, in essence, not sensitive to these new values. Intra-national border effects continue to be statistically non-significant, while the international border effect arises as a relevant source of price dispersion.

Specifically, regardless of the threshold considered, crossing the Portugal–Spain border adds between 3.60% and 3.70% to the dispersion of prices.

[Please insert Table 5 about here]

As an alternative to the White correction for general forms of heteroskedasticity, we also use a weighted generalized least squares (GLS) estimator, where a proxy variable for the relative average size of retailers (by cities) has been employed as a weighting factor. Specifically, the proxy variable has been defined as the number of inhabitants divided by the number of petrol stations in each city.[12] Table 6 displays the corresponding results based on Eq. (2) for different threshold values of driving time. As can be seen, our conclusions related to border effects remain unaffected.

[Please insert Table 6 about here]

Finally, we also ask ourselves whether our results are robust to the use of geographical distances to proxy the transportation cost. We think that it is important to perform the corresponding robustness check because it is the typical option in this research area, even in the most modern papers (e.g. Bergstrand et al., 2015; Borraz et al., 2016; Chen et al., 2016; Elberg, 2016; Kashiha, et al., 2016; Hayakawa, 2017). With this purpose in mind, we employ the conventional great-circle geodesic distance, which has been calculated from our coordinates by using the Vincenty (1975) ellipsoid method via the *geodist* module available in Stata (Picard, 2012). We selected 14, 14.5, 15, 15.5 and 16 kilometres as the threshold values, since this implies a number of observations for the experimental group comparable to that considered in the analysis based on driving time. Results are presented in Table 7.[13] Price dispersion is affected by a distance of separation between stations within 10 kilometres of each other. A longer distance has no additional effect on the observed price differences. Findings concerning boundaries are also quite robust to the consideration of geographical distances. We obtained that crossing the Portugal–Spain border adds about 3.1% to the average price dispersion between petrol stations.

---

[12] While data on population for each city have been obtained from the corresponding National Statistical Offices of Portugal (https://www.ine.pt/) and Spain (http://www.ine.es/), the number of petrol stations in each territory has been calculated from the dataset described in Section 2.

[13] Because the standard approach based on Eq. (1) has not yet been reported using the great-circle distance variable as proxy, it has also been included in Appendix B. Findings are robust to those obtained by the use of driving time.

[Please insert Table 7 about here]

8. Conclusions

A great part of the empirical literature that assesses the relevance of border frictions has often been an important source of concern as regards the degree of market integration reached among countries. It has frequently been suggested that the efforts to remove tariff and non-tariff barriers to trade might not be sufficient, and that the mere presence of borders between countries implies a strong preference for consumption of home goods and significant deviations from the Law of One Price. However, since the paper by Gorodnichenko and Tesar (2009), the usual empirical strategy consisting in estimating border effects on trade flows or existing prices between pairs of locations has been widely questioned. In fact, the border effect measured from a simple introduction of a dummy variable in regressions could often be contaminated with other spatial factors unrelated to borders, referred to as heterogeneity effects. The proposed solution requires credible theory-based restrictions to build a structural model or the observation of an extraordinary situation of elimination (creation) of borders to apply a natural experiment. Here we have shown that, when there are enough sellers spatially disseminated along borders, it is possible to implement a simple quasi-experimental design as an alternative to estimate the international border effects.

We found that the existence of the Portugal–Spain border has a significant albeit modest impact on the price dispersion from petrol stations. Specifically, the friction generated by the international border can be considered at most equivalent to an extra round trip by car of about ten minutes for consumers. The estimated importance of this border seems rather more reasonable than that obtained from the typical empirical strategy. In fact, we have shown that the estimated friction from our Engel and Rogers-type coefficient would be equivalent to an extra round trip for consumers of more than forty minutes.

The empirical results would suggest that the existence of borders in itself does not seem to be an important limitation to further progress on market integration in the European Union. However, we recognize that it is necessary to carry out more research work on other relevant sectors and countries. Specifically, it is reasonable to think that there will be more arbitrage activity by consumers as products are more valuable, they are easier to transport and store, and can be transported by more alternative modes. Moreover, for some products, a greater number of land borders for each country could also be expected to increase the degree of arbitrage. Regardless of the products and countries analysed, we hope that the estimation strategy displayed here allows more reliable evidence on the effect of the borders across countries to be obtained in order to better evaluate the success of international integration policies.

References

Aker, J.C., Klein, W., O'Connell S.A., Yang, M., 2014. Borders, ethnicity and trade. Journal of Development Economics 107, 1-16.

Anderson, M., Smith, S., 1999. Do national borders really matter? Canada–US regional trade reconsidered. Review of International Economics 7, 219-227.

Anderson, J.E., Van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. The American Economic Review 93, 170-192.

Balaguer, J., Ripollés, J., 2017. Revisiting the importance of border effect in sub-national regions. Evidence from a quasi-experimental design. Papers in Regional Science. Forthcoming (doi:10.1111/pirs.12298).

