CENTER FOR FINANCIAL STUDIES

CFS Working Paper Series

No. 536

Günter W. Beck, Hans-Helmut Kotz, Natalia Zabelina

Prices and consumer purchasing preferences at the border: evidence from a multi-country household scanner data set

Center for Financial Studies | Goethe University | Theodor-W.-Adorno-Platz 3 | 60323 Frankfurt am Main | Germany
Phone +49 69 798 30050 | Fax +49 69 798 30077 | info@ifk-cfs.de | www.ifk-cfs.de

The CFS Working Paper Series

presents ongoing research on selected topics in the fields of money, banking and finance. The papers are circulated to encourage discussion and comment. Any opinions expressed in CFS Working Papers are those of the author(s) and not of the CFS.

The Center for Financial Studies, located in Goethe University Frankfurt's House of Finance, conducts independent and internationally oriented research in important areas of Finance. It serves as a forum for dialogue between academia, policy-making institutions and the financial industry. It offers a platform for top-level fundamental research as well as applied research relevant for the financial sector in Europe. CFS is funded by the non-profit-organization Gesellschaft für Kapitalmarktforschung e.V. (GfK). Established in 1967 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia. GfK members comprise major players in Germany's financial industry. The funding institutions do not give prior review to CFS publications, nor do they necessarily share the views expressed therein.

Center for Financial Studies | Goethe University | Theodor-W.-Adorno-Platz 3 | 60323 Frankfurt am Main | Germany
Phone +49 69 798 30050 | Fax +49 69 798 30077 | info@ifk-cfs.de | www.ifk-cfs.de

Prices and consumer purchasing preferences at the border: evidence from a multi-country household scanner data set[1]

Guenter W. Beck[2] Hans-Helmut Kotz[3] Natalia Zabelina[4]

This draft: August 2016

Abstract

Studies employing micro price data to examine the extent of international goods market integration tend to find that borders induce arbitrage-impeding transaction costs which contribute to segment national markets. Analyzing household scanner price data from the three euro area countries Belgium, Germany and Netherlands, we document that Belgian households living in the vicinity of the border to Netherlands pay almost 10% more for the same good as their Dutch counterparts. German consumers on the other hand face prices that are on average up to around 3% smaller than those in the neighboring Netherlands. Counterfactual evidence for within-country price discontinuities provides no evidence of any existing border effects. The induced costs of crossing national borders amount to at least 13%. We also find evidence on border discontinuities in various household preference characteristics (such as demand elasticities and goods valuation) and household shopping patterns such as shopping frequencies.

JEL Classification: D12, D40, F40

Keywords: *Goods market integration, international price setting, border effects, price discrimination, demand elasticities, habit formation, scanner price data.*

[1]We are very grateful to Aimark and Europanel for providing us with the data used in this paper. We are in particular thankful to Alfred Dijs and Bernadette van Ewijk for their invaluable assistance in dealing with the data and their very useful suggestions. Guenter Beck furthermore would like to thank the Institute for Monetary and Financial Stability (IMFS) for its hospitality while part of this research was carried out. The usual disclaimer applies.

[2]University of Siegen and IMFS. Contact: guenter.beck@uni-siegen.de.

[3]Harvard University and SAFE PC. Contact: kotz@fas.harvard.edu.

[4]University of Siegen. Contact: natalia.zabelina@uni-siegen.de.

1 Introduction and literature overview

Starting with the seminal work by Engel and Rogers (1996) (ER), numerous attempts have been made to quantify the effect of borders on the integration of international goods markets by employing price data. The basic starting point of these studies is the law of one price (LOOP) which implies that differences in prices of identical goods between two locations are limited by the amount of transaction costs which consumers have to incur when purchasing a good.[1]

The more recent literature on the topic has progressed relatively to earlier studies basically with respect to two dimensions.[2] First, some studies have been concerned with refining and extending the econometric methods and identification strategies used to determine the border effect. A prominent example is Gorodnichenko and Tesar (2009), who argue that by neglecting to account for differences in within-country heterogeneity in the variability of prices between cities, ER considerably overestimate the effect of borders on deviations from the law of one price. Once accounting for this heterogeneity, the authors find that the border effect between the U.S. and Canada not only becomes significantly smaller but almost completely vanishes. Gopinath et al. (2011) (GGHL) - who also address the second, data-related issue discussed below - demonstrate that the identification strategy of ER rests on the assumption that transaction costs within countries are homogenous across households and markets. Doubting the general validity of this assumption, GGHL conduct an empirical, micro-price-data based study which is motivated by a spatial model of retailer price competition and which employs a regression discontinuity approach to identify the border effect. They find that the U.S.-Canadian border induces a discontinuous change of (at least) 24% in retail and wholesale prices.

Other more recent studies on international price dispersion advocate employing micro-price rather than price-index data since the underlying goods baskets are not identical across countries.[3] Consequently, nearly all more recent studies on the LOOP such as Broda and Weinstein (2008) or the above mentioned work by Gopinath et al.

[1] A classical reference on a study of the LOOP is Isard (1977), corresponding references to its version as pertaining to the general price level, i.e., the purchasing power parity (PPP), are Officer (1976), Mussa (1979) and Rogoff (1996).

[2] An overview of early studies on the importance of borders is given by Engel and Rogers (2004).

[3] The problem that can arise from working with prices of not identical goods baskets across countries is very clearly illustrated by Broda and Weinstein (2008). Using a large data set of retail prices these authors show that *i*) even narrowly defined categories (like "milk") contain many differentiated goods and *ii*) that neglecting to account for this heterogeneity can lead to estimation biases. A second study that shows the pitfalls of employing aggregate data and neglecting problems arising from aggregating micro data is Imbs, Mumtaz, et al. (2005). These authors demonstrate that the time series properties of LOOP deviations are heavily influenced by the level of aggregation one considers.

(2011) use price data at a highly disaggregated level, optimally at a level where goods can be identified by their GTIN (global trade item number).[4] While these studies have in common that they tend to find smaller border effects than that generally documented in earlier studies their overall findings are somewhat mixed. An extreme is given by Cavallo, Neiman, and Rigobon (2014) who, employing online prices of identical goods sold by four large global retailers in a large set of countries, show that the LOOP holds perfectly within currency unions but does not do so outside a currency union even if exchange rates are nominally fixed. As noted above, GGHL on the other hand find evidence in favor of sizeable border effects.[5]

Our paper extends the existing literature on intra- versus international price dispersion with respect to several dimensions. First, to the best of our knowledge, our multi-country price data set is unique in the literature on international price dispersion across European countries in the sense that it is characterized by a very high level of disaggregation at both the goods and geographical dimension.[6] This allows us, amongst others, to employ the approach by Gopinath et al. (2011) and provide credibly identified estimates of border effects for our European sample including Belgium (BE), Germany (GE) and the Netherlands (NL). These numbers are interesting since the involved countries represent a very interesting comparison group to the U.S.-Canadian case: our sample countries are not only members of a highly integrated economic area (the European Union) and share close historical and cultural ties (partly including a common language) but are also not impacted by nominal exchange rate dynamics which dominates short-run real exchange rate movements. Considering the variables which McAfee (2008) lists as factors preventing or limiting resale capabilities[7] we see that the establishment of the Common Market for Goods has largely eliminated those of them being classified as formal barriers to trade. Thus, any observed significant price discontinuity must be due to informal barriers to trade for which we therefore provide a quantitative assessment.

Secondly, unlike most other scanner-data-based studies in the field, we employ demand- rather than supply-side data. More specifically, our data set comprises

[4]The GTIN is a number that uniquely identifies a given trade item. In the literature, alternative terms for the GTIN such as the UPC (unique product code, USA) or the EAN (European article number, Europe) are very often used.

[5]Recent evidence on price differences at the micro level across European countries is provided by Imbs, Mumtay, et al. (2010), Reiff and Rumler (2014) and Kulikov (2014)

[6]Existing studies using similarly granular micro price data have focussed on the U.S. and Canada, whereas studies employing European data are confined to very few goods (see, e.g., Fischer, 2012 or Imbs, Mumtay, et al., 2010). In contrast, the other existing, micro-data based studies build on prices at a smaller disaggregation level and are, moreover, confined to very few locations (see, e.g., Engel and Rogers, 2004, Reiff and Rumler, 2014, Kulikov, 2014).

[7]These include legal impediments to resale, thin markets or matching problems, informational problems, and contracts and warranties or high transportation costs.

comprehensive information about purchases of around 40,000 households (at any given point in time). Since purchases also include cross-border shopping trips, we know the prices that households located on two different sides of a border actually paid for an identical good. This enables us to infer, whether two border-separated markets are integrated according to the definition by GGHL, i.e., whether significant cross-border shopping equalizing prices paid at the border occurs. Our demand side perspective might moreover provide for another advantage: if large, country-wide operating retailers follow a national rather than a local pricing strategy (as suggested by Gordon, Li, and Netzer, 2013), then prices at outlets close to the border would not necessarily respond to cross-border shopping by households but would represent averages of optimal prices across local markets within the respective countries. In this case, nationally active retailers might react to relatively lower sales in border regions by either "subsidizing" them (in the sense of Gordon, Li, and Netzer, 2013), reducing their size or shutting down the ones closest to the border (as documented for the U.S.-Canadian border by Campbell and Lapham, 2004). If such a phenomenon is present in our European data, a supply-side based analysis of price differences in the spirit of GGHL would tend to over-estimate border effects.[8]

Our third major contribution concerns a quantitative assessment of systematic differences in household preferences across our sample countries. When arbitrage or reselling of goods is impossible or possible only to a limited sense price differences across markets can result from price discrimination. For price discrimination to be feasible across markets in addition to missing/limited arbitrage opportunities two conditions must be satisfied: (i) first, firms must have some market power; and (ii) secondly, consumers must differ in their demands for a given good or service.[9] Employing the rich set of household purchasing information in our data sample we construct proxies for variables regularly used by retailers for discrimination purposes and examine whether households located on two different sides of a border systematically differ with respect to these measures. Moreover, we examine to which extent shopping behavior exploiting price advantages across retailers is distinct across countries.

Our results are as follows: in line with previous findings, we show that price dispersions for identical goods are sizeable across most of our country pairs including

[8]Referring to evidence by Campbell and Lapham (2004) who show that consumers in the U.S. and Canada indeed conduct cross-border shopping, GGHL argue that their finding of completely segmented U.S.-Canadian goods market is not inconsistent with this result by pointing out that the price setting decision by stores in their sample are obviously not significantly affected by these consumers.

[9]See Stole (2007); McAfee (2008) for an extended discussion on price discrimination under imperfect competition.

the seemingly very highly integrated country pair Belgium and Netherlands. In other words, we find significant border effects across most country pairs, the sizes of which differ substantially though. Border effects are largest for Belgium-Germany where prices discontinuously drop by around 20% when crossing the border from Belgium to Germany. For Belgium-Netherlands, this border effect amounts to around 13%. Interestingly, our results suggest only relatively moderate discontinuous price changes between the German-Dutch border. Overall, our estimates indicate larger border within-EMU border effects than those found by Cavallo, Neiman, and Rigobon (2014) for online prices but somewhat smaller values than those documented by GGHL for the U.S.-Canadian border. Our results concerning household goods preferences indicate significant differences across countries. Households located in countries where prices tend to be higher exhibit smaller demand elasticities and stronger valuation of purchased goods. Households' hopping patterns also differ across countries. Our results indicate that households with higher shopping costs and thus stronger tendencies for one-stop shopping tend to pay higher prices. However, results are somewhat mixed in this respect.

The rest of this paper is organized as follows. In Section 2, we describe our data set and provide some descriptive statistics. Section 3 motivates and describes our estimation approach whose results are presented and discussed in Section 4. Section 5 examines the households' purchasing behavior at the border and conducts a counterfactual analysis. Section 6 summarizes and concludes.

2 Data and descriptive statistics

2.1 Data

To conduct our study we employ a unique and very rich set of European scanner-price data which, to our best knowledge, has not been employed by the literature on international price dispersion so far. The data is provided by AiMark (Advanced International Marketing Knowledge) which is a non-profit cooperation that promotes research in the area of retail markets.[10] As will be outlined in more detail below, our data is best comparable to the IRI data used by Coibion, Gorodnichenko, and Hong (2015) and the household scanner data employed by Broda and Weinstein (2008) with the major difference that their data set covers the U.S. (Coibion, Gorodnichenko, and Hong, 2015) or the U.S. and Canada (Broda and Weinstein, 2008), whereas our

[10]The data which AiMark provides is originally compiled by Europanel and its partners Gesellschaft für Konsumforschung (GfK), Kantar Worldpanel and IRI to enable academic studies which aim at improving our understanding of the functioning of these markets. More detailed information on AiMark can be found at: `https://www.aimark.org`.

data comprises information from the three euro area countries Belgium, Germany and the Netherlands.

In each country included in the data set, the data provider (GfK) operates a panel of households (of differing sizes across countries) where panelists are chosen to constitute a representative sample of the consumers in the respective retail market. Each household is endowed with a scanning technology which it uses to scan all the products belonging to the categories of *fast-moving consumer goods* it buys at retail outlets including, e.g., all major supermarket chains (such as Rewe or Aldi in Germany or Albert Heijn and C1000 in the Netherlands), drugstores, small corner shops and also internet stores. This scanning technology is similar to the one underlying the Nielsen HomeScan database. For each product bought the household scans the barcode which uniquely identifies the product via the Global Trade Item Number (GTIN)[11] and enters the volume and the price it paid for this product.[12] We have also available a description of each product and a classification system of the goods into different (more aggregate) product categories which are internationally comparable. The household also provides a description of the type of store where it bought the product - including the name of the retail chain. As mentioned above, the products generally belong to the categories of fast-moving consumer goods, which include (among others) grocery products, home and personal care products and beverages.

For all three countries included, our data covers the period January 2005 to December 2008. As Table 1, employing information for the forth quarter of 2008 as an example, shows, the number of available observations (purchases) is enormous: It ranges from somewhat more than one million (per quarter) for Belgium to more than 6 millions for Germany. These observations include the purchases of more than 80 000 unique products in Belgium and the Netherlands and more than 150 000 unique products in Germany. The (average) number of households included in the respective country panels corresponds to around 4 000 for Belgium, 20 000 for Germany and 6 500 for the Netherlands. Additionally, we have information on household characteristics which comprises the location of the household, the income group it belongs to as well as its age group.

[11]The GTIN-12 code corresponds to the Universal Product Code (UPC) which is used in the U.S. and Canada. In Europe, the GTIN was formerly known as European Article Number (EAN).

[12]In case the product does not have a barcode, the household enters this information manually.

2.2 Matching results for common goods across countries

The nature of our data set allows us to make inference on the size of border costs based on gaps in prices of identical goods purchased by households located on different sides of a given national frontier. Following the literature we use as our basic measure of price dispersion, denoted by q_i^{hk}, the percentage difference between the price that a household located in location h and a household located in location k pays for the same good i, i.e., our measure of price dispersion across markets h and k is given by[13]

$$q_i^{hk} = \ln\left(p_i^h\right) - \ln\left(p_i^k\right), \tag{1}$$

where all prices are denominated in euro.

We declare a purchased product to be identical across two locations (that are located either within a country or across two countries) if the purchased good has the same GTIN/EAN.[14] The lower panel of Table 1 shows that while the number of matched goods is generally fairly sizeable (reaching from around 14 800 goods for Belgium-Netherlands to almost 4 500 for Belgium-Germany) these goods nevertheless represent only a relatively small share of all goods purchased by our sample households. This observation is in line with Broda and Weinstein (2008) who report similar results for their U.S.-Canadian data sample and who conclude that most price comparisons of consumption baskets across countries (i.e. purchasing power parity analyses) suffer from a compositional bias effect because they ignore this fact. The number of commonly sold goods is comparable to the figures reported by Broda and Weinstein (2008) who also use demand-side data and smaller by a factor of around four compared to GGHL who employ supply-side data.

The figures reported in Tables 2 and 4 show that we have a rather comprehensive, fairly balanced and sufficiently deep coverage of the available goods categories. The reported numbers of goods included in each subcategory confirm the findings by Broda and Weinstein (2008) that even narrowly defined product categories contain a substantive amount of different products which might lead to the potentially erroneous conclusion of large deviations from the LOOP, simply as an impact of purely compositional effects.

[13] Please note that in line with Broda and Weinstein (2008) who also use household scanner data we set the location of the household who buys the good equal to the location of the market we consider.

[14] In the following, the terms GTIN and EAN will be used interchangeably where most often the shorter form EAN will be used.

2.3 Descriptive evidence on within- and cross-country price dispersion

Tables 5 to 7 show that there exist considerable differences between within- and cross-country price differences. When computing differences by first aggregating (averaging) the prices of a good observed in a given sample period (2005Q1 or 2008Q4) at the regional level and then computing price gaps for each regional pair (upper panels of Tables 5 and 6) we observe that both mean and median price gaps are very close to zero for all within-country combinations. The corresponding cross-country figures are considerably larger corresponding to around 4% for Belgium-Netherlands, around 3% for Germany-Netherlands and around 8% for Belgium-Germany (median value) in 2005Q1 and to around 6.5% for Belgium-Netherlands, around 3% for Germany-Netherlands and around 8% for Belgium-Germany in 2008Q4. When absolute values of price gaps are considered (lower panels of Tables 5 and 6) a similar picture emerges. Median values for within-country price differences range from 0% to around 5%, whereas cross-country figures are at least three times larger and move around 15% (and even higher for BE-GE).

A graphical illustration of the reported price differences is provided by Figure 1. They show that mean/median values of international price gaps have been fairly stable for Germany-Netherlands but have increased in recent years somewhat for Belgium-Netherlands. Moreover, they clearly illustrate that there exists a considerable dispersion in observed price gaps across goods. Interestingly, this is true for all considered country pairs, including those for which the average of the observed mean values are relatively small. Absolute price differences of 20% or more are by far no exception.

A last interesting observation results from comparing the extent of with-country dispersion in prices across countries. Gorodnichenko and Tesar (2009) show that neglecting the considerable differences in within-country price dispersion between the U.S. and Canada can lead to erroneous conclusions with respect to the importance of borders for price dispersion when employing the ER approach. Our obtained numbers show that also in our sample the within-country price dispersion numbers are differ across countries (ranging from somewhat more than 1% for Belgium to somewhat more than 5% for the Netherlands) but are considerably smaller than those observed U.S.-Canadian data samples.

3 Estimation approach

3.1 Theoretical considerations

The motivation underlying our empirical approach as well as the interpretation of the obtained results rest on the following theoretical considerations:[15] we consider a two-country, multi-market setup where local retailers are located on both sides of the border which separates the two countries. We assume that both countries are inhabited by a large number of households, evenly spread across space in both countries. These households not only differ with respect to their locations, but also with respect to other characteristics such as income, age, preferences, ... which influence their transaction costs of buying a good at a specific retailer. When deciding from whom to buy a particular good, households take into account the price at which the good is offered and the transaction costs they incur from doing so. Transaction costs arise from travelling but are also caused by other factors such as information collection. We assume these transaction costs to be household specific.

