UNIVERSITÄT SIEGEN

Essays in macroeconomics using scanner price data

DISSERTATION

zur Erlangung des Grades einer
Dr. rer. pol.
der Fakultät III – Wirtschaftswissenschaften, Wirtschaftsinformatik und Wirtschaftsrecht der Universität Siegen

vorgelegt von

M.Sc. Natalia Zabelina

Erstgutachter: Univ.-Prof.. Dr. Günter W. Beck
Zweitgutachter: Univ.-Prof.. Dr. Sebastian G. Kessing

Datum der Disputation: 10.12.2021
Dekan der Fakultät III: Univ.-Prof. Dr. Marc Hassenzahl

Declaration of Authorship

I, M.Sc. Natalia Zabelina, declare that this thesis titled "Essays in macroeconomics using scanner price data" has been composed by me and is based on my own research work, unless mentioned otherwise. This work was done mainy while in candidature for a doctoral degree at the University of Siegen. Where parts of the thesis are based on jointly work with others, it has been clearly stated. This thesis has not been previously submitted for a adegre or any other quantili cation at the University of Siegen or any other institution. All references and verbatim extracts have been quoted, and all sources of information have been speci cally acknowledged.

Signature:

_____

Date:

_____

Acknowledgements

I would like to thank my supervisor, Univ.-Prof.. Dr. Günter W. Beck, for the introduction to the exicitng eld of scanner price data research, valuable advices, and unwavering guidance during all stages of this dissertation. Your expertise and insightful feedback were invaluable in formulating the research questions and sharpening my thinking allowing me to bring my work to a higher level. I would also like to express my deepest appreciation to the members of the defence committee, Prof. Dr. Sebastian G. Kessing and Prof. Dr. Matthias Hunold, who have kindy agreed to support the nal steps of the dissertaion and examine my work.

I am extremely grateful to Aimark and Europanel for providing me with scanner price data used in this dissertaion. I am also very grateful to Alfred Dijs and Bernadette van Ewijk for organising the access to the data, their insightfull comments, and practical recommendations in dealing with the data. Without your support and insightful suggestions the completion of my dissertation would not have been possible.

I would also like to extend my deepest gratitude to the Chair of Macroeconomic Studies and Faculty III: School of Economic Disciplines for providing me with a great opportunity to gain invaluable insights into research and teaching activities of the University of Siegen while working on my dissertation. My sincere thanks to my colleages for helpfull advices and unwavering support.

My greatest thanks to my husband, family, and friends for profound belief in my abilities and the greatest support one could ever wish for.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The number of studies in macroeconomics employing scanner price data has increased signi cantly over the last two decades. Scanner price data possess several attractive properties that motivate researchers to exploit the data in empirical analyses. First, the granularity of scanner price data provides a unique opportunity to analyse prices of the same good bought in different locations and in different time periods. Previously, researchers studying price dynamics across space were using highly-aggregated price data, such as city consumer price index, thus the price comparisons always re ected not only the pure price difference but also the differences in the baskets underlying the computation of these price indexes. Scanner data allow to avoid this problem and to study the pure price differences that are not caused by differences in quality or characteristics of compared goods. Second, scanner data provide information on actual transactions, quantity bought, and prices paid. This information is extremely valuable to the analysis of consumers behaviour because with these data researchers can observe and study any changes in the market shares of goods, retailers, or markets. Third, the coverage of goods in scanner price data presents the whole universe of goods that consumers buy, which is much bigger than statistical agencies survey. Additionally, the new goods entering the market, which are not yet identi ed by a statistical of ce as signi cant enough, can be already observed in scanner price data.

This dissertation employs household scanner price data for Belgium, Germany, the Netherlands, and Spain. The scanner price data are provided by Aimark[1] and collected

[1] Aimark is a non-pro t foundation that supports academic research by providing access to various price data on behalf of its data partners (GfK, Kantar, and others). More information about the organisation can be obtained on the Aimark s website: https://aimark.net/.

from a representative panel of households operated by GfK or Kantar. In each country, the households are provided with a scanning device and scan all their purchases in the category of the fast moving consumer goods (FMCG). More speci cally, for each purchased good the households provide the date of purchase, price paid, quantity bought, and the name of the retailer where the good was purchased. Additionally, the data providers supply the data with sociodemographic characteristics of the households and detailed product characteristics, e.g., brand, product category, product volume, manufacturer, etc. Using these data the empirical analysis in the three subsequent chapters of this dissertation investigates the following questions: (i) are the markets in the Euro area countries segmented or integrated and how big are the cost generated by national borders, (ii) how do regional in ation rates develop in Germany and Spain and how signi cant are the biases in in ation rates computed by statistical of ces, and (iii) how sensitive are regional in ation rates computed with the use of different price indexes to local economic condition?

Chapter 2 presents the study, co-authored with Günter W. Beck and Hans-Helmut Kotz, which answers the rst question that is to what extent does a national border generates additional costs segmenting international good markets. A big surge of the literature in international macroeconomics starting from the seminal study of Engel and Rogers (1996) aim to measure the costs preventing the arbitrage of price differences across international markets. The existing evidence for European countries is inconclusive and is limited to a narrow set of goods (see, e.g. Imbs et al. (2010), Fischer (2012), Dvir and Strasser (2018)), online prices of a few retailers (see Cavallo, Neiman, and Rigobon (2014)), higher level of goods aggregation or a few locations (see, e.g., Engel and Rogers (2004), Kulikov (2014), Reiff and Rumler (2014)). Our study adds to this literature by employing household scanner price data for the three highly-integrated Euro area countries: Belgium, Germany, and the Netherlands and adopting the estimation approach of Gopinath et al. (2011) to identify the border cost. The data allow us to identify a substantial number of identical goods bought in different countries, make an inference on whether the markets are segmented or integrated by using the information about the cross-border shopping, and estimate the border cost empirically by applying a regression discontinuity design using the data on prices that households living close to the

border pay for the same good in different countries.

Chapter 3 addresses the question of regional in ation rates computation. Statistical of ces compute in ation rates using the measures of in ation that are biased due to ignoring consumers substitution patterns, product turnover, and shifts in consumers preferences. The "substitution bias" arises when a price index does not take into account shifts in expenditure shares due to the changes in the relative prices of goods, the "variety bias" arises due to the fact that new goods are not introduced to the basket immediately after they appear at the market, and the "consumer valuation bias" arises when a price index does not account for expenditure shifts caused by the changes in consumers valuation of goods. In this study, I use household scanner price data for Germany and Spain to compute regional in ation rates using different types of price indexes and to quantify the aforementioned biases using the approach developed in Stephen J. Redding and David E. Weinstein (2020).

Chapter 4 presents the study that investigates the sensitivity of regional in ation rates to local economic conditions. The relationship between in ation and economic activity usually investigated using the of cial " xed basket" price indexes. These price indexes do not allow to fully understand the effects of economic uctuations on consumer welfare though, since they do not take into account changes in consumer preferences or changes in the consumer utility due to the changes in the variety of available products. Empirical literature shows that the creation of new products is procyclical (see, e.g. Broda and D. Weinstein (2010)) and, therefore, can potentially amplify the effects of business cycles on consumer welfare. Additionally, it is documented in the literature that the shopping behaviour of consumers also changes as economic conditions deteriorate, such that consumers switch to the less expensive shops, increase their purchases on sales and buy goods of worse quality (see Argente and M. Lee (2017), Nevo and Wong (2019), Coibion, Gorodnichenko, and Hong (2015), and Stroebel and Vavra (2019)). Hence, a measure of in ation that takes into account changes in consumer preferences towards different goods and incorporates changes in product variety would provide with a better understanding how business cycles affect consumer welfare. To explore this question, in this study, I use regional in ation rates computed using scanner price data to investigate how different measures of in ation react to local economic conditions in Germany and Spain.

Chapter 2

Price gaps at the border: evidence from multi-country household scanner data

Guenter W. Beck, Hans-Helmut Kotz, and Natalia Zabelina (published as: Beck, G. W., Kotz, H. H., & Zabelina, N. (2020). Price gaps at the border: Evidence from multi-country household scanner data. Journal of International Economics, 127, 103368.)[1]

2.1 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.[2]

The more recent literature on the topic has progressed relatively to earlier studies basically with respect to two dimensions.[3] First, some studies have been concerned with

[1]We would like to thank the editor and two anonymous referees for their comments and suggestions which provided considerable value-adde for the paper. We are 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.

[2]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 Of cer (1976), Mussa (1979) and Rogoff (1996).

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

re ning and extending the econometric methods and identi cation strategies used to determine a 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 or regions, ER is prone to considerably overestimate the effect of borders. Accounting for this heterogeneity, the authors nd that the border effect between the U.S. and Canada not only becomes substantially smaller but almost completely vanishes. Gopinath et al. (2011) (GGHL) - who also address the second extension of the original literature to be discussed in the next paragraph - demonstrate that the identi cation strategy of ER rests on the assumption that transaction costs within countries are homogeneous across households as well as 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 a possible border effect. They nd 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.[4] Consequently, nearly all more recent studies on the LOOP, such as Broda and D. Weinstein (2008) or the above mentioned work by Gopinath et al. (2011) use price data at a highly disaggregated level, optimally at a level where goods can be identi ed by their GTIN (global trade item number).[5] While these studies have in common that they tend to detect smaller border effects than those regularly uncovered in earlier studies, their overall ndings are somewhat mixed. An extreme is given by Cavallo, Neiman, and Rigobon (2014) who, employing online prices of identical goods sold by four major global retailers in a large set of countries, show that LOOP holds perfectly

[4]The problem that can arise from working with prices of not identical goods baskets across countries is very clearly illustrated by Broda and D. Weinstein (2008). Using a large data set of retail prices these authors show that *i*) even narrowly de ned 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 et al. (2005). These authors demonstrate that the time series properties of LOOP deviations are heavily in uenced by the level of aggregation one considers.

[5]A GTIN uniquely identi es 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 often used.

within currency unions but does not do so outside them, even if exchange rates are nominally fixed (allegedly). As noted above, GGHL, on the other hand, find evidence in favor of sizable border effects.[6]

Our paper extends the existing literature on intra- versus international price dispersion with respect to two major 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 as well as the geographical dimension.[7] Moreover, our data sample comprises price information from all retailers, generally present in a given market. Our data allow us, inter alia, to employ the approach by Gopinath et al. (2011), providing credibly identified estimates of border effects for our European sample, comprised of Belgium (BE), Germany (GE) and the Netherlands (NL). Our results are interesting, given that the countries involved represent a highly pertinent comparison (control) group to the U.S.-Canadian case: our sample countries are not only members of a deeply integrated economic area (the European Union), sharing close historical and cultural ties, partly including a common language. Moreover, as members of the eurozone (i.e. sharing a common currency), they are also not faced with any expectations of changing nominal exchange rates, dominating short-run real exchange rate dynamics when national currencies exist. In addition, considering the factors listed in McAfee (2008) preventing or limiting resale capabilities[8], ever since the European Single Act (of 1986), such non-tariff barriers have almost completely vanished. Thus, any (statistically as well as economically) significant price discontinuity must be due to other barriers to arbitrage, for which we provide a quantitative assessment. The multi-retailer dimension in turn allows us to assess the role that differences in the composition of retailers across national markets have on the obtained border estimates. This is novel, given that comparable previous

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

[7] Existing studies, using similarly granular micro price data, have instead focused on the U.S. and Canada, whereas studies employing European data are confined to very few goods (see, e.g., Fischer, 2012, Imbs et al., 2010 or Dvir and Strasser, 2018). In contrast, other existing, micro-data based studies build on prices at a higher level of aggregation and are, moreover, confined to few locations (see, e.g., Engel and Rogers, 2004, Reiff and Rumler, 2014, Kulikov, 2014).

[8] These include legal impediments to resale, thin markets and/or matching problems, informational issues, differences in contract features and warranties or significant costs of transportation.

studies generally employing data from one retailer only focussed on the role of differences in marginal costs as a source of international price differences.

Secondly, unlike most other scanner-data-based studies in the eld, we employ demand- rather than supply-side data. More speci cally, our data set comprises comprehensive information about purchases by some 30,000 households (in any given quarter). Since our data set also includes information on the location of households, we know the prices which households, located on different sides of a border, actually paid for an identical good and can therefore analyze potential discontinuous price drops at the border employing a fairly large number of geographic observations in the close neighborhood of that border. Using a somewhat con ned but nevertheless instructive set of information on cross-border shopping enables us to assess whether two border-separated markets are integrated according to the de nition by GGHL for a majority of goods included in our data sample. Moreover, the cross-border shopping information allows us to pin down the cost of crossing the border for a signi cant subsample of goods. For goods for which no cross-border shopping is observed, we can unfortunately not make any conclusive inference about the question of integration versus segmentation. However, we are at least able to provide a lower bound for border costs in the overwhelming number of cases.[9]

Our results are as follows: In line with previous ndings, we show that price dispersions for (in all pertinent attributes) identical goods are sizeable across all of our country pairs, including the seemingly very highly integrated country-pair of Belgium and the Netherlands. Our regressions results reveal that around 75% of all sample goods exhibit signi cant price gaps at the border. By contrast, counterfactual evidence for price

[9]Our demand side perspective might moreover provide a further advantage: if large, country-wide operating retailers follow a national rather than a local pricing strategy (as suggested by Li, Gordon, and Netzer, 2018 and DellaVigna and Gentzkow, 2019), then prices at their outlets close to the border would not necessarily respond to cross-border shopping by households. Instead, their price-setting would represent an optimizing approach averaging across conditions in local markets within the respective country. In this case, nationally active retailers might react to relatively lower sales in border regions by either "subsidizing" them (in the sense of Li, Gordon, and Netzer, 2018), 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). Hence, if such phenomena were present in our European data, a supply-side perspective of price differences, in the spirit of GGHL, would tend to over-estimate border effects. Referring to evidence by Campbell and Lapham (2004), showing that consumers in the U.S. and Canada indeed conduct cross-border shopping, GGHL argue that their nding of completely segmented U.S.-Canadian goods markets is not inconsistent with this result. They observe that price setting decisions by stores in their sample are not signi cantly affected by these consumers.

differences across within-country regions yields only negligible evidence of such discontinuities. Also in line with existing micro-price based studies, the estimated border effects exhibit considerable heterogeneity across goods, the distribution of which is comparable across country pairs.

The information on cross-border purchases shows that these activities are present for a sizeable amount of goods, implying that the markets for these goods are integrated in the sense of GGHL. Descriptive statistics suggest that cross-border shopping primarily takes place for goods that are purchased relatively often and which are, on average, relatively cheaper. Border estimates for those goods for which an exact measure of border costs can be derived reveal median values of these costs in the range between 15% and 20% (of the price of a good). Inspecting the distribution of estimates suggests that most values lie in a range of $\pm 40\%$. The estimates of the lower bounds of border costs obtained for goods without cross-shopping activities exhibit median values between 18% and 20%, while the distribution of the estimates is comparable to that obtained for goods with cross-border shopping information. Overall, these estimates indicate substantially and signi cantly bigger within-EMU border effects than those found by Cavallo, Neiman, and Rigobon (2014) for online prices, but somewhat smaller than those documented by GGHL for the U.S.-Canadian border.

An analysis of co-movements of prices in border and non-border regions reveals that shocks in one country are generally not transmitted to prices in another country. This is true even in the case of integrated markets, suggesting that cross-border shopping does not in uence retailer price setting within a country.

Our analysis of the role of retailer composition con rms the ndings by Li, Gordon, and Netzer (2018) and DellaVigna and Gentzkow (2019) that individual retailers seem to follow a uniform pricing strategy within a country. Across countries, they charge different prices, with the distribution of these prices having comparable features to the one derived using all observations. Analyzing prices from different retailers reveals that although the dispersion of prices increases, the average border estimate is not sizeably affected.

The rest of the paper is organized as follows. In Section 2.2, we describe our data set

and provide some descriptive statistics. Section 2.3 motivates and describes our estimation approach. the results of which are presented and discussed in Section 2.4. Moreover, this Section introduces the cross-border shopping information and analyzes the question of integration vs. segmentation of markets and the costs of crossing the border. The role of retailer composition is examined in Section 2.5. Section 2.6 looks at different sample characteristics, and studies their relationship with the obtained results. Section 4.6 summarizes and concludes.

2.2 Data and descriptive statistics

2.2.1 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 in the literature on international price dispersion so far. The data was provided by AiMark (Advanced International Marketing Knowledge), a non-pro t cooperation, promoting research in the area of retail markets and covers the three euro area countries Belgium, Germany and the Netherlands.[10] Given its nature as a consumer/household (as opposed to a retailer) panel, and as outlined in more detail below, our data is comparable to the AC Nielsen Homescan and the IRI Consumer Network data widely used in US-related research.

In each country included in the data set, the data provider (GfK) operates a panel of households (of differing sizes across countries), with panelists being chosen to constitute a representative sample of the consumers in the respective retail market. Each household is equipped with a scanning device, which it uses to scan all the products belonging to the categories of *fast-moving consumer goods* (FMCG) it buys at retail outlets. FMCG include (among others) grocery products, home and personal care products and beverages, whereas retailer information comprises all major supermarket chains (such as Rewe or Aldi in Germany or Albert Heijn and C1000 in the Netherlands), drugstores, small corner shops as well as internet stores. The scanning device is similar to the one underlying the

[10] AiMark s data is originally compiled by Europanel and its partners, i.e. Gesellschaft für Konsumforschung (GfK), Kantar 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.

TABLE 2.1: 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	5 873	90 959	730 723	3 699	20 449
BE-NL	14 809	273 237	1 037 444	3 844	6 253
GE-NL	8 097	1 277 824	154 248	20 595	6 133

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 identi ed 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 (de ned by the same GTIN) commonly purchased by households in the two countries indicated in the rst 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 rst and second country of the respective country pair.

Nielsen Homescan database. For each product bought, a household scans the barcode, uniquely identifying the product via the Global Trade Item Number (GTIN),[11] and enters volume and price paid for the product.[12] We also have a description of each product and a classi cation system of the goods into different (more aggregate) product categories which are internationally comparable. Households also provide a description of the type of store where they bought the product - including the name of the retail chain.

For all three countries included, our panel data covers the period from January 2005 to

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

[12] In [the rare] case a product does not have a barcode, households enter this information manually.

December 2008. As, for example, Table 2.1, portraying data for 2008Q4, shows, the number of observations (purchases) available in the total sample is enormous: It ranges from somewhat more than one million (per quarter) for Belgium, to more than 6 millions for Germany. These observations include purchases of more than 80,000 unique products in Belgium and the Netherlands, and more than 150,000 unique products in Germany. The number of households included in the respective country panels in 2008Q4 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 a household, its income group as well as its age category.