Banfi, S., Filippini, M., Hunt, L.C., 2005. Fuel tourism in border regions: The case of Switzerland. Energy Economics 27, 689-707.

Becker, S.O., Egger, P.H., Seidel, T., 2009. Common political culture: evidence on regional corruption contagion. European Journal of Political Economy 25, 300-310.

Berkowitz, D., DeJong, D.N., 1999. Russia's internal border. Regional Science and Urban Economics 29, 633-649.

Bergstrand, J.H., Larch, M., Yotov, Y.V., 2015. Economic integration agreements, border effects, and distance elasticities in the gravity equation. European Economic Review 78, 307-327.

Borraz, F., Cavallo, A., Rigobon, R., Zipitria, L., 2016. Distance and political boundaries: Estimating border effects under inequality constraints. International Journal of Finance and Economics 21, 3-35.

Connelly, R.T., Goel, R.K., Ram, R., 2009. Demand for cigarettes in the United States: effects of prices in bordering states and contiguity with Mexico and Canada. Applied Economics 41, 2255-2260.

Ceglowski, J., 2003. The law of one price: intranational evidence for Canada. Canadian Journal of Economics 36, 373-400.

Chen, Z., Rus, H.A., Sen, A., 2016. Border effects before and after 9/11: Panel data evidence across industries. The World Economy 39, 1456-1481.

Connelly, R. T., Goel, R. K., Ram, R., 2009. Demand for cigarettes in the United States: effects of prices in bordering states and contiguity with Mexico and Canada. Applied Economics 41, 2255-2260.

Daumal, M., Zignago, S., 2010. Measure and determinants of border effects of Brazilian states. Papers in Regional Science 89, 735-758.

De Sousa, J., Lochard, J., 2005. Do currency barriers solve the border effect puzzle? Evidence from the CFA Franc Zone. Review of World Economics 141, 422-441.

Deardorff, A., 2014. Terms of trade: glossary of international economics. World Scientific Publishing Co Pte Ltd.

Crucini, M. J., Shintani, M., Tsuruga, T., 2010. The law of one price without the border: the role of distance versus sticky prices. The Economic Journal 120, 462-480.

Elberg, A., 2016. Sticky prices and deviations from the Law of One Price: Evidence from Mexican micro-price data. Journal of International Economics 98, 191-203.

Engel C., Rogers J.H., 1996. How wide is the border? American Economic Review 86, 1112-1125.

Engel C., Rogers J.H. 2001. Deviations from purchasing power parity: causes and welfare costs. Journal of International Economics 55, 29-57.

Engel C., Rogers J.H., Wang S.Y., 2003. Revisiting the border: an assessment of the law of one price using very disaggregated consumer price data. Board of Governors of the Federal Reserve System.

Feenstra, R.C., 2002. Border effects and the gravity equation: consistent methods for estimation. Scottish Journal of Political Economy 49, 491-506.

Gil-Pareja, S., Llorca-Vivero, R., Martínez-Serrano, J., Oliver-Alonso, J., 2005. The border effect in Spain. The World Economy 28, 1617-1631.

Goel, R.K., Saunoris, J.W., 2014. Global corruption and the shadow economy: spatial aspects. Public Choice 161, 119-139.

Gorodnichenko Y., Tesar L.L., 2009. Border effect or country effect? Seattle may not be so far from Vancouver after all. American Economic Journal: Macroeconomics 1, 219-241.

Hayakawa, K., 2017. Domestic and international border effects: The cases of China and Japan. China Economic Review 43, 118-126.

Head K., Mayer T., 2000. Non-Europe: the magnitude and causes of market fragmentation in the EU. Weltwirtschaftliches Archiv 136, 284-314.

Helliwell J.F., 1996. Do borders matter for social capital? Economic growth and civic culture in US states and Canadian provinces. (No. w5863). National Bureau of Economic Research.

Helliwell J.F., 1997. National borders, trade and migration. Pacific Economic Review 2, 165-185.

Helliwell, J.F., Verdier, G., 2001. Measuring internal trade distances: a new method applied to estimate provincial border effects in Canada. Canadian Journal of Economics 34, 1024-1041.

Hillberry R.H., 2002. Aggregation bias, compositional change, and the border effect. Canadian Journal of Economics 35, 517-530.

Huber, S., Rust, C., 2016. Calculate travel time and distance with OpenStreetMap data using the Open Source Routing Machine (OSRM). The Stata Journal 16, 416-423.

Ishise, H., Matsuo, M., 2015. US–Canada border effect between 1993 and 2007: smaller, less asymmetrical, and declining. Review of World Economics 151, 291-308.

Kashiha, M., Depken, C., Thill, J.C., 2016. Border effects in a free-trade zone: Evidence from European wine shipments. Journal of Economic Geography 17, 411-433.