Following GGHL and other authors, we assume transaction costs to depend on the distance that the household has to travel to a given retailer and whether it has to cross a border. Denoting the overall transaction costs by λ and the retailer from which household, h, considers to buy a given good by A we have $\lambda_A^h = \tau_A^h + bI_A^h$, where τ_A^h denotes the transaction costs the household faces as a consequence of the distance between its residence and the retailer A. I_A^h is an indicator variable being equal to one if the household has to cross a border when buying from A and zero otherwise. The parameter b therefore denotes additional transaction costs caused by the border.

The number of consumers - and thus its sales - which a given retailer attracts depends on the price it charges for a given good. Given the existence of transaction costs, it can make profits even if it offers goods at a price above that of neighboring retailers. The reason is that due to the distance-related transaction costs, households living very close to a particular retailer will very likely decide to buy from this supplier. Likewise, households living very far away will not buy from this retailer. In between these two cases, there will be some households that are just indifferent between buying from the considered retailer or its neighbor. The number of customers that decide to buy a good from a particular retailer represent the market for this retailer.

For illustrational purposes, let us now consider the situation of two retailers,

[15] Our framework largely rests on the ideas presented by GGHL. However, it modifies their considerations where our different data dimension (demand-side rather than supply-side data) implies deviating conclusions.

denoted by A and B, located in one of the two countries, denoted as the reference (domestic) country, and two retailers, denoted by A^* and B^*, that are located in the other, counterpart (foreign) country. We assume that retailers A and A^* are located next to the border, whereas retailers B and B^* are located somewhat further away. In line with GGHL, we differentiate between two scenarios: markets are either internationally integrated or segmented. This might imply differing price patterns to prevail which are discussed in the following.

Following GGHL, we define two markets to be integrated if *"equilibrium prices in these two markets are such that at least one consumer h in one of the markets is indifferent between buying in the market she lives in or paying [the transaction costs] to buy in the other market"*. A graphical illustration of the situation, which will result under this assumption, is given in the upper panel of Figure 3 where we assumed for simplicity that the distance between retailers is symmetric. We furthermore assumed that retailer A^* charges a lower price than retailer A (0.85 rather than 1) but that retailer B^* charges a higher price than A^* (1 rather than 0.85) and B charges a lower price than A (0.85 rather than 1). To understand the situation at the border (bold black line), consider households located between (the expensive) retailer A and (the cheaper) retailer A^*. It is obvious that all households located in the counterpart country will decide to buy from A^*, i.e., household will pay (and report) p_{A^*}. Concurrently, households located in the reference country and living close to retailer A will buy at price p_A. However, the closer we move to the border the more likely it is that for some household, h, $p_{A^*} + \lambda^h_{A^*} = p_A + \lambda^h_A \Leftrightarrow p_{A^*} + \tau^h_{A^*} + b = p_A + \tau^h_A$ holds, thus making this household indifferent between buying at A and A^*. Households with similar transaction costs but smaller distance to the border will then no longer buy from A but cross the border and buy the good at price p_{A^*} in the neighboring country. The closer we come to the border, the more households will do so and thus will report p_{A^*}. This is captured by the decreasing thickness of the line at the y value of 1 (the price of retailer A) and the continuously increasing thickness of the line at 0.85, the price of retailer A^*. As a consequence, no discontinuity will be observed in prices reported by households living close to each side of the border.

In other words, if we find that prices paid by households living in different countries, but (very) close to the respective border are equal we can conclude that cross-country retails markets are integrated and cross-border shopping takes place. However, a precise determination of the costs of crossing the border, i.e., an estimate of b is not possible in this case. Given that under these circumstances $p_{A^*} + \tau^h_{A^*} + b < p_A + \tau^h_A$ holds, the price difference between p_A and p_{A^*} represents an upper bound on the border costs provided that we get $b < p_A - p_{A^*}$ if we let

$\tau_A^h \to \tau_{A^*}^h$, i.e., if we move closer and closer to the border.

To wrap up, if we observe prices paid by households living on two sides of a border are very similar we can conclude that these markets are integrated and that the transaction costs generated by the border are sufficiently small not to prevent cross-border shopping from taking place.

The alternative case is that markets are segmented. Again following GGHL, we define two markets as segmented when the transaction costs of crossing them *"are large enough relative to the price gap between the two markets such that all consumers in either market are better off purchasing the good in the market where they live."* Making the same assumptions about the market structure and households as for the integrated case a graphical illustration of the resulting situation is depicted in the lower panel of Figure 3. As a consequence of the prohibitively high transaction costs, no cross-border shopping occurs: thus, a discontinuity in prices reported by households living on the two sides of the border results. These price differences in turn reflect the prices retailers charge on either side of the border. Considering the situation of a household located close to the border in the reference country we have: $p_{A^*} + \tau_{A^*}^h + b > p_A + \tau_A^h$. Letting again $\tau_A^h \to \tau_{A^*}^h$ yields $b > p_A - p_{A^*}$, i.e., the reported price discontinuity by households at the border would represents a lower bound on the costs of crossing the border.

Summarizing, the two cases of integrated and segmented markets differ considerably as a consequence of the shopping behavior of household living close to the border: in the case of integrated markets we should observe an equalization of prices households pay on both sides of the border whereas in the case of segmented markets we would observe price differences, providing a lower bound on the border costs.

3.2 Estimation approach

Observing the purchasing behavior of households living on both sides of a given border the theoretical considerations from the previous subsection suggest that if we let the limit of the households' distances to the border go to zero we should either find an equalization of paid prices observed (integrated markets) or a discontinuous gap between them (segmented markets). Employing this intuition we follow GGHL and conduct a regression discontinuity (RD) analysis to empirically evaluate the importance of borders for goods markets segmentation across our sample countries.

RD designs are used to evaluate treatment/causal effects in situations in which the assignment to the treatment is (at least partly) determined by the value of an observed covariate, also denoted as forcing variable.[16] More specifically, RD

[16]Imbens and Lemieux (2008); Lee and Lemieux (2010); Skovron and Titiunik (2015) provide for

approaches exploit discontinuities in the policy assignment at a given threshold in the sense that - under a so-called sharp regression discontinuity design which is considered in this paper - no treatment occurs for values of the forcing variable below the threshold and treatment occurs for value equal or above the threshold. In most applications, the determination of the threshold (and thus the assignment of the treatment) results from administrative decisions. Under the identifying assumption that all unobserved determinants on the outcome variable are continuously related to the forcing variable the difference in the average outcomes just above and below the threshold provides an estimate of the treatment effect.

GGHL apply the reasoning underlying RD designs to the situation of price-setters located on either side of the U.S.-Canadian border Following GGHL, we interpret the treatment as being located in a different country (compared to the respectively chosen reference country). The forcing variable is given by the distance of households' residence to the border where distances are measured by positive values for one side of the border and by negative values for the other side of the border. The threshold value chosen for selecting households to a treatment corresponds to a value of zero for the distance to the border. Figure 4 provides for an graphical illustration of this approach by potting the price of an exemplary good in the vicinity of the considered borders.

The application of the RD approach provides for three major benefits when it comes to estimating the effects of borders on transaction costs. Firstly, it allows to overcome problems associated with estimating border effects in the presence of within-country heterogeneities in price dispersion.[17] Secondly, as GGHL argue, regional prices very likely depend on a variety of not observable factors such as demographic factors or elasticities across retailers. Given the validity of the assumptions underlying the RD approach, it can control for the effects of these unobservable variables. Thirdly, the approach allows to directly observe whether markets are integrated or segmented, given the above outlined theoretical considerations. Moreover, it provides for a direct measure of transaction costs given the price data used.

recent overviews of the literature on RD designs.

[17] Gorodnichenko and Tesar (2009) show that there are potential identification problems in the approach to estimate border effects which was first employed by Engel and Rogers (1996) and which has been used by most subsequent studies in these fields. More specifically, Gorodnichenko and Tesar (2009) argue that by neglecting to account for differences in within-country heterogeneity in the variability of prices between cities, ER considerably overestimate the effect of borders on deviations from the law of one price. Once accounting for this heterogeneity, the authors find that the border effect between the U.S. and Canada not only becomes significantly smaller but almost completely vanishes.

Formally, to estimate border effects the following regression equation is run:

$$\ln p_i^h = \alpha_i + \beta_i B_i + \theta_i D^h + \phi_i D^h B_i + \varepsilon_i^h, \tag{2}$$

where $\ln p_i^h$ denotes the (log) price of good i paid by household h, α_i is a good-specific dummy variable, X^h captures household h specific variables (income and age) and ε_i^h denotes household and goods-specific unobserved characteristics. B_i is a dummy variable equal to one if the household is located in the country which is defined as the counterpart (foreign) country in a given regression. D^h denotes the distance of household h to the border with D^h being positive for the reference (domestic) country and being negative for the other (foreign) country. Taking into account the fact that the functional relationship between distance and price might differ across the two countries, we allow the distance coefficient to change its size at the threshold point by including an interaction term between distance and the border dummy.

The coefficient of interest is β_i. If some of the unobserved characteristics contained in ε_i^h are not independent from the location of household i, i.e., if $E\left[\varepsilon_i^h | B_i\right] \neq 0$, an OLS estimate of the border coefficient, β_i, obtained without the inclusion of the distance variable would be biased. However, if the unobservable covariates become more "similar" the closer we get to the border, i.e., if they are a continuous function of the distance of households to the border, then including the distance from the border as an additional regressor allows to control for the effects of these variables by introducing distance from the border.[18]

Unfortunately, a formal test of the identification assumption is not possible. However, considering various observable features of our data sample we can provide for some indirect graphical evidence on the plausibility of its validity.

Figures 5 to 7 illustrate that the locations of our sample households (as of 2008Q4, similar pictures emerge for the other sample periods) are - with the exception of some fairly thinly populated rural and some very densely populated metropolitan areas - generally evenly spread within and across the countries including the border regions.[19] The resulting high availability of observations in the vicinity of the threshold, i.e., the border favors the applicability of the RD approach.

Figure 8 shows that the density of our households is subject to some fluctuations when plotted against the distance from the border. These generally reflect variations in the ratio of rural/metropolitan areas considered at a given distance. There are also

[18] Formally, the identifying assumption is given by:

$$\lim_{\epsilon \uparrow 0} E\left[\varepsilon_i^h | B_i\right] = \lim_{\epsilon \downarrow 0} E\left[\varepsilon_i^h | B_i\right].$$

[19] In the Belgian-German case, regions on both sides of the border are very thin.

some differences in population density close to the border. However, the degree of differences is relatively moderate. Moreoever, given that densities tend to be higher on the high-price side of a given border we don't consider the observed differences as indications of a potentially "strategic behavior" of individuals aiming at receiving a positive treatment by moving to the lower-price side of a given border.

The discontinuity plots of the distributions of age and income of our sample households in the proximity of the border - presented in Figures 9 to 11 - generally show no indications of major discontinuities at the border. In this context, it is to be noted that the selection of households is not intended to create comparable samples across countries but to generate representatives samples for a given country. In sum, we take the missing evidence of discontinuities in observed variables as a confirmation for the validity of the RD approach.

To estimate the border effect we use a local linear regression which - using a rectangular kernel (see Lee and Lemieux, 2010) - amounts to estimating a standard linear regression over a band of width h on both sides of the border. Lee and Lemieux (2010) emphasize that the choice of the bandwidth in general involves finding an - optimal - balance between precision and bias. To choose the bandwidth we apply data-driven bandwidth selection methods as outlined and implemented by Calonico et al. (2016). Given the relatively widely and tightly spread locations of the households in all our sample countries, the bandwidths chosen are fairly small, ranging from 20 to 60 km. The results reported are based on a bandwith of 60 km.[20] Moreover, to ensure a minimum level of estimation precision we only run regressions for goods for which we have available at least 10 observations on either sides of the border.

4 Price gaps at the border

The regression results presented in Table 8 indicate significant and partly sizeable discontinuities of most prices at the border in all considered country pairs. For Belgium-Netherlands, around 76% of all border coefficients are significant for 20005Q1 and around 80% for 2008Q2. For Germany-Netherlands, the corresponding numbers are around 60% and 58%. Price gaps are fairly large for Belgium-Netherlands indicating 6% (9%) higher prices (median values) in Belgium relative to the Netherlands. Numbers are smaller (in absolute terms) for Germany-Netherlands. However, in this case price gaps have experienced a interesting dynamics within our sample period changing from around positive 0% (median) in 2005Q1 to around negative

[20] In the appendix, we also report results for the respectively chosen optimal bandwidths. The obtained numbers are qualitatively and quantitatively similar.

3% (median) in 2008Q2. In other words, Dutch households tended to pay the same prices in 2005 but paid 3% higher prices four years later. For Belgium-Germany, the results are even more pronounced in this respect: Belgian households paid around around the same as their German counterparts in 2005Q1 but faced around 9% higher prices in 2008Q4 (all median values). In light of the small number of goods included (particularly for 2005) together with the comparably small proportion of significant estimates the Belgian-German numbers have to be treated with some caution though.

Together with the more broadly based, but with respect to their dynamics similar results for Germany-Netherlands our findings imply that between the considered countries a noteworthy change in price gaps occurred. This is somewhat remarkable given that normally within an international environment the dynamics of relatively large (9% in the case of BE-GE) short- to medium-term LOOP deviations are driven by nominal exchange rate movements which are absent within the euro area. Figure 12 shows that the results for the two discussed sample periods very well fit into the pattern obtained for the overall sample of each country pair. For BE-NL we can observe a continuous increase in median values from around 6% to 9%. The results for the two German related country pairs indicate a steady increase in border effects with respect to this country. Taken together this finding suggests that German prices have declined with respect to both Belgium and the Netherlands whereas Belgian prices have increased with respect to both other countries. Results for estimated mean values basically exhibit the similar pattern.

Considering absolute values of estimated price gaps (Table 9) we see that both obtained mean and median values are fairly similar across county pairs: median values range from around 13% for BE-NL (2005Q1) to somewhat more than 15% for BE-NL (2008Q4). Over time, the results remained more or less stable. Based on the theoretical considerations outlined above, these numbers suggest that border costs are at least 14% for GE-NL and BE-GE and more than 15% for BE-NL.[21]

The standard deviations of the reported estimates indicate considerable dispersion of border effects across goods. This result is nicely illustrated in plots of the kernel density functions of the estimates obtained (Figure 14). Whilst mean/median values

[21]Including covariates, i.e., controlling for age and income of households has no major effects on the results for the border coefficients (see (Table 18, Table 19). The age groups are defined as follows: Households are classified as young if their age is lower than 34 years, medium-aged households are those between 35 and 64 years and old households are those older than 65 years. The income groups are defined as follows: Low-income households are those with an income of less than 1240,- (BE)/1249,- (GE)/1300,- (NL). High-income households are those with an income higher than 2726,- (BE)/2750,- (DE)/2700,- (NL). Medium-income households are those with income between low- and high-income households.

have partly changed strongly over time the dispersions have remained fairly stable (at a high level).

To conclude, our estimates imply border costs in the range of 14% to 15%. As outlined in our theoretical considerations, these numbers represent lower limits. Moreover, these numbers are larger than those one might have expected in light of the findings by Cavallo, Neiman, and Rigobon (2014) who document that within the euro area basically no discrimination in online prices occurs. Comparing our results to those found by GGHL for the U.S.-Canadian border we see that they are smaller by around 1/3. However, in light of the very deep (and partly very long (BE-NL)) history of integration between the considered countries including the existence of a monetary union and considering the fact that some of the border regions share the same language, the documented border costs of around 14% appear to be remarkably high.

Counterfactual evidence: border effects across within-country regions

Similarly to GGHL, we conduct "counterfactual experiments" to examine to which extent the documented price discontinuities at national borders also occur across states/regions within a country. To this end, we select for each country two comparable (and not too small) regions and estimate whether a price discontinuity can be observed across them. Given the relatively frequent political turmoils occurring between them, for Belgium the Flemisch-Wallonian regions represent excellent Belgian candidates for this exercise. Another, related reason why this region pair is particularly well suited results from the fact that different languages are spoken in both regions. For Germany, we have chosen the state pair Lower-Saxony and North-Rhine Westphalia which are amongst the largest states in Germany (in terms of geographical size) and which are located at the Dutch and Belgian (North-Rhine Westphalia) border. For the Netherlands, we haven chosen the two NUTS 1 regions located next to Germany.

Table 10 shows, that cross-region price differences are very moderate for all region pairs in all considered sample periods. Both mean and median values of actual price gaps are are always close to zero. Absolute values are somewhat higher (in the range of 1% to 4% (median values)) but nevertheless around 80% lower than cross-country price differences. The kernel density plots (Figure 21) confirm both the more or zero mean/median values and the very small variation across these values.

Not surprisingly therefore, the RD regression results basically indicate no border effects. This is true for both actual (Table 11) and absolute (Table 12) values. The quasi non-existing or very tiny distribution of estimates is illustrated in Figure 23.

Overall, the results from the counterfactual experiments confirm that national borders provide a significant impediment to integration across the considered goods markets.

5 Borders and household shopping preferences

In the previous section, we documented the existence of partly sizeable price gaps across European retail markets which represent minimum estimates of transaction costs according to our theoretical considerations. The existence of these costs is a necessary, however not sufficient condition for the observed price differences used to derive them. For prices to differ across markets at least one out of several additional potential reasons must be given. One obvious factor which can cause a price wedge between markets are distinct marginal costs resulting, e.g., from differences in local labor or rental costs.[22] A similar effect can arise from distinct goods' taxes such as excise or value-added taxes.

A second important explanation traces price differences back to price discrimination. For price discrimination to be feasible across markets two conditions must be satisfied (in addition to arbitrage-preventing forces):[23] (i) first, firms must have some market power; and (ii) secondly, consumers must differ in their demands for a given good or service. The Belgian, Dutch and German national and regional concentration measures presented in European System of Central Banks (2011, Section 1.3) indicate some degree of market concentration in these countries' retail markets suggesting some pricing power of the existing firms. Concerning (ii), Armstrong (2006) provides a good overview of how differences between consumers with respect to their preference and shopping characteristics such as their valuation of a given good, their preference in favor of a given good and their "choosiness", i.e., the degree to which they dislike buying their preferred good, can be employed to discriminate between consumers under both monopolistic and oligopolistic price-setting. McAfee (2008) points out that firms attempt to trace back these differences in tastes to observable characteristics of consumers and as a consequence segment markets by characteristics such as geography, nationality, age or purchasing history. Fortunately, the specifities of our data set allow us to provide some suggestive evidence on these aspects. Given our detailed information on the location and purchasing behavior of Belgian, Dutch and German consumers we can construct a variety of measures approximating the characteristics usually employed by retailers to price discriminate.

[22] Of course, if there are no barriers for arbitrage prices can equalize despite differences in marginal costs if the former force suppliers of the good to lower their price-cost margins in the high-cost market.

[23] See McAfee (2008).

To examine the relationship between borders and household shopping preferences potentially giving rise to price discrimination we construct variables which provide information either at the household-goods level or at the more aggregated household level. Employing this data we run - depending on the aggregation level - either of the following modifications of our regression equation (2):

$$Y_i^h = \alpha_i + \beta_i B_i + \theta_i D^h + \phi_i D^h B_i + \gamma_i X_i^h + \varepsilon_i^h, \tag{3}$$

when the dependent variable is available at the household-goods level or

$$Y^h = \alpha_h + \beta_h B + \theta_i D^h + \phi_h D^h B + \gamma_h X^h + \varepsilon^h, \tag{4}$$

when the dependent variable is available at the household level only. The variables D^h and B_i are defined as above whereas B in the second equation is analogously defined to B_i.