2.2.2 Matching results for goods across countries

Our data set allows us to make inferences on border effects using price information for identical goods, purchased by households located on different sides of a given national frontier. We assume a purchased product to be identical between two places (located either within a country or across two countries) if the purchased good has the same GTIN/EAN.[13] The lower panel of Table 2.1 shows that, while the number of matched goods is generally fairly sizeable (reaching from almost 5 800 for Belgium-Germany to around 14 800 goods for Belgium-Netherlands), these goods nevertheless represent only a relatively small share of all goods purchased by our sample households. This observation also holds for Broda and D. Weinstein (2008), who report similar results for their U.S.-Canadian data sample, concluding that most price comparisons of consumption baskets across countries (i.e. purchasing power parity analyses) suffer from a serious compositional bias effect because they (by de nition) ignore this fact. The number of commonly sold goods is comparable to the gures reported by Broda and D. Weinstein (2008), who also use demand-side data, and is at least 1.5 times bigger (for BE-GE) compared to GGHL, using supply-side data.

The gures reported in Tables B.1.1 to B.1.3 of Section 2.B.1 document that we can rely on a relatively comprehensive, fairly balanced and suf ciently deep coverage of goods categories. Reported numbers of goods included in each category con rm ndings by

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

Broda and D. Weinstein (2008), according to which even narrowly de ned product categories contain a relevant amount of different products which might lead to the potentially erroneous conclusion of large deviations from LOOP, simply as an upshot of purely compositional effects.

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

Following the existing literature, as our basic measure of price dispersion, denoted by q_i^{hk}, we use 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[14]

$$q_i^{hk} = \ln\left(p_i^h\right) - \ln\left(p_i^k\right), \tag{2.1}$$

where all prices are denominated in euro. Table 2.2 shows that there exist considerable differences between within- and cross-country price differences for all country pairs. Recorded gures are obtained by rst averaging all available prices for a given good in a given market within the considered sample period and then computing relative prices across all regions for this good.[15]

The reported data re ect the mean, median and standard deviations of all relative prices of all goods considered within countries or on a cross-country basis. We observe that both the mean and median price gaps are very close to zero for all within-country combinations. However, interestingly, the corresponding cross-country gures

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

[15] Our de nition of a local market makes use of the so-called NUTS classi cation system of the Statistical Of ce of the European Union, Eurostat. NUTS is the short form for "Nomenclature of territorial units for statistics". It represents Eurostat s of cial classi cation system for dividing up the economic territory of the EU for the purpose of the collection, development and harmonization of European regional statistic and socio-economic analyses of the regions. The NUTS classi cation system is made up of three hierarchical levels: NUTS-1, NUTS-2 and NUTS-3 regions. We characterize local markets to correspond to the so-called NUTS 2 regions. NUTS-2 regions usually have between 800,000 and 3 million inhabitants. Belgium comprises 11 NUTS-2 regions, the Netherlands comprises 12 NUTS-2 regions and Germany is composed of 38 NUTS-2 regions. Our data is available at the ZIP code level. The ZIP code systems differ considerably across countries. For each ZIP code area, we use the latitude/longitude information of its center. The number of ZIP code areas included in our sample corresponds to around 830 for Belgium, somewhat more than 6500 for the Netherlands and more than 5500 for Germany.

TABLE 2.2: Within- and cross-country price dispersion

2005Q1		Price gap			Absolute price gap		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	63464	0.26	0.00	10.86	5.07	1.93	9.61
NL-NL	158177	0.22	0.00	11.88	6.03	2.13	10.24
BE-NL	145704	7.11	3.45	28.52	19.85	14.43	21.68
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	862541	0.22	0.00	9.93	5.08	1.11	8.54
NL-NL	39475	0.11	0.00	12.81	6.25	1.54	11.18
GE-NL	278886	-5.90	-4.20	24.68	18.00	13.53	17.88
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	29652	0.42	0.00	11.49	5.14	1.99	10.29
GE-GE	538349	0.16	0.00	10.72	5.30	1.10	9.32
BE-GE	208621	9.79	7.63	26.82	20.35	15.25	20.02
2008Q4		Price gap			Absolute price gap		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	99412	-0.07	0.00	13.22	4.40	1.29	12.47
NL-NL	281911	0.25	0.00	10.78	5.40	1.85	9.34
BE-NL	234810	7.36	3.39	28.85	18.75	12.76	23.12
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	1527189	0.03	0.00	10.45	5.88	2.32	8.64
NL-NL	59609	0.29	0.00	11.08	5.07	0.96	9.86
GE-NL	427319	-7.25	-4.10	26.62	19.02	13.45	19.98
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	42637	0.03	0.00	14.05	4.11	0.74	13.44
GE-GE	1052441	-0.02	0.00	10.46	5.83	2.26	8.68
BE-GE	325217	12.86	11.55	27.25	21.81	17.30	20.80

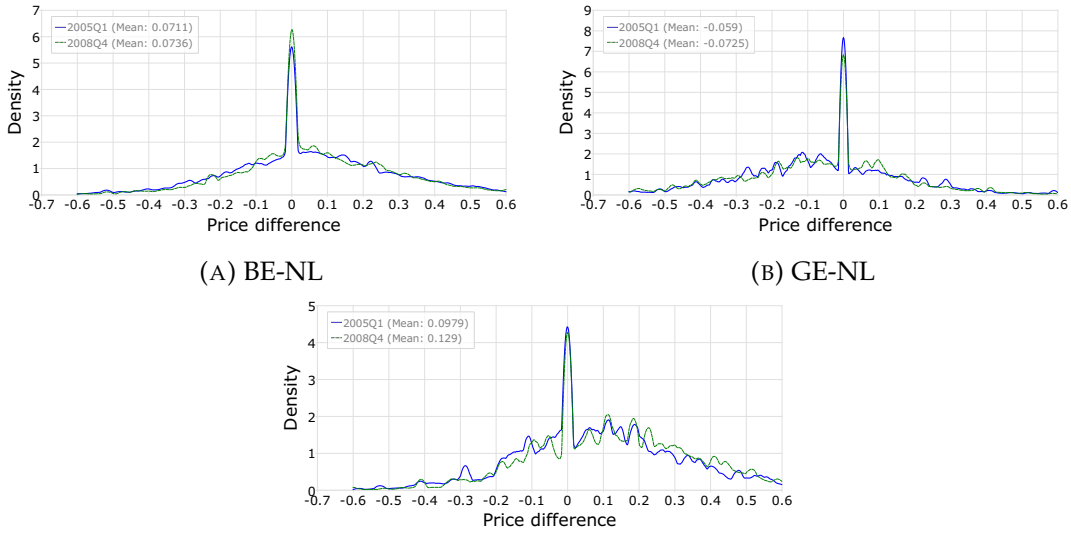
Notes: Table 2.2 reports descriptive statistics on within and between-country price dispersion employing regional goods-level price gaps. To compute the statistics, we proceed as follows: First, all recorded prices of a given good within a given (NUTS2) region are averaged for the considered time period (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 subsample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (2.1). 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. The latter are reported in percentage terms.

are considerably bigger, with median (mean) values of around 3/3% (7/7%) for Belgium-Netherlands, about -4/-4% (-6/-7%) for Germany-Netherlands and 8/12% (10/12%) for

Belgium-Germany in 2005Q1/2008Q4. When absolute values of price gaps are considered (columns 6 to 8), a similar picture emerges. Median values for within-country price differences range from around 1% to somewhat more than 3%, whereas cross-country gures are at least ve times bigger, ranging from more than 12% to around 17%.

A graphical illustration of the reported cross-country price differences is provided in Figure 2.1. It shows that the mean/median values of international price gaps have been fairly stable for all country pairs. Moreover, the gure clearly illustrates 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 no exception.[16]

FIGURE 2.1: Distribution of mean price differences



(A) BE-NL (B) GE-NL

(C) BE-GE

Notes: Figure 2.1 plots the kernel density estimates of regional goods-level price gaps of all matched goods of the country pair indicated below each panel for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). To compute regional goods-level price gaps, 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 cross country bi-regional price gaps are computed. In each panel, the base country is the country indicated rst in the subtitle. A positive value indicates that prices are higher in the base country than in the reference country (mentioned secondly).

[16]Similar evidence emerges if cross-regional differences are computed by rst aggregating (averaging) the prices of a good observed in a given sample period (2005Q1 or 2008Q4) at the regional level and then computing mean and median price gaps of commonly sold goods for each regional pair, see Table B.3.1 and Table B.3.2 of Section 2.B.3.

A nal interesting observation results from comparing the extent of within-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, as an approach following ER is liable to do, can lead to erroneous conclusions with respect to the importance of borders for price dispersion. Our results in Table 2.2 also show a within-country price dispersion in our sample, differing at most by 1.5 percentage points between countries for a given country pair.[17] These differences are considerably smaller than those observed in the U.S.-Canadian data samples.

2.3 Estimation approach

2.3.1 Theoretical considerations

The motivation underlying our empirical approach - as well as the interpretation of the results obtained - rests on the following theoretical considerations:[18] 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 distributed 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 in uence their transaction costs of buying a good at a speci c retailer. When deciding from whom to buy a particular good, households take into account the price at which the good is offered as well as any transaction costs they incur from shopping at a particular place. Transaction costs are a metric for costs arising not only from traveling (tickets, cost of fuel, time), but also re ect other factors, such as collecting and assessing information. We assume these transaction costs to be household-speci c.

Following GGHL and other authors, we further suppose transaction costs to depend on the distance that a household has to travel to a given retailer and whether it has to cross a border in doing so. Denoting overall transaction costs by λ and the retailer from

[17] The numbers for within country dispersions differ across country pairs since in each case, only goods across the respectively considered countries are taken into account.

[18] Our framework heavily builds on the model and ideas presented by GGHL. It basically modi es their considerations only where our different data dimension (demand-side rather than supply-side data) requires adjustments and/or allows for different insights.

which a household, h, considers buying a given good by A, we have $\lambda_A^h = \tau_A^h + bI_A^h$, where τ_A^h denotes the transaction costs a household faces, capturing 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 is otherwise zero. The parameter b therefore denotes additional transaction costs caused by crossing the border.

The number of consumers which a given retailer attracts depends on the price it charges for a given good. Due to the existence of distance-related transaction costs, it can make pro ts even if it offers goods at a price above that of neighboring retailers. This is because the transaction costs will very likely induce households living very close to a particular retailer to always buy from this supplier (unless the price gap to the nearest competitor becomes very big). Likewise, households living far away from a retailer will very unlikely buy from it. In between these two cases, there will be some households being 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, we consider two retailers, A and A^*, where A is located in one of the two countries, denoted as the reference or domestic country and A^* is located in the other - counterpart or foreign - country. We assume that A and A^* are those retailers of their respective countries situated closest to the border in their respective country. For the ease of exposition, we assume that the distance between retailers and the border is symmetric. We furthermore suppose that the foreign retailer A^* charges a lower price than the domestic retailer A.

Following GGHL, we differentiate between two scenarios: markets are either internationally integrated or segmented. According to GGHL, two markets are characterized as 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"*. To illustrate which conclusions our empirical analysis allow us to draw in such a case, we consider the situation of households being located between (the expensive) retailer A and (the cheaper) retailer A^*. It is obvious that all households residing in the foreign country will decide to buy from A^* and will therefore pay (and report) the price of this retailer, denoted by p_{A^*}. At the

same time, domestic households living far away from the border and very close to retailer A will likely buy at price p_A.[19] However, the further we move away from A in the direction of the border, the more likely it becomes 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^*. When markets are integrated, we assume that at least one such household indeed exists. How far away from retailer A and how close to the border this household lives depends on the relative sizes of the price gap between retailers A and A^* compared to the transaction and border costs. Generally, one can say that the larger the price gap, for given transaction and border costs, the further away this household will be located. In an extreme case, this household could live directly at the border, which would imply that the price gap corresponds to the border costs. If, on the other hand, this household lives some distance away from the border, then households with similar transaction costs but relatively a smaller distance to the border will no longer buy from A, but cross the border and purchase the good at price p_{A^*}. The closer we get to the border, the more households would likely act accordingly and thus will report p_{A^*}.

Taking into account the fact that our data set comprises information on cross-border shopping, these considerations allow us to derive a number of interesting insights about the integration of markets and the size of border costs. First, if we observe cross-border shopping, we can conclude that the markets are integrated. Secondly, for goods for which we obtain a signi cant border estimate and observe some cross-shopping, we are able to derive an estimate of the exact border costs. Referring to the above theoretical deliberations, such an outcome corresponds to a situation where there is a price gap between countries and where some households partake in cross-border shopping, while at the same time a considerable number of households buys at home. Denoting households buying domestically with h and households conducting cross-border shopping with i, we have in this case: $p_A + \tau^h_A \leq p_{A^*} + \tau^h_{A^*} + b$ and $p_{A^*} + \tau^i_{A^*} + b \leq p_A + \tau^i_A$. Letting differences in distance go to zero and rearranging expressions, we obtain $p_A - p_{A^*} \leq b$ and $p_A - p_{A^*} \geq b$, which is only satis ed in the case of equality, i.e., if $p_A - p_{A^*} = b$ holds.

[19] This argument implicitly assumes that for domestic households living further and further away from the border (and thus relatively closer and closer to the domestic retailer) transaction costs increase with distance in a manner that the sum of the transaction and border costs ultimately exceed the price gap. A less strict assumption would be that it holds at least for households being located further away from the border than the domestic retailer. If neither of the two assumption held, the retailer would not make any sales.

More generally, i.e., if we do not make any assumption about which country is cheaper, we obtain the result that the absolute price gap between two countries corresponds to the border costs in this case, i.e., $|p_A - p_{A^*}| = b$. Practically, to obtain the estimate of b, one needs to take into account that the inclusion of the cheaper cross-border purchase prices into the regression sample would result in a downward biased estimate of b. To correct for this bias, one thus needs to exclude the cross-border purchases from the regressions to get b.[20]

In the case of cross-border shopping (integration) and a non-signi cant border estimate, two cases can be differentiated as regards the possibility of being able to identify border costs. If it turns out that signi cant border estimates prevail after removing cross-border price observations from the relatively more expensive country, the border coef - cient obtained in this way provides an estimate of the border costs. In all the other cases, no inference on the size of border costs can be made.

Considering the de nition of integration by GGHL, integration could also prevail, even if we don t observe any cross-border shopping. This could be the case, if there is at least one household in the high-price country which is indifferent between buying the good abroad or at home, but which decides to buy at home. Such a scenario is most likely, if the indifferent household is located close to or directly at the border (otherwise, one would probably observe cross-border shopping activities by households living closer to the border). Should the border estimate turn out to be signi cant in such a case, it would re ect the border costs. On the other hand, if the border estimate is not signi cant, no conclusion about the size of border costs can be drawn.

The alternative case to integration is that markets are segmented. Again following GGHL, we de ne 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."* The above theoretical considerations imply that in this situation, the following lessons can be drawn from our empirical analysis. First, market segmentation can only prevail for goods for which we don t observe any cross-border shopping activities. Secondly, for segmented markets for which we observe signi cant price gaps, the latter represent a lower bound of the

[20] In some rare cases, the removal of the cross-border shopping data points implies that the minimum number of observations required is no longer given. In this case, the observation is dropped.

border costs for the following reasons: Considering the situation of a household located close to the border in the (high-price) domestic country we have: $p_{A^*} + \tau^h_{A^*} + b > p_A + \tau^h_A$. Letting $\tau^h_A \to \tau^h_{A^*}$ yields $b > p_A - p_{A^*}$, showing that the reported price discontinuity by households at the border would de ne a lower bound on the costs of crossing the border. More generally, a signi cant absolute price gap at a border represents a lower bound of border costs when markets are segmented.

Thirdly, if markets are segmented but no (signi cant) price gap can be observed at the border, we cannot make any inference on the border costs. The rst thing to notice in this context is that the existence of equal prices across markets is fully compatible with the prevailance of segmented markets. This would occur when the factors underlying the price setting of retailers induce them to choose the same (or very similar) prices on both sides of the border. As for integration, in this case no inference on the cost of borders can be made, unfortunately.

Concluding this discussion, our above considerations imply that we are able to answer the question of the integration of markets unambiguously when we observe cross-border shopping activities. In these cases, signi cant border price gaps provide an estimate of borders costs, while for goods with non-signi cant border coef cients, no inference on border costs can be made. When no cross-border purchases are made, we cannot unaninomously distinct between integrated and segmented markets. Moreover, and related to this point, signi cant border price gaps can only be seen as a lower bound of borders costs. Again, for non-signi cant border estimates, no implications on border costs can be drawn.

2.3.2 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, estimate any potential price discontinuity present at the border, and combine the results with our information on cross-border shopping, we can draw useful lessons on the question of whether markets for a given good are integrated or segmented and how costly borders are. To estimate the

price gap at the border, we conduct a regression discontinuity (RD) analysis which is, as shown below, well suited to our data.

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 or running variable.[21] More speci cally, RD 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 only occurs for values equal or above the threshold. In most applications, the determination of this 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 average outcomes just above and below the threshold (or cut-off) provides an estimate of the treatment effect (and of causality).

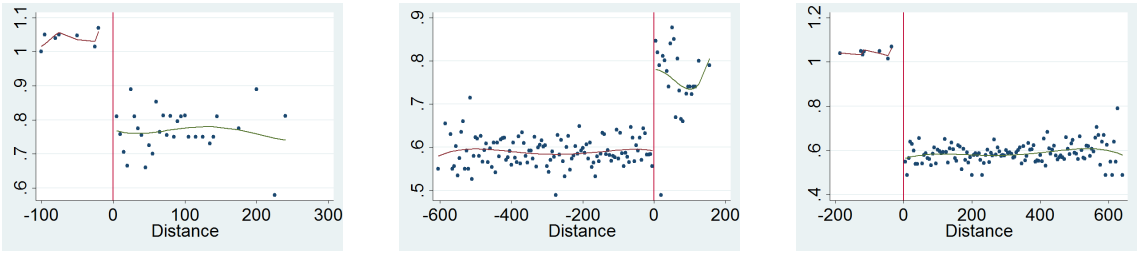
GGHL apply the reasoning underlying RD designs to the situation of price-setters located on either side of the U.S.-Canadian border. Following GGHL, we interpret the treatment as being located in a different country (compared to the respectively chosen reference country). The forcing variable is given by the distance of households location to the border. Distances are measured by positive values for one side of the border and by negative ones for the other side. The threshold value chosen for selecting households to a treatment corresponds to a value of zero for the distance to the border. Figure 2.2 provides a graphical illustration of this approach by plotting the price of an exemplary good in the vicinity of the borders considered.