Kolmogorov A.N., 1933. Sulla determinazione empirica delle leggi di probabilita. Giornale dell'Istituto Italiano degli Attuari 4, 1-11.

Liu, X, Whalley, J., Xin X., 2010. Non-tradable goods and the border effect puzzle. Economic Modelling 27, 909-914.

Magalhães, F., 2012. Landscape and regionalism in Portugal: The case of the Leiria region. Journal of Contemporary European Studies 20, 55-76.

McCallum J., 1995. National borders matter: Canada-US regional trade patterns. American Economic Review 85, 615-623.

Nitsch V., 2000. National borders and international trade: evidence from the European Union. Canadian Journal of Economics 33, 1091-1105.

Obstfeld, M., Rogoff, K., 2000. The six major puzzles in international macroeconomics: is there a common cause?. NBER Macroeconomics Annual 15, 339-390.

Picard, R., 2012. GEODIST: Stata module to compute geodetic distances. Statistical Software Components.

Parsley D.C., Wei S.J., 2001. Explaining the border effect: the role of exchange rate variability, shipping costs, and geography. Journal of International Economics, 55, 87-105.

Persyn D., Torfs, W., 2016. A gravity equation for commuting with an application to estimating regional border effects in Belgium. Journal of Economic Geography 16, 155-175.

Requena, F., Llano, C., 2010. The border effects in Spain: an industry-level analysis. Empirica 37, 455-476.

Rietveld, P., Bruinsma, F.R., van Vuuren, D.J., 2001. Spatial graduation of fuel taxes; consequences for cross-border and domestic fuelling. Transportation Research Part A 35, 433-457.

Smirnov N.V., 1939. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Moscow University Mathematics Bulletin 2.2.

Syrett, S., 1997. The politics of partnership: the role of social partners in local economic development in Portugal. European Urban and Regional Studies 4, 99-114.

Umber, M. P., Grote, M. H., Frey, R., 2014. Same as it ever was? Europe's national borders and the market for corporate control. Journal of International Money and Finance 40, 109-127.

Vincenty, T., 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. Survey Review 23, 88-93.

Yi, K.M., 2010. Can multistage production explain the home bias in trade? The American Economic Review 100, 364-393.

Appendix A. Diesel fuel taxes in Portugal and Spain

European retail prices for diesel fuel are subject to a general tax on consumption that is known as VAT (Value Added Tax), and several excise duties that can be heterogeneous across countries and even sub-national regions. On the one hand, the rate of VAT on diesel fuel in Portugal is currently 23%, while the excise duties in the country are composed of a Tax on Oil Products (0.31841 €/litre), the Road Service Contribution (0.111 €/litre) and a Carbon Tax (0.01651 €/litre). On the other hand, diesel fuel in Spain is subject to a VAT rate of 21%, and the excise duties are composed of three Special Hydrocarbon Taxes: a General section (0.307 €/litre), a State section (0.024 €/litre) and a Regional section (which ranges from 0 to 0.048 €/litre, depending on the individual criteria of each Spanish autonomous community). Table A.1 summarises the current taxation framework in both countries.

[Please insert Table A.1 about here]

In view of the previous considerations, we can represent the linkage between retail prices (p_i) and prices net-of-tax ($\ddot{p}_i$) as:

$$p_i = (1 + VAT_i)\cdot(\ddot{p}_i + T_i) \qquad (A.1)$$

where T_i denotes the total excise duties levied on diesel fuel for petrol station i. Therefore, as a result of retailers i and j being located in different countries and/or regions, certain tax differences can arise between them, as follows:

$$Tax_i - Tax_j = (1 + VAT_i)\cdot T_i - (1 + VAT_j)\cdot T_j \quad (A.2)$$

Appendix B

[Please insert Table B.1 about here]

Table 1. Kolmogorov-Smirnov test based on the pairwise price comparisons between neighbouring regions

		Portugal	Spanish Autonomous Communities											
		(II) Simulated region (East)	Andalusia	Aragon	Asturias	Basque Country	Catalonia	Castile-La-Mancha	Castile-Leon	Extremadura	Galicia	Madrid	Murcia	Navarre
Portugal	(I) Simulated region (West)	0.045***												
	Whole country		0.733***						0.081***	0.119***	0.120***			
Spanish autonomous communities	Andalusia							0.446***		0.375***				
	Aragon						0.055***	0.091***	0.115***					0.098***
	Asturias								0.031***		0.064***			
	Basque Country								0.044***					0.107***
	Cantabria				0.090***	0.060***			0.073***					
	Catalonia													
	Castile-La-Mancha								0.058***			0.074***		
	Castile-Leon									0.145***	0.076***	0.064***		
	Murcia		0.464***					0.050***						
	Navarre													
	Rioja			0.081***		0.068***			0.133***					0.120***
	Valencian Community			0.075***			0.069***	0.072***					0.067***	

We use *** to indicate the rejection of the null hypothesis of homogeneity at the 1% level.