The term Y_i^h in the first of these equation represents household-goods-specific variables which are related to the willingness of a given household to buy good i and thus can - if known - in principle be used by retailers to discriminate between consumers/consumer groups. Since the factors directly determining the preferences of consumers towards a given good are unobservable we rely in our analysis on proxies for those factors which are in many cases also employed by retailers for differentiating their consumers base. These variables are presented and discussed in Section 5.1.

Y^h in Equation (4) intends to capture household-specific purchasing patterns that reflect households' tendencies to exploit price differences across retailers. These variables are presented and discussed in Section 5.2.

5.1 Household-goods-level related preference characteristics

The availability of fairly comprehensive and detailed purchase data allows us to construct some measures at the household-goods level at which also paid prices are available. These include - for a sizeable subsample of goods - an estimate of the demand elasticity and proxies for the strength of preference for a given good by households.

Consumer demand elasticity

Since our data sample comprises information on the price of a good and the quantity bought at this price we are able to estimate demand elasticities which play an eminent role in pricing under imperfect competition. However, this possibility is constrained to a subsample of available goods only. This is due to the fact that

a reliable estimation of elasticities imposes restrictions on the minimum numbers of observations available per good in the given data sample. To estimate demand elasticities we employ a standard nested multinomial mixed logit model similar to Gordon, Goldfarb, and Li (2013). The upper nest contains the categories whereas the lower nest represents the various EAN-level goods comprising the respective category. Details of the conducted data pruning and the estimation approach are provided in Appendix A.

For BE-NL, our results (left panel of Table 13) suggest that demand elasticities are around 0.15% (median) lower for Belgian households compared to their Dutch counterparts. Under non-perfectly competitive market structures (as, e.g., in the case of the very widely used model of monopolistic competition) lower elasticities are generally associated with larger price markups such that our results would imply Belgian households to face higher prices than their Dutch counterparts. Above, we saw that this is exactly what we observe. The number of significant values is smaller than for prices but with around 30% fairly sizeable. Moreover, results are stable over time.

The German-Dutch data are due to the above mentioned data constraints only available for 2008Q4. The median value suggests that German households are more price-elastic, a finding which is again consistent with the results from the previous section indicating significant lower prices in Germany compared to the Netherlands. However, it is to be noted that the sample size is fairly. Due to data constraints no results for BE-GE could be obtained.

Consumer preference for the good

One major aspect that many firms consider as an elementary input factor for their pricing decision, in particular if they follow a value-based pricing strategy (see Forbis and Mehta, 1981), is the economic value that customer ascribe to the good. To proxy for a consumer's valuation of or preference for a good we take two different approaches both of which aim at gauging the importance of a good for the customer. First, we compute a variable indicating how regularly the household purchased a given good in the respectively considered sample period (PurchFreq). Secondly, we construct a habit-formation/state dependence variable indicating whether the given good was purchased by the household in the preceding period (StateDep).

The results for the Belgian-Dutch sample in Table 13 indicate that Belgian households tend to buy a given good more often than Dutch ones. This finding would imply Belgian prices to be relatively higher if firms in both countries employ customer valuation (as measured by the purchasing frequency) as an input factor for their

pricing decision. This implication is in line with the empirical observations obtained in Section 5.1. It is to be noted though that the proportion of significant values is somewhat smaller than in the case of demand elasticities. Similar conclusions apply to the German-Dutch sample. Germans tend to purchase a given good less frequently, a finding consistent with lower prices paid in Germany when value-based pricing plays a role. For Belgian-German households the empirical results are also in line with the documented cross-country price patterns, even though the it is to be noted that the number of available goods is fairly small.

Our findings for the occurrence of state dependence in demand on both sides of a given border reveal that Belgian household exhibit only fairly small differences in their degree of state dependence compared to Dutch households. Germans on the other hand always exhibit a lower value of this variable, a finding consistent with the lower prices they tend to be for the same good. The share of significant coefficients is of comparable size to that of the purchasing frequency. For the BE-GE country pair, we find higher degrees of goods preferences for Belgian households which also tend to pay higher prices.

5.2 Household-level related preference characteristics

In this subsection we provide between-country evidence on purchasing habits of households influencing the prices they pay. More specifically, we provide for proxies for one-stop vs. multi-stop shopping behavior.

In the consumer research literature, two extreme types of consumers with respect to their shopping frequency play an important role. On the one side, there exist so-called "bargain hunters" or "cherry pickers", i.e., households that tend to go from store to store to pick the best-priced items and leave the rest and "one-stop shoppers" who tend to make all of their purchases in one store. In between these two cases, so-called multi-stop shoppers are located. The frequency at which households shop might also be linked to the price they pay for a given good. While households shopping very frequently or at several retailers are more likely to pick lower prices than shoppers who decide to conduct their planned purchase at one time, probably to save transaction costs.

Employing the available purchase records we construct four variables to proxy for the extent to which households engage in multi-shop/one-stop shopping. First, we calculate the average purchase amount of each household per shopping trip (PurchAmount). Secondly, we provide for a measure of the average number of goods bought at a shopping trip (PurchItems). Thirdly, we construct a variable capturing the number of shopping trips per period (PurchFreq). Finally, we measure the share

of purchases of the most favorite retailer per household. All four of these measures can be related to the phenomenon of multi-stop/one-stop-shopping to differing degrees. So, the average purchase amount can be expected to be larger for one-stop shoppers than for multi-stop shoppers. The same applies for the average number of goods bought at a shopping trip. On the other hand, purchasing frequency should be high for multi-stop shoppers and is likely to be lower for one-stop shoppers. Lastly, the share of sales of the respectively most favorite retailers should be positively related to a tendency for one-stop shopping.

The results presented in Tables 15 and 17 exhibit consistent patterns across GE-NL and BE-GE households whereas the picture emerging for BE-NL is somewhat mixed. For the GE-NL country pair our results indicate that German households on average purchase smaller volumes and smaller baskets per trip, shop more often and less frequent at their most favorite retailer than their Dutch counterparts. Overall, this indicates that German households likely exhibit smaller shopping costs which enables them to pick smaller prices. This implication is coherent with the findings of a negative border gap between Germany and the Netherlands.

Likewise, Belgian households on average purchase larger volumes and larger baskets per trip, shop less often and more frequent at their most favorite retailer than their German counterparts indicating that these households have higher shopping costs. As a consequence, they enjoy less frequently lower prices, a finding consistent with positive border gap between Belgium and Germany.

Concerning BE-NL, we find that Belgian households buy larger volumes than their Dutch counterparts indicating larger shopping costs and implying a larger tendency to one-stop shopping for the former. The remaining three variables indicate the opposite though. The numbers of good purchased by a Belgian household at an average shopping trip is on average smaller by 3 items and purchase frequency tends to be somewhat higher. Similarly, the tendency to buy at the most favorite retailers is somewhat smaller. Overall, these latter three results would indicate that Belgian households tend to shop more often and thus should be able to pick lower prices on average. However, as we have seen above the opposite is the case. Given that retail prices a household faces also depend on the local market structure and taking into account that local retail market structures can vary considerably across regions[24] an explanation of this somewhat puzzling result requires a deeper analysis of local supply side structures which is unfortunately beyond the scope of this paper.

[24]See, e.g., the study by Hottman (2014) in this context.

6 Summary and conclusions

Employing a rich and unique set of EAN level price data this study has examined the importance of borders for goods market integration across the three fairly homogenous and seemingly very well integrated European countries Belgium, Germany and the Netherlands. To our knowledge, this is the first study of this issue for European countries employing micro price data set of this volume and particularly at this level of disaggregation.

Our empirical results show that across these three European countries average border estimates are indeed generally much smaller than suggested by early findings by Engel and Rogers (2004) and Crucini, Telmer, and Zachariadis (2005). They amount to around 14% to 15%. It is to be noted though that the theoretical considerations outlined in the paper imply that these numbers represent lower limits of the additional costs implied by the borders. Moreover, these numbers are larger than those one might have expected in light of the findings by Cavallo, Neiman, and Rigobon (2014) who document that within the euro area basically no discrimination in online prices occurs. Moreover, in light of the very deep (and partly long) history of integration between the considered countries including the existence of a common currency, the documented border costs of at least 14% are remarkably high, also when compared to the not so much higher value of 24% for the U.S. and Canada.

Constructing variables which proxy for factors that are regularly used by retailer for price discrimination purposes or reflect purchasing patterns of households we find that household shopping preferences and behavior exhibit marked discontinuities at borders. In line with our intuition goods-level specific household preferences differ by countries in such a manner that they are in line with the observed cross-country price patterns. When considering household purchasing preferences we similarly find strong indications of discontinuities. Again, a majority of observed patterns is consistent with observed price patterns, however, also some counter-intuitive findings emerge.

References

Armstrong, Mark (2006). "Recent developments in the economics of price discrimination". In: *Advances in Economics and Econometrics: Theory and Applications*. Ed. by Richard Blundell, Whitney K. Newey, and Torsten Persson. Cambridge University Press, pp. 97–141.

Broda, Christian and David E. Weinstein (Nov. 2008). "Understanding international price differences using barcode data". NBER Working Paper Series No. 14017.

Calonico, Sebastian et al. (2016). "rdrobust: Software for Regression Discontinuity Designs". Working paper, University of Michigan.

Campbell, Jeffrey R and Beverly Lapham (Sept. 2004). "Real exchange rate fluctuations and the dynamics of retail trade industries on the US-Canada border". In: *The American Economic Review* 94.4, pp. 1194–1206.

Cavallo, Alberto, Brent Neiman, and Roberto Rigobon (2014). "Currency unions, product introductions, and the real exchange rate". In: *The Quarterly Journal of Economics* 2.

Coibion, Olivier, Yuriy Gorodnichenko, and Gee Hee Hong (2015). "The Cyclicality of Sales, Regular and Effective Prices: Business Cycle and Policy Implications". In: *American Economic Review* 105.3, pp. 993–1029.

Crucini, Mario J., Chris I. Telmer, and Marios Zachariadis (2005). "Understanding European Real Exchange Rates". In: *American Economic Review* 95, pp. 724–738.

Engel, Charles and John H. Rogers (1996). "How Wide is the Border?" In: *American Economic Review* 86, pp. 1112–1125.

— (2004). "European product market integration after the euro". In: *Economic Policy* 19, pp. 347–384.

European System of Central Banks, Task Force of the Monetary Policy Committee of the European System of Central Banks (2011). "Structural Features of Distributive Trades and Their Impact on Prices in the Euro Area". ECB Occasional Paper Series No. 128.

Fischer, Christoph (2012). "Price convergence in the EMU? Evidence from micro data". In: *European Economic Review* 56, pp. 757–776.

Forbis, John L. and Nitin T. Mehta (1981). "Value-based strategies for industrial products". In: *Business horizons* 3, pp. 32–41.

Gopinath, Gita et al. (2011). "International prices, costs and markup differences". In: *American Economic Review* 101, pp. 2450–2486.

Gordon, Brett, Avi Goldfarb, and Yang Li (Feb. 2013). "Does price elasticity vary with economic growth? A cross-category analysis". In: *Journal of Marketing Research* 50.1, pp. 4–23.

Gordon, Brett, Yang Li, and Oded Netzer (2013). "An Empirical Study of National vs. Local Pricing under Multimarket Competition". Manuscript. Kellogg School of Management.

Gorodnichenko, Yuriy and Linda Tesar (2009). "Border effect or country effect? Seattle may not be so far from Vancouver after all". In: *American Economic Journal: Macroeconomics* 1.1, pp. 219–241.

Hottman, Colin (2014). “Retail Markups, Misallocation, and Store Variety in the US”. In: *Unpublished Manuscript*.

Imbens, Guido W and Thomas Lemieux (2008). “Regression discontinuity designs: A guide to practice”. In: *Journal of econometrics* 142.2, pp. 615–635.

Imbs, Jean, Haroon Mumtay, et al. (2010). “One TV, One Price?” In: *The Scandinavian Journal of Economics* 112, pp. 753–781.

Imbs, Jean, Haroon Mumtaz, et al. (2005). “PPP strikes back: Aggregation and the real exhange rate”. In: *Quarterly Journal of Economics* 120, pp. 1–44.

Isard, Peter (1977). “How Far Can we Push the Law of One Price?” In: *American Economic Review* 67, pp. 942–948.

Kulikov, Dmitry (2014). “Law of One Price in the euro area: an empirical investigation using Nielsen disaggregated price data”. Bank of Estonia Working Paper Series No. 10/2014.

Lee, David and Thomas Lemieux (2010). “Regression discontinuity designs in economics”. In: *Journal of Economic Literature* 48, pp. 281–355.

McAfee, R Preston (2008). “Price discrimination”. In: *Issues in Competition Law and Policy* 1, pp. 465–484.

Mussa, Michael (1979). “Empirical regularities in the behavior of exchange rates and theories of the foreign exchange market”. In: *Carnegie-Rochester Conference Series on Public Policy*. Vol. 11. Elsevier, pp. 9–57.

Officer, Lawrence H. (1976). “The Purchasing-Power-Parity Theory of Exchange Rates: A Review Article”. IMF Staff Papers.

Reiff, Adam and Fabio Rumler (Nov. 2014). “Within-and cross-country price dispersion in the euro area”. ECB Working Paper Series No. 1742.

Rogoff, Kenneth (1996). “The Purchasing Power Parity Puzzle”. In: *Journal of Economic Literature* 86, pp. 647–668.

Skovron, Christopher and Rocío Titiunik (2015). “A Practical Guide to Regression Discontinuity Designs in Political Science”. Working paper, University of Michigan.

Stole, Lars A. (2007). “Price discrimination and competition”. In: *Handbook of Industrial Organizatoin*. Ed. by Mark Armstrong and Robert H. Porter. Vol. 3. Elsevier, pp. 2221–2299.

7 Tables

Table 1: Summary information on the data sample

Country	Unique EANs	Purchases		Panelists	
BE	81 841	1 338 016		3 923	
GE	155 108	6 460 842		20 750	
NL	86 694	2 653 998		6 262	
Country pair	Matched EANs	Purchases		Panelists	
		1country	2country	1country	2country
BE-GE	4 477	49 322	443 391	3 593	20 208
BE-NL	14 809	273 237	1 037 444	3 844	6 253
GE-NL	6 426	728 010	84 417	20 479	6 055

Notes: (1) All numbers are reported for the forth quarter of 2008. Similar numbers apply to the other periods of the data sample. (2) The short name "BE" denotes Belgium, "GE" Germany and "NL" the Netherlands. (3) In the upper panel, the column "Unique EANs" reports the number of unique goods available where goods are identified by their GTIN (barcode). The column "Purchases" reports the number of transactions conducted by the households of the respective country. The number of panelists reported (column "Panelists") corresponds to the number of households included in the panel of the respective country. (4) In the lower panel, the column "Matched EANs" reports the number of goods (defined by the same GTIN) commonly purchased by households in the two countries indicated in the first column ("Country"). Columns 3 and 4 and 5 and 6 report the total number of purchases available for the matched EANs and the number of households which conducted these purchases in the first and second country of the respective country pair.

Table 2: Number of EANs and purchases by product category, BE - GE

Catergory	Matched EANs Freq.	%	BE: Purchases Freq.	%	GE: Purchases Freq.	%
AlcholfreeCO2	70	1.56	2673	5.42	15699	3.54
AlcholfreeNoCO2	82	1.83	877	1.78	4527	1.02
Alcohol	26	0.58	153	0.31	1697	0.38
Animalcare	314	7.01	3000	6.08	10348	2.33
Babyproducts	57	1.27	531	1.08	2373	0.54
Basicfood	123	2.75	1216	2.47	14480	3.27
Beer	3	0.07	10	0.02	433	0.10
Bodycare	610	13.63	3214	6.52	19205	4.33
Candy	542	12.11	6385	12.95	86734	19.56
Cereals	40	0.89	405	0.82	1984	0.45
Champagne	21	0.47	126	0.26	1028	0.23
DairyWhite	222	4.96	4443	9.01	43708	9.86
DairyYellow	177	3.95	3457	7.01	18500	4.17
Delicasees	151	3.37	1290	2.62	17378	3.92
Fatoils	19	0.42	74	0.15	18538	4.18
Frische	57	1.27	397	0.80	9354	2.11
Frozenproducts	143	3.19	3135	6.36	8671	1.96
HotDrinks	116	2.59	909	1.84	12871	2.90
HouseholdCleansers	111	2.48	955	1.94	5109	1.15
Hygieneproducts	104	2.32	547	1.11	7407	1.67
Laundry	73	1.63	464	0.94	4767	1.08
Liquor	92	2.05	1211	2.46	2712	0.61
Meat	139	3.10	544	1.10	22860	5.16
MouthTooth	52	1.16	288	0.58	2741	0.62
PreservedFood	141	3.15	1290	2.62	13298	3.00
Readymade	285	6.37	1794	3.64	28372	6.40
Rest	67	1.50	1125	2.28	4503	1.02
Snacks	330	7.37	3672	7.44	25830	5.83
Spreads	40	0.89	1024	2.08	5796	1.31
Vegetables	206	4.60	3783	7.67	31299	7.06
Wine	64	1.43	330	0.67	1169	0.26
Total	4477	100	49322	100	443391	100

Notes: (1) All numbers are reported for the forth quarter of 2008. Similar numbers apply to the other sample periods. (2) The short name "BE" denotes Belgium and "GE" Germany. (3) The column "Matched EANs" reports the number of goods (defined by the same GTIN) commonly purchased by households in the two countries indicated in the first column ("Country"). Columns 3 and 4 and 5 and 6 report the total and relative number of purchases available for the matched EANs in the first and second country of the considered country pair.

Table 3: Number of EANs and purchases by product category, BE - NL

Catergory	Matched EANs Freq.	%	BE: Purchases Freq.	%	NL: Purchases Freq.	%
AlcholfreeCO2	54	0.36	2405	0.88	4693	0.45
AlcholfreeNoCO2	111	0.75	1941	0.71	8283	0.80
Animalcare	423	2.86	7729	2.83	13346	1.29
Babyproducts	218	1.47	2046	0.75	2633	0.25
Basicfood	296	2.00	10592	3.88	27148	2.62
Beer	102	0.69	3481	1.27	2681	0.26
Bodycare	1029	6.95	7553	2.76	12633	1.22
Candy	1066	7.20	17415	6.37	43329	4.18
Cereals	72	0.49	1818	0.67	5826	0.56
Champagne	28	0.19	233	0.09	402	0.04
DairyWhite	710	4.79	23245	8.51	160615	15.48
DairyYellow	48	0.32	2824	1.03	6987	0.67
Delicasees	506	3.42	9299	3.40	33296	3.21
Fatoils	150	1.01	3807	1.39	39921	3.85
Frische	486	3.28	10653	3.90	56577	5.45
Frozenproducts	531	3.59	11148	4.08	29754	2.87
HotDrinks	375	2.53	5383	1.97	38299	3.69
HouseholdCleansers	451	3.05	5159	1.89	14760	1.42
Hygieneproducts	174	1.17	5959	2.18	16048	1.55
Laundry	160	1.08	2204	0.81	6336	0.61
Liquor	138	0.93	2033	0.74	1180	0.11
Meat	396	2.67	11087	4.06	58517	5.64
MouthTooth	162	1.09	750	0.27	3777	0.36
PreservedFood	428	2.89	10419	3.81	22824	2.20
Readymade	523	3.53	7112	2.60	38086	3.67
Rest	4333	29.26	64175	23.49	208247	20.07
Snacks	970	6.55	25934	9.49	90370	8.71
Spreads	237	1.60	5853	2.14	21604	2.08
Vegetables	409	2.76	8356	3.06	62115	5.99
Wine	223	1.51	2624	0.96	7157	0.69
Total	14809	100	273237	100	1037444	100

Notes: (1) All numbers are reported for the forth quarter of 2008. Similar numbers apply to the other sample periods. (2) The short name "BE" denotes Belgium and "NL" Netherlands. (3) The column "Matched EANs" reports the number of goods (defined by the same GTIN) commonly purchased by households in the two countries indicated in the first column ("Country"). Columns 3 and 4 and 5 and 6 report the total and relative number of purchases available for the matched EANs in the first and second country of the considered country pair.