Applying an RD approach offers three major bene ts in terms of estimating the effects of borders on transaction costs. Firstly, it overcomes problems associated with estimating border effects in the presence of within-country heterogeneities in price dispersion.[22]

[21] Imbens and Lemieux (2008), D. Lee and Lemieux (2010), and Skovron and Titiunik (2015) provide a recent overview of the literature on RD designs.

[22] Gorodnichenko and Tesar (2009) highlight potential identi cation problems in the approach to estimating border effects as it was rst employed by Engel and Rogers (1996) and which has been used by most subsequent studies in this eld. More speci cally, Gorodnichenko and Tesar (2009) argue that by neglecting differences in within-country heterogeneity with regard to 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 nd that the border effect between the U.S. and Canada not only becomes signi cantly smaller but almost completely vanishes.

FIGURE 2.2: Discontinuity plots for the price of a selected good



(A) BE-NL (B) GE-NL (C) BE-GE

Notes: The panels of Figure 2.2 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.

Secondly, regional prices very likely depend on a variety of not observable factors such as, for instance, demographics or differential elasticities across retailers. Given the validity of the assumptions underlying the RD approach, it can control for the effects of these unobservable variables. Thirdly, combining the outcomes of the approach with our cross-border purchase data or employing an alternative indirect approach allows us to learn important lessons, not only on whether markets are integrated or segmented, but also how large the costs of crossing the border (at least) are in a large number of cases.

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

$$\ln p_i^h = \alpha_i + \beta_i B_i + \theta_i D^h + \phi_i D^h B_i + \gamma_h X^h + \varepsilon_i^h, \tag{2.2}$$

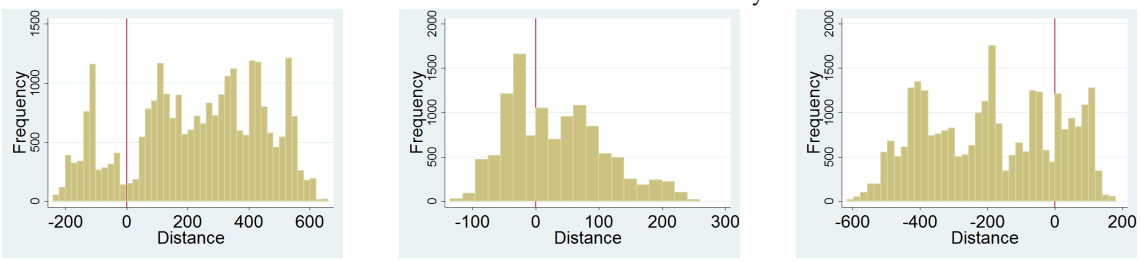
where $\ln p_i^h$ denotes the (log) price of good i paid by household h, α_i is a good-specific dummy variable, X^h captures household h specific variables (income and age) and ε_i^h denotes household and goods-specific unobserved characteristics. B_i is a dummy variable equal to one if the household 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 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, by including an interaction term between distance and the border dummy, we allow the distance coefficient to change its size at the threshold point.

The coef cient 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 coef cient, 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.[23]

Unfortunately, a formal test of the identi cation 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.

Figure 2.4 (and Figures B.4.1 to B.4.3 of Section 2.B.4) illustrate that the locations of our sample households (as of 2008Q4, similar pictures hold true 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, most importantly, the border regions.[24] The resulting high availability of observations in the vicinity of the threshold, i.e., the border, naturally favors the applicability of the RD approach.

FIGURE 2.3: 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 rst 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.

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

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

.

[24] In the Belgian-German case, regions on both sides of the border are very thinly populated.

Figure 2.3 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 some differences in population density close to the border. However, the degree of difference is relatively moderate. Moreover, given that in some cases densities tend to be higher on the high-price side of a given border, it would be very difficult to make the case that observed differences indicate "strategic behavior" of individuals, aiming at receiving a positive treatment by moving to the lower-price side of a given border.

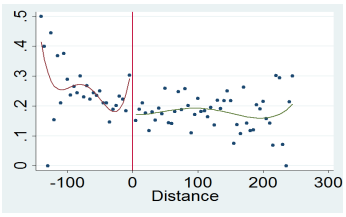
The discontinuity plots of the distributions of age and income of our sample households in the proximity of the border - presented in Figure 2.4 - do not generally hint at major discontinuities at the border. In this context, it is to be noted that the selection of households is not intended to create comparable samples across countries but, instead, to generate representatives samples for a given country. Overall, 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 D. Lee and Lemieux, 2010) - amounts to estimating a standard linear regression over a bandwidth h on both sides of the border. D. Lee and Lemieux (2010) emphasize that the choice of the bandwidth - in our case, essentially a measure of distance - generally involves finding an - optimal - balance between precision and bias.[25] To choose the bandwidth, we apply data-driven bandwidth selection methods as outlined and implemented by Calonico et al. (2017). 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 100 km. The reported results are based on a bandwidth of 80 km. 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 each side of the border.

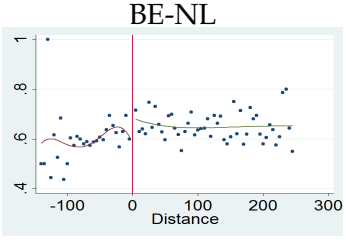
[25] See also Imbens and Kalyanaraman (2012).

FIGURE 2.4: Structure of households in the vicinity of the border

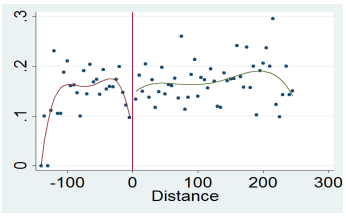
BE-NL



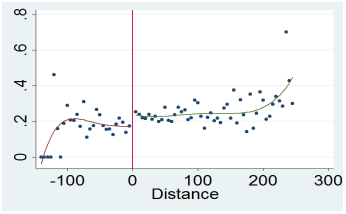
(A) Young



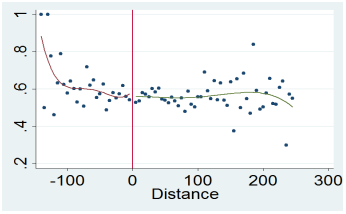
(B) Medium age



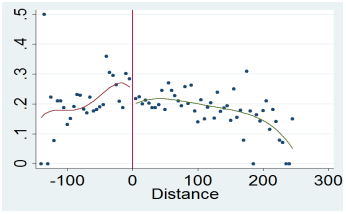
(C) Old



(D) Low income

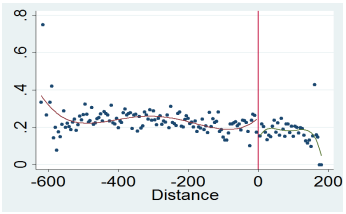


(E) Medium income

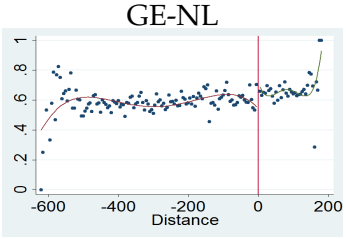


(F) High income

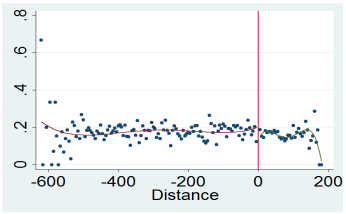
GE-NL



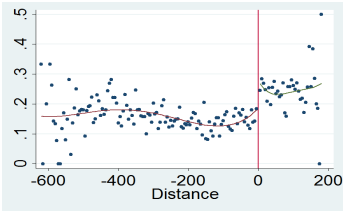
(G) Young



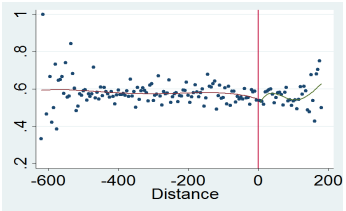
(H) Medium age



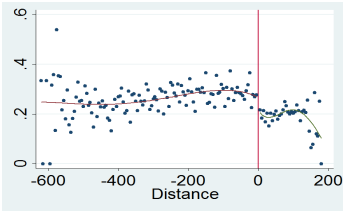
(I) Old



(J) Low income

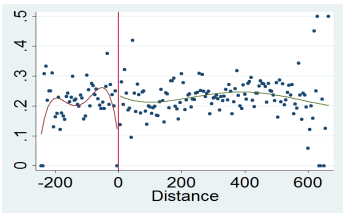


(K) Medium income

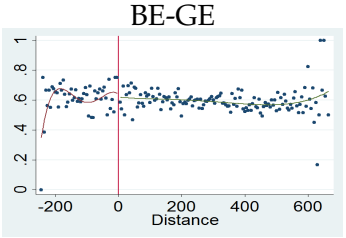


(L) High income

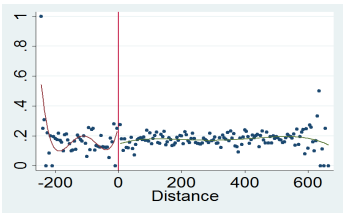
BE-GE



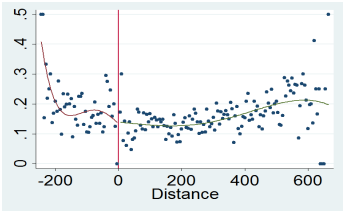
(M) Young



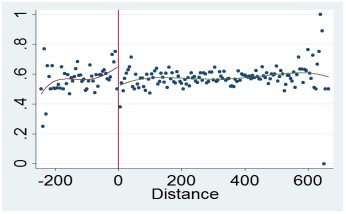
(N) Medium age



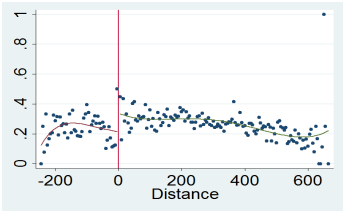
(O) Old



(P) Low income



(Q) Medium income



(R) High income

Notes: Figure 2.4 plots densities of the age and income structure of Belgian-Dutch (upper panel), German-Dutch (medium panel) and Belgian-German (lower panel) households (year 2008) against the distances of the respective households homes to the border. In each case, the locations of the households of the rst country from the country pair are plotted with negative distances, locations of households from the second country are plotted with positive distances. Bins have a width of 5 km.

To obtain an idea how the characteristics of the data in the overall sample and in the sample used for estimation compare, Table 2.3 presents selected statistics on three major features of Dutch goods across samples. Comparing the gures for the three panels, we observe that the estimation samples share certain properties relative to the overall sample. First, the distribution of prices in the estimation samples is shifted towards relatively cheaper goods, extremely high-priced goods are not included in these samples. Secondly, goods in the estimation samples are, on average, purchased (partly much) more often and by (again partly many) more households. The patterns of goods across samples are similar for the other countries.[26]

TABLE 2.3: Major characteristics of goods (example: Netherlands) in the total and the estimation samples. (Q4 2008).

Dutch goods: total sample									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	91.77	446.11	0.00	0.12	0.36	0.80	4.33	499.00	43,900.00
Purchases	29.96	129.22	1.00	1.00	1.00	4.00	17.00	124.00	9,499.00
Households	17.84	58.33	1.00	1.00	1.00	3.00	12.00	78.00	3,130.00
Observations	84690								
Dutch goods: Belgian-Dutch estimation sample									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	6.79	40.02	0.02	0.09	0.24	0.53	0.98	6.64	546.24
Purchases	130.26	178.75	12.00	23.00	41.00	76.00	153.00	394.00	2,851.00
Households	80.50	89.76	11.00	18.00	30.00	51.00	97.00	232.00	1,180.00
Observations	1021								
Dutch goods: German-Dutch estimation sample									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	2.61	21.15	0.02	0.09	0.24	0.48	0.80	2.14	366.43
Purchases	120.56	181.18	11.00	21.00	38.00	65.00	122.00	399.00	2,210.00
Households	75.54	96.89	10.00	17.00	28.00	44.00	75.50	233.00	1,180.00
Observations	524								

Notes: Table 2.3 presents summary statistics for purchases by Dutch households being included in the overall and the two estimation samples. "Price" refers to the average price paid by a Dutch household for a given good, "Purchases" re ects the number of purchases of the good and "Households" contains information on the number of households that bought the good. The sample period is 2008Q4. Comparable data on the other samples is provided in Section 2.B.2.

[26] Statistics for the other countries are presented in Section 2.B.2.

2.4 Integration vs. segmentation and the cost of borders

Guided by the theoretical discussion in Section 2.3.1, our empirical approach to analyze the issue of integration/segmentation of markets and the cost of crossing the border occurs in two steps. First, we document the extent to which we observe signi cant discontinuities at the border and provide evidence from some counterfactual experiments. Then, employing the ndings from the rst step, we draw lessons about the integration of markets, and provide estimates of border costs (where possible).

2.4.1 Price gaps at the border: Base results

The regression results presented in Table 2.4 indicate signi cant and sizeable discontinuities of most prices at the border for all considered country pairs. For Belgium-Netherlands, around 77% of all border coef cients are signi cant for 2005Q1 and around 79% for 2008Q4. For Germany-Netherlands, the corresponding numbers are 76% and 77%, for Belgium-Germany 81% and 74%. Given the theoretical deliberations above, we are thus able to provide at least a lower bound on the border costs in the overwhelming number of cases included in our sample.

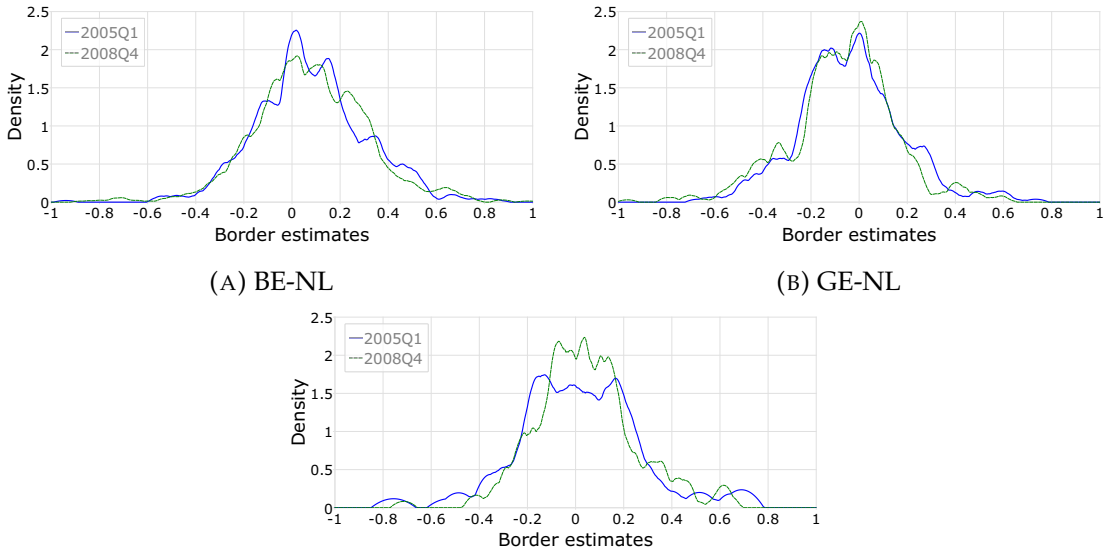
TABLE 2.4: Regression discontinuity results: baseline speci cation

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.000	0.098	16.9	0.000	-0.012	0.112	24.0	0.000	-0.002	0.070	23.5
Border	0.063	0.082	0.246	76.6	-0.043	-0.038	0.218	76.3	0.005	0.014	0.257	80.9
Border x Distance	0.000	0.010	0.145	12.4	0.000	0.014	0.146	22.0	0.000	-0.010	0.111	8.8
Constant	-0.884	-0.770	1.417	91.0	-0.894	-0.978	1.140	88.7	-0.788	-1.057	1.172	80.9
Observations	680.000	.	.	.	354.000	.	.	.	68.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	-0.001	0.097	12.8	0.000	0.000	0.101	15.1	0.000	0.016	0.128	27.0
Border	0.070	0.075	0.292	79.3	-0.054	-0.067	0.237	76.7	0.025	0.041	0.213	73.6
Border x Distance	-0.001	-0.013	0.169	13.4	0.000	0.000	0.144	16.4	0.000	-0.032	0.153	12.6
Constant	-0.652	-0.565	1.467	95.9	-0.717	-0.738	1.160	96.9	-0.732	-0.796	1.026	88.1
Observations	1021.000	.	.	.	524.000	.	.	.	159.000	.	.	.

Notes: Table 2.4 reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2.2). For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. Positive values of the border coef cient 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 coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

Considering the reported aggregate statistics, the price gaps are the biggest for Belgium-Netherlands indicating prices that are around 6% (7%) higher (median values) in Belgium relative to the Netherlands in 2005Q1 (2008Q4). Price discontinuities are somewhat smaller (in absolute terms) for Germany-Netherlands, with prices being around 4% (5%) lower in Germany than in the Netherlands. For Belgium-Germany, the price wedges are the smallest: Belgian households paid around 1% (3%) more than their German counterparts. While the median/mean values of price gaps are comparably small, standard deviations of the estimates reported in (2.4) indicate a considerable dispersion of border effects across goods. This observation is further illustrated in Figure 2.5, plotting kernel density functions of the border estimates obtained. Whereas the general form of all three distributions is similar, their respective horizontal positions differ somewhat, as implied by the mean/median results reported in Table 2.4. However, they all clearly indicate that large price gaps of 20% are no exception but are relatively common in all cases. Referring to our theoretical considerations, this finding suggests that transaction costs as a consequence of the border are sizeable for many goods.