Table 2. Results from a simulated border based on Eq. (1). Dependent variable: log ratio of prices

$Border^{(Simulated)}_{ij}$	0.343***	(0.044)
$D\_TC_{ij}[0,5)$	25.444***	(0.391)
$D\_TC_{ij}[5,10)$	24.981***	(0.228)
$D\_TC_{ij}[10,15)$	26.772***	(0.186)
$D\_TC_{ij}[15,20)$	27.945***	(0.171)
$D\_TC_{ij}[20,25)$	30.286***	(0.160)
$D\_TC_{ij}[25,30)$	32.598***	(0.155)
$D\_TC_{ij}[30,35)$	33.793***	(0.150)
$D\_TC_{ij}[35,maximum)$	33.585***	(0.039)
$Brand_{ij}$	6.427***	(0.043)
R^2	0.595	
Total observations	2,764,776	

White's (1980) heteroskedasticity-robust standard errors are presented in parentheses. We use *** to indicate statistical significance at the 1% level. We cannot reject the null hypothesis of equality between $D\_TC_{ij}[30, 35)$ and $D\_TC_{ij}[35, maximum)$ even at the 10% level. Estimated coefficients and standard errors are multiplied by 10^3.

Table 3. Results from a simulated border in Portugal based on Eq. (2). Dependent variable: log ratio of prices

$Border_{ij}^{(simulated)}\big\|_{TC_{ij} \leq \delta}$	0.201	(1.330)
$Border_{ij}^{(simulated)}\big\|_{TC_{ij} > \delta}$	0.343^{***}	(0.044)
$D\_TC_{ij}[0,5)$	25.445^{***}	(0.391)
$D\_TC_{ij}[5,10)$	24.983^{***}	(0.229)
$D\_TC_{ij}[10,15)$	26.774^{***}	(0.186)
$D\_TC_{ij}[15,20)$	27.945^{***}	(0.171)
$D\_TC_{ij}[20,25)$	30.286^{***}	(0.160)
$D\_TC_{ij}[25,30)$	32.597^{***}	(0.155)
$D\_TC_{ij}[30,35)$	33.793^{***}	(0.150)
$D\_TC_{ij}[35,maximum)$	33.585^{***}	(0.039)
$Brand_{ij}$	6.427^{***}	(0.043)
R^2	0.595	
Obs. in $Border_{ij}^{(simulated)}\big\|_{TC_{ij} \leq \delta}$	401	
Total observations	2,764,776	

A threshold driving time (δ) of 16 minutes has been considered. White's (1980) heteroskedasticity-robust standard errors are presented in parentheses. We use ***, ** and * to indicate statistical significance at the 1%, 5% and 10% levels, respectively. We cannot reject the null hypothesis of equality between D_TC_{ij} $[30,35)$ and $D\_TC_{ij}[35, maximum)$ even at the 10% level. Estimated coefficients and standard errors are multiplied by 10^3.

Table 4. Results based on real borders in the Iberian Peninsula. Dependent variable: log ratio of prices

	Eq. (1)		Eq. (2)	
$Border_{ij}^{(International)}$	40.742***	(0.052)		
$Border_{ij}^{(International)}\|_{TC_{ij} \leq \delta}$			36.890***	(3.000)
$Border_{ij}^{(International)}\|_{TC_{ij} > \delta}$			40.742***	(0.052)
$Border_{ij}^{(Intra-national)}$	1.258***	(0.023)		
$Border_{ij}^{(Intra-national)}\|_{TC_{ij} \leq \delta}$			0.937	(0.856)
$Border_{ij}^{(Intra-national)}\|_{TC_{ij} > \delta}$			1.259***	(0.023)
$D\_TC_{ij}[0,5)$	36.343***	(0.228)	36.347***	(0.228)
$D\_TC_{ij}[5,10)$	38.023***	(0.143)	38.028***	(0.143)
$D\_TC_{ij}[10,15)$	39.361***	(0.116)	39.371***	(0.116)
$D\_TC_{ij}[15,20)$	39.868***	(0.103)	39.871***	(0.103)
$D\_TC_{ij}[20,25)$	40.914***	(0.097)	40.914***	(0.097)
$D\_TC_{ij}[25,maximum)$	40.846***	(0.022)	40.845***	(0.022)
$Brand_{ij}\ [Portugal]$	0.023	(0.031)	0.023	(0.031)
$Brand_{ij}\ [Spain]$	11.895***	(0.021)	11.895***	(0.021)
$Brand_{ij}\ [Portugal - Spain]$	4.143***	(0.043)	4.143***	(0.043)
Tax_{ij}	157.687***	(0.144)	157.687***	(0.144)
R^2	0.754		0.754	
Obs. in $Border_{ij}^{(International)}\|_{TC_{ij} \leq \delta}$			351	
Obs. in $Border_{ij}^{(Intra-national)}\|_{TC_{ij} \leq \delta}$			2,406	
Total observations	26,699,957		26,699,957	

A threshold driving time (δ) of 16 minutes has been considered. White's (1980) heteroskedasticity-robust standard errors and p-values are presented in parentheses. We use ***, ** and * to indicate statistical significance at the 1%, 5% and 10% levels, respectively. We cannot reject the null hypothesis of equality between $??\_TC_{ij}[20,25)$ and $D\_TC_{ij}[25,maximum)$ even at the 10% level. Estimated coefficients and standard errors are multiplied by 10^3.