Table 4: Number of EANs and purchases by product category, GE - NL

Catergory	Matched EANs Freq.	%	GE: Purchases Freq.	%	NL: Purchases Freq.	%
AlcholfreeCO2	99	1.54	38377	5.27	38377	5.27
AlcholfreeNoCO2	123	1.91	8793	1.21	8793	1.21
Alcohol	30	0.47	2214	0.30	2214	0.30
Animalcare	433	6.74	18270	2.51	18270	2.51
Babyproducts	54	0.84	2224	0.31	2224	0.31
Basicfood	122	1.90	21793	2.99	21793	2.99
Beer	32	0.50	6108	0.84	6108	0.84
Bodycare	1040	16.18	26931	3.70	26931	3.70
Candy	711	11.06	124727	17.13	124727	17.13
Cereals	40	0.62	2049	0.28	2049	0.28
Champagne	15	0.23	2131	0.29	2131	0.29
DairyWhite	310	4.82	78656	10.80	78656	10.80
DairyYellow	164	2.55	26083	3.58	26083	3.58
Delicasees	280	4.36	42946	5.90	42946	5.90
Fatoils	37	0.58	26073	3.58	26073	3.58
Frische	59	0.92	14257	1.96	14257	1.96
Frozenproducts	207	3.22	17018	2.34	17018	2.34
HotDrinks	148	2.30	14333	1.97	14333	1.97
HouseholdCleansers	148	2.30	6981	0.96	6981	0.96
Hygieneproducts	119	1.85	15242	2.09	15242	2.09
Laundry	95	1.48	6563	0.90	6563	0.90
Liquor	164	2.55	5917	0.81	5917	0.81
Meat	246	3.83	43923	6.03	43923	6.03
MouthTooth	69	1.07	7866	1.08	7866	1.08
PreservedFood	246	3.83	22061	3.03	22061	3.03
Readymade	459	7.14	51042	7.01	51042	7.01
Rest	80	1.24	3766	0.52	3766	0.52
Snacks	419	6.52	41581	5.71	41581	5.71
Spreads	61	0.95	9943	1.37	9943	1.37
Vegetables	287	4.47	37224	5.11	37224	5.11
Wine	129	2.01	2918	0.40	2918	0.40
Total	6426	100	728010	100	728010	100

Notes: Table 4 (1) All numbers are reported for the forth quarter of 2008. Similar numbers apply to the sample periods. (2) The short name "GE" denotes Germany and "NL" Netherlands. (3) The column "Matched EANs" reports the number of goods (defined by the same GTIN) commonly purchased by households in the two countries indicated in the first column ("Country"). Columns 3 and 4 and 5 and 6 report the total and relative number of purchases available for the matched EANs in the first and second country of the considered country pair.

Table 5: Within and cross-country price dispersion: First quarter of 2005

		Mean			Median		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	0.28	0.41	0.53	0.00	0.00	0.00
NL-NL	66	0.25	0.34	0.85	0.00	0.00	0.00
BE-NL	132	6.85	6.83	1.26	3.95	4.13	1.93
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	741	0.43	0.38	1.36	0.01	0.00	0.19
NL-NL	66	0.42	0.48	1.24	0.26	0.00	0.50
GE-NL	468	-3.82	-3.73	2.61	-2.96	-2.84	2.41
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	0.50	0.50	1.09	0.10	0.00	0.27
GE-GE	741	0.40	0.40	1.45	0.07	0.00	0.39
BE-GE	429	10.71	10.75	2.37	8.22	8.32	2.85
		Absolute Mean			Absolute Median		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	4.88	4.87	0.22	1.83	1.85	0.15
NL-NL	66	6.23	6.23	0.40	2.32	2.28	0.30
BE-NL	132	20.67	20.74	1.30	15.01	15.48	2.12
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	741	7.62	7.61	0.65	4.29	4.32	0.61
NL-NL	66	9.57	9.63	0.73	5.43	5.33	0.68
GE-NL	468	20.02	20.24	1.89	14.96	15.25	1.78
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	0.50	0.50	1.09	0.10	0.00	0.27
GE-GE	741	0.40	0.40	1.45	0.07	0.00	0.39
BE-GE	429	10.71	10.75	2.37	8.22	8.32	2.85

Notes: Table 5 reports descriptive statistics on within and between-country price dispersion. To compute the statistics, we proceed as follows: First, all recorded prices of a given good within a given (NUTS2) region are averaged for the considered time period (2005Q1). Then, for each region pair, the mean (absolute) price gap is obtained by first computing the (absolute) price gaps of all goods that are commonly sold in both regions and then calculating the mean value of these price gaps. Price gaps are computed according to Equation (1). For R regions in a given sample, we obtain $R * (R + 1)/2$ price gaps. The numbers reported in the table correspond to the number of available region pairs (N) and the mean, median and standard deviation of computed (absolute) price gaps.

Table 6: Within and cross-country price dispersion: Fourth quarter of 2008

		Mean			Median		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	0.08	0.05	0.32	0.00	0.00	0.00
NL-NL	66	0.33	0.39	0.73	0.00	0.00	0.01
BE-NL	132	9.02	9.13	1.59	7.04	6.45	1.93
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	741	0.10	0.12	0.99	0.00	0.00	0.03
NL-NL	66	0.43	0.54	1.19	0.17	0.00	0.44
GE-NL	468	-4.91	-4.88	1.94	-3.16	-3.09	2.21
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	0.11	0.06	0.69	0.00	0.00	0.01
GE-GE	741	0.13	0.12	1.03	0.01	0.00	0.07
BE-GE	429	18.50	19.06	2.98	17.56	18.68	4.34
		Absolute Mean			Absolute Median		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	4.37	4.33	0.24	1.36	1.36	0.14
NL-NL	66	5.79	5.81	0.33	2.24	2.20	0.22
BE-NL	132	20.56	20.94	1.96	14.06	15.11	2.81
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	741	7.68	7.72	0.66	4.30	4.36	0.61
NL-NL	66	9.06	9.07	0.61	5.26	5.28	0.53
GE-NL	468	20.78	21.27	2.04	15.59	16.13	2.48
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	6.31	6.22	0.57	3.20	3.18	0.31
GE-GE	741	7.84	7.88	0.60	4.58	4.60	0.55
BE-GE	429	25.64	25.96	2.25	20.92	21.55	2.91

Notes: Table 6 reports descriptive statistics on within and between-country price dispersion based on comparing regional average price gaps. To compute the statistics, we proceed as follows: First, all recorded prices of a given good within a given (NUTS2) region are averaged for the considered time period (2008Q4). Then, for each region pair, the mean (absolute) price gap is obtained by first computing the (absolute) price gaps of all goods that are commonly sold in both regions and then calculating the mean value of these price gaps. Price gaps are computed according to Equation (1). For R regions in a given sample, we obtain $R*(R+1)/2$ price gaps. The numbers reported in the table correspond to the number of available region pairs (N) and the mean, median and standard deviation of computed (absolute) price gaps.

Table 7: Within- and cross-country price dispersion: goods-level comparisons

2005Q1		Price gap			Absolute price gap		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	143999	0.00	0.00	0.11	0.05	0.02	0.10
NL-NL	402686	0.00	0.00	0.12	0.06	0.02	0.10
BE-NL	161587	6.98	3.55	30.33	20.47	14.60	23.45
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	1980476	0.00	0.00	0.13	0.08	0.04	0.10
NL-NL	37209	0.00	0.00	0.15	0.09	0.05	0.12
GE-NL	163921	-4.18	-2.82	28.29	19.71	14.64	20.72
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	32793	0.00	0.00	0.13	0.07	0.04	0.11
GE-GE	1131650	0.00	0.00	0.13	0.08	0.05	0.11
BE-GE	126586	11.05	8.57	29.44	22.69	17.61	21.77
2008Q4		Price gap			Absolute price gap		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	192791	0.00	0.00	0.13	0.04	0.01	0.12
NL-NL	582299	0.00	0.00	0.11	0.06	0.02	0.09
BE-NL	264283	8.63	6.21	30.28	19.94	13.10	24.37
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	2974238	0.00	0.00	0.12	0.08	0.04	0.10
NL-NL	48161	0.00	0.00	0.15	0.09	0.05	0.12
GE-NL	261595	-4.85	-2.40	28.68	20.17	14.58	20.96
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	40667	0.00	0.00	0.16	0.06	0.03	0.14
GE-GE	1715411	0.00	0.00	0.12	0.08	0.04	0.10
BE-GE	194684	17.60	16.35	30.07	25.06	20.20	24.21

Notes: Table 7 reports descriptive statistics on within and between-country price dispersion employing regional goods-level price gaps. To compute the statistics, we proceed as follows: First, all recorded prices of a given good within a given (NUTS2) region are averaged for the considered time period (2005Q1, 2008Q4). Then, in the spirit of Engel and Rogers (1996), for each good all possible bi-regional price gaps are computed. Finally, for each considered sub-sample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (1). The numbers reported in the table correspond to the number of available goods-level regional price gaps (N) and the mean, median and standard deviation of computed (absolute) price gaps.

Table 8: Regression discontinuity results: Baseline model

	BE-NL				GE-NL				BE-GE			
2005Q1	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.000	0.000	0.099	15.9	-0.033	-0.023	0.156	19.6	-0.012	-0.023	0.135	22.2
Border	0.076	0.065	0.301	76.3	0.002	0.000	0.218	60.1	0.003	0.005	0.242	50.0
Border x Distance	0.002	0.000	0.248	11.3	0.035	0.023	0.215	17.5	0.026	0.016	0.166	11.1
Constant	-0.730	-0.863	1.456	90.0	-0.775	-0.597	1.200	90.9	-0.588	-0.445	1.256	72.2
Observations	763.000	.	.	.	143.000	.	.	.	18.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.000	0.000	0.094	14.7	-0.009	-0.002	0.156	18.9	0.038	0.005	0.227	8.0
Border	0.103	0.095	0.288	81.2	-0.035	-0.032	0.259	58.2	0.092	0.093	0.196	42.0
Border x Distance	-0.014	-0.002	0.163	13.8	0.026	0.026	0.218	19.4	-0.082	-0.043	0.254	12.0
Constant	-0.594	-0.697	1.469	95.2	-0.402	-0.462	1.146	94.5	-0.340	-0.341	0.937	88.0
Observations	1165.000	.	.	.	201.000	.	.	.	50.000	.	.	.

Notes: Table 8 reports regression discontinuity results for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2). For each country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. Positive values of the border coefficient indicate that prices are higher in the reference country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level. The chosen bandwidth in all specifications is 80km.

Table 9: Regression discontinuity results, absolute values: Baseline model

	BE-NL				GE-NL				BE-GE			
2005Q1	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.055	0.024	0.082	15.9	0.102	0.058	0.122	19.6	0.094	0.057	0.095	22.2
Border	0.205	0.149	0.234	76.3	0.170	0.141	0.137	60.1	0.176	0.130	0.160	50.0
Border x Distance	0.093	0.045	0.230	11.3	0.148	0.101	0.159	17.5	0.119	0.048	0.115	11.1
Constant	1.276	1.044	1.011	90.0	1.054	0.685	0.963	90.9	0.874	0.498	1.065	72.2
Observations	763.000	.	.	.	143.000	.	.	.	18.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.054	0.024	0.077	14.7	0.101	0.064	0.119	18.9	0.126	0.080	0.191	8.0
Border	0.209	0.155	0.224	81.2	0.187	0.130	0.183	58.2	0.169	0.139	0.134	42.0
Border x Distance	0.093	0.047	0.135	13.8	0.159	0.119	0.151	19.4	0.163	0.112	0.210	12.0
Constant	1.206	0.986	1.027	95.2	0.845	0.546	0.871	94.5	0.698	0.438	0.707	88.0
Observations	1165.000	.	.	.	201.000	.	.	.	50.000	.	.	.

Notes: Table 9 reports absolute vlaues of regression discontinuity results for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2). For each country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level. The chosen bandwidth in all specifications is 80km.

Table 10: Counterfactual within- and cross-country price dispersion: goods-level comparisons

2005Q1		Price gap			Absolute price gap		
FL-WL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
FL-FL	87877	0.00	0.00	0.15	0.06	0.02	0.14
WL-WL	53701	0.00	0.00	0.16	0.06	0.02	0.15
FL-WL	167688	0.50	0.00	16.73	6.13	1.78	15.58
LSax-NRW	N	Mean	Median	St.Dev	Mean	Median	St.Dev
LSax-LSax	70916	0.00	0.00	0.17	0.08	0.02	0.15
NRW-NRW	143423	0.01	0.00	0.17	0.08	0.02	0.15
LSax-NRW	248915	-0.61	0.00	17.84	8.74	2.67	15.57
NLE-NLS	N	Mean	Median	St.Dev	Mean	Median	St.Dev
NLE-NLE	819	0.01	0.00	0.12	0.07	0.04	0.10
NLS-NLS	327	-0.02	-0.01	0.09	0.05	0.03	0.08
NLE-NLS	1785	-0.02	0.00	10.63	6.12	3.24	8.69
2008Q4		Price gap			Absolute price gap		
FL-WL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
FL-FL	119692	0.00	0.00	0.16	0.06	0.02	0.15
WL-WL	77894	0.00	0.00	0.17	0.05	0.01	0.16
FL-WL	232267	0.46	0.00	17.98	5.91	1.67	16.98
LSax-NRW	N	Mean	Median	St.Dev	Mean	Median	St.Dev
LSax-LSax	103814	0.00	0.00	0.15	0.07	0.01	0.13
NRW-NRW	211577	0.00	0.00	0.15	0.07	0.02	0.13
LSax-NRW	363348	-0.48	0.00	16.00	7.89	2.40	13.93
NLE-NLS	N	Mean	Median	St.Dev	Mean	Median	St.Dev
NLE-NLE	1712	0.01	0.00	0.11	0.06	0.04	0.09
NLS-NLS	597	-0.02	-0.01	0.08	0.05	0.02	0.06
NLE-NLS	3507	0.25	0.35	10.17	6.25	3.68	8.02

Notes: Table 10 reports descriptive statistics on within and between-country price dispersion employing regional goods-level price gaps. To compute the statistics, we proceed as follows: First, all recorded prices of a given good within a given (NUTS2) region are averaged for the considered time period (2005Q1, 2008Q4). Then, in the spirit of Engel and Rogers (1996), for each good all possible bi-regional price gaps are computed. Finally, for each considered sub-sample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (1). The numbers reported in the table correspond to the number of available goods-level regional price gaps (N) and the mean, median and standard deviation of computed (absolute) price gaps.

Table 11: Regression discontinuity results for within-country regions: Baseline model

2005Q1	BEFL-BEWL				LSax-NRW				NLE-NLS			
	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	-0.006	0.000	0.251	15.7	-0.008	0.000	0.143	11.1	0.010	0.000	0.225	13.8
Border	-0.004	0.000	0.103	12.0	0.000	0.000	0.114	8.0	0.018	0.002	0.111	14.0
Border x Distance	-0.014	0.000	0.276	12.2	0.012	0.000	0.222	9.1	0.015	0.000	0.281	11.0
Constant	-0.580	-0.882	1.912	90.6	-1.110	-1.152	1.267	65.7	-0.994	-1.219	1.748	84.3
Observations	2731.000	.	.	.	2564.000	.	.	.	2472.000	.	.	.
2008Q4	BEFL-BEWL				LSax-NRW				NLE-NLS			
	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	-0.004	0.000	0.444	13.4	-0.010	0.000	0.143	13.4	-0.003	0.000	0.179	14.0
Border	0.002	0.000	0.141	10.9	-0.007	0.000	0.109	9.6	0.016	0.002	0.097	14.3
Border x Distance	0.003	0.000	0.491	11.1	0.012	0.000	0.219	11.0	0.039	0.005	0.239	13.9
Constant	-0.399	-0.709	1.882	88.4	-0.895	-0.958	1.414	73.0	-0.895	-1.063	1.536	92.3
Observations	4286.000	.	.	.	4022.000	.	.	.	4228.000	.	.	.

Notes: Table 8 reports regression discontinuity results for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2). For each hypothetical country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level. The chosen bandwidth in all specifications is 80km.

Table 12: Regression discontinuity results for within-country regions, absolute values: Baseline model

2005Q1	BEFL-BEWL				LSax-NRW				NLE-NLS			
	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.102	0.041	0.229	15.7	0.069	0.014	0.125	11.1	0.112	0.048	0.195	13.8
Border	0.040	0.017	0.095	12.0	0.054	0.011	0.101	8.0	0.064	0.031	0.093	14.0
Border x Distance	0.130	0.060	0.244	12.2	0.110	0.026	0.193	9.1	0.154	0.072	0.236	11.0
Constant	1.552	1.213	1.258	90.6	1.405	1.280	0.928	65.7	1.654	1.444	1.142	84.3
Observations	2731.000	.	.	.	2564.000	.	.	.	2472.000	.	.	.
2008Q4	BEFL-BEWL				LSax-NRW				NLE-NLS			
	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.100	0.032	0.433	13.4	0.070	0.021	0.125	13.4	0.096	0.040	0.151	14.0
Border	0.039	0.014	0.136	10.9	0.054	0.016	0.094	9.6	0.055	0.026	0.082	14.3
Border x Distance	0.127	0.047	0.474	11.1	0.114	0.036	0.188	11.0	0.137	0.063	0.200	13.9
Constant	1.445	1.061	1.270	88.4	1.342	1.153	1.000	73.0	1.439	1.231	1.044	92.3
Observations	4286.000	.	.	.	4022.000	.	.	.	4228.000	.	.	.

Notes: Table 12 reports regression discontinuity results for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2). For each hypothetical country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level. The chosen bandwidth in all specifications is 80km.