FIGURE 2.5: Kernel density estimates of border effects



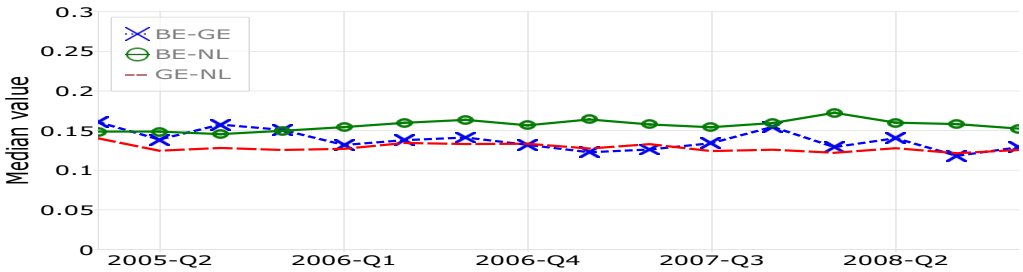
(A) BE-NL (B) GE-NL

(C) BE-GE

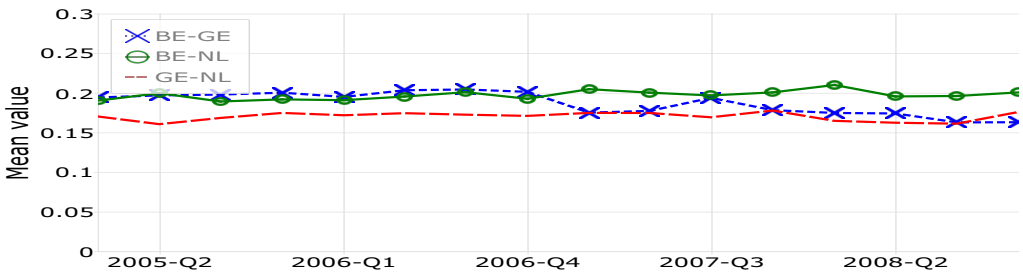
Notes: Figure 2.5 plots kernel density functions of the estimated border coefficients. The considered sample periods are 2005Q1 and 2008Q4. BE-NL denotes the Belgian-Dutch, GE-NL the German-Dutch and BE-GE the Belgian-Dutch values. The border dummy takes the value one if the household is located in the 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.

This issue is examined more closely in Table 2.5, where we report statistics for the distribution of absolute values of estimated price gaps whose signi cant cases represent estimates of (the lower bound of) border costs. The reported numbers suggest a large distribution of these values, with most numbers in a range of up to 40%. Looking at the "median good", crossing the border adds at least 12.5% of its price to transaction costs in the case of GE-NL (2008Q4), around 13% (2008Q4) in the case of BE-GE, and around 15% in the case of BE-NL (2008Q4). The comparison of ndings between 2005Q1 and 2008Q4 indicates some changes over time whose sizes, however, are not considerable and whose signs are ambiguous.

FIGURE 2.6: Dynamics of border coef cients, absolute values



(A) Median values of the border estimates



(B) Mean values of the border estimates

Notes: Figure 2.6 plots median and mean values of estimated quarterly border coef cients(absolute values). 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 rst in the legend.

The dynamics of the estimated absolute price gaps throughout our sample period are illustrated in Figure 2.6. We see that both the mean and the median values are fairly similar across county pairs. Moreover, the two values are quite stable over time. Median values range from around 12% to 14% for GE-NL, 12% to 16% for BE-GE, and from 14.5% to around 17% for BE-NL. Based on our theoretical considerations outlined above, these results suggest that, depending on whether integration or segmentation prevails in the

case of signi cant border estimates. the median border costs correspond to 14% of the price of a good for GE-NL, 16% for BE-GE and 17% for BE-NL or more,[27]

To check for robustness, we include (categorial) variables capturing households age and income. The age groups are de ned as follows: Households are classi ed as young if their age is lower than 34 years, medium-aged households are those between 35 and 64 years, and old households are older than 65 years. The income groups are de ned 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,- (GE)/2700,- (NL). Medium-income households are those with an income between low- and high-income households. As Table C.2.1, Table C.2.3 of Section 2.C.2 show, including these covariates, i.e., controlling for age and income of households, has no major effects on the results for the border coef cients.[28]

2.4.2 Counterfactual evidence from within-country regions

To examine to which extent the documented price discontinuities at national borders also occur across states/regions within a country, we conduct a series of "counterfactual experiments". To this end, for each country, we select two comparable (and not too small) states/regions and estimate whether a price discontinuity can be observed across them. For Belgium, this choice is pretty straightforward: Given their fairly frequent, persistent and sometimes deep political tensions, and given that two different languages are spoken, we decided to examine potential border effects across the Flemish-Wallonian region pair. For Germany, we chose the state pair Lower-Saxony/North-Rhine Westphalia, which are amongst the biggest states in Germany (in terms of geographic area), and which are also located at the Dutch and Belgian (North-Rhine Westphalia) border.

[27] To complete the picture, Figure C.1.1 of Section 2.C.1 plots the dynamics of medians and means of the estimated border estimates reported in Table 2.4 for the overall sample period. It reveals that differences in average price gaps are fairly stable for GE-NL whereas they show a slight upward trend for BE-NL and, particularly, BE-GE. Overall, the results suggest that German households enjoy a persistent average price advantage compared to Dutch and Belgian households of which the latter face the relatively highest prices.

[28] Additionally, to control for observable market characteristics, we also employed of cial data on population, degree of urbanisation (Degurba) and average income at ZIP-level areas. Data on population and degree of urbanisation are collected from Eurostat (2011 Census database). Income data for 2010 are collected from national statistical of ces, income in Germany and Belgium de ned as total income, reported to the tax authorities, divided by the number of declaration, while for the Netherlands we used data on average income per person. The results are reported in Tables C.2.2 and C.2.4 of Section 2.C.2. As can be seen from these tables, estimates are also robust with respect to the inclusion of these variables.

TABLE 2.5: Regression discontinuity results: absolute values

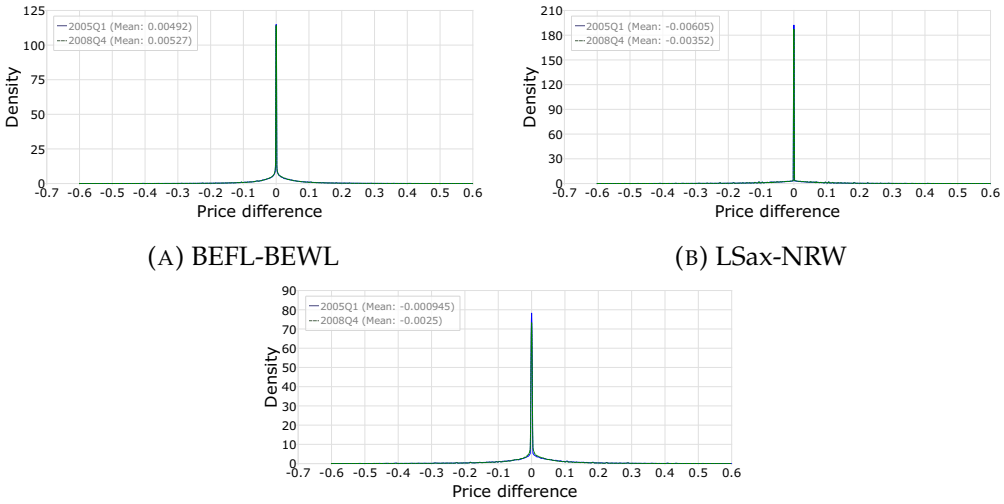
	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.023	0.053	0.082	16.9	0.022	0.058	0.097	24.0	0.002	0.028	0.064	23.5
Border	0.149	0.191	0.176	76.6	0.140	0.171	0.141	76.3	0.160	0.195	0.167	80.9
Border x Distance	0.044	0.085	0.117	12.4	0.035	0.080	0.123	22.0	0.026	0.058	0.095	8.8
Constant	1.074	1.272	0.991	91.0	0.954	1.171	0.939	88.7	0.924	1.208	1.012	80.9
Observations	680.000	.	.	.	354.000	.	.	.	68.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.021	0.052	0.082	12.8	0.017	0.049	0.088	15.1	0.008	0.049	0.119	27.0
Border	0.152	0.201	0.224	79.3	0.125	0.176	0.172	76.7	0.129	0.163	0.143	73.6
Border x Distance	0.043	0.092	0.143	13.4	0.036	0.081	0.119	16.4	0.027	0.073	0.138	12.6
Constant	0.929	1.187	1.031	95.9	0.910	1.062	0.873	96.9	0.799	0.997	0.831	88.1
Observations	1021.000	.	.	.	524.000	.	.	.	159.000	.	.	.

Notes: Table 2.5 reports absolute vlaues of regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2.2). For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

For the Netherlands, we selected Eastern Netherlands and Southern Netherlands - the two NUTS 1 regions located next to Germany.

A graphical illustration of our findings for 2005Q1 and 2008Q4 is given in Figure 2.7, summary statistics are presented in Table D.1 of Section 2.D. The results show that cross-region price differences are very moderate for all region pairs in all the sample periods considered. Both the mean and median values of the actual price gaps are always close to zero. The absolute values are somewhat higher (in the range of 1% to slightly bigger than 2% (median values)) but nevertheless around 80% lower than the comparable cross-country price differences. The kernel density plots (Figure 2.7) confirm the close-to-zero mean/median values, and show that the variation across these values is tiny.

FIGURE 2.7: Counterfactual evidence: distribution of mean price differences across within-country regions



(A) BEFL-BEWL

(B) LSax-NRW

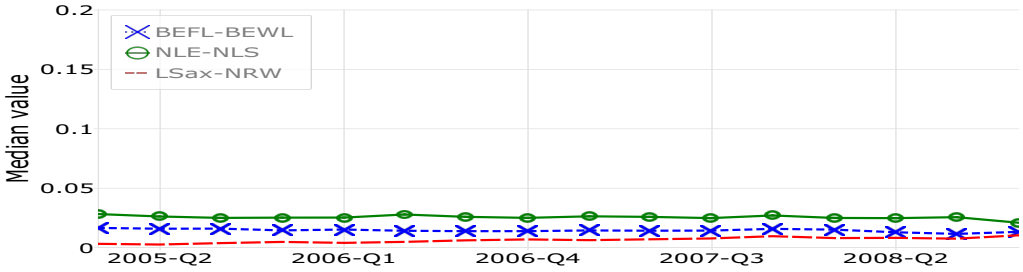
(C) NLE-NLS

Notes: Figure 2.7 plots the kernel density estimates of regional goods-level price gaps of all matched goods of the sub-national states pair indicated below each panel for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). To compute regional goods-level price gaps, 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 cross states bi-regional price gaps are computed. 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).

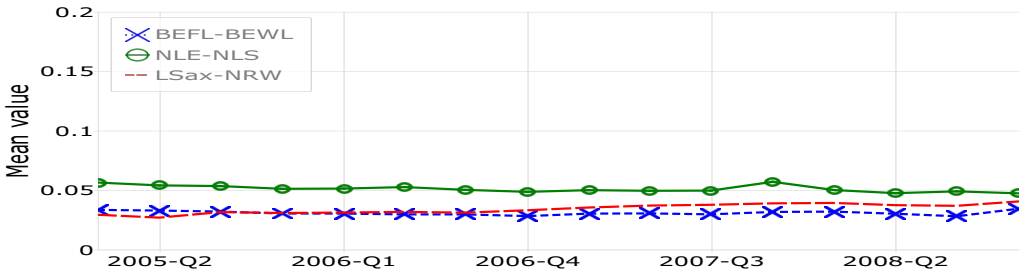
Not surprisingly, the RD regression results effectively indicate no or only very tiny border effects between our artificial within-country states (see Table D.2 of Section 2.D).

The share of signi cant border estimates is only about 10% to 12%, the estimated border price gaps (absolute) median values range between 3% and slightly more than 5%, with standard deviations being considerably smaller than in the cross-country cases. The quasi non existent or very tiny distribution of estimates is illustrated in Figure D.1 of Section 2.D. Figure 2.8, moreover, shows that the nding of very small border effects is valid throughout our sample period.[29] Overall, the results from the counterfactual experiments con rm that signi cant price discontinuities tend to exist between markets separated by a national border, whereas no such discontinuities exist between markets separated by an intra-state regional border.

FIGURE 2.8: Counterfactual evidence: dynamics of estimates for border coef cients across within-country regions, absolute values



(A) Median values of the border estimates



(B) Mean values of the border estimates

Notes: Figure 2.8 plots median and mean values of estimated quarterly border estimates (absolute values). 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 rst in the legend.

[29] A plot of median and mean values of estimated quarterly border estimates, presented in Figure D.2 of Section 2.D, conveys the same message.

2.4.3 Direct evidence from cross-border shopping data

The results from the previous subsection show that there exist signi cant price gaps at the border for the majority of goods. Referring to the theoretical considerations of Section 2.3.1, this section derives lessons that can be learned on the questions of integration and segmentation and the size of border costs from these ndings. This takes place in two ways. First, we relate the estimated border results to information on cross-border shopping for the considered goods. Secondly, we compare the dynamic correlation patterns of within-country and cross-border prices.

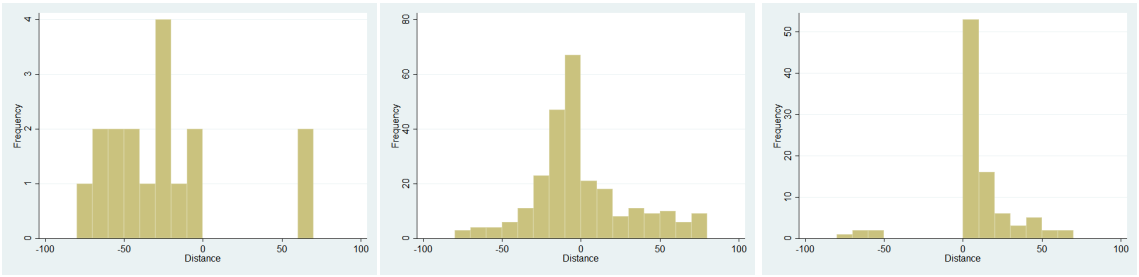
As stated in the data description, our sample includes a variable containing information about the retailer at which the respective household bought a given good. In addition to domestic retailers, this variable also has entries on some foreign retailers and/or a category capturing purchases from abroad. Our data thus comprises direct information about cross-border shopping. Given our theoretical considerations, any documented cross-border purchase for a given good implies that the market for this good is integrated. It is to be noted, however, that according to our data supplier the coverage rate of purchases made abroad is unfortunately not as complete as that for domestic purchases. In other words, the actual scope of cross-border shopping trips is likely to be higher than the one documented, implying that our ndings for the proportion of integrated markets is probably a lower bound on this number.

The histograms presented in Figure 2.10 show the number of households that reported cross-border purchases in 2008Q4 as a function of the distance of their residences to the border. In line with the descriptive statistics on price differences for the countries presented above, the gures show that households in the relatively more expensive countries tend to conduct relatively more cross-border shopping trips. Moreover, more distantly located households tend to shop abroad less frequently.

Combining the cross-border shopping information with the regression results reported in the above section and differentiating between signi cant and non-signi cant border estimates as well as between including cross-border and not including such cases, the goods in our estimation sample can be assigned to one out of four different groups. Table 2.6 provides summary results for these groups. In each case, it reports the number of goods (N) for which we observe cross-border shopping and statistics about absolute

border estimates for 2005Q1 and 2008Q4.[30] A rst observation is that our sample comprises a non-negligible number of cases in all four groups. Concerning the cases with signi cant border estimates, we see that the group of goods for which no cross-border shopping takes places is considerably bigger than that for which cross-border shopping. Whilst these results could be expected, the relatively large number of goods for which some cross-shopping takes place in this group appears to be somewhat surprising.

FIGURE 2.10: Frequency of cross-border shopping



(A) BE-GE (B) BE-NL (C) GE-NL

Notes: Figure 2.10 illustrate the number of households conducting cross-border shopping as a function of the distance of the households residence to the border.

Given our theoretical discussion outlined in Section 2.3.1, we know that markets are integrated in this case, and the estimates obtained for the price gaps represent a measure of border costs. The reported numbers (columns ve to seven) suggest a considerable variation in these costs with median values lying in the range of 15% to 20%. However, these values are likely to be biased downward for the following reason: Given the inclusion of cross-border shopping observations, i.e., of the - lower - prices abroad, the true value of the price discontinuity at the border is likely bigger than the one reported. To obtain an unbiased measure, we remove the cross-border shopping observations. The results are presented in the lower panel of Table 2.6. As expected, we obtain bigger estimates of border costs, with the differences being in the range of 1 to 2 percentage points.[31]

[30] Results for the other sample periods and actual (non-absolute) border estimates are presented in Section 2.E.

[31] For BE-GE, differences partly have the "incorrect" sign. This is due to the fact that few observations were added from the group of insigni cant estimates with cross-border shopping activities. Given the small number of EANs included in this country group these few additional observation strongly in uence the summary statistics.

For cases, in which we observe no cross-border shopping but obtain a signi cant border estimate, our theoretical considerations do not allow us to differentiate between integration and segmentation. The signi cant price gap at the border presents a lower bound on the border costs. The obtained numbers (columns nine to eleven) suggest that lower bounds on border costs tend to be somewhat bigger than those obtained for integrated markets and signi cant border estimates. The median values range from around 18% for GE-NL to around 20% for BE-NL. The numbers for BE-GE range between around 15% to 18%, however, the results for this country-pair are based on very few observations though. The standard deviations show that differences in the lower bound of border costs are also considerable in this case, with values up to 40% again not being exceptional.

Turning to the group of goods with not signi cant border estimates and comparing the outcomes between goods for which cross-border shopping takes place with those for which only domestic purchases are recorded, we notice that median border estimates tend to be somewhat smaller in the group with cross-border shopping. However, the differences are fairly small. As stated in Section 2.3.1, we can conclude that markets with cross-border shopping in this group are integrated, whereas no inference on this issue can be made for markets for which such purchases are not conducted. Estimates of border costs can also generally not be derived in the group of not signi cant border estimates. An exception is given when re-running regressions for goods for which cross-border purchases made leads to signi cant outcomes after these purchases are excluded. The obtained border coef cient then represents an estimate of border costs. The comparison of results with and without cross-border shopping information prevails that for some goods this is indeed the case. A particular large number of cases is obtained for the BE-NL country pair, where around 20% of border estimates become signi cant after dropping the cross-border observations. In all remaining cases, which represent around 20% of all goods considered, no direct evidence on integration and border costs is possible.