Table 5. Robustness check considering reasonable alternative threshold values in Eq. (2). Dependent variable: log ratio of prices

	$\delta = 16.5$ minutes		$\delta = 17$ minutes		$\delta = 17.5$ minutes		$\delta = 18$ minutes	
$Border_{ij}^{(International)}\|_{TC_{ij} \leq \delta}$	35.998***	(2.851)	36.289***	(2.754)	36.984	(2.625)	36.216***	(2.475)
$Border_{ij}^{(International)}\|_{TC_{ij} > \delta}$	40.742***	(0.052)	40.742***	(0.052)	40.743	(0.052)	40.743***	(0.052)
$Border_{ij}^{(Intra-national)}\|_{TC_{ij} \leq \delta}$	0.741	(0.814)	0.586	(0.771)	0.171	(0.726)	-0.116	(0.691)
$Border_{ij}^{(Intra-national)}\|_{TC_{ij} > \delta}$	1.259***	(0.023)	1.259***	(0.023)	1.260***	(0.023)	1.260***	(0.023)
$D\_TC_{ij}[0,5)$	36.348***	(0.228)	36.348***	(0.228)	36.348***	(0.228)	36.349***	(0.228)
$D\_TC_{ij}[5,10)$	38.03***	(0.143)	38.031***	(0.143)	38.033***	(0.143)	38.035***	(0.143)
$D\_TC_{ij}[10,15)$	39.374***	(0.116)	39.376***	(0.116)	39.380***	(0.116)	39.384***	(0.116)
$D\_TC_{ij}[15,20)$	39.874***	(0.103)	39.876***	(0.103)	39.881***	(0.103)	39.889***	(0.103)
$D\_TC_{ij}[20,25)$	40.914***	(0.097)	40.914***	(0.097)	40.914***	(0.097)	40.914***	(0.097)
$D\_TC_{ij}[25,maximum)$	40.845***	(0.022)	40.845***	(0.022)	40.845***	(0.022)	40.845***	(0.022)
$Brand_{ij}\ [Portugal]$	0.023	(0.031)	0.023	(0.031)	0.023	(0.031)	0.024	(0.031)
$Brand_{ij}\ [Spain]$	11.895***	(0.021)	11.895***	(0.021)	11.895***	(0.021)	11.895***	(0.021)
$Brand_{ij}\ [Portugal-Spain]$	4.143***	(0.043)	4.143***	(0.043)	4.143***	(0.043)	4.143***	(0.043)
Tax_{ij}	157.687***	(0.144)	157.687***	(0.144)	157.687***	(0.144)	157.687***	(0.144)
R^2	0.754		0.754		0.754		0.754	
Obs. in $Border_{ij}^{(Internatioal)}\|_{TC_{ij} \leq \delta}$	377		403		432		473	
Obs. in $Border_{ij}^{(Intra-national)}\|_{TC_{ij} \leq \delta}$	2,663		2,955		3,277		3,617	
Total observations	26,699,957		26,699,957		26,699,957		26,699,957	

White's (1980) heteroskedasticity-robust standard errors are presented in parentheses. We use ***, ** and * to indicate statistical significance at the 1%, 5% and 10% levels, respectively. Transportation costs between each pair of petrol stations are measured in minutes of driving time. We cannot reject the null hypothesis of equality between $D\_TC_{ij}[20,25)$ and $D\_TC_{ij}[25,maximum)$ even at the 10% level. Estimated coefficients and standard errors are multiplied by 10^3.

Table 6. Robustness check by using weighted GLS method in Eq. (2). Dependent variable: log ratio of prices