Table 13: Regression discontinuity results: Household-goods preferences

	BE-NL				GE-NL				BE-GE			
2005Q1	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
PurchFreq	0.758	0.597	2.513	15.5	-0.497	-0.253	1.536	25.2	3.369	2.565	3.839	27.778
Observations	763.000	.	.	.	143.000	.	.	.	18.000	.	.	.
Elasticity	-0.227	-0.183	0.793	27.3	-0.001	-0.001	0.000	0.0	.	.	.	.
Observations	66.000	.	.	.	1.000	.	.	.	.	.	.	.
StateDep	0.002	0.000	0.276	16.4	-0.076	-0.063	0.299	15.5	0.029	0.074	0.442	16.667
Observations	834.000	.	.	.	148.000	.	.	.	18.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
PurchFreq	0.303	0.263	1.750	15.6	-0.285	-0.102	1.846	17.9	0.244	0.522	3.041	8.000
Observations	1165.000	.	.	.	201.000	.	.	.	50.000	.	.	.
Elasticity	-0.121	-0.155	0.912	28.9	0.013	0.173	0.976	27.3	0.203	0.452	0.562	0.000
Observations	405.000	.	.	.	33.000	.	.	.	3.000	.	.	.
StateDep	-0.027	-0.007	0.260	14.1	-0.075	-0.066	0.327	26.4	0.048	0.042	0.456	18.000
Observations	1165.000	.	.	.	201.000	.	.	.	50.000	.	.	.

Notes: Table 13 reports regression discontinuity results for households-goods preference variables for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). For the state dependence variable (StateDep) the first sample period is 2005Q2. Results are based on estimating Equation (3). For each country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. Positive values of the border coefficient indicate that prices are higher in the reference country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level.

Table 14: Regression discontinuity results, absolute values: Household-goods preferences

	BE-NL				GE-NL				BE-GE			
2005Q1	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
PurchFreq	1.228	0.854	1.283	15.6	1.151	0.656	1.469	17.9	1.934	1.141	2.344	8.000
Observations	1165.000	.	.	.	201.000	.	.	.	50.000	.	.	.
Elasticity	0.581	0.390	0.713	28.9	0.771	0.680	0.583	27.3	0.496	0.452	0.088	0.000
Observations	405.000	.	.	.	33.000	.	.	.	3.000	.	.	.
StateDep	0.190	0.141	0.180	14.1	0.247	0.189	0.227	26.4	0.345	0.286	0.298	18.000
Observations	1165.000	.	.	.	201.000	.	.	.	50.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
PurchFreq	1.228	0.854	1.283	15.6	1.151	0.656	1.469	17.9	1.934	1.141	2.344	8.000
Observations	1165.000	.	.	.	201.000	.	.	.	50.000	.	.	.
Elasticity	0.581	0.390	0.713	28.9	0.771	0.680	0.583	27.3	0.496	0.452	0.088	0.000
Observations	405.000	.	.	.	33.000	.	.	.	3.000	.	.	.
StateDep	0.190	0.141	0.180	14.1	0.247	0.189	0.227	26.4	0.345	0.286	0.298	18.000
Observations	1165.000	.	.	.	201.000	.	.	.	50.000	.	.	.

Notes: Table 14 reports regression discontinuity results for households-goods preference variables for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). For the state dependence variable (StateDep) the first sample period is 2005Q2. Results are based on estimating Equation (3). For each country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of the estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level.

Table 15: Regression discontinuity results: Household preferences, BE-NL

2005Q1

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-4.128***	-2.027***	0.588**	0.0526***
	(-3.65)	(-3.29)	(2.29)	(3.39)
Border	8.626***	-2.937***	0.585***	-0.0331***
	(6.82)	(-7.24)	(3.77)	(-3.43)
Border × Distance	-22.03***	-3.539***	1.737***	-0.143***
	(-6.34)	(-3.55)	(4.84)	(-6.14)
Constant	21.14***	11.64***	4.593***	0.419***
	(39.37)	(39.52)	(41.52)	(59.23)
Observations	4603	4603	4603	6588
R^2	0.163	0.016	0.022	0.010

2008Q4

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-6.108***	-2.601***	0.576***	0.0647***
	(-6.51)	(-5.63)	(3.23)	(4.71)
Border	10.45***	-3.265***	0.588***	-0.0538***
	(10.05)	(-10.32)	(4.98)	(-6.48)
Border × Distance	-11.67***	-1.597**	1.290***	-0.143***
	(-4.22)	(-2.11)	(4.79)	(-7.11)
Constant	25.14***	12.22***	4.102***	0.442***
	(53.96)	(53.07)	(50.42)	(70.21)
Observations	7016	7016	7016	8689
R^2	0.163	0.015	0.015	0.024

Notes: Table 15 reports regression discontinuity results for households preference variables for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (3). The reference country is Belgium. Positive values of the border coefficient indicate that values of the respective variable are higher in the reference country. The chosen bandwidth in all specifications is 80 km. Parentheses report t statistics, signifcance of the coefficients is indicated as following: $*$ ($p < 0.10$), $**$ ($p < 0.05$), $***$ ($p < 0.01$).

Table 16: Regression discontinuity results: household preferences, GE-NL

2005Q1

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-1.397	-0.103	0.755***	0.00737
	(-1.29)	(-0.17)	(3.08)	(0.49)
Border	-5.194***	-3.277***	0.806***	-0.0934***
	(-8.25)	(-9.38)	(4.63)	(-9.76)
Border × Distance	2.135	0.179	-0.859**	-0.0194
	(1.59)	(0.24)	(-2.29)	(-0.95)
Constant	20.17***	11.03***	4.408***	0.436***
	(42.47)	(42.55)	(44.37)	(64.77)
Observations	4042	4042	4042	5344
R^2	0.058	0.082	0.015	0.061

2008Q4

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-2.102**	-0.526	0.401**	0.0461***
	(-2.11)	(-1.10)	(2.28)	(3.42)
Border	-5.079***	-2.907***	0.238*	-0.0722***
	(-8.21)	(-10.07)	(1.89)	(-8.28)
Border × Distance	1.465	0.454	-0.592**	-0.0598***
	(1.11)	(0.73)	(-2.22)	(-3.22)
Constant	23.50***	11.51***	4.114***	0.446***
	(52.02)	(54.42)	(54.25)	(76.11)
Observations	6925	6925	6925	8054
R^2	0.026	0.051	0.002	0.048

Notes: Table 16 reports regression discontinuity results for households preference variables for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (3). The reference country is Germany. Positive values of the border coefficient indicate that values of the respective variable are higher in the reference country. The chosen bandwidth in all specifications is 80 km. Parentheses report t statistics, signifcance of the coefficients is indicated as following: $*$ ($p < 0.10$), $**$ ($p < 0.05$), $***$ ($p < 0.01$).

Table 17: Regression discontinuity results: household preferences, BE-GE

2005Q1

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-2.802**	-1.105*	0.684	0.0380*
	(-2.20)	(-1.77)	(1.54)	(1.75)
Border	31.64***	5.266***	-1.174***	0.121***
	(7.85)	(5.92)	(-3.35)	(5.40)
Border × Distance	16.97**	4.978***	-1.195*	0.0221
	(2.28)	(3.01)	(-1.86)	(0.51)
Constant	16.71***	8.410***	4.726***	0.330***
	(21.68)	(22.03)	(18.23)	(26.62)
Observations	1363	1363	1363	1829
R^2	0.274	0.120	0.075	0.045

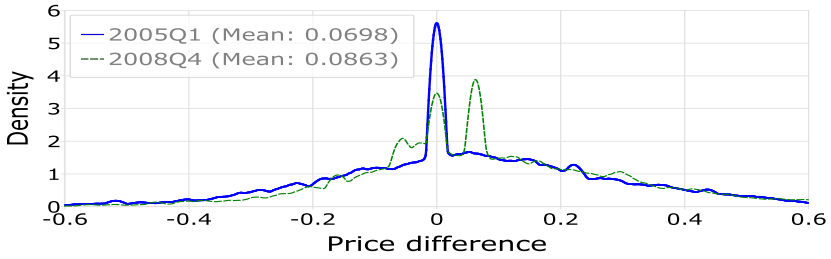
2008Q4

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	0.951	0.468	-0.150	0.0267
	(0.72)	(0.82)	(-0.48)	(1.35)
Border	28.14***	4.393***	-1.003***	0.0671***
	(10.99)	(6.40)	(-3.93)	(3.66)
Border × Distance	2.974	2.591**	-0.633	0.0217
	(0.59)	(1.97)	(-1.30)	(0.64)
Constant	18.96***	8.588***	4.352***	0.369***
	(24.59)	(25.48)	(24.05)	(31.61)
Observations	2492	2492	2492	3163
R^2	0.276	0.053	0.015	0.008

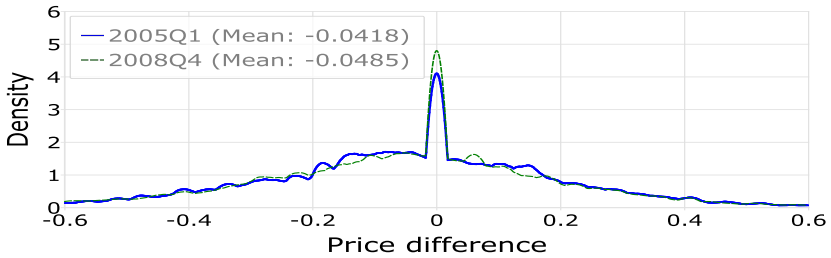
Notes: Table 17 reports regression discontinuity results for households preference variables for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (3). The reference country is Belgium. Positive values of the border coefficient indicate that values of the respective variable are higher in the reference country. The chosen bandwidth in all specifications is 80 km. Parentheses report t statistics, signifcance of the coefficients is indicated as following: $*$ ($p < 0.10$), $**$ ($p < 0.05$), $***$ ($p < 0.01$).

8 Figures

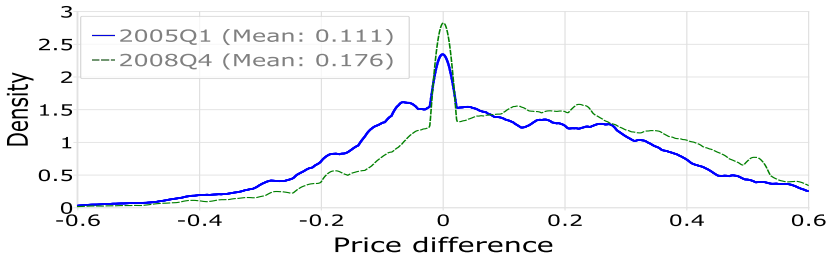
Figure 1: Distribution of mean price differences: goods-based



(a) BE-NL



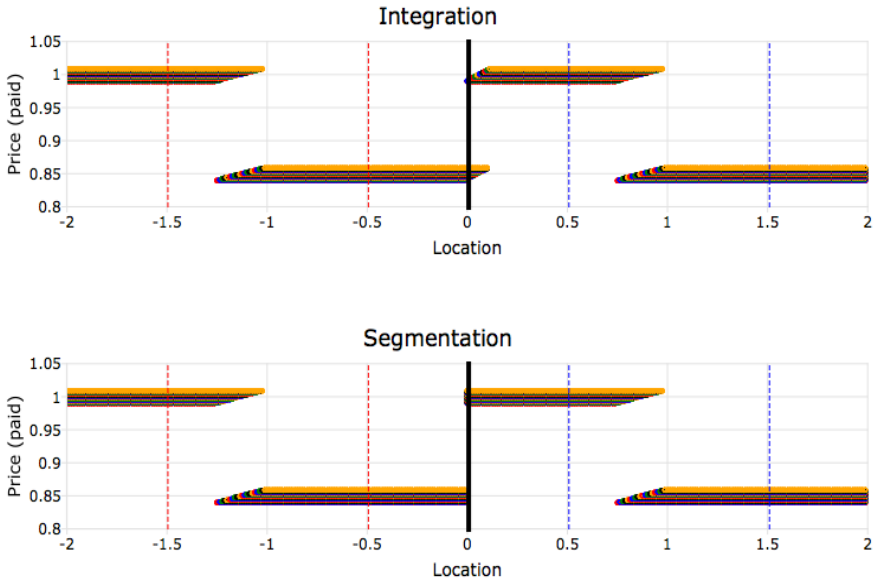
(b) GE-NL



(c) BE-GE

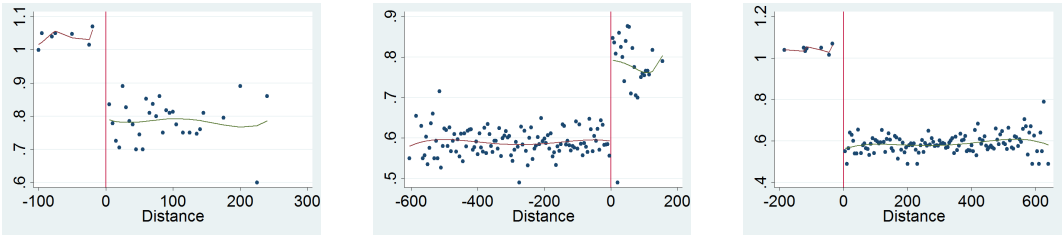
Notes: Figure 1 plots the kernel density estimates of the mean (log) cross-country price differences of all matched goods of the country pair indicated below each panel for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). In each panel, the base country is the country indicated first in the subtitle. A positive value indicates that prices are higher in the base country than in the reference country (mentioned secondly).

Figure 3: The case of integrated and segmented markets



Notes: Figure 3 illustrates which prices household would pay if markets were integrated (upper pannel) or segmented (lower panel). On the x axis distance is plotted, on the y axis the prices households pay. It is assumed that there exist two retailers in the reference country (whose locations are indicated by the blue dotted lines) and two retailers in the counterpart country (whose locations are indicated by the red dotted lines). We assume that the retailer closest to the border in the reference country charges a price of 1, whereas the corresponding retailer in the counterpart country charges a price of 0.85. The second retailer in the reference country charges a price of 0.85, the second retailer in the counterpart country charges of price of 1. The vertical lines indicate the prices which household pay.

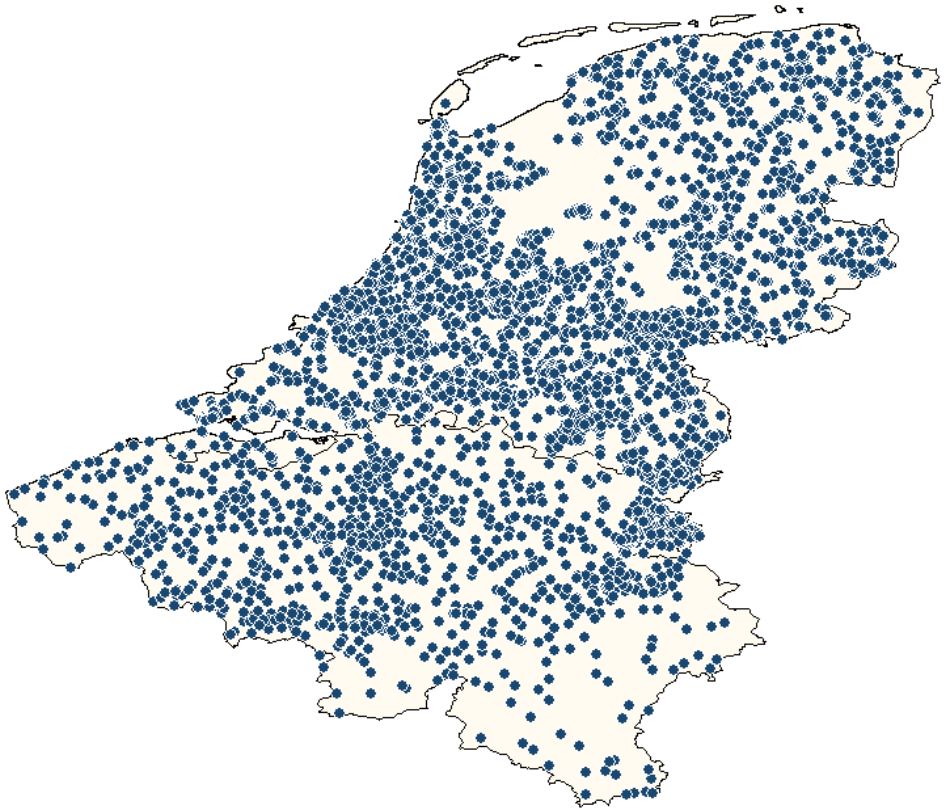
Figure 4: Discontinuity plots for the price of a selected good



(a) BE-NL (b) GE-NL (c) BE-GE

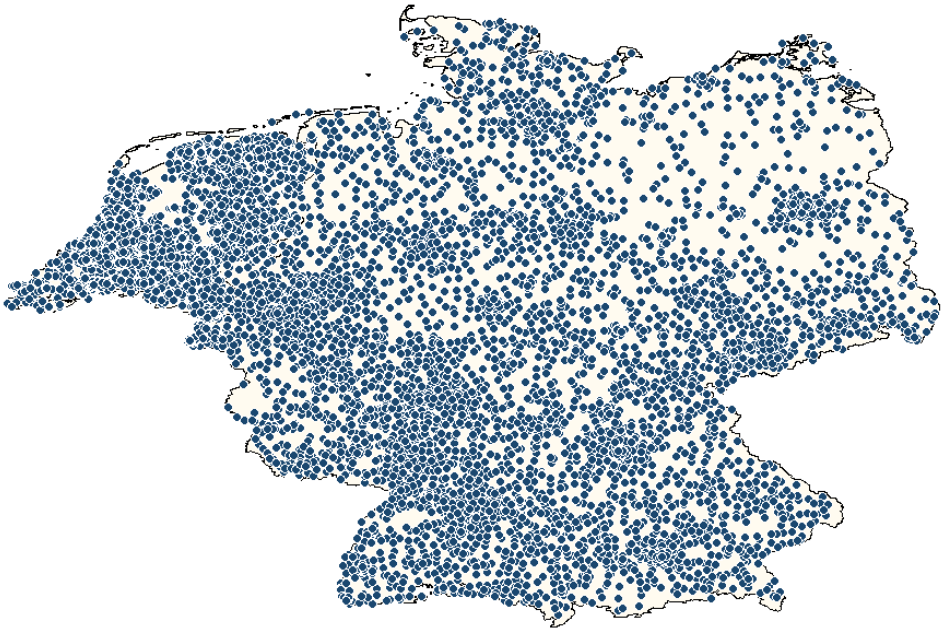
Notes: The panels of Figure 4 plot the price of a selected good (chocolate, 100 gramm) in the forth quarter of 2008 on both sides of the Belgian-Dutch (left panel), German-Dutch (middle) and Belgian-German (right) panel. The prices of the country mentioned first are associated with negative distance values whereas the prices of the other country are associated with positive distance values. The selected bin width is 5 km.