TABLE 2.6: Border estimates and cross border shopping: absolute numbers

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
	All observations																	
	BE-NL																	
2005q1	680	77	142	0.161	0.216	0.202	379	0.197	0.239	0.164	45	0.025	0.037	0.036	114	0.036	0.060	0.094
2008q4	1,021	79	336	0.194	0.223	0.169	474	0.203	0.253	0.272	107	0.030	0.044	0.045	104	0.045	0.058	0.056
	GE-NL																	
2005q1	354	77	80	0.141	0.165	0.105	192	0.189	0.225	0.147	27	0.020	0.044	0.057	55	0.034	0.050	0.051
2008q4	524	77	112	0.135	0.182	0.184	291	0.180	0.225	0.174	36	0.038	0.045	0.038	85	0.033	0.058	0.063
	BE-GE																	
2005q1	68	81	3	0.169	0.190	0.104	52	0.181	0.231	0.169	1	0.043	0.043	.	12	0.022	0.050	0.067
2008q4	159	73	11	0.207	0.220	0.162	105	0.147	0.194	0.148	12	0.083	0.101	0.083	31	0.060	0.065	0.062
	Excluding cross-border shoppings																	
	BE-NL																	
2005q1	669	78	140	0.175	0.218	0.195					36	0.013	0.028	0.031				
2008q4	993	81	331	0.201	0.231	0.170					84	0.026	0.041	0.046				
	GE-NL																	
2005q1	348	76	74	0.169	0.188	0.112					27	0.031	0.042	0.049				
2008q4	517	77	108	0.146	0.196	0.191					33	0.039	0.054	0.076				
	BE-GE																	
2005q1	67	81	2	0.107	0.107	0.039					1	0.037	0.037	.				
2008q4	158	76	14	0.194	0.226	0.165					8	0.050	0.088	0.091				

Notes: Table 2.6 provides summary results for the estimated border coef cients of all goods split by the prevalence of cross-border shopping (CB shopping vs. no CB shopping) and the signi cance of the estimated border coef cient (Signi cant estimates vs. Not signi cant estimates). The upper panel is based on all available purchase data, whereas the lower panel excludes the cross-border transactions.

To obtain an idea of how goods subject to cross-border shopping compare to those which are not, Table 2.7 provides summary statistics for the two groups of goods employing data from the biggest data sample, i.e., from BE-NL, for illustration purpose.[32] The reported numbers show that goods that are bought abroad are relatively cheaper, and are purchased (much) more often by (many) more households than goods which are bought domestically only. These patterns are also found for the other two country groups. As regards goods categories, tables presented at the end of Section 2.E.4 show that goods that are purchased abroad are present in basically all categories.

TABLE 2.7: Summary statistics on goods subject to (non) cross-border shopping, 2008Q4

Belgium - the Netherlands										
	N	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Goods with cross-border purchases										
Price	443	5.22	32.00	0.02	0.08	0.23	0.51	0.82	4.36	371.47
Purchases	443	192.23	183.56	27.00	40.00	77.00	129.00	230.00	570.00	1,264.00
Households	443	113.97	96.42	22.00	31.00	52.00	82.00	139.00	303.00	709.00
Abs. ln. price gap	443	0.19	0.17	0.00	0.01	0.06	0.16	0.27	0.48	1.57
Goods without cross-border purchases										
Price	578	8.25	44.92	0.03	0.09	0.25	0.56	1.17	11.64	571.66
Purchases	578	97.88	78.84	21.00	29.00	45.00	70.50	127.00	264.00	538.00
Households	578	64.20	45.42	20.00	24.00	33.00	48.00	81.00	157.00	310.00
Abs. ln. price gap	578	0.20	0.18	0.00	0.01	0.08	0.17	0.28	0.53	1.55
All goods										
Price	1,021	6.93	39.85	0.02	0.09	0.24	0.54	0.97	7.18	571.66
Purchases	1,021	138.82	142.50	21.00	31.00	54.00	90.00	170.00	425.00	1,264.00
Households	1,021	85.79	76.19	20.00	25.00	38.00	61.00	103.00	232.00	709.00
Abs. ln. price gap	1,021	0.20	0.17	0.00	0.01	0.07	0.16	0.27	0.51	1.57
Observations	1021									

Notes: Table 2.7 provides summary statistics on goods for which cross-border shopping information is available (upper panel) and those for which no such information is available (medium panel). Summary statistics on all goods are provided in the lower panel. The variable "Price" corresponds to the average price of the good, the variable "Purchases" provides information on the number of purchases of the good, similarly the variable "Households" captures the number of households that a given good bought. "Abs. ln. price gap" denotes the percentage difference in the average price between the two countries.

2.4.4 Indirect evidence from (inter-)national price co-movements

As outlined above, the ndings for the group of goods with no cross-border shopping activities, i.e., the majority of cases, do not allow us to make any inference on whether the markets for these goods are integrated or segmented. Following an idea by GGHL,

[32] The statistics for the BE-GE and GE-NL sample goods are presented in Tables E.4.1 and E.4.2 of Section 2.E.

we provide indirect evidence on this question in the following. We do so by examining the extent to which prices in border and non-border regions move together across and within countries. In case of integration and the presence of strong arbitrage forces, any change in a given price in one region should induce arbitrage movements, implying that prices should move fairly closely together over time. To implement the approach, we first define a border and a non-border region for each country, and then compute the average price of a given good in each region and each time period. Finally, we calculate simple correlation coefficients for the four price time series that are obtained in this way. The time frequency chosen is monthly and regions are defined by their distance to the border. Border regions comprise all observations within a distance of 80 km or less to the border, whereas non-border regions are observations which are more than 80 km but less than or equal to 160 km away from the border.[33]

Table 2.8 presents summary statistics on the correlation coefficients for the price series (measured in terms of monthly percentage changes) of the goods for which we were able to compute complete monthly time series throughout our sample period. The results for all observations (upper panel) are very revealing: Correlation coefficients tend to be positive and sizeable for within-country regions and distinctly lower for all cross-country pairs. This is a strong indicator that border costs prevent shocks to the price of a good in given country to propagate across borders.

Splitting goods based on whether cross-border shopping took place and whether border estimates are significant and re-computing correlation coefficients for each group (lower four panels of Table 2.8), we obtain the result that the correlation patterns for the goods without cross-border shopping mirror those of the overall sample. This suggests that international markets are segmented while national ones are integrated. A somewhat surprising finding is that results are more or less equal for markets for which cross-border shopping is observed, given that one would have expected that cross-border shopping might lead to a somewhat stronger synchronization of price movements. This finding suggests that price-setting of retailers appears not to be strongly affected by extent of cross-border shopping and that retailers, following a national pricing strategy, take potentially suboptimal outcomes in certain regions into account.

[33] Results obtained when employing a width of 40 or 60 km for a region are reported in Section 2.F.

TABLE 2.8: Co-movement of prices within and across countries

	All observations											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.52	1.00			0.54	1.00			0.65	1.00		
p_b^*	0.15	0.14	1.00		0.13	0.16	1.00		0.12	0.12	1.00	
p_{nb}^*	0.13	0.17	0.64	1.00	0.14	0.17	0.66	1.00	0.11	0.12	0.58	1.00
	Signi cant border estimates, cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.50	1.00			0.54	1.00			0.68	1.00		
p^*	0.15	0.12	1.00		0.13	0.16	1.00		0.13	0.13	1.00	
p_{nb}^*	0.10	0.15	0.61	1.00	0.14	0.17	0.66	1.00	0.12	0.12	0.59	1.00
	Signi cant border estimates, no cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.53	1.00			0.54	1.00			0.66	1.00		
p^*	0.16	0.15	1.00		0.13	0.16	1.00		0.12	0.13	1.00	
p_{nb}^*	0.14	0.17	0.64	1.00	0.13	0.17	0.67	1.00	0.12	0.13	0.60	1.00
	Not signi cant border estimates, cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.37	1.00			0.53	1.00			0.66	1.00		
p^*	0.13	0.10	1.00		0.12	0.18	1.00		0.13	0.15	1.00	
p_{nb}^*	0.09	0.12	0.51	1.00	0.14	0.19	0.61	1.00	0.12	0.13	0.55	1.00
	Not signi cant border estimates, no cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.49	1.00			0.49	1.00			0.60	1.00		
p^*	0.16	0.12	1.00		0.08	0.21	1.00		0.14	0.16	1.00	
p_{nb}^*	0.10	0.14	0.52	1.00	0.13	0.22	0.58	1.00	0.13	0.14	0.54	1.00

Notes: Table 2.8 presents summary statistics of simple correlation coef cients of four price time series. These time series measure the price of a given good measured over the sample period in four regions, namely the border regions of the country pairs indicated on top of each panel and one non-border region of each country. Border regions comprise all observations within a distance of 80 km or less to the border, whereas non-border regions are observations which are more than 80 km but less or equal to 160 km away from the border. The time frequency chosen is monthly.

2.5 Retailer variation and border effects

Assuming some price-setting power, the widely used model of monopolistic composition implies that the (regular) price a retailer sets depends on the marginal costs of a good and the preferences of its customers, where the latter determine the markup over marginal costs. For national-brand goods, a major part of the marginal costs are given by the wholesale price of the good, which is determined in a negotiation process between the retailer and the manufacturer. Marginal costs for private-label goods - which are of considerable importance in the retail markets of our sample markets - on the other hand are given by the marginal production costs of the good.

These considerations have the following implications for the prices that we should observe for a given retailer across markets: Within a country, marginal costs are likely to be more or less the same across locations. For national-brand goods, this is the case, since it can be assumed that contracts between a retailer and a manufacturer are normally signed at a country-wide level, implying identical costs of the goods across markets. Likewise, the production of private-label goods can be assumed to follow a nationally standardized process, again leading to equal marginal costs for stores within a country. On the other hand, consumer preferences might differ across geographic areas. An optimizing retailer would then "price-to-market" such that we should observe differences in prices across regions. DellaVigna and Gentzkow (2019) show that this is generally not the case, i.e., a given retailer tends to follow a uniform-pricing strategy within a given country.

However, a different outcome might apply for the price of a good supplied by a given retailer in two distinct countries. GGHL show that differences in the price of a given good sold by an identical retailer both in the US and Canada are mostly driven by differences in marginal costs. Differences in markups across countries do not seem to play a major role. However, it might well be that this result is speci c to the US/Canadian situation.

When we consider the situation of a market in which several retailers offer the same good, we would expect a larger dispersion in prices both within and in particular across countries for a variety of reasons. On the one hand, prices for the same good - which conceptionally can only be a national-brand good - might be different across retailers because they face different wholesale prices due to different negotiation powers with respect to

the manufacturer. On the other hand, different retailers might target different customers with different preferences, such that optimal markups might differ as well. In the likely - and empirically relevant - case that retailer composition differs across countries, these considerations imply that the observed price differences would be particularly big across countries.

Given that our data contains purchase information from basically all of the relevant retailers operating in a given market, we can directly address the question of whether price differences are, on average, greater across markets characterized by larger retailer heterogeneity. The compositions of retailers in our sample differ considerably across countries, with most of the retailers operating only in one market. There are, however, a non-negligible number of retailers which operate in at least two markets. For BE-NL there are fteen common retailers, while for both GE-NL and BE-GE our data sample comprises six common retailers. Four (very large) retailers operate in all three countries.

Table 2.9 compares within- and cross-country price dispersion for the goods included in our estimation sample when all retailers are considered and when the prices of a common retailer (across two given markets) are considered only. The presented statistics are from 2008Q4.[34] The gures support the considerations outlined above. We observe that both within and across countries, price gaps are bigger in the multi-retailer setting than in the common-retailer case. Our within-country evidence for the common-retailer case, moreover, clearly supports the ndings by DellaVigna and Gentzkow (2019), given that observed price gaps are really small within each country in this case.

[34] Additional results for 2005Q1 are provided in Section 2.G reports summary statistics for the case that all possible goods (not only the ones for which we have border estimates) are considered. The results are comparable in all cases.

TABLE 2.9: Within-/Cross-country price dispersion - All retailers and common retailers

2008Q4		Price gap			Absolute price gap		
			All retailers				
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	38357	0.04	0.00	12.35	3.42	0.88	11.86
NL-NL	50619	0.20	0.00	8.48	4.05	1.19	7.45
BE-NL	95447	7.53	6.54	26.46	19.99	16.08	18.89
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	290942	-0.11	0.00	8.02	4.00	0.58	6.95
NL-NL	26453	0.23	0.00	7.71	3.28	0.51	6.98
GE-NL	182126	-6.78	-5.42	24.06	18.32	13.40	17.01
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	7624	0.29	0.00	4.60	2.08	0.37	4.11
GE-GE	104222	0.06	0.00	6.57	3.20	0.61	5.74
BE-GE	59391	4.21	0.66	21.99	17.00	12.79	14.57
			Common retailers				
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	19090	0.03	0.00	3.14	0.98	0.00	2.98
NL-NL	25965	0.00	0.00	4.69	1.28	0.11	4.51
BE-NL	48313	-0.41	-0.13	23.97	18.00	14.26	15.84
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	201210	-0.19	0.00	5.47	2.16	0.00	5.03
NL-NL	18753	0.08	0.00	3.68	1.16	0.11	3.49
GE-NL	128072	-7.80	-6.48	22.69	17.49	12.82	16.42
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	5673	-0.01	0.00	2.73	1.10	0.02	2.50
GE-GE	79560	-0.02	0.00	4.61	2.00	0.00	4.15
BE-GE	44740	2.52	0.00	22.47	16.83	11.90	15.11

Notes: The table reports descriptive statistics on within and between-country price dispersion. Only goods for which we have a border estimate are considered. Figures in the upper panel are computed for the case that purchases from all retailers are considered whereas for the lower panel the prices of goods purchased from a common retailer (across two given markets) are considered only. 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. Then, for each good all possible bi-regional price gaps are computed. Finally, for each considered subsample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (2.1). 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 2.10: Regression discontinuity results: absolute values, only goods sold by the same retailer

Period	N	Median	Mean	Std.	Sign.(%)
			BE-NL		
2005Q1	293	0.142	0.181	0.165	85.7
2008Q4	479	0.143	0.182	0.160	87.7
			GE-NL		
2005Q1	233	0.144	0.173	0.142	85.0
2008Q4	354	0.134	0.173	0.162	86.7
			BE-GE		
2005Q1	56	0.160	0.201	0.173	85.7
2008Q4	119	0.116	0.164	0.149	84.9

Notes: The table reports absolute vlaues of regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Only goods that are sold by the same retailer in two different countries are considered. Results are based on estimating Equation (2.2). For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

Table 2.10 reports summary statistics on border estimates for goods which are sold by a common retailer.[35] The results do not exhibit a consistent pattern across country groups. For BE-NL, we nd that numbers are somewhat smaller than in the overall sample. The median value, e.g., in 2005Q1 is 0.142 compared to 0.149 in the overall sample. For 2008Q4 the difference is almost 1 percentage point. The dispersion of estimates is also smaller (0.165 compared to 0.171 in 2005Q1 and 0.160 compared to 0.224 in 2008Q4). For GE-NL, median border estimates are slightly (0.4/0.9 percentage points in 2005Q1/2008Q4) larger than those obtained in the overall sample. For the dispersion we also nd a slight increase. Mixed results are obtained for BE-GE with medians being equal (2005Q1) or slightly smaller (2008Q4) than in the overall sample and standard deviations being slightly larger. Summarizing, the results of this section suggest, that the retailer composition does not seem to play a major role for observed border price gaps. This nding might be explained by the fact that in the presence of different retailers

[35] Further regression results are provided in Section 2.G.3.

the dispersion of prices might increase, however, not necessarily, differences in average prices across markets.

2.6 Sample characteristics and border estimates

2.6.1 The role of the composition of sample goods

As illustrated in Figure 2.5 and discussed in detail above, our results suggest an appeciable degree of heterogeneity in border estimates across goods. Moreover, we have seen that summary statistics on the estimated border coef cients differ somewhat across country pairs. One potential explanation for differences in border price gaps could be that factors in uencing the levels of prices such as preferences or marginal costs vary across countries (or only border regions, see the next subsection on this point), implying price differences across them in the presence of border costs/segmentations. However, price gaps could also arise solely for statistical reasons. This could be the case, if, e.g, the the data samples considered for two country groups differ with respect to the patterns of the border estimates of goods included. In the rst part of this subsection, we examine this issue considering only goods commonly purchased in all three country pairs and used in our estimations. In a second step, we group goods by some of their characteristics (price, purchase frequency, category (composition)) and explore whether we systematic relationships between these groups and border estimates can be found.

Restricting the sample to goods that are sold in all three countries in a given period, Table 2.11 (column N) shows that we have a fairly small number of those goods in 2005Q1. However, their quantity has increased, amounting to 41 in 2008Q4. These gures suggest that we have to treat ndings for 2005Q1 with some caution, given that results can be strongly in uenced by a few particular values in such small samples. As a consequence, the fairly high median/mean values obtained for 2005Q1 and the sizeable decrease of these values by more than 10 percentage points between 2005Q1 and 2008Q4 for BE-NL and GE-NL is due to the small sample size in the rst period. More specifically, it seems likely that the fairly large values of the descriptive statistics and border estimates obtained for BE-NL and GE-NL in 2005Q1 are the result of the presence of a few relatively highly priced goods in Belgium and Germany compared to the Netherlands.

TABLE 2.11: Descriptive statistics and regression results: common goods across country pairs

	Price gap				Border est.				Border est. (abs.)			
Period	N	Med.	Mean	Std.	N	Med.	Mean	Std.	Med.	Mean	Std.	Sign.
BE-NL												
2005Q1	1874	12.28	14.04	25.73	16	0.132	0.124	0.263	0.235	0.242	0.152	100
2008Q4	5147	1.12	0.10	25.37	41	0.003	0.008	0.265	0.169	0.196	0.175	90.2
GE-NL												
2005Q1	7129	12.42	9.28	20.16	16	0.090	0.123	0.200	0.156	0.180	0.148	93.8
2008Q4	18643	1.45	1.92	17.26	41	0.022	0.009	0.168	0.105	0.128	0.107	87.8
BE-GE												
2005Q1	6173	0.19	-4.35	23.84	16	-0.075	-0.010	0.237	0.138	0.176	0.152	87.5
2008Q4	16648	-0.40	-5.26	23.42	41	-0.052	-0.016	0.232	0.097	0.106	0.160	87.8

Notes: Table 2.11 reports absolute vlaues of regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2.2). For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km. More detailed descriptive statistics and estimation results are provided in Tables H.1.2 to H.1.4 of Section 2.H.

Comparing the obtained results within a sample period we observe that price gaps exhibit consistent/transitive pattern across country pairs, at least when median values are considered: For 2005Q1, median values of relative price gaps suggest that Belgium is (very) slightly more expensive than Germany and both are considerably more expensive than the Netherlands. For 2008Q4, Germany is the most expensive country, followed by Belgium and the Netherlands. For mean values transitivity doesn t hold which is likely due to some outliers in the BE-GE sample. For border estimates, we don t get any transitive results suggesting that border results might be due to factors speci c to border regions. This issue is examined in more detail in the next subsection, employing a broader sample of data.