	$\delta = 16$ minutes		$\delta = 16.5$ minutes		$\delta = 17$ minutes		$\delta = 17.5$ minutes		$\delta = 18$ minutes	
$Border_{ij}^{(International)}\vert_{TC_{ij} \leq \delta}$	36.796***	(2.395)	35.900***	(2.310)	36.209***	(2.233)	36.928***	(2.155)	36.114***	(2.060)
$Border_{ij}^{(International)}\vert_{TC_{ij} > \delta}$	40.813***	(0.050)	40.813***	(0.050)	40.813***	(0.050)	40.814***	(0.050)	40.814***	(0.050)
$Border_{ij}^{(Intra-national)}\vert_{TC_{ij} \leq \delta}$	0.953	(0.909)	0.751***	(0.864)	0.597	(0.820)	0.180	(0.779)	-0.107***	(0.742)
$Border_{ij}^{(Intra-national)}\vert_{TC_{ij} > \delta}$	1.253***	(0.024)	1.253***	(0.024)	1.253***	(0.024)	1.254***	(0.024)	1.254***	(0.024)
$D\_TC_{ij}[0,5)$	36.355***	(0.253)	36.356***	(0.253)	36.356***	(0.253)	36.356***	(0.253)	36.357***	(0.253)
$D\_TC_{ij}[5,10)$	38.026***	(0.157)	38.028***	(0.156)	38.028***	(0.156)	38.03***	(0.156)	38.032***	(0.156)
$D\_TC_{ij}[10,15)$	39.368***	(0.128)	39.371***	(0.128)	39.373***	(0.128)	39.377***	(0.128)	39.381***	(0.128)
$D\_TC_{ij}[15,20)$	39.879***	(0.115)	39.881***	(0.115)	39.884***	(0.115)	39.889***	(0.115)	39.896***	(0.115)
$D\_TC_{ij}[20,maximum)$	39.887***	(0.025)	39.886***	(0.025)	39.886***	(0.025)	39.886***	(0.025)	39.886***	(0.025)
$Brand_{ij}\ [Portugal]$	0.013	(0.039)	0.013	(0.039)	0.013	(0.039)	0.013	(0.039)	0.013	(0.039)
$Brand_{ij}\ [Spain]$	11.907***	(0.024)	11.907***	(0.024)	11.907***	(0.024)	11.907***	(0.024)	11.907***	(0.024)
$Brand_{ij}\ [Portugal - Spain]$	4.069***	(0.037)	4.069***	(0.037)	4.069***	(0.037)	4.069***	(0.037)	4.069***	(0.037)
Tax_{ij}	157.501***	(0.139)	157.501***	(0.139)	157.501***	(0.139)	157.501***	(0.139)	157.501***	(0.139)
Obs. in $Border_{ij}^{(International)}\vert_{TC_{ij} \leq \delta}$	351		377		403		432		473	
Obs. in $Border_{ij}^{(Intra-national)}\vert_{TC_{ij} \leq}$	2,406		2,955		3,277		3,617		3,636	
Total observations	26,699,957		26,699,957		26,699,957		26,699,957		26,699,957	

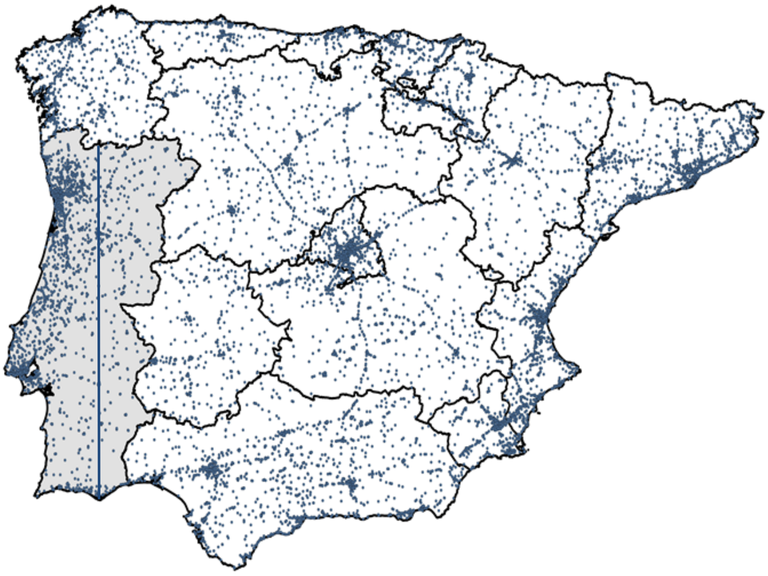
White's (1980) heteroskedasticity-robust standard errors are presented in parentheses. We use ***, ** and * to indicate statistical significance at the 1%, 5% and 10% levels, respectively. Transportation costs between each pair of petrol stations are measured in minutes of driving time. We cannot reject the null hypothesis of equality between $D\_TC_{ij}[15,\ 20)$ and $D\_TC_{ij}[20,\ maximum)$ even at the 10% level. Estimated coefficients and standard errors are multiplied by 10^3.