Figure 5: Locations of Belgian and Dutch households



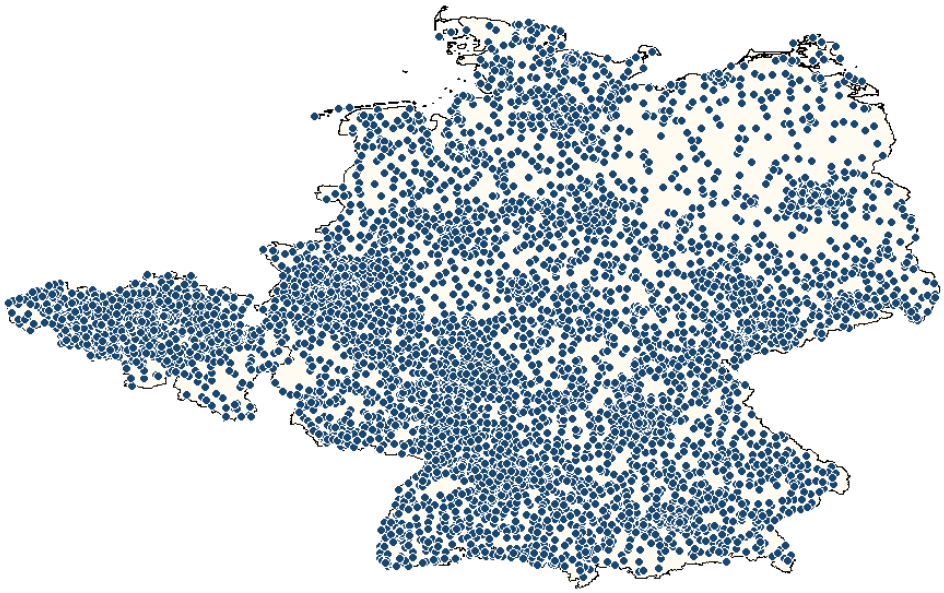
Notes: Figure 7 plots locations of the Belgian and Dutch households included in the data sample in the forth quarter of 2008.

Figure 6: Locations of German and Dutch households



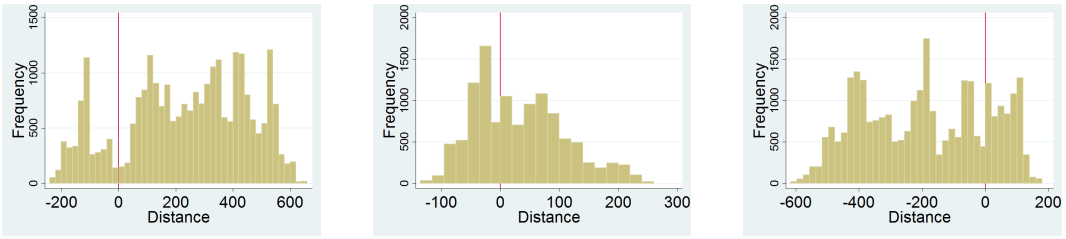
Notes: Figure 7 plots locations of the German and Dutch households included in the data sample in the forth quarter of 2008.

Figure 7: Locations of Belgian and German households



Notes: Figure 7 plots locations of the Belgian and German households included in the data sample in the forth quarter of 2008.

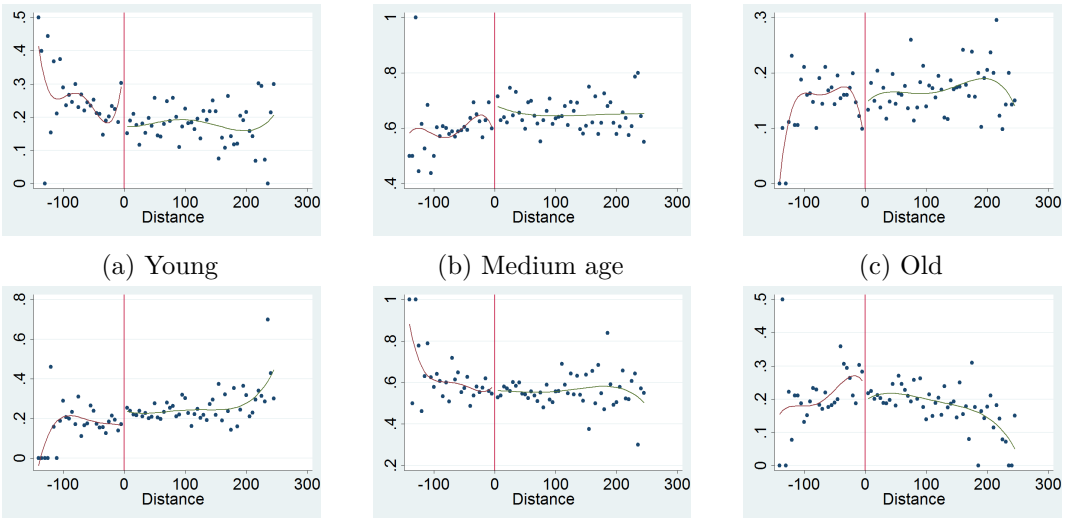
Figure 8: Household densities in the vicinity of the border



(a) BE-GE (b) BE-NL (c) GE-NL

Notes: Panels (a) to (c) plot densities of household locations in dependence of their distances to the border between the two countries indicated below the respective panel. In each case, the locations of households of the country mentioned first below the panel are plotted with negative distances, locations of households of the country mentioned secondly are plotted with positive distances. Bins have a width of 20 km. Densities are plotted for the households included in the data sample in the 2008.

Figure 9: Age and income structure of households in the vicinity of the border: BE-NL

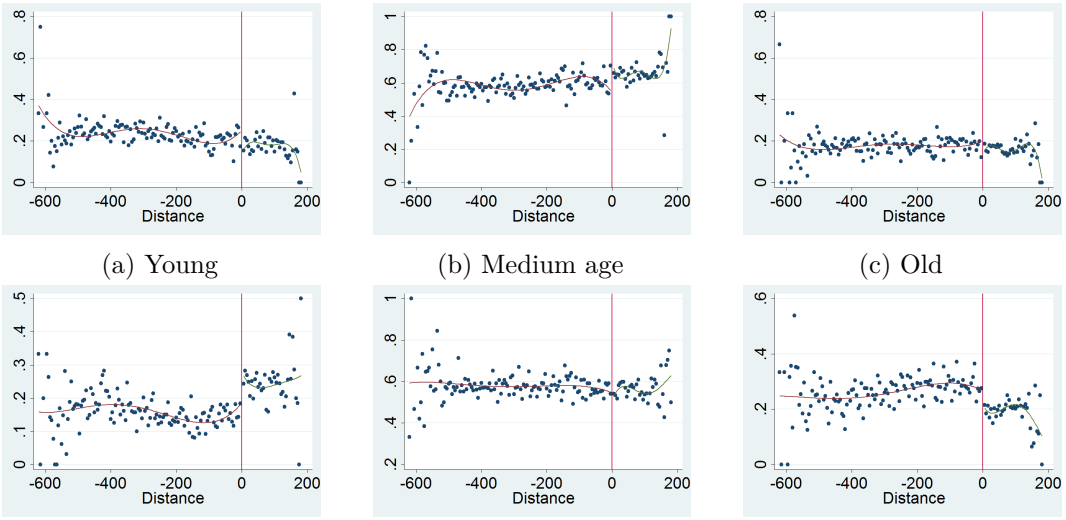


(a) Young (b) Medium age (c) Old

(d) Low income (e) Medium income (f) High income

Notes: Figure 9 plots densities of the age (panels (a) to (c)) and income (panels (d) to (f)) structure of Belgian-Dutch households in dependence to the distances of the respective households homes to the border. In each case, the locations of Belgian households are plotted with negative distances, locations of Dutch households are plotted with positive distances. Bins have a width of 5km. Densities are plotted for the households included in the data sample in the 2008.

Figure 10: Age and income structure of households in the vicinity of the border: GE-NL

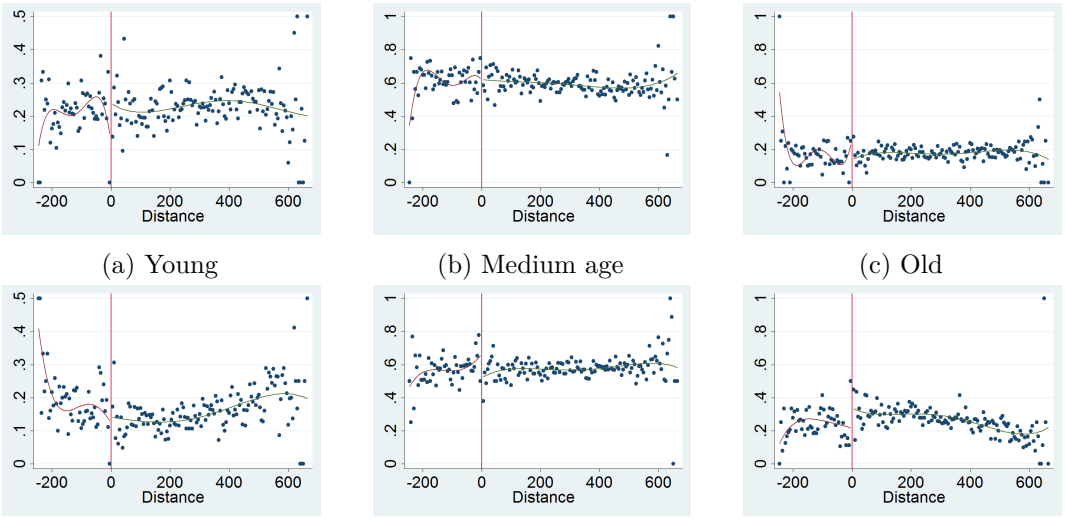


(a) Young (b) Medium age (c) Old

(d) Low income (e) Medium income (f) High income

Notes: Figure 10 plots densities of the age (panels (a) to (c)) and income (panels (d) to (f)) structure of German-Dutch households in dependence to the distances of the respective households homes to the border. In each case, the locations of German households are plotted with negative distances, locations of Dutch households are plotted with positive distances. Bins have a width of 5km. Densities are plotted for the households included in the data sample in the 2008.

Figure 11: Age and income structure of households in the vicinity of the border: BE-GE

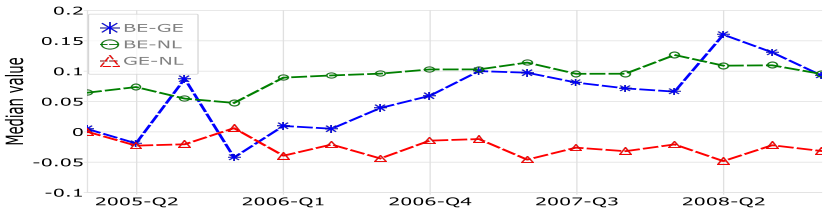


(a) Young (b) Medium age (c) Old

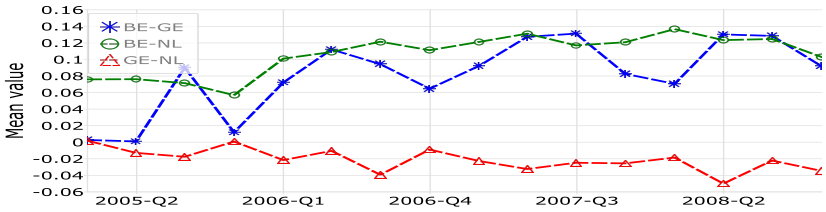
(d) Low income (e) Medium income (f) High income

Notes: Figure 11 plots densities of the age (panels (a) to (c)) and income (panels (d) to (f)) structure of Belgian-German households in dependence to the distances of the respective households homes to the border. In each case, the locations of Belgian households are plotted with negative distances, locations of German households are plotted with positive distances. Bins have a width of 5km. Densities are plotted for the households included in the data sample in the 2008.

Figure 12: Dynamics of border coefficients



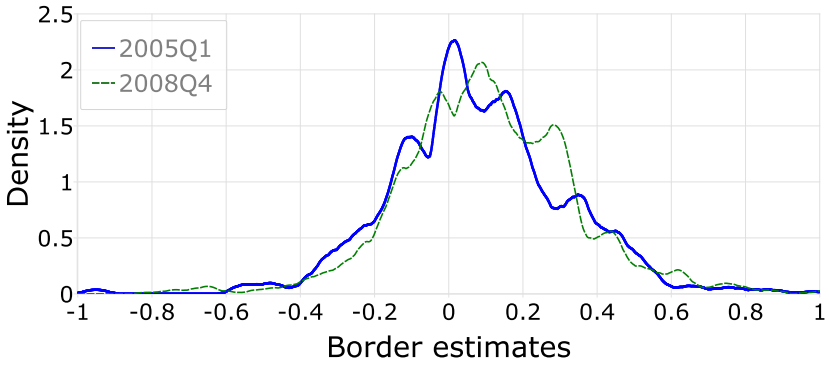
(a) Median values of the border estimates



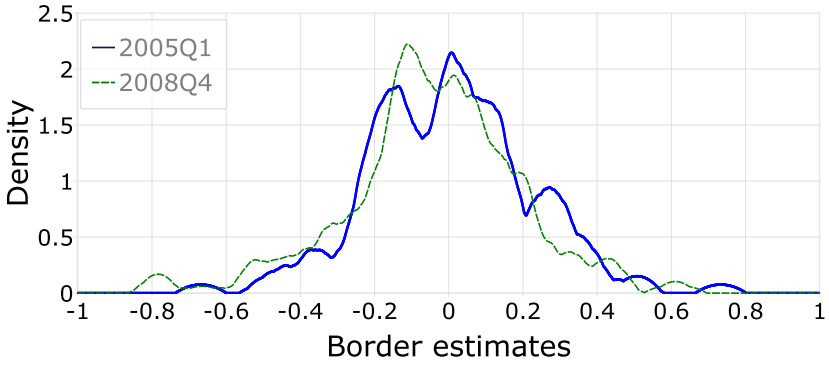
(b) Mean values of the border estimates

Notes: Figure 12 plots median and mean values of estimated quarterly border coefficietns. BE-NL denotes the Belgian-Dutch, GE-NL the German-Dutch and BE-GE the Belgian-Dutch values. The border dummy takes the value one if the household is located in the country mentioned first in the legend. Thus, positive values for the BE-NL border coefficient indicate that prices are higher in Belgium than in the Netherlands.

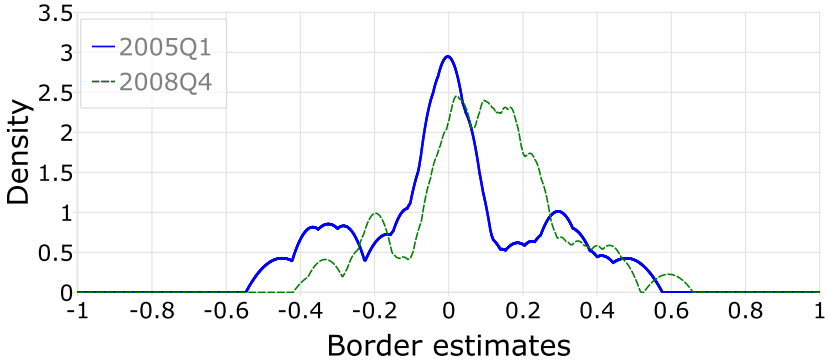
Figure 14: Kernel density estimates of border effects



(a) BE-NL



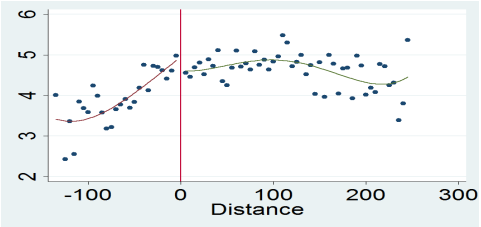
(b) GE-NL



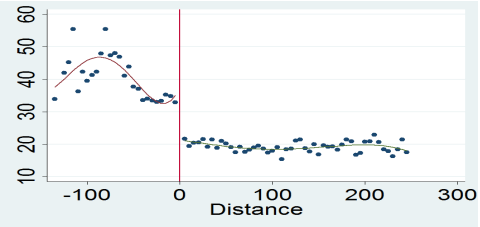
(c) BE-GE

Notes: Figure 14 plots kernel density functions of the estimated border coefficients. The considered sample periods are 2005Q1 and 2008Q4. BE-NL denotes the Belgian-Dutch, GE-NL the German-Dutch and BE-GE the Belgian-Dutch values. The border dummy takes the value one if the household is located in the first country of a given country pair. Thus, positive values for the BE-NL border coefficient indicate that prices are higher in Belgium than in the Netherlands.

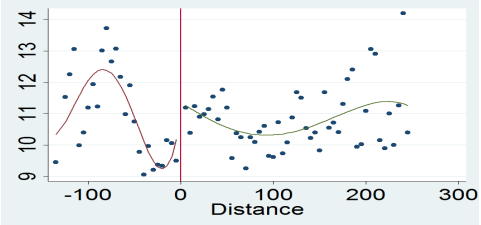
Figure 16: Household shopping patterns in the vicinity of the border: BE-NL



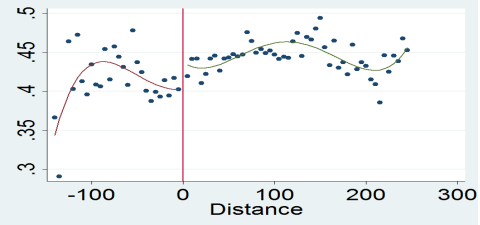
(a) Frequency of shopping trips



(b) Average purchase amount



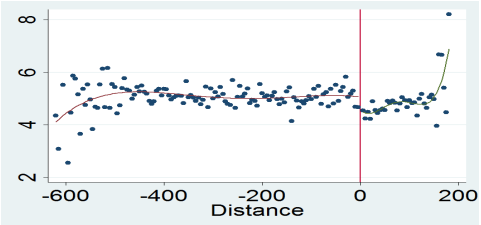
(c) Average number of items purchased per shopping trip



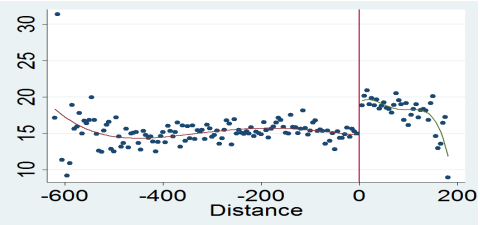
(d) Share of shopping trips of preferred retailer

Notes: Figure 9 plots densities of the households variables indicated below each panel in dependence to the distances of the respective household's home to the border. In each case, the locations of Belgian households are plotted with negative distances, locations of Dutch households are plotted with positive distances. Bins have a width of 5km. Densities are plotted for the households included in the data sample in the 2008.

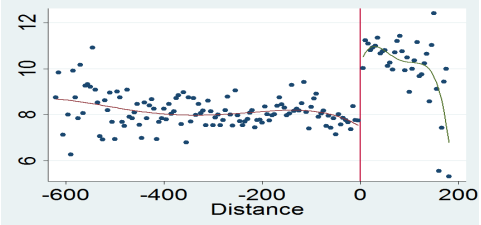
Figure 17: Household shopping patterns in the vicinity of the border: GE-NL



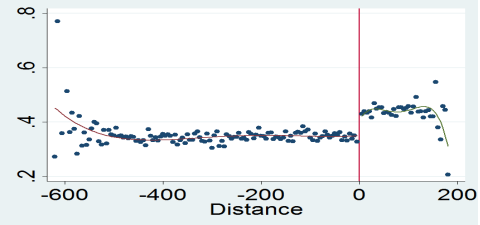
(a) Frequency of shopping trips



(b) Average purchase amount



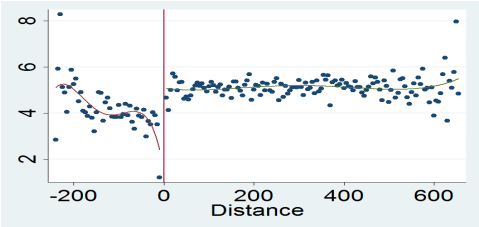
(c) Average number of items purchased per shopping trip



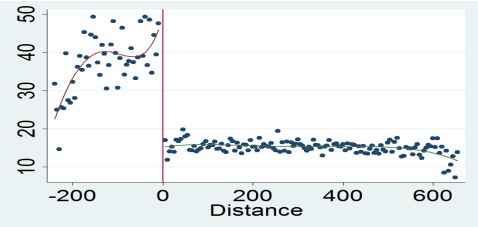
(d) Share of shopping trips to the most preferred retailer

Notes: Figure 10 plots densities of the household variables indicated below each panel in dependence to the distances of the respective household's home to the border. In each case, the locations of German households are plotted with negative distances, locations of Dutch households are plotted with positive distances. Bins have a width of 5km. Densities are plotted for the households included in the data sample in the 2008.