If one compares sizes and distributions of border estimates across country groups one nds differences which are of similar size (or even larger size) than that obtained not controlling for compositional heterogeneity in the estimation sample. This nding suggests that country-speci c factors induce differences in prices of identical goods.

To shed some light on the question whether there exist some systematic relationships between major goods characteristics and border estimates, we group goods by features

which we can compute using our data sample or which are included in it. More speci - cally, we classify goods according to their price into low- and high-price goods (with the median price acting as the separation value) and according to their purchase frequency into low- and high-frequency goods (again using the median frequency as the separating value). Finally, we employ a variable provided by the data supplier which assigns goods into different categories.

TABLE 2.12: Price and purchase frequency of goods and border estimates

		Border coeff.			Absolute border coeff.		
	N	Median	Mean	Std	Median	Mean	Std
				Low-price goods			
BE-NL	497	0.032	0.051	0.343	0.151	0.214	0.273
GE-NL	259	-0.056	-0.078	0.218	0.132	0.170	0.157
BE-GE	79	-0.043	0.007	0.236	0.132	0.172	0.161
				High-price goods			
BE-NL	496	0.109	0.103	0.240	0.162	0.200	0.167
GE-NL	258	-0.053	-0.061	0.262	0.137	0.191	0.189
BE-GE	79	0.074	0.078	0.191	0.134	0.160	0.128
				Goods with a low purchase frequency			
BE-NL	498	0.060	0.070	0.274	0.164	0.209	0.191
GE-NL	260	-0.076	-0.095	0.247	0.154	0.193	0.180
BE-GE	81	0.015	0.028	0.222	0.134	0.167	0.148
				Goods with a high purchase frequency			
BE-NL	495	0.095	0.084	0.319	0.152	0.205	0.258
GE-NL	257	-0.024	-0.044	0.233	0.118	0.167	0.167
BE-GE	77	0.039	0.058	0.211	0.127	0.165	0.143

Notes: Table 2.12 reports summary statistics of border estimates obrtained when grouping goods according to their price (into low and high price goods) or their purchasing frequency (into low and high frequency goods).

Concerning the price of a good, the results in Table 2.12 do not reveal a consistent relationship between this variable and its border estimates. While in most cases, reported values are fairly similar across goods categories, we observe some marked differences (e.g., for the BE-NL case) which appear to be country-group speci c though.

When considering the purchase frequency of a good we observe consistent patterns across country groups. Here, median/mean values tend to be smaller by up to around 5 percentage points (GE-NL) for low-frequency goods, whereas the summary statistics

for the absolute values suggest border price gaps are up to somewhat more than 3% points larger (again GE-NL) for this group of goods. Taking into the account the considerable standard deviations of each group which all tend to be at least as large as that for the overall sample we have to note though that enormous heterogeneities exist between goods within a frequency group.

TABLE 2.13: Border estimates by goods categories, all country pairs, absolute values)

Category	2005Q1				2008Q4			
	N	Median	Mean	Std.	N	Median	Mean	Std.
AlcholfreeCO2	8	0.110	0.164	0.178	5	0.107	0.292	0.278
AlcholfreeNoCO2	35	0.147	0.174	0.139	48	0.127	0.150	0.129
Alcohol					1	0.703	0.703	.
Animalcare	19	0.145	0.152	0.100	59	0.119	0.144	0.109
Babyproducts	5	0.069	0.224	0.355	7	0.099	0.144	0.110
Basicfood	23	0.215	0.298	0.233	35	0.177	0.305	0.269
Beer	4	0.261	0.229	0.112	6	0.065	0.057	0.039
Bodycare	40	0.173	0.214	0.156	75	0.171	0.212	0.160
Candy	150	0.126	0.153	0.118	211	0.159	0.196	0.195
Cereals	13	0.075	0.140	0.136	15	0.060	0.135	0.137
DairyWhite	95	0.142	0.164	0.134	168	0.124	0.150	0.128
DairyYellow	26	0.202	0.185	0.110	35	0.164	0.195	0.163
Delicasees	35	0.186	0.212	0.274	46	0.300	0.354	0.329
Fatoils	13	0.267	0.284	0.176	13	0.147	0.166	0.098
Frische	17	0.167	0.145	0.106	20	0.064	0.131	0.113
Frozenproducts	80	0.140	0.157	0.126	109	0.143	0.213	0.327
HotDrinks	11	0.224	0.218	0.127	49	0.183	0.217	0.157
HouseholdCleansers	54	0.147	0.217	0.201	62	0.176	0.236	0.198
Hygieneproducts	14	0.209	0.262	0.265	16	0.171	0.221	0.161
Laundry	22	0.154	0.163	0.121	19	0.115	0.182	0.159
Liquor	2	0.117	0.117	0.080	7	0.068	0.058	0.037
Meat	51	0.117	0.191	0.196	109	0.119	0.147	0.112
MouthTooth	7	0.294	0.312	0.215	12	0.343	0.346	0.165
PreservedFood	51	0.150	0.181	0.152	58	0.134	0.157	0.120
Readymade	14	0.212	0.240	0.159	41	0.144	0.198	0.177
Rest	305	0.145	0.189	0.181	459	0.152	0.212	0.275
Snacks	129	0.166	0.210	0.164	177	0.152	0.195	0.175
Spreads	23	0.068	0.092	0.077	25	0.096	0.105	0.095
Vegetables	24	0.189	0.225	0.174	35	0.137	0.194	0.184
Wine	8	0.113	0.179	0.189	22	0.196	0.237	0.203

Notes: Table 2.13 reports summary statistics of border estimates grouped by categories.

To group goods by category, we make use of a common classi cation scheme for

the products in our data-set for all countries which rests on the one employed by the national data providers. However, given that the grouping systems of the individual providers slightly differ across countries, we constructed comparable categories of goods by using the classi cation scheme of Germany as a basis and assigning the categories of the other countries to their German counterpart. This classi cation was done using both the assistance by country representatives of GfK and the extensive documentation of the different classi cation schemes, to which we had access at the data providers of ces. The results for the grouping of all goods from the three sample by category (Table 2.13) show that for a given group of goods results can considerably differ across time. This is generally the case when the number of goods is relatively small. Comparing numbers across categories we nd sizeable differences which can be due to sample size issues though. For categories for which fairly many observations are available numbers are not only fairly constant over time but also lie in the range observed for the overall sample.[36]

2.6.2 Border region characteristics and border effects

The RD approach rests on the assumption that all factors other than the border in u-encing the price of a given good evolve "smoothly" in the neighborhood of the border. Further away from the border, these factors can differ considerably in the two respectively considered countries, however. If that is the case, the estimated border effect represents a LATE, speci c to the conditions in the border regions. To examine the extent to which this applies, we redo our RD estimations employing different bandwidths. We moreover compute cross-border differences in the average price of a good for each bandwidth. Should local conditions further away from the border imply different price gaps for two markets being situated in different countries than those found at the border, we should observe a change in the relationship between price gaps and border estimates for different bandwidths. If, e.g., a positive price gap existing at border regions increases for markets further away we should observe that reported price gaps increase by more than the border estimates.

Results for the case that we keep the sample of goods constant across bandwidth are

[36] Results of border estimates grouped by goods categories are presented in Section 2.H.2 of Section 2.H.

TABLE 2.14: Bandwidth and border estimate

	BE-NL						
	N Eans	N	Price Gap	Border	Abs Price Gap	Abs Border	Sign
20	310	41	0.057	0.057	0.148	0.140	76
40	310	72	0.064	0.055	0.151	0.138	81
60	310	100	0.068	0.059	0.153	0.142	85
80	310	126	0.063	0.060	0.152	0.143	86
100	310	148	0.067	0.061	0.152	0.152	88
120	310	158	0.067	0.061	0.152	0.152	88
140	310	165	0.067	0.063	0.152	0.154	89
160	310	170	0.066	0.065	0.152	0.156	89
	GE-NL						
	N Eans	N	Price Gap	Border	Abs Price Gap	Abs Border	Sign
20	74	46	-0.009	-0.024	0.115	0.110	82
40	74	76	-0.006	-0.013	0.114	0.116	84
60	74	123	-0.007	-0.007	0.113	0.113	85
80	74	175	-0.012	-0.017	0.114	0.112	85
100	74	214	-0.007	-0.007	0.116	0.111	86
120	74	255	-0.010	-0.009	0.114	0.112	86
140	74	277	-0.011	-0.009	0.115	0.109	88
160	74	289	-0.013	-0.006	0.113	0.111	89
	BE-GE						
	N Eans	N	Price Gap	Border	Abs Price Gap	Abs Border	Sign
60	72	34	0.003	0.029	0.111	0.114	75
80	72	56	0.002	0.016	0.114	0.115	79
100	72	72	0.005	0.001	0.113	0.114	82
120	72	113	0.005	0.006	0.115	0.113	86
140	72	137	0.001	0.002	0.114	0.115	86
160	72	163	0.003	0.006	0.114	0.115	85

Notes: Table 2.14 presents the median and absolute median values of raw price gaps and border estimates (estimated without cross border shopping observations) for different bandwidth for the 2008Q4. Raw price gaps are computed as the log difference between means of the prices observed at each side of the border. Only goods present at all bandwidths are included. "N Eans" represents the total number of goods while column "N" presents the median number of observation per good. "Sign" reports the percentage of the signi cant estimates.

presented in Table 2.14.[37] Considering the (median) number of observations available for a given bandwidth (column N), we see that this gure increases notably with the bandwidth (by a factor of at least four between the smallest and largest bandwidth). Moreover, a comparison between price gaps and border estimates reveals some differences providing some evidence that the dispersion of prices indeed is larger further away from the

[37] Results employing increasing samples for larger bandwidths (which is due to a larger availability of observations for larger geographic areas) are presented in Section 2.H.3. The ndings are similar.

border. However, differences are small. Moreover, they don t tend to rise with the bandwidth. This is true for all three country pairs. This nding is probably not too surprising if one takes into account the evidence in favor of uniform pricing within a country documented above and inspecting the dynamics of household characteristics living close and further away from the border (Figure 2.4). This gure demonstrates that at least with respect to household characteristics border regions generally do not show patterns of the presented variables being strikingly different from the regions in the rest of the country.

2.7 Summary and conclusions

Employing a rich and unique set of barcode-level price data, this study has examined the importance of borders for goods market integration across three fairly homogeneous and seemingly deeply integrated European countries: Belgium, Germany and the Netherlands. To our knowledge, this is the rst study of this issue for European countries employing such a rich set of micro price data, especially at this level of disaggregation.

Our empirical results show that across these three European countries generally three out of four or even more goods exhibit signi cant price discontinuities at the border. Combining the outcomes of the regression analysis with information on cross-border shopping allows us to derive interesting insights into the question of whether markets are integrated and how costly crossing the border is. Our cross-border shopping data suggest that markets for up to 40% of all goods are integrated. Based on theoretical considerations inspired by GGHL, we can, moreover, provide estimates of border costs for those integrated markets for which we observe signi cant border effects. The results suggests that these costs differ widely across goods with medians or means taking values between 15% and 20%.

For goods for which no cross-border shopping trips occur, we are not able to draw de nite conclusions about the questions of integration vs. segmentation. However, we can provide lower bounds on border costs for those goods that exhibit signi cant price discontinuities at the border. Again, estimates differ considerably with median values ranging between 18% to 20% in the majority of cases. Comparing our ndings with those of the literature, we rst observe that our estimates are larger than those one might have probably expected in the light of the ndings by Cavallo, Neiman, and Rigobon

(2014). They hold that within the euro area basically no dispersion in online prices exists. Moreover, taking into account the very deep (and particularly long) history of integration between the countries considered, including the existence of a common currency, with median values of 15% or above the border costs uncovered are remarkably high. This also holds when compared with the median value of 24% found by GGHL for the U.S. and Canada. Since our data set comprises purchase information from all major retailers present in the considered markets, we are in addition able to shed light on the role of retailer variation across markets for the international price dispersion. When considering patterns of prices of a common retailer across markets, we find that prices are fairly similar within a country. This result confirms previous evidence on uniform pricing by retailers within a country. International evidence shows that price gaps at the border tend to be - if at all - slightly less pronounced than found for the overall sample. We thus conclude that retailer variation does not seem to be a major source of cross-country price dispersions.

Restricting the sample to only goods which are sold in all three countries results in patterns of border estimates across countries which show some pronounced differences. This suggests again that border costs and/or price-setting factors differ across countries, implying potentially distinct border effects for the same good. When we relate characteristics of goods to border estimates, we find that the price does not seem to play a role, while purchase frequency seems to be negatively related to border estimates. In terms of goods categories, we observe substantial differences. However, in many cases estimates might be influenced by small sample issues.

Finally, we show that the obtained border estimates do not reflect particular conditions in border regions but are they seem to be generic, holding for most of the respective country.

Appendices

2.A The role of household income

Previous studies have shown that the shopping behavior of a household is associated with its income, probably because it impacts differentially household s transaction costs.[38] If different income groups indeed experience different transaction costs (including those of crossing the border), then results reported for goods that are bought by both low and high income households might represent average effects with the actual price gaps experienced by the income group with higher transaction costs being potentially markedly higher.

There are a variety of channels that can induce a link between a household s income and its transaction costs. On the one hand, low income household can be expected to have lower transaction costs, given that their opportunity costs (in terms of spending time on searching for the lowest available price) are likely to be lower than those of their high income counterparts. If this difference also applies for the costs of obtaining information from across the border, we would expect to observe signi cant differences in observed border price gaps between these categories of households [for those goods where the transaction costs of the high income households are above and those of low income household below the differences in prices charged by the retailers on different sides of the borders]. On the other hand, there are factors giving rise to transaction costs being higher for low compared to high income households. Firstly, low income households tend to live in poorer districts (or rural areas) that might have an underdeveloped public (private) infrastructures which comes with relatively higher search costs. Relatedly, high income households might have available better transportation and storage capabilities, enabling them to buy larger volumes of a given good which in turn could considerably

[38] For an overview of factors that can explain why households with different incomes might systematically pay different prices for the same good see e.g. Broda, Leibtag, and D. Weinstein (2009) or Mendoza (2011).

reduce transaction costs per unit of the good. There are basically three factors that might disadvantage relatively poorer households in this respect: (i) they might not have a suitable environment to store it, (ii) they might not have a car to transport it or (iii) they might not have an opportunity to spend a bigger share of their wallet for a cheaper product.

To investigate if there exist noticeable differences in border price gaps across income groups, we split the sample into a subsample containing the purchases of low income and one containing the purchases of high income households only. Since the income categories provided by the data supplier slightly differ across country, the criteria to assign households to either of the group vary slightly between countries: A household in Belgium is considered to be a low income household if it earns less than 1984 euros per month. The respective gures for Germany and the Netherlands are 2000 and 1900. We then re-estimate equation 2 of the main text for each good for each subsample separately. To examine the extent to which we observe heterogeneities in the treatment effect induced by differences in transaction costs we focus only on those goods that are bought at a suf cient frequency by both types of households..[39]

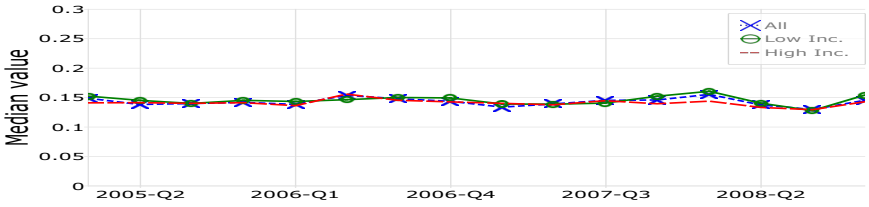
A graphical illustration of obtained border estimates (in absolute values) is given in Figure A.1.[40] In all three cases, the median and mean price gaps at the border are very similar for low and high income groups. There are small differences in some quarters, however, these differences do not exhibit any systematic pattern. Moreover, the proportion of signi cantly estimated price discontinuities (not reported) does likewise not systematically differ across income groups. Overall, our ndings thus suggest that the considerations discussed above do not give rise to differences in border price gaps across households with different incomes.

[39] We apply the same requirement of having at least 10 households on each side of the border as in the baseline estimation. Therefore, we can not obtain estimates for all goods we previously used for both subsamples.

[40] Detailed result statistics are available from the authors upon request.

FIGURE A.1: Dynamics of border coeff. by income groups, absolute values

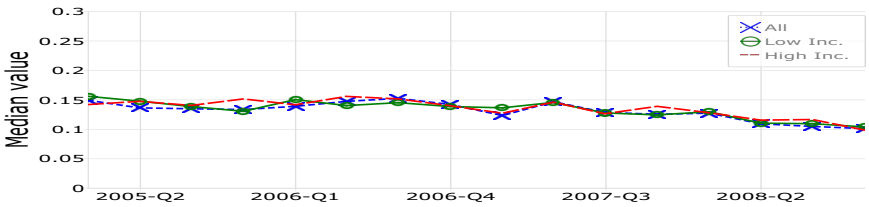
BE-NL



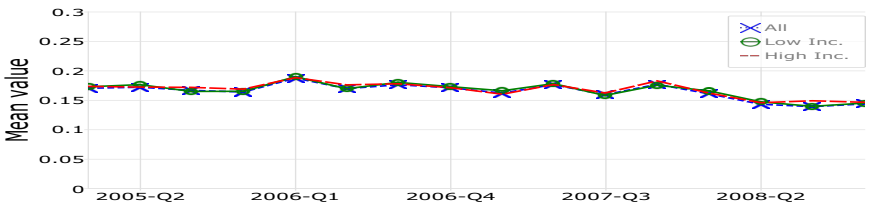
(A) Median values of the border estimates

(B) Mean values of the border estimates

GE-NL

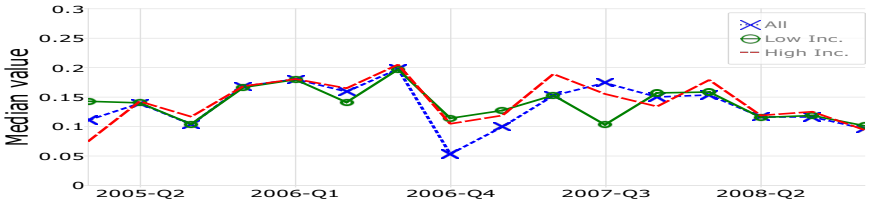


(C) Median values of the border estimates

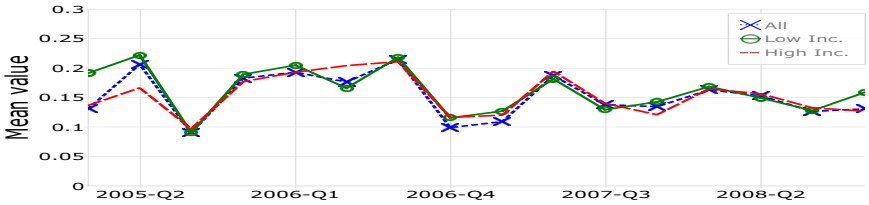


(D) Mean values of the border estimates

BE-GE



(E) Median values of the border estimates



(F) Mean values of the border estimates

Notes: Figure A.1 plots median and mean values of estimated quarterly border coef cients(absolute values). All denotes the the coef cient was estimated using the complete sample, Low Inc. - only data for low income household were used, High Inc. - only data for high income household were used. The border dummy takes the value one if the household is located in the country mentioned rst in the title.