Table 7. Robustness check based on geographical distance as a proxy of transportation cost in Eq. (2). Dependent variable: log ratio of prices

	$\delta = 14$ km		$\delta = 14.5$ km		$\delta = 15$ km		$\delta = 15.5$ km		$\delta = 16$ km	
$Border_{ij}^{(International)}\|_{TC_{ij} \leq \delta}$	30.067***	(2.641)	30.778***	(2.587)	30.579***	(2.523)	31.153***	(2.437)	31.612***	(2.365)
$Border_{ij}^{(International)}\|_{TC_{ij} > \delta}$	40.735***	(0.052)	40.730***	(0.052)	40.730***	(0.052)	40.730***	(0.052)	40.730***	(0.052)
$Border_{ij}^{(Intra-national)}\|_{TC_{ij} \leq \delta}$	0.705	(0.852)	0.468	(0.795)	0.202	(0.751)	0.030	(0.714)	-0.010	(0.683)
$Border_{ij}^{(Intra-national)}\|_{TC_{ij} > \delta}$	1.251***	(0.023)	1.246***	(0.023)	1.246***	(0.023)	1.246***	(0.023)	1.246***	(0.023)
$D\_TC_{ij}[0,5)$	37.739***	(0.155)	37.739***	(0.155)	37.740***	(0.155)	37.740***	(0.155)	37.739***	(0.155)
$D\_TC_{ij}[5,10)$	39.070***	(0.118)	39.071***	(0.118)	39.074***	(0.118)	39.075***	(0.118)	39.074***	(0.118)
$D\_TC_{ij}[10,15)$	40.015***	(0.107)	40.020***	(0.107)	40.028***	(0.107)	40.030***	(0.107)	40.029***	(0.107)
$D\_TC_{ij}[15,maximum)$	40.091***	(0.101)	39.859***	(0.021)	39.858***	(0.021)	39.858***	(0.021)	39.858***	(0.021)
$Brand_{ij}\ [Portugal]$	0.020	(0.031)	0.018	(0.031)	0.018	(0.031)	0.018	(0.031)	0.018	(0.031)
$Brand_{ij}\ [Spain]$	11.895***	(0.021)	11.896***	(0.021)	11.896***	(0.021)	11.896***	(0.021)	11.896***	(0.021)
$Brand_{ij}\ [Portugal-Spain]$	4.143***	(0.043)	4.143***	(0.043)	4.143***	(0.043)	4.143***	(0.043)	4.143***	(0.043)
Tax_{ij}	157.686***	(0.144)	157.686***	(0.144)	157.686***	(0.144)	157.686***	(0.144)	157.686***	(0.144)
R^2	0.754		0.754		0.754		0.754		0.754	
Obs. in $Border_{ij}^{(International)}\|_{TC_{ij} \leq \delta}$	388		411		438		466		492	
Obs. in $Border_{ij}^{(Intra-national)}\|_{TC_{ij} \leq \delta}$	2,393		2,692		2,996		3,313		3,636	
Total observations	26,699,957		26,699,957		26,699,957		26,699,957		26,699,957	

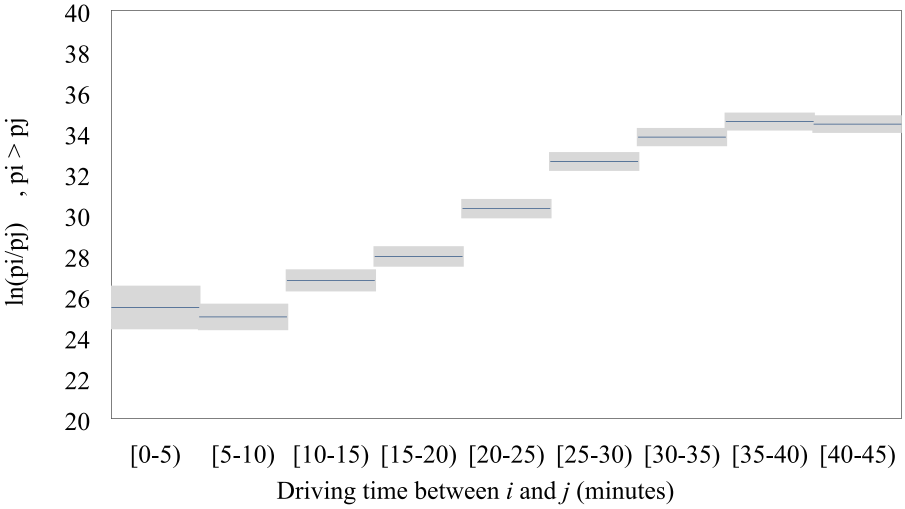
White's (1980) heteroskedasticity-robust standard errors are presented in parentheses. We use ***, ** and * to indicate statistical significance at the 1%, 5% and 10% levels, respectively. Transportation costs between each pair of petrol stations are measured in kilometres. We cannot reject the null hypothesis of equality between $D\_TC_{ij}[10, 15)$ and $D\_TC_{ij}[15, maximum)$ even at the 10% level. Estimated coefficients and standard errors are multiplied by 10^3.