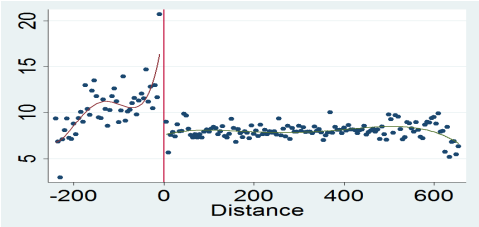
Figure 18: Household shopping patterns in the vicinity of the border: BEGE



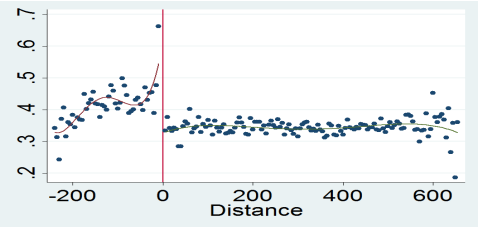
(a) Frequency of shopping trips



(b) Average purchase amount



(c) Average number of items purchased per shopping trip

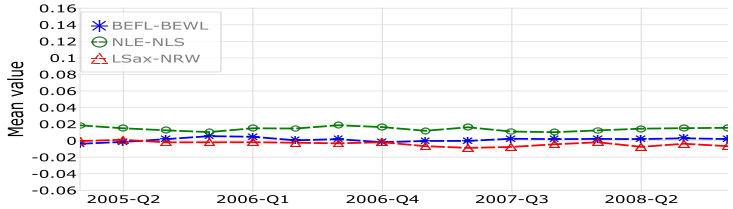


(d) Share of shopping trips to the most preferred retailer

Notes: Figure 11 plots densities of the household variables indicated below each panel in dependence to the distances of the respective household's home to the border. In each case, the locations of Belgian households are plotted with negative distances, locations of German households are plotted with positive distances. Bins have a width of 5km. Densities are plotted for the households included in the data sample in the 2008.

Figure 19: Dynamics of border coefficients

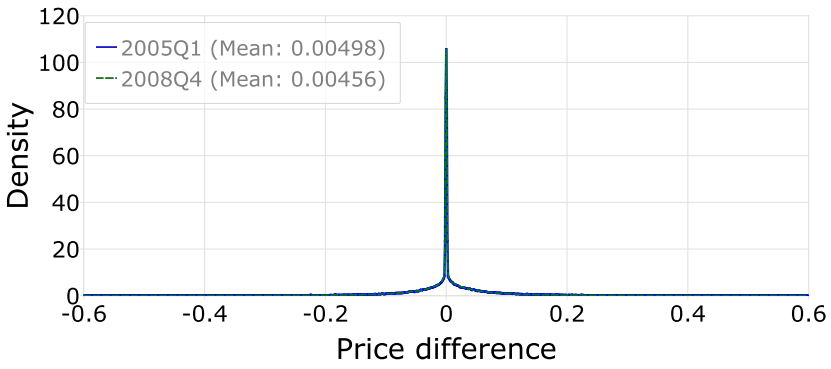
(a) Median values of the boder estimates



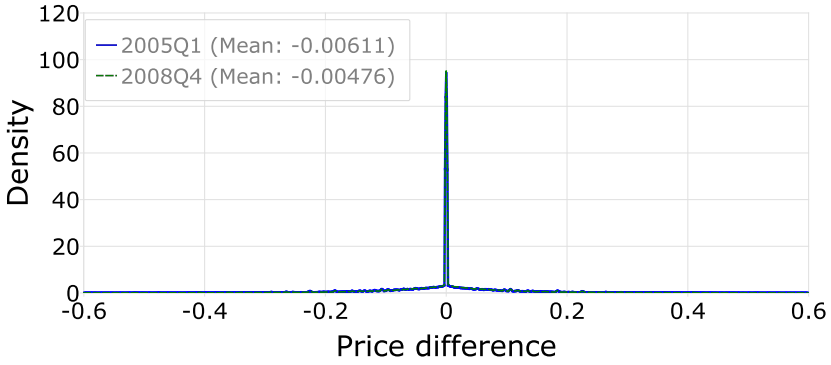
(b) Mean values of the border estimates

Notes: Figure 19 plots median and mean values of estimated quarterly border estimates. BEFL-BEWL denotes the Flanders-Wallonia, LSax-NRW the Lower Saxony-North Rhine-Westphalia and NLE-NLS the Eastern Netherlands-Southern Netherlands values. The border dummy takes the value one if the household is located in the country mentioned first in the legend. Thus, positive values for the BEFL-BEWL border coefficient indicate that prices are higher in Flanders than in Wallonia.

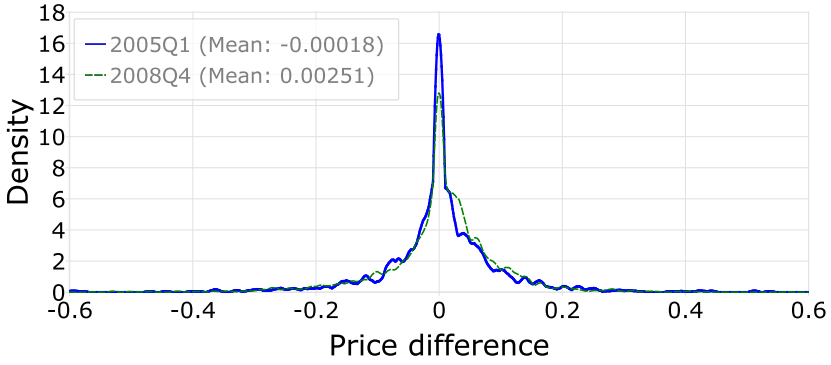
Figure 21: Distribution of mean price differences: goods-based



(a) BEFL-BEWL



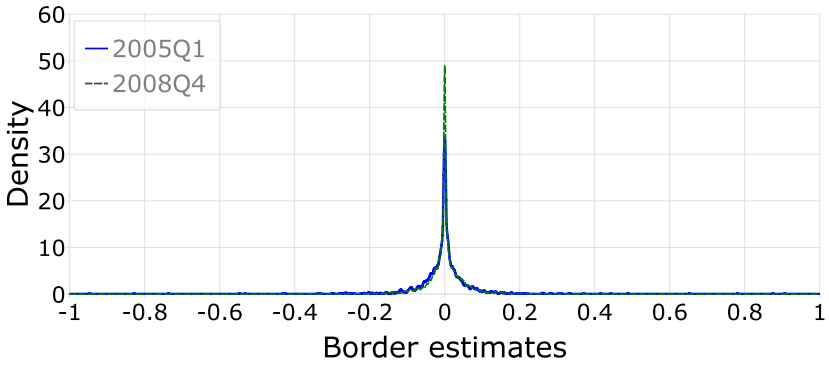
(b) LSax-NRW



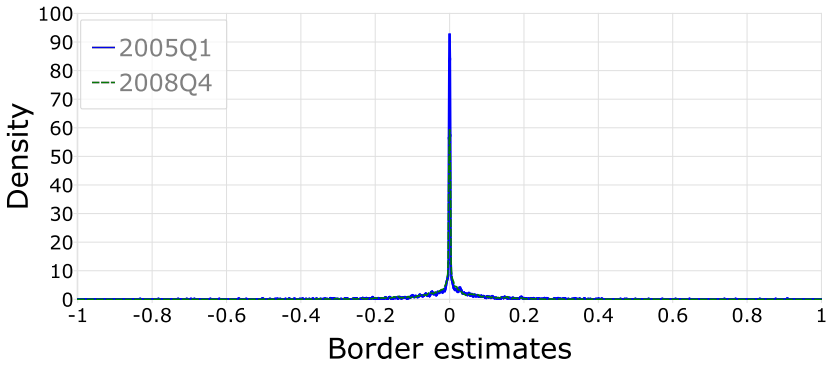
(c) NLE-NLS

Notes: Figure 21 plots the kernel density estimates of the mean (log) cross-country price differences of all matched goods of the country pair indicated below each panel for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). In each panel, the base country is the country indicated first in the subtitle. A positive value indicates that prices are hifher in the base country than in the reference country (mentioned secondly).

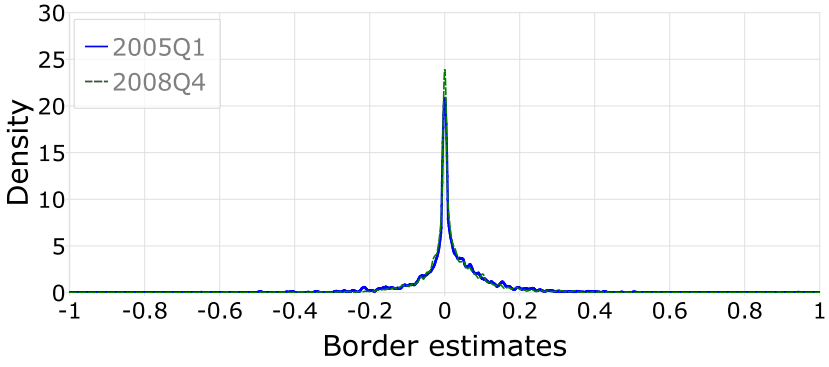
Figure 23: Kernel density estimates of border effects



(a) BEFL-BEWL



(b) LSax-NRW



(c) NLE-NLS

Notes: Figure 23 plots kernel density functions of the estimated border coefficients. The considered sample periods are 2005Q1 and 2008Q4. EFL-BEWL denotes the Flanders-Wallonia, LSax-NRW the Lower Saxony-North Rhine-Westphalia and NLE-NLS the Eastern Netherlands-Southern Netherlands values. The border dummy takes the value one if the household is located in the first country of a given country pair. Thus, positive values for the BE-NL border coefficient indicate that prices are higher in Belgium than in the Netherlands.

A Estimation of demand elasticities

A.1 Consumer choice model

Our estimates of households' demand elasticities are based on a nested multinomial mixed logit model similar to that used by Gordon, Goldfarb, and Li (2013). The upper nest contains the categories whereas the lower nest represents the various EAN-level goods of a given category. As Briesch, Dillon, and Blattberg (2008) note the estimation of demand elasticities at the individual goods (EAN) level has the advantage to avoid estimation biases caused by aggregation.

Having decided about the goods category, the utility of household j from choosing alternative i of category s is given by:

$$U_{ijt} = \alpha_{ij} - \beta_{ij} p_{jt} + \gamma I_{ij,t-1} + \delta x_{ijt} + \varepsilon_{ijt}. \tag{5}$$

p_{it} denotes the price of good i in period t and $I_{ij,t-1}$ is a dummy variable indicating whether household j bought good i in period $t-1$, i.e., $I_{ij,t-1}$ is an indicator for state dependence. x_{ijt} represents household-goods specific variables. To model consumer heterogeneity we assume the coefficients for the price and state dependence to be normally distributed.

Utility from choosing category s is given by

$$u_{it} = \phi IV_{it} + \nu_{it}, \tag{6}$$

where IV_{it} denotes the inclusive value of upper nest. ν_{it} is assumed to be i.i.d. extreme value.

A.2 Data pruning

To prune our data we apply procedures standardly used in the literature of estimating demand elasticities based on household scanner data (see, e.g., Gordon, Goldfarb, and Li, 2013 or Sun, Neslin, and Srinivasan, 2003). More specifically, we proceed as follows (unless anything to the contrary is stated the individual data steps are applied at the category level):

1. First, we remove uncommon package sizes. To this end, we compute relative frequency shares of each available volume size, sort them by frequency share and select all packages that yield a cumulative market share of at least 90% or until we have included all package sizes with a market share of bigger than 3%.

2. Thirdly, we focus on most regular buyers and drop households that make very rare purchases. We require a household to make at least five purchases per year (see also M. P. Keane and Wasi, 2012).

3. To identify outliers, we drop all prices (measured in cents per mg/ml) that deviate by more than two standard deviations by a good's mean price. (XXX Reference)

4. Finally, we require each good to have a minimum number of observations available in each year. We drop all EANs which have been purchased less than 10 times per year.

A.3 Constructing households' choice sets

To estimate the nested multinomial mixed logit model we need the complete choice set of all goods of a given category that a purchasing household faced. However, for each transaction we only have available the price of the good that the household bought. To fill in the prices of the missing goods we proceed as follows:

1. First, we check whether we can find households that bought the alternative goods in the same region (NUTS2) and same time period (quarter). If we find several price observations we take the (geographical and time) average.

2. If we can't find a price for an alternative good in the same region and time period we try the preceeding and following period (of the same region).

3. If this step is not successful we look for prices in the two (of four) closest regions in the same period.

4. If this is still not successful we employ the good's price averaged over all periods in the same region (as in Ching, Erdem, and M. Keane (2009)).

A.4 Instruments

To address endogeneity problems we employ the control function approach developed by Petrin and Train (2010). As instruments, we employ the quarterly mean of normalised prices of the region, which is furthest (second furthest) away from the region we consider. If we can't find a price in these regions we employ a good's average price (over time).

A.5 Estimation procedure

To estimate the multinomial mixed logit model we proceed as follows:

1. First, for each purchase we construct a full choice set as described in Appendix A.3). We use real price that the household paid for the alternative bought. Prices for other alternatives we impute using regional prices.

2. Given the normally large size of choice sets we then follow the suggestion by M. P. Keane and Wasi (2012), randomly select 9 EANs of each choice set and add these to the actually made choice.

3. To control for endogeneity we compute the residual from regressing a household's actually paid price on the instrument described in Appendix A.4, i.e., we compute an estimate of η_{uijt} from

$$p_{uijt} = \theta Z_{uijt} + \eta_{jt} \tag{7}$$

, where Z_{uijt} corresponds to the price of good u in the furthest (second furthest) region.

4. We then estimate the following brand-choice model:

$$U_{ijt} = \alpha_{ij} - \beta_{ij} p_{jt} + \gamma I_{ij,t-1} + \delta x_{ijt} + \lambda \eta_{uijt} + \varepsilon_{ijt} \tag{8}$$

5. Finally, we estimate the standard logit model for the upper nest.

To compute elasticities for each EAN we aggregate elasticities over households elasticities. In doing so we drop the elasticities that are in 1st% percentile of price distribution or above the 99th% percentile.

A.6 References - Appendix A

Briesch, Richard A, William R Dillon, and Robert C Blattberg (2008). "Treating zero brand sales observations in choice model estimation: Consequences and potential remedies". In: *Journal of Marketing Research* 45.5, pp. 618–632.

Ching, Andrew, Tülin Erdem, and Michael Keane (2009). "The price consideration model of brand choice". In: *Journal of Applied Econometrics* 24.3, pp. 393–420.

Gordon, Brett, Avi Goldfarb, and Yang Li (Feb. 2013). "Does price elasticity vary with economic growth? A cross-category analysis". In: *Journal of Marketing Research* 50.1, pp. 4–23.

Keane, Michael P. and Nada Wasi (2012). "Estimation of discrete choice models with many alternatives using random subsets of the full choice set. With an application to demand for frozen pizza". Nuffield College Economics Working Papers. Oxford. URL: `http://www.nuffield.ox.ac.uk/economics/papers/2012/Random%20Choice%20sets%20-%20Final%20-%20Oct%203,%202012.pdf`.

Petrin, Amil and Kenneth Train (2010). “A control function approach to endogeneity in consumer choice models”. In: *Journal of Marketing Research* 47.1, pp. 3–13.

Sun, Baohong, Scott A Neslin, and Kannan Srinivasan (2003). “Measuring the impact of promotions on brand switching when consumers are forward looking”. In: *Journal of Marketing Research* 40.4, pp. 389–405.

B Additional tables

Table 18: Regression discontinuity results: Model including additional covariates

	BE-NL				GE-NL				BE-GE			
2005Q1	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.003	0.000	0.113	13.2	-0.032	-0.025	0.165	18.9	-0.018	-0.019	0.142	16.7
Border	0.074	0.057	0.302	74.7	0.001	0.003	0.219	60.8	-0.005	0.011	0.256	50.0
Border x Distance	-0.002	0.000	0.247	13.0	0.037	0.036	0.223	18.9	0.018	-0.014	0.190	11.1
Middle aged	-0.005	0.000	0.053	11.6	-0.017	-0.009	0.070	14.7	0.009	0.012	0.082	16.7
Old	0.004	0.000	0.070	11.2	-0.003	-0.002	0.080	14.7	0.024	0.011	0.080	5.6
Medium inc.	0.002	0.000	0.071	9.4	0.000	0.003	0.058	11.2	0.018	0.011	0.079	22.2
High inc.	0.006	0.000	0.070	8.5	0.007	0.006	0.069	13.3	0.011	-0.001	0.095	22.2
Constant	-0.741	-0.866	1.450	94.8	-0.765	-0.610	1.200	90.9	-0.611	-0.450	1.311	83.3
Observations	744.000	.	.	.	143.000	.	.	.	18.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.002	0.000	0.103	12.8	0.000	-0.004	0.160	19.8	0.043	0.011	0.211	8.5
Border	0.106	0.102	0.288	81.0	-0.027	-0.025	0.262	56.1	0.100	0.078	0.191	36.2
Border x Distance	-0.015	-0.001	0.175	12.6	0.022	0.028	0.232	20.3	-0.075	-0.046	0.274	12.8
Middle aged	-0.001	0.000	0.052	11.6	-0.009	-0.010	0.050	10.2	0.006	0.004	0.067	21.3
Old	0.005	0.000	0.067	11.7	0.009	0.008	0.080	12.3	0.022	0.011	0.100	25.5
Medium inc.	0.000	0.000	0.051	11.9	0.001	0.004	0.073	18.7	-0.001	-0.002	0.065	12.8
High inc.	0.004	0.000	0.054	9.4	0.005	0.009	0.082	17.6	0.004	-0.003	0.093	21.3
Constant	-0.615	-0.724	1.453	95.8	-0.435	-0.475	1.144	91.4	-0.352	-0.345	0.970	80.9
Observations	1132.000	.	.	.	187.000	.	.	.	47.000	.	.	.

Notes: Table 18 reports regression discontinuity results for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2). For each country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. Positive values of the border coefficient indicate that prices are higher in the reference country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level. The chosen bandwidth in all specifications is 80km.