2.B Additional results on the data and descriptive statistics (Section 2.2)

2.B.1 Additional tables on matching results for goods across countries

TABLE B.1.1: 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	100.00	273237	100.00	1037444	100.00

Notes:Table B.1.1 (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 (de ned by the same GTIN) commonly purchased by households in the two countries indicated in the rst 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 rst and second country of the considered country pair.

TABLE B.1.2: Number of EANs and purchases by product category, GE-NL

Catergory	Matched EANs Freq.	%	GE: Purchases Freq.	%	NL: Purchases Freq.	%
AlcholfreeCO2	108	1.33	39900	3.12	2702	1.75
AlcholfreeNoCO2	164	2.03	30039	2.35	6280	4.07
Alcohol	46	0.57	4752	0.37	1014	0.66
Animalcare	472	5.83	23159	1.81	5173	3.35
Babyproducts	61	0.75	2807	0.22	615	0.40
Basicfood	161	1.99	50571	3.96	1777	1.15
Beer	33	0.41	6111	0.48	275	0.18
Bodycare	1148	14.18	31414	2.46	6389	4.14
Candy	874	10.79	161659	12.65	18573	12.04
Cereals	64	0.79	4694	0.37	1522	0.99
Champagne	25	0.31	2939	0.23	66	0.04
DairyWhite	415	5.13	171258	13.40	14444	9.36
DairyYellow	210	2.59	53263	4.17	7124	4.62
Delicasees	340	4.20	52866	4.14	4561	2.96
Fatoils	57	0.70	37941	2.97	682	0.44
Frische	80	0.99	32262	2.52	438	0.28
Frozenproducts	325	4.01	36504	2.86	11104	7.20
HotDrinks	195	2.41	33039	2.59	6351	4.12
HouseholdCleansers	245	3.03	19361	1.52	3438	2.23
Hygieneproducts	159	1.96	45139	3.53	3197	2.07
Laundry	117	1.44	8681	0.68	961	0.62
Liquor	212	2.62	11050	0.86	1126	0.73
Meat	351	4.33	98308	7.69	7589	4.92
MouthTooth	81	1.00	9073	0.71	620	0.40
PreservedFood	384	4.74	73704	5.77	11839	7.68
Readymade	549	6.78	76728	6.00	6264	4.06
Rest	96	1.19	4640	0.36	1552	1.01
Snacks	555	6.85	88353	6.91	16003	10.37
Spreads	84	1.04	17176	1.34	1862	1.21
Vegetables	293	3.62	41051	3.21	8099	5.25
Wine	193	2.38	9382	0.73	2608	1.69
Total	8097	100.00	1277824	100.00	154248	100.00

Notes: Table B.1.2 (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 (de ned by the same GTIN) commonly purchased by households in the two countries indicated in the rst 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 rst and second country of the considered country pair.

TABLE B.1.3: Number of EANs and purchases by product category, BE-GE

Catergory	Matched EANs Freq.	%	BE: Purchases Freq.	%	GE: Purchases Freq.	%
AlcholfreeCO2	80	1.36	3666	4.03	16608	2.27
AlcholfreeNoCO2	111	1.89	2637	2.90	16295	2.23
Alcohol	33	0.56	307	0.34	1871	0.26
Animalcare	345	5.87	3755	4.13	13586	1.86
Babyproducts	65	1.11	899	0.99	3001	0.41
Basicfood	157	2.67	2685	2.95	27721	3.79
Beer	3	0.05	10	0.01	433	0.06
Bodycare	752	12.80	4342	4.77	23921	3.27
Candy	680	11.58	9582	10.53	109550	14.99
Cereals	66	1.12	1156	1.27	5728	0.78
Champagne	32	0.54	197	0.22	1587	0.22
DairyWhite	292	4.97	7583	8.34	82887	11.34
DairyYellow	213	3.63	6926	7.61	36569	5.00
Delicasees	190	3.24	2132	2.34	20255	2.77
Fatoils	33	0.56	408	0.45	24666	3.38
Frische	70	1.19	786	0.86	15947	2.18
Frozenproducts	251	4.27	6009	6.61	22964	3.14
HotDrinks	154	2.62	2008	2.21	20097	2.75
HouseholdCleansers	196	3.34	2401	2.64	11341	1.55
Hygieneproducts	133	2.26	1377	1.51	19740	2.70
Laundry	91	1.55	772	0.85	5667	0.78
Liquor	118	2.01	1448	1.59	4006	0.55
Meat	236	4.02	4930	5.42	63758	8.73
MouthTooth	67	1.14	478	0.53	4119	0.56
PreservedFood	223	3.80	4077	4.48	30588	4.19
Readymade	362	6.16	3754	4.13	39886	5.46
Rest	81	1.38	1512	1.66	5485	0.75
Snacks	465	7.92	7712	8.48	56798	7.77
Spreads	60	1.02	1716	1.89	10904	1.49
Vegetables	211	3.59	4973	5.47	31465	4.31
Wine	103	1.75	721	0.79	3280	0.45
Total	5873	100.00	90959	100.00	730723	100.00

Notes: Table B.1.3 (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 (de ned by the same GTIN) commonly purchased by households in the two countries indicated in the rst 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 rst and second country of the considered country pair.

2.B.2 Comparing goods composition between overall and the estimation sample

The tables in this subsection compare major characteristics of goods included in our data set to those included in our estimation sample. Statistics reported are the average goods prices, number of good purchases in the considered quarter and number of households buying the good.

TABLE B.2.1: Descriptive statistics on goods in the full data samples, 2008Q4

Belgium									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	92.75	422.50	0.00	0.11	0.37	0.90	3.33	500.00	28,017.00
Purchases	16.07	67.85	1.00	1.00	1.00	3.00	9.00	63.00	4,780.00
Households	9.84	30.43	1.00	1.00	1.00	2.00	7.00	40.00	1,209.00
Observations	79557								
Germany									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	16.52	109.76	0.00	0.08	0.25	0.58	1.31	39.75	4,200.00
Purchases	39.57	328.78	1.00	1.00	2.00	6.00	24.00	151.00	61,225.00
Households	27.17	95.42	1.00	1.00	2.00	5.00	20.00	113.00	6,852.00
Observations	151071								
The Netherlands									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	91.77	446.11	0.00	0.12	0.36	0.80	4.33	499.00	43,900.00
Purchases	29.96	129.22	1.00	1.00	1.00	4.00	17.00	124.00	9,499.00
Households	17.84	58.33	1.00	1.00	1.00	3.00	12.00	78.00	3,130.00
Observations	84690								

Notes: Table 3.1 reports descriptive statistics (as indicated in the column headings) on prices paid, the number of purchases and the number of households for the three countries in our sample.

TABLE B.2.2: Descriptive statistics on goods selected for estimation, Q4 2008.

Belgium - Germany									
Belgium									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	1.01	3.02	0.03	0.08	0.25	0.54	0.89	1.85	33.47
Purchases	115.99	100.27	29.00	36.00	60.00	88.00	145.00	267.00	932.00
Households	72.18	50.69	21.00	28.00	41.00	56.00	87.00	172.00	413.00
Observations	159								
Germany									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	0.92	2.74	0.02	0.08	0.25	0.49	0.82	1.63	31.50
Purchases	572.56	678.52	78.00	130.00	236.00	382.00	589.00	1,606.00	6,304.00
Households	389.69	375.34	67.00	116.00	171.00	293.00	438.00	1,058.00	3,159.00
Observations	159								
Belgium - The Netherlands									
Belgium									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	7.16	40.29	0.02	0.09	0.25	0.57	1.04	7.89	594.36
Purchases	73.31	93.87	10.00	14.00	25.00	43.00	83.00	237.00	1,297.00
Households	44.92	49.39	10.00	11.00	18.00	29.00	51.00	131.00	640.00
Observations	1021								
The Netherlands									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	6.79	40.02	0.02	0.09	0.24	0.53	0.98	6.64	546.24
Purchases	130.26	178.75	12.00	23.00	41.00	76.00	153.00	394.00	2,851.00
Households	80.50	89.76	11.00	18.00	30.00	51.00	97.00	232.00	1,180.00
Observations	1021								
Germany - The Netherlands									
Germany									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	2.26	17.54	0.02	0.08	0.23	0.47	0.77	2.01	321.57
Purchases	367.84	645.63	22.00	58.00	107.00	188.50	361.00	1,271.00	7,028.00
Households	250.65	354.88	18.00	50.00	87.00	143.00	267.00	853.00	3,432.00
Observations	524								
The Netherlands									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	2.61	21.15	0.02	0.09	0.24	0.48	0.80	2.14	366.43
Purchases	120.56	181.18	11.00	21.00	38.00	65.00	122.00	399.00	2,210.00
Households	75.54	96.89	10.00	17.00	28.00	44.00	75.50	233.00	1,180.00
Observations	524								

Notes: Table B.2.2 reports descriptive statistics (as indicated in the column headings) on prices paid, the number of purchases and the number of households for the various cross-country samples indicated.

TABLE B.2.3: Descriptive statistics on goods selected for estimation, Q4 2008 (only observations observed at the 80 km bandwidth).

Belgium - Germany									
Belgium									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	1.01	3.03	0.03	0.08	0.25	0.54	0.88	1.79	33.55
Purchases	35.28	33.10	11.00	13.00	17.00	26.00	42.00	74.00	300.00
Households	21.43	15.12	10.00	10.00	12.00	16.00	26.00	48.00	123.00
Observations	159								
Germany									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	0.92	2.74	0.02	0.08	0.26	0.48	0.84	1.67	31.50
Purchases	44.50	50.93	10.00	13.00	19.00	29.00	48.00	116.00	370.00
Households	30.30	27.69	10.00	11.00	15.00	22.00	34.00	76.00	195.00
Observations	159								
Belgium - The Netherlands									
Belgium									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	7.17	40.30	0.02	0.10	0.25	0.57	1.03	7.93	590.40
Purchases	64.13	83.32	10.00	13.00	22.00	36.00	70.00	206.00	1,114.00
Households	39.42	43.86	10.00	10.00	16.00	25.00	45.00	117.00	554.00
Observations	1021								
The Netherlands									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	6.80	40.10	0.02	0.09	0.24	0.53	0.99	6.64	546.67
Purchases	74.69	93.32	10.00	14.00	24.00	44.00	88.00	225.00	1,144.00
Households	46.42	49.18	10.00	11.00	17.00	30.00	58.00	131.00	608.00
Observations	1021								
Germany - The Netherlands									
Germany									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	2.25	17.41	0.02	0.08	0.23	0.47	0.78	2.03	318.47
Purchases	66.70	112.27	10.00	12.00	20.00	34.00	65.50	224.00	1,171.00
Households	44.98	60.15	10.00	10.00	16.00	26.00	48.00	133.00	494.00
Observations	524								
The Netherlands									
	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Price	2.63	21.40	0.02	0.09	0.24	0.48	0.81	2.13	369.92
Purchases	70.73	111.87	10.00	13.00	21.00	36.00	68.00	235.00	1,342.00
Households	44.00	57.66	10.00	10.00	15.00	25.00	44.00	138.00	678.00
Observations	524								

Notes: Table B.2.3 reports descriptive statistics (as indicated in the column headings) on prices paid, the number of purchases and the number of households for the various cross-country samples indicated.

2.B.3 Within and cross-country price dispersion at a regional level: differences in regional mean prices

TABLE B.3.1: Within and cross-country price dispersion at a regional level (2008Q4)

		Distribution of mean values			Distribution of median values		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	-0.01	-0.09	0.46	0.00	0.00	0.00
NL-NL	66	0.30	0.28	0.76	0.00	0.00	0.01
BE-NL	132	7.58	7.73	1.36	4.69	4.95	2.54
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	741	0.00	0.03	0.79	0.00	0.00	0.00
NL-NL	66	0.31	0.36	0.54	0.00	0.00	0.00
GE-NL	468	-7.38	-7.37	1.39	-4.86	-5.28	2.19
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	0.05	0.12	0.53	0.00	0.00	0.00
GE-GE	741	-0.04	0.00	0.71	0.00	0.00	0.00
BE-GE	429	12.88	13.07	1.63	11.60	11.74	2.08
		Distr. of abs. mean values			Distr. of abs. median values		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	4.37	4.42	0.36	1.25	1.27	0.18
NL-NL	66	5.43	5.42	0.30	1.86	1.83	0.19
BE-NL	132	19.47	19.92	2.24	13.91	15.12	3.10
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	741	5.95	5.91	0.64	2.41	2.35	0.71
NL-NL	66	5.00	5.00	0.57	0.93	0.94	0.31
GE-NL	468	19.31	19.75	1.75	14.02	14.49	1.96
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	4.08	3.94	0.57	0.76	0.75	0.20
GE-GE	741	5.89	5.90	0.54	2.32	2.37	0.60
BE-GE	429	21.92	21.97	1.28	17.45	17.87	1.37

Notes: Table B.3.1 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 (2008Q4). Then, for each region pair, the mean/median (absolute) price gap is obtained by rst 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) of the main text. 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.

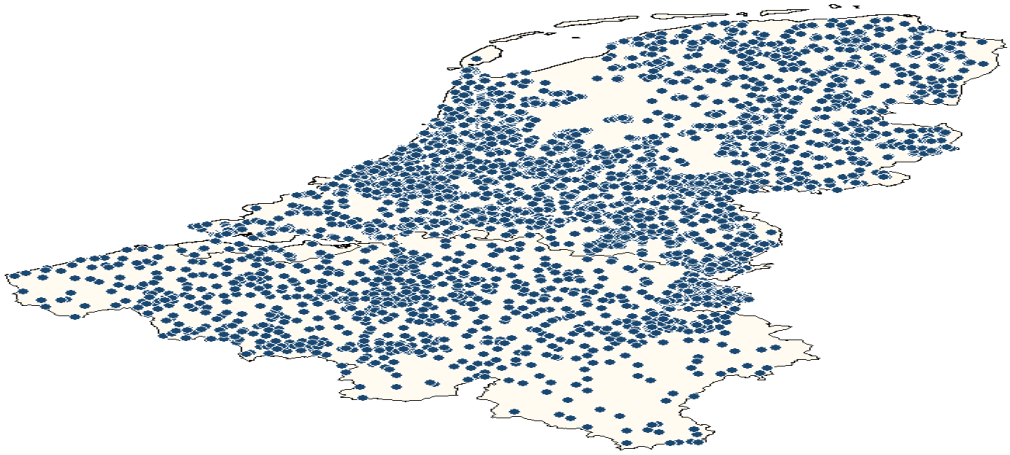
TABLE B.3.2: Within and cross-country price dispersion at a regional level (2005Q1)

		Mean			Median		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	0.30	0.33	0.64	0.00	0.00	0.00
NL-NL	66	0.26	0.29	0.89	0.00	0.00	0.00
BE-NL	132	6.95	7.18	1.36	3.89	4.12	1.99
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	741	0.22	0.21	0.99	0.00	0.00	0.00
NL-NL	66	0.18	0.24	0.72	0.00	0.00	0.00
GE-NL	468	-5.79	-5.71	1.62	-4.69	-5.34	2.42
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	0.48	0.47	0.79	0.00	0.00	0.00
GE-GE	741	0.15	0.14	0.83	0.00	0.00	0.00
BE-GE	429	9.52	9.63	1.66	7.44	7.58	1.72
		Absolute Mean			Absolute Median		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	5.08	5.06	0.29	1.91	1.95	0.21
NL-NL	66	6.07	6.04	0.35	2.13	2.13	0.26
BE-NL	132	20.06	20.28	1.33	14.80	15.39	2.12
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	741	5.15	5.12	0.72	1.26	1.07	0.82
NL-NL	66	6.24	6.23	0.50	1.46	1.53	0.52
GE-NL	468	18.30	18.85	1.78	13.96	14.47	2.21
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	55	5.28	5.27	0.65	2.05	2.07	0.39
GE-GE	741	5.31	5.32	0.67	1.17	1.19	0.72
BE-GE	429	20.29	20.36	1.00	15.22	15.26	0.88

Notes: Table B.3.2 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 (2005Q1). Then, for each region pair, the mean (absolute) price gap is obtained by rst 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 to Equation (1) of the main text. 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.