Figure 1. Location of petrol stations and borders



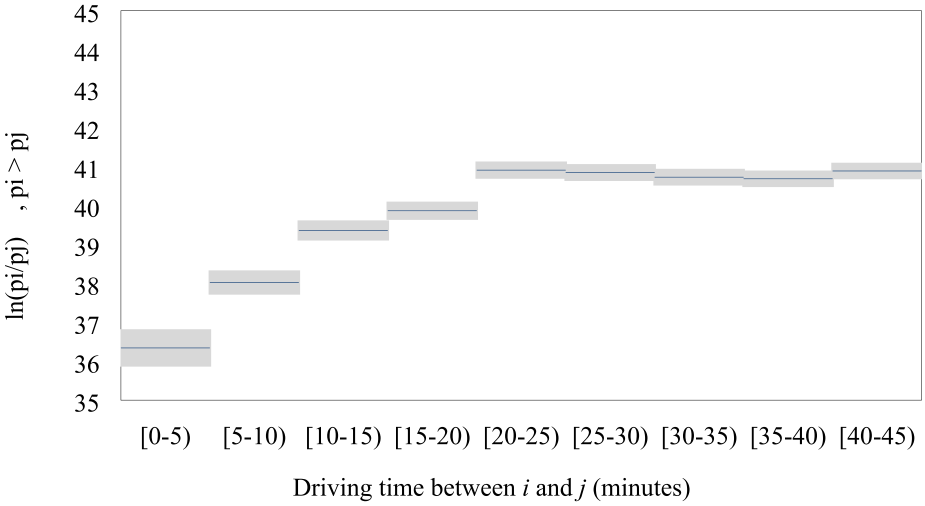
Each of the 11,513 dots denotes the location of one petrol station, and the lines represent the borders. The vertical straight line symbolizes a simulated boundary, and the black lines indicate the administrative boundaries of Portugal (grey area) and Spain (white area).

Figure 2. Relationship between price dispersion and driving time in Portugal based on Eq. (1)



Dark lines represent the estimated coefficients for the staggered dummy variables $D\_TC_{ij}[a,b)$, and the shaded areas show their corresponding 95% confidence intervals. They correspond to an auxiliary regression that includes additional staggered dummies up a maximum of 45 minutes. Estimates are multiplied by 10^3.

Figure 3. Relationship between price dispersion and driving time in the Iberian Peninsula based on Eq. (2)



Dark lines represent the estimated coefficients for the staggered dummy variables $D\_TC_{ij}[a,b)$, and the shaded areas show their corresponding 95% confidence intervals. They correspond to an auxiliary regression that includes additional staggered dummies up a maximum of 45 minutes. Estimates are multiplied by 10^3.

Table A.1. Taxes on diesel motor fuels

VAT rate (%)	
Portugal	23
Spain	21
Excise duties (expressed in €/litre)	
Portugal	
Tax on Oil Products	0.31841
Road Service Contribution	0.11100
Carbon Tax	0.01651
Spain	
Special General Tax	0.307
Special State Tax	0.024
Special Regional Tax	
(1) Andalusia	1.070
(2) Aragon	1.048
(3) Asturias	1.081
(4) Basque Country	1.050
(5) Cantabria	1.038
(6) Catalonia	1.058
(7) Castile-La Mancha	1.078
(8) Castile-Leon	1.036
(9) Extremadura	1.062
(10) Galicia	1.096
(11) Madrid	1.058
(12) Murcia	1.057
(13) Navarre	1.040
(14) Rioja	1.037
(15) Valencian Community	1.057

From data provided by the “Asociación Española de Operadores de Productos Petrolíferos”, and by the “Associação Portuguesa de Empresas Petrolíferas” (November 2016).

Table B.1. Robustness check based on geographical distance as proxy of transportation cost in Eq. (1). Dependent variable: log ratio of prices

$Border^{(International)}_{ij}$	40.728***	(0.052)
$Border^{(Intranational)}_{ij}$	1.244***	(0.023)
$D\_TC_{ij}[0,5)$	37.728***	(0.155)
$D\_TC_{ij}[5,10)$	39.054***	(0.118)
$D\_TC_{ij}[10,15)$	39.997***	(0.107)
$D\_TC_{ij}[15, maximum)$	39.859***	(0.021)
$Brand_{ij}\ [Portugal]$	0.017	(0.031)
$Brand_{ij}\ [Spain]$	11.896***	(0.021)
$Brand_{ij}\ [Portugal - Spain]$	4.143***	(0.043)
Tax_{ij}	157.688***	(0.144)
R^2	0.754	
Obs. in $Border^{(Intern.)}_{ij}\big\|_{TC_{ij} \leq \delta}$		
Obs. in $Border^{(Intran..)}_{ij}\big\|_{TC_{ij} \leq \delta}$		
Total observations	26,699,957	

White's (1980) heteroskedasticity-robust standard errors are presented in parentheses. We use ***, ** and * to indicate statistical significance at the 1%, 5% and 10% levels, respectively. Transportation costs between each pair of petrol stations are measured in kilometres. Estimated coefficients and standard errors are multiplied by 10^3.