Table 19: Regression discontinuity results, absolute values: Model including additional covariates

	BE-NL				GE-NL				BE-GE			
2005Q1	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.062	0.029	0.095	13.2	0.111	0.070	0.126	18.9	0.094	0.046	0.106	16.7
Border	0.204	0.147	0.235	74.7	0.173	0.148	0.133	60.8	0.190	0.158	0.165	50.0
Border x Distance	0.100	0.049	0.226	13.0	0.161	0.120	0.158	18.9	0.126	0.064	0.140	11.1
Middle aged	0.030	0.013	0.044	11.6	0.048	0.032	0.053	14.7	0.043	0.016	0.070	16.7
Old	0.039	0.018	0.058	11.2	0.059	0.044	0.054	14.7	0.058	0.061	0.058	5.6
Medium inc.	0.029	0.012	0.065	9.4	0.041	0.028	0.040	11.2	0.058	0.039	0.054	22.2
High inc.	0.033	0.015	0.062	8.5	0.049	0.033	0.049	13.3	0.065	0.038	0.069	22.2
Constant	1.277	1.047	1.010	94.8	1.050	0.669	0.958	90.9	0.916	0.474	1.107	83.3
Observations	744.000	.	.	.	143.000	.	.	.	18.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.059	0.026	0.084	12.8	0.105	0.066	0.120	19.8	0.127	0.073	0.173	8.5
Border	0.210	0.155	0.224	81.0	0.188	0.139	0.185	56.1	0.166	0.128	0.136	36.2
Border x Distance	0.098	0.045	0.145	12.6	0.169	0.118	0.160	20.3	0.178	0.099	0.220	12.8
Middle aged	0.025	0.011	0.046	11.6	0.038	0.030	0.035	10.2	0.049	0.038	0.045	21.3
Old	0.035	0.016	0.057	11.7	0.055	0.040	0.058	12.3	0.071	0.049	0.073	25.5
Medium inc.	0.026	0.011	0.044	11.9	0.049	0.033	0.054	18.7	0.042	0.026	0.049	12.8
High inc.	0.027	0.012	0.047	9.4	0.055	0.036	0.062	17.6	0.060	0.043	0.070	21.3
Constant	1.209	0.986	1.014	95.8	0.851	0.574	0.878	91.4	0.711	0.437	0.742	80.9
Observations	1132.000	.	.	.	187.000	.	.	.	47.000	.	.	.

Notes: Table 19 reports regression discontinuity results for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2). For each country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of the estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level. The chosen bandwidth in all specifications is 80km.

Table 20: Regression discontinuity results, model including additional covariates: Household preferences, BE-NL

2005Q1

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-4.214***	-1.983***	0.554**	0.0533***
	(-3.82)	(-3.32)	(2.19)	(3.43)
Border	8.496***	-2.915***	0.503***	-0.0323***
	(6.62)	(-7.26)	(3.26)	(-3.29)
Border × Distance	-22.45***	-3.769***	1.722***	-0.143***
	(-6.45)	(-3.81)	(4.83)	(-5.97)
Middle aged	-2.058*	-1.582***	1.297***	-0.0241***
	(-1.92)	(-4.62)	(13.04)	(-3.59)
Old	-8.191***	-4.086***	1.319***	-0.0230***
	(-7.19)	(-11.40)	(11.22)	(-2.68)
Medium inc.	5.205***	1.438***	0.480***	-0.00693
	(6.84)	(6.61)	(5.29)	(-1.12)
High inc.	9.791***	2.373***	0.568***	-0.0184**
	(8.34)	(7.14)	(4.43)	(-2.37)
Constant	19.65***	12.28***	3.129***	0.444***
	(16.84)	(28.11)	(20.80)	(43.29)
Observations	4485	4485	4485	6151
R^2	0.199	0.068	0.054	0.013

Notes: Table 21 reports regression discontinuity results for households preference variables for the first quarter of 2005 (2005Q1). Results are based on estimating Equation (3). The reference country is Belgium. Positive values of the border coefficient indicate that values of the respective variable are higher in the reference country. The chosen bandwidth in all specifications is 80 km. Parentheses report t statistics, signifcance of the coefficients is indicated as following: $*$ ($p < 0.10$), $**$ ($p < 0.05$), $***$ ($p < 0.01$).

Table 21: Regression discontinuity results, model including additional covariates: Household preferences, BE-NL

2008Q4

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-5.865***	-2.586***	0.706***	0.0556***
	(-6.36)	(-5.66)	(3.97)	(4.04)
Border	10.85***	-3.131***	0.552***	-0.0566***
	(10.28)	(-9.91)	(4.67)	(-6.72)
Border × Distance	-11.95***	-1.482**	1.078***	-0.124***
	(-4.28)	(-1.97)	(4.01)	(-6.16)
Middle aged	-0.932	-1.096***	1.219***	-0.0412***
	(-1.22)	(-4.79)	(19.38)	(-7.16)
Old	-6.868***	-3.679***	1.640***	-0.0589***
	(-8.25)	(-15.06)	(19.33)	(-8.10)
Medium inc.	4.912***	1.273***	0.527***	-0.0209***
	(8.26)	(7.07)	(7.43)	(-3.64)
High inc.	10.77***	2.556***	0.696***	-0.0315***
	(13.09)	(10.53)	(7.78)	(-4.70)
Constant	21.86***	12.38***	2.530***	0.497***
	(24.29)	(38.44)	(23.75)	(54.61)
Observations	6685	6685	6685	8058
R^2	0.203	0.066	0.066	0.037

Notes: Table 21 reports regression discontinuity results for households preference variables for the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (3). The reference country is Belgium. Positive values of the border coefficient indicate that values of the respective variable are higher in the reference country. The chosen bandwidth in all specifications is 80 km. Parentheses report t statistics, signifcance of the coefficients is indicated as following: $*$ ($p < 0.10$), $**$ ($p < 0.05$), $***$ ($p < 0.01$).

Table 22: Regression discontinuity results, model including additional covariates: Household preferences, GE-NL

2005Q1

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-1.695	-0.232	0.758***	0.0160
	(-1.62)	(-0.40)	(3.12)	(1.05)
Border	-5.546***	-3.243***	0.592***	-0.0844***
	(-8.98)	(-9.48)	(3.41)	(-8.79)
Border × Distance	2.430*	0.323	-0.926**	-0.0284
	(1.88)	(0.45)	(-2.49)	(-1.39)
Middle aged	-0.234	-0.547**	1.045***	-0.0178***
	(-0.46)	(-2.00)	(10.02)	(-2.76)
Old	-4.363***	-2.926***	1.416***	-0.0217***
	(-8.47)	(-10.66)	(11.68)	(-2.77)
Medium inc.	3.359***	1.507***	0.317***	-0.00901
	(9.06)	(7.09)	(3.03)	(-1.40)
High inc.	5.595***	1.873***	0.499***	-0.0218***
	(11.41)	(7.20)	(3.92)	(-2.90)
Constant	18.52***	10.86***	3.200***	0.454***
	(27.47)	(28.78)	(21.69)	(48.69)
Observations	4042	4042	4042	5271
R^2	0.125	0.137	0.042	0.064

Notes: Table 23 reports regression discontinuity results for households preference variables for the first quarter of 2005 (2005Q1). Results are based on estimating Equation (3). The reference country is Germany. Positive values of the border coefficient indicate that values of the respective variable are higher in the reference country. The chosen bandwidth in all specifications is 80 km. Parentheses report t statistics, signifcance of the coefficients is indicated as following: $*$ ($p < 0.10$), $**$ ($p < 0.05$), $***$ ($p < 0.01$).

Table 23: Regression discontinuity results, model including additional covariates: Household preferences, GE-NL

2008Q4

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-2.777***	-0.815*	0.433**	0.0476***
	(-2.83)	(-1.72)	(2.46)	(3.51)
Border	-5.946***	-3.333***	0.289**	-0.0771***
	(-9.61)	(-11.48)	(2.26)	(-8.82)
Border × Distance	2.329*	0.648	-0.514*	-0.0720***
	(1.78)	(1.05)	(-1.90)	(-3.87)
Middle aged	0.256	-0.713***	1.207***	-0.0423***
	(0.62)	(-3.62)	(19.01)	(-7.58)
Old	-4.280***	-3.104***	1.866***	-0.0609***
	(-9.60)	(-14.82)	(21.12)	(-8.80)
Medium inc.	4.569***	1.812***	0.308***	-0.0215***
	(12.99)	(10.88)	(3.95)	(-3.69)
High inc.	8.788***	2.871***	0.418***	-0.0215***
	(19.24)	(13.67)	(4.52)	(-3.22)
Constant	20.04***	11.09***	2.724***	0.498***
	(35.56)	(40.66)	(27.93)	(60.89)
Observations	6499	6499	6499	7545
R^2	0.102	0.108	0.059	0.066

Notes: Table 23 reports regression discontinuity results for households preference variables for the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (3). The reference country is Germany. Positive values of the border coefficient indicate that values of the respective variable are higher in the reference country. The chosen bandwidth in all specifications is 80 km. Parentheses report t statistics, signifcance of the coefficients is indicated as following: $*$ ($p < 0.10$), $**$ ($p < 0.05$), $***$ ($p < 0.01$).

Table 24: Regression discontinuity results, model including additional covariates: Household preferences, BE-GE

2005Q1

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	-3.218***	-1.200**	0.642	0.0409*
	(-2.68)	(-2.05)	(1.47)	(1.90)
Border	33.22***	5.443***	-0.976***	0.113***
	(8.34)	(6.26)	(-2.82)	(4.97)
Border × Distance	19.08***	5.306***	-0.953	0.0136
	(2.63)	(3.26)	(-1.48)	(0.31)
Middle aged	-1.339	-1.458**	0.830***	-0.00738
	(-0.63)	(-2.25)	(5.16)	(-0.65)
Old	-5.820***	-3.386***	1.243***	-0.00564
	(-2.85)	(-5.26)	(6.62)	(-0.40)
Medium inc.	6.106***	1.487***	0.500***	-0.0249**
	(3.09)	(3.71)	(2.97)	(-2.01)
High inc.	7.640***	1.735***	0.565***	-0.0400***
	(3.46)	(3.74)	(2.74)	(-2.99)
Constant	13.61***	8.958***	3.419***	0.361***
	(5.73)	(12.48)	(11.22)	(19.68)
Observations	1350	1350	1350	1769
R^2	0.295	0.167	0.102	0.050

Notes: Table 21 reports regression discontinuity results for households preference variables for the first quarter of 2005 (2005Q1). Results are based on estimating Equation (3). The reference country is Belgium. Positive values of the border coefficient indicate that values of the respective variable are higher in the reference country. The chosen bandwidth in all specifications is 80 km. Parentheses report t statistics, signifcance of the coefficients is indicated as following: $*$ ($p < 0.10$), $**$ ($p < 0.05$), $***$ ($p < 0.01$).

Table 25: Regression discontinuity results, model including additional covariates: Household preferences, BE-GE

2008Q4

	PurchAmount	PurchItems	PurchFreq	RetLoyalty
Distance	1.258	0.652	-0.247	0.0253
	(0.92)	(1.10)	(-0.78)	(1.25)
Border	29.49***	4.522***	-0.916***	0.0592***
	(11.56)	(6.50)	(-3.52)	(3.13)
Border × Distance	3.647	2.218*	-0.273	0.0155
	(0.72)	(1.65)	(-0.55)	(0.44)
Middle aged	-1.199	-1.207***	0.938***	-0.0111
	(-1.08)	(-3.72)	(9.18)	(-1.39)
Old	-6.620***	-3.357***	1.516***	-0.0281***
	(-5.44)	(-9.49)	(10.56)	(-2.60)
Medium inc.	6.465***	1.286***	0.412***	-0.0389***
	(5.52)	(3.97)	(3.17)	(-3.96)
High inc.	9.506***	1.626***	0.628***	-0.0537***
	(7.62)	(4.50)	(4.17)	(-5.09)
Constant	14.38***	8.687***	3.139***	0.415***
	(9.73)	(16.97)	(14.19)	(26.41)
Observations	2305	2305	2305	2881
R^2	0.302	0.094	0.062	0.020

Notes: Table 21 reports regression discontinuity results for households preference variables for the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (3). The reference country is Belgium. Positive values of the border coefficient indicate that values of the respective variable are higher in the reference country. The chosen bandwidth in all specifications is 80 km. Parentheses report t statistics, signifcance of the coefficients is indicated as following: $*$ ($p < 0.10$), $**$ ($p < 0.05$), $***$ ($p < 0.01$).

Table 26: Regression discontinuity results: Model including additional covariates

	BEFL-BEWL				LSax-NRW				NLE-NLS			
2005Q1	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	-0.005	0.000	0.262	12.2	-0.006	0.000	0.151	9.4	0.008	0.000	0.241	11.3
Border	-0.003	0.000	0.102	12.2	0.000	0.000	0.120	7.6	0.018	0.001	0.117	13.6
Border x Distance	-0.011	0.000	0.296	10.8	0.009	0.000	0.234	8.5	0.016	0.000	0.311	11.2
Middle aged	0.001	0.000	0.082	14.0	-0.003	0.000	0.106	12.6	-0.003	0.000	0.085	11.7
Old	0.000	0.000	0.093	12.9	0.007	0.000	0.117	11.6	0.005	0.000	0.110	12.0
Medium inc.	-0.001	0.000	0.054	10.6	-0.003	0.000	0.093	10.9	0.002	0.000	0.070	10.6
High inc.	-0.002	0.000	0.070	10.3	0.001	0.000	0.107	10.2	0.008	0.000	0.102	11.7
Constant	-0.579	-0.895	1.923	92.1	-1.109	-1.152	1.269	65.0	-0.995	-1.222	1.753	85.2
Observations	2627.000	.	.	.	2564.000	.	.	.	2472.000	.	.	.
2008Q4	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	-0.006	0.000	0.508	10.0	-0.012	0.000	0.158	11.2	-0.002	0.000	0.190	11.2
Border	0.003	0.000	0.155	10.6	-0.008	0.000	0.122	9.1	0.017	0.003	0.104	13.3
Border x Distance	0.006	0.000	0.589	9.3	0.012	0.000	0.241	10.8	0.040	0.005	0.261	13.5
Middle aged	0.000	0.000	0.103	12.4	-0.005	0.000	0.100	10.7	-0.005	0.000	0.077	14.0
Old	0.004	0.000	0.149	12.6	0.005	0.000	0.112	11.5	0.008	0.000	0.110	14.8
Medium inc.	-0.003	0.000	0.116	11.3	0.003	0.000	0.091	13.0	0.003	0.000	0.065	11.4
High inc.	-0.004	0.000	0.134	10.6	0.009	0.000	0.101	11.5	0.009	0.000	0.077	11.9
Constant	-0.393	-0.711	1.896	89.5	-0.897	-0.958	1.419	73.0	-0.903	-1.071	1.536	91.5
Observations	4170.000	.	.	.	3708.000	.	.	.	4135.000	.	.	.

Notes: Table 26 reports regression discontinuity results for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2). For each country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. Positive values of the border coefficient indicate that prices are higher in the reference country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level. The chosen bandwidth in all specifications is 80km.

Table 27: Regression discontinuity results: Model including additional covariates

	BEFL-BEWL				LSax-NRW				NLE-NLS			
2005Q1	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.112	0.046	0.237	12.2	0.073	0.016	0.133	9.4	0.122	0.050	0.207	11.3
Border	0.042	0.018	0.093	12.2	0.056	0.011	0.106	7.6	0.067	0.031	0.097	13.6
Border x Distance	0.139	0.062	0.262	10.8	0.112	0.025	0.206	8.5	0.166	0.079	0.264	11.2
Middle aged	0.032	0.012	0.075	14.0	0.048	0.008	0.094	12.6	0.044	0.018	0.073	11.7
Old	0.040	0.016	0.084	12.9	0.054	0.010	0.104	11.6	0.061	0.027	0.092	12.0
Medium inc.	0.026	0.011	0.048	10.6	0.043	0.008	0.083	10.9	0.039	0.018	0.057	10.6
High inc.	0.033	0.014	0.061	10.3	0.049	0.010	0.095	10.2	0.053	0.023	0.088	11.7
Constant	1.560	1.209	1.264	92.1	1.406	1.287	0.929	65.0	1.657	1.465	1.146	85.2
Observations	2627.000	.	.	.	2564.000	.	.	.	2472.000	.	.	.
2008Q4	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)	Mean	Median	Std.	Sign.(%)
Distance	0.107	0.034	0.497	10.0	0.076	0.023	0.139	11.2	0.101	0.044	0.162	11.2
Border	0.043	0.014	0.149	10.6	0.060	0.018	0.107	9.1	0.058	0.026	0.089	13.3
Border x Distance	0.140	0.048	0.572	9.3	0.121	0.037	0.208	10.8	0.145	0.063	0.221	13.5
Middle aged	0.030	0.009	0.099	12.4	0.043	0.011	0.090	10.7	0.038	0.017	0.066	14.0
Old	0.039	0.013	0.144	12.6	0.054	0.015	0.099	11.5	0.056	0.023	0.094	14.8
Medium inc.	0.030	0.010	0.112	11.3	0.045	0.012	0.080	13.0	0.036	0.016	0.054	11.4
High inc.	0.035	0.012	0.129	10.6	0.050	0.014	0.088	11.5	0.043	0.020	0.065	11.9
Constant	1.449	1.057	1.285	89.5	1.349	1.147	1.000	73.0	1.443	1.238	1.045	91.5
Observations	4170.000	.	.	.	3708.000	.	.	.	4135.000	.	.	.

Notes: Table 27 reports regression discontinuity results for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2). For each country pair (indicated in the first row of each sample period), the reference country is the respectively firstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of the estimated coefficients given in the first column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coefficient is significant at the 10% level. The chosen bandwidth in all specifications is 80km.

CENTER FOR FINANCIAL STUDIES

CFS Working Paper Series

Recent Issues

All CFS Working Papers are available at *www.ifk-cfs.de*.

No.	Authors	Title
535	Christian Rühl	*Size and welfare costs of price differences across European countries*
534	Günter Beck, Sarah M. Lein	*Microeconometric evidence on demand-side real rigidity and implications for monetary nonneutrality*
533	Söhnke M. Bartram, Gregory Brown, and René M. Stulz	*Why does idiosyncratic risk increase with market risk?*
532	Ludger Schuknecht	*The supply of "safe" assets and fiscal policy*
531	Volker Brühl	*Die Kosten der Flüchtlingskrise in Deutschland – eine Investition in die Zukunft?*
530	Andrea Zaghini	*Fragmentation and Heterogeneity in the Euro – Area Corporate Bond Market: Back to normal?*
529	Qing He, Liping Lu, Steven Ongena	*Who Gains from Credit Granted between Firms? Evidence from Inter-corporate Loan Announcements Made in China*
528	Christina E. Bannier and Milena Neubert	*Actual and perceived financial sophistication and wealth accumulation: The role of education and gender*
527	Dae Woong Kang, Nick Ligthart, and Ashoka Mody	*The European Central Bank: Building a Shelter in a Storm*
526	Peter A. Zadrozny	*Extended Yule-Walker - Identification Of Varma Models With Single- Or Mixed Frequency Data*

Center for Financial Studies | Goethe University | Theodor-W.-Adorno-Platz 3 | 60323 Frankfurt am Main | Germany
Phone +49 69 798 30050 | Fax +49 69 798 30077 | info@ifk-cfs.de | www.ifk-cfs.de