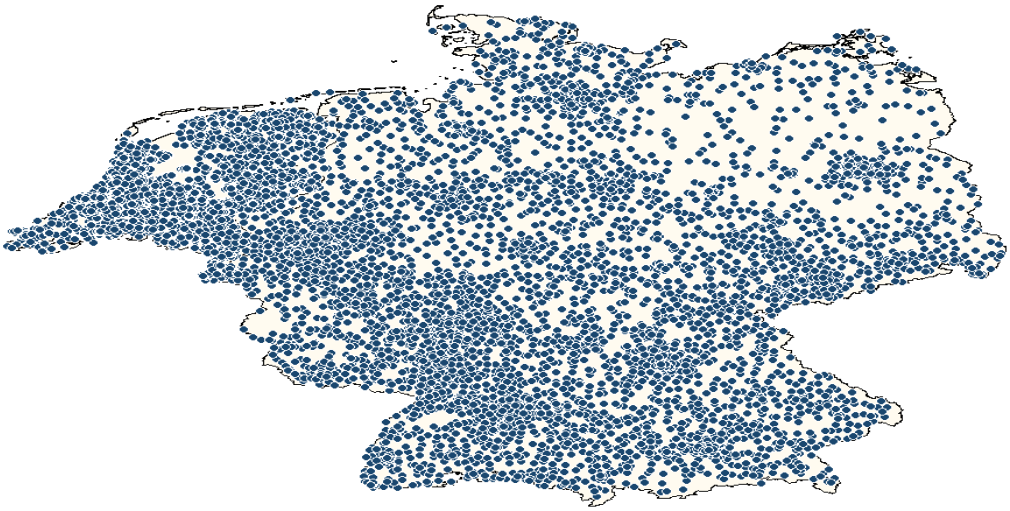
2.B.4 Additional gures illustrating the data sample and descriptive statistics

FIGURE B.4.1: Locations of Belgian and Dutch households



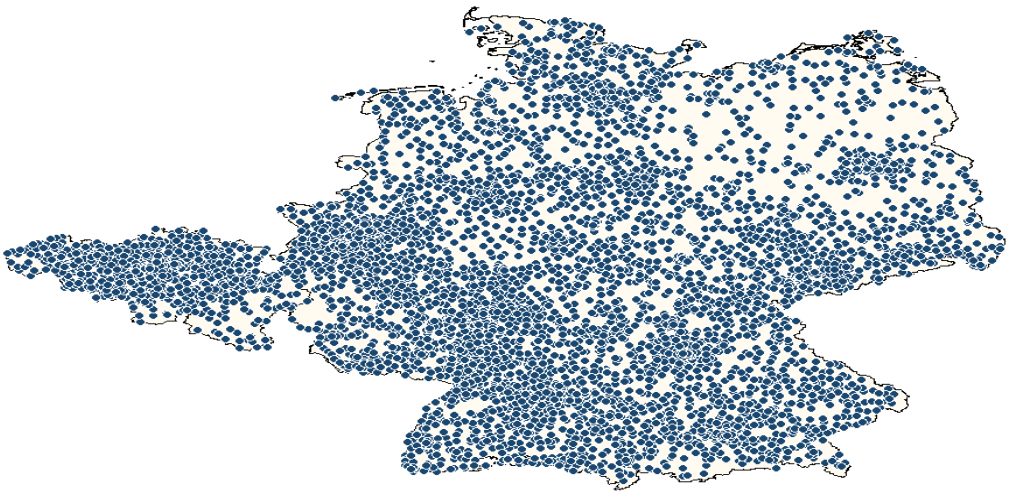
Notes: Figure B.4.1 plots locations of the Belgian and Dutch households included in the data sample in the forth quarter of 2008.

FIGURE B.4.2: Locations of German and Dutch households



Notes: Figure B.4.2 plots locations of the German and Dutch households included in the data sample in the forth quarter of 2008.

FIGURE B.4.3: Locations of Belgian and German households

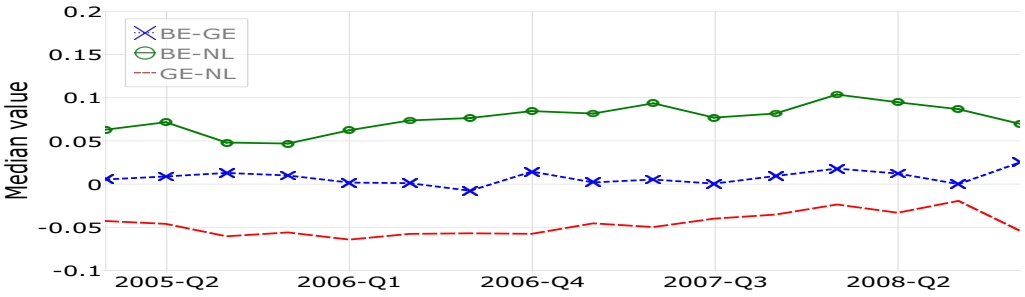


Notes: Figure B.4.3 plots locations of the Belgian and German households included in the data sample in the forth quarter of 2008.

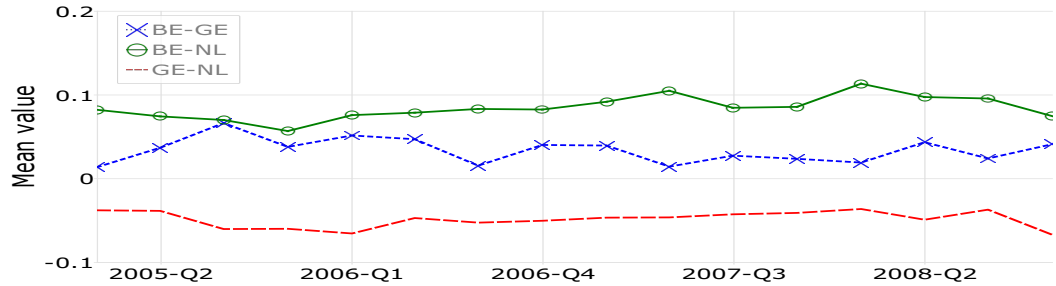
2.C Additional results for the basic regression analysis (Section 2.4.1)

2.C.1 Additional gures illustrating basic regression results

FIGURE C.1.1: Dynamics of border coef cients



(A) Median values of the border estimates



(B) Mean values of the border estimates

Notes: Figure C.1.1 plots median and mean values of estimated quarterly border coef cients. 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 rst in the legend. Thus, positive values for the BE-NL border coef cient indicate that prices are higher in Belgium than in the Netherlands.

2.C.2 Testing for the robustness of the RD baseline regression speci cation

TABLE C.2.1: Regression discontinuity results: model including household income and age as additional covariates

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.004	0.113	13.5	0.000	-0.012	0.118	18.9	0.000	-0.004	0.073	8.8
Border	0.058	0.082	0.250	74.7	-0.040	-0.038	0.218	75.1	0.006	0.010	0.262	76.5
Border x Distance	0.000	0.004	0.158	14.4	0.000	0.016	0.150	22.0	0.000	-0.018	0.135	8.8
Middle aged	0.000	-0.004	0.052	12.0	0.000	-0.007	0.046	9.3	0.000	0.000	0.037	5.9
Old	0.000	0.004	0.070	11.9	0.000	-0.003	0.051	9.3	0.000	-0.001	0.057	8.8
Medium inc.	0.000	0.003	0.073	9.3	0.000	0.000	0.038	8.5	0.000	0.005	0.039	7.4
High inc.	0.000	0.006	0.072	9.5	0.000	0.003	0.045	8.8	0.000	0.001	0.048	11.8
Constant	-0.888	-0.779	1.411	94.6	-0.891	-0.973	1.141	92.9	-0.807	-1.061	1.172	89.7
Observations	665.000	.	.	.	354.000	.	.	.	68.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.001	0.102	12.6	0.000	0.003	0.102	12.9	0.000	0.017	0.121	12.4
Border	0.077	0.077	0.292	78.1	-0.056	-0.066	0.240	76.5	0.038	0.045	0.217	69.9
Border x Distance	-0.001	-0.014	0.175	13.2	0.000	-0.003	0.150	15.3	0.000	-0.028	0.165	11.8
Middle aged	0.000	0.000	0.050	11.3	0.000	-0.004	0.035	9.3	0.000	0.002	0.040	9.8
Old	0.000	0.004	0.065	11.7	0.000	0.002	0.053	8.9	0.000	0.005	0.061	11.8
Medium inc.	0.000	0.000	0.051	11.2	0.000	0.001	0.046	10.7	0.000	-0.001	0.041	13.1
High inc.	0.000	0.002	0.055	9.6	0.000	0.004	0.052	10.5	0.000	0.003	0.055	11.8
Constant	-0.676	-0.582	1.458	95.5	-0.745	-0.757	1.144	95.4	-0.747	-0.804	1.041	93.5
Observations	991.000	.	.	.	503.000	.	.	.	153.000	.	.	.

Notes: Table C.2.1 reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. Positive values of the border coef cient 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 coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

TABLE C.2.2: Regression discontinuity results: model including market (ZIP level) characteristics as covariates

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.003	0.101	12.5	0.000	-0.007	0.098	19.7	0.000	0.008	0.048	4.8
Border	0.064	0.082	0.244	74.2	-0.057	-0.039	0.221	77.7	-0.006	0.032	0.262	83.3
Border x Distance	0.000	0.007	0.147	13.6	0.000	0.004	0.134	18.8	-0.007	-0.028	0.062	4.8
Populaion	0.000	0.000	0.002	11.7	0.000	0.000	0.004	10.5	0.000	-0.001	0.004	4.8
Degurba	0.000	0.003	0.032	8.8	0.000	0.003	0.032	7.4	0.000	-0.005	0.038	9.5
Average Income	0.000	-0.001	0.051	10.2	0.000	0.001	0.037	15.3	0.000	-0.006	0.025	11.9
Constant	-0.884	-0.771	1.433	88.5	-1.002	-1.121	1.087	87.3	-1.083	-1.273	1.299	92.9
Observations	678.000	.	.	.	229.000	.	.	.	42.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.000	0.112	11.0	0.000	0.003	0.107	18.4	0.000	0.004	0.138	16.5
Border	0.069	0.074	0.295	78.1	-0.052	-0.073	0.230	73.1	0.056	0.065	0.220	70.9
Border x Distance	-0.004	-0.016	0.176	14.4	0.000	-0.007	0.145	13.4	0.000	-0.006	0.158	17.5
Populaion	0.000	0.000	0.002	12.8	0.000	0.000	0.004	8.4	0.000	-0.001	0.006	7.8
Degurba	0.000	0.001	0.032	9.7	0.000	0.003	0.033	8.1	0.000	-0.008	0.030	10.7
Average Income	0.000	-0.003	0.046	11.0	0.000	0.002	0.044	13.4	0.000	-0.001	0.052	16.5
Constant	-0.680	-0.560	1.481	89.5	-0.904	-0.820	1.223	86.2	-0.761	-0.905	1.042	82.5
Observations	1021.000	.	.	.	320.000	.	.	.	103.000	.	.	.

Notes: Table C.2.2 reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. Positive values of the border coef cient 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 coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

TABLE C.2.3: Regression discontinuity results, absolute values: model including household income and age as additional covariates

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.028	0.061	0.095	13.5	0.023	0.062	0.101	18.9	0.004	0.031	0.067	8.8
Border	0.145	0.191	0.180	74.7	0.143	0.171	0.140	75.1	0.163	0.198	0.170	76.5
Border x Distance	0.048	0.093	0.127	14.4	0.037	0.084	0.125	22.0	0.030	0.071	0.116	8.8
Middle aged	0.013	0.030	0.043	12.0	0.007	0.023	0.040	9.3	0.005	0.016	0.033	5.9
Old	0.018	0.039	0.058	11.9	0.011	0.028	0.043	9.3	0.006	0.026	0.051	8.8
Medium inc.	0.012	0.029	0.067	9.3	0.010	0.021	0.032	8.5	0.006	0.020	0.034	7.4
High inc.	0.015	0.033	0.064	9.5	0.010	0.024	0.038	8.8	0.006	0.023	0.042	11.8
Constant	1.054	1.270	0.990	94.6	0.953	1.169	0.938	92.9	0.899	1.208	1.018	89.7
Observations	665.000	.	.	.	354.000	.	.	.	68.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.023	0.057	0.085	12.6	0.017	0.051	0.089	12.9	0.009	0.051	0.111	12.4
Border	0.158	0.203	0.223	78.1	0.133	0.178	0.173	76.5	0.121	0.166	0.147	69.9
Border x Distance	0.043	0.096	0.147	13.2	0.037	0.085	0.124	15.3	0.022	0.078	0.148	11.8
Middle aged	0.010	0.024	0.045	11.3	0.007	0.020	0.029	9.3	0.007	0.022	0.034	9.8
Old	0.015	0.033	0.056	11.7	0.011	0.029	0.045	8.9	0.010	0.031	0.053	11.8
Medium inc.	0.010	0.025	0.045	11.2	0.007	0.023	0.040	10.7	0.008	0.021	0.035	13.1
High inc.	0.012	0.027	0.048	9.6	0.008	0.027	0.045	10.5	0.007	0.026	0.048	11.8
Constant	0.931	1.190	1.022	95.5	0.898	1.061	0.869	95.4	0.842	1.007	0.844	93.5
Observations	991.000	.	.	.	503.000	.	.	.	153.000	.	.	.

Notes: Table C.2.3 reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of the estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

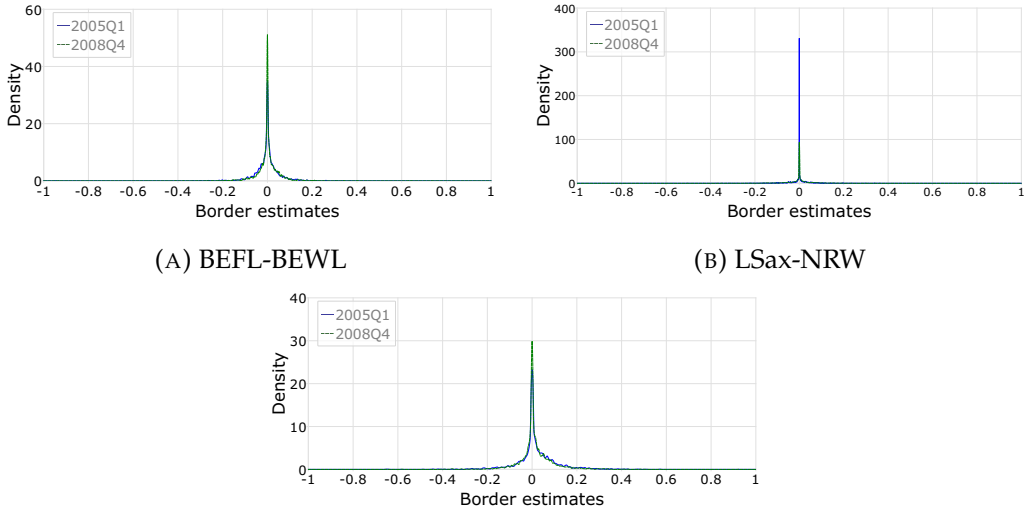
TABLE C.2.4: Regression discontinuity results, absolute values: model including market (ZIP level) characteristics as covariates

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.031	0.059	0.082	12.5	0.024	0.051	0.084	19.7	0.004	0.022	0.043	4.8
Border	0.149	0.190	0.174	74.2	0.135	0.174	0.141	77.7	0.150	0.200	0.168	83.3
Border x Distance	0.045	0.088	0.118	13.6	0.035	0.076	0.111	18.8	0.019	0.043	0.053	4.8
Populaion	0.000	0.001	0.002	11.7	0.001	0.002	0.003	10.5	0.001	0.002	0.003	4.8
Degurba	0.009	0.019	0.026	8.8	0.006	0.016	0.028	7.4	0.004	0.014	0.036	9.5
Average Income	0.016	0.030	0.042	10.2	0.011	0.022	0.030	15.3	0.004	0.014	0.022	11.9
Constant	1.050	1.281	1.004	88.5	1.030	1.237	0.951	87.3	1.232	1.462	1.076	92.9
Observations	678.000	.	.	.	229.000	.	.	.	42.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.025	0.058	0.096	11.0	0.024	0.056	0.092	18.4	0.016	0.069	0.120	16.5
Border	0.154	0.201	0.228	78.1	0.127	0.173	0.168	73.1	0.134	0.171	0.152	70.9
Border x Distance	0.047	0.099	0.146	14.4	0.036	0.083	0.119	13.4	0.033	0.086	0.133	17.5
Populaion	0.000	0.001	0.002	12.8	0.001	0.002	0.004	8.4	0.001	0.003	0.005	7.8
Degurba	0.007	0.016	0.027	9.7	0.006	0.018	0.028	8.1	0.007	0.017	0.026	10.7
Average Income	0.013	0.027	0.038	11.0	0.010	0.026	0.036	13.4	0.009	0.028	0.044	16.5
Constant	0.933	1.193	1.041	89.5	0.927	1.125	0.950	86.2	0.773	1.044	0.901	82.5
Observations	1021.000	.	.	.	320.000	.	.	.	103.000	.	.	.

Notes: Table C.2.4 reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of the estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

2.D Additional results on the countercyclical evidence analysis (Section 2.4.2)

FIGURE D.1: Counterfactual evidence: kernel density estimates of borders across within-country regions



(A) BEFL-BEWL

(B) LSax-NRW

(C) NLE-NLS

Notes: Figure D.1 plots kernel density functions of the estimated border coefficients. The considered sample periods are 2005Q1 and 2008Q4. 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 first country of a given country pair. Thus, positive values for the BEFL-BEWL border coefficient indicate that prices are higher in Flanders than in Wallonia.

TABLE D.1: Counterfactual evidence: within-country price dispersion

2005Q1			Price gap			Absolute price gap	
FL-WL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
FL-FL	82396	0.08	0.00	12.35	4.88	1.50	11.34
WL-WL	50993	0.19	0.00	13.51	4.91	1.54	12.58
FL-WL	159169	0.49	0.00	14.37	5.26	1.64	13.38
LSax-NRW	N	Mean	Median	St.Dev	Mean	Median	St.Dev
LSax-LSax	59962	0.35	0.00	12.28	5.61	0.47	10.93
NRW-NRW	119684	0.31	0.00	11.90	5.56	1.28	10.52
LSax-NRW	209079	-0.61	0.00	13.39	6.36	1.41	11.80
NLE-NLS	N	Mean	Median	St.Dev	Mean	Median	St.Dev
NLE-NLE	26136	0.56	0.00	13.81	6.63	1.96	12.13
NLS-NLS	12469	-1.07	0.00	12.80	6.01	1.95	11.35
NLE-NLS	64368	-0.09	0.00	14.26	6.94	2.25	12.45
2008Q4			Price gap			Absolute price gap	
FL-WL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
FL-FL	113178	0.03	0.00	13.80	4.68	1.41	12.99
WL-WL	74167	-0.09	0.00	15.48	4.73	1.37	14.74
FL-WL	221305	0.53	0.00	16.27	5.15	1.56	15.44
LSax-NRW	N	Mean	Median	St.Dev	Mean	Median	St.Dev
LSax-LSax	91614	-0.23	0.00	11.25	5.36	0.61	9.90
NRW-NRW	185320	0.16	0.00	10.48	5.13	1.12	9.14
LSax-NRW	318574	-0.35	0.00	12.03	6.07	1.52	10.39
NLE-NLS	N	Mean	Median	St.Dev	Mean	Median	St.Dev
NLE-NLE	38164	0.25	0.00	12.32	5.98	1.74	10.78
NLS-NLS	17755	-0.43	0.00	12.17	5.46	1.56	10.88
NLE-NLS	92229	-0.25	0.00	13.49	6.43	1.96	11.86

Notes: Table D.1 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 subsample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (1) of the main text. 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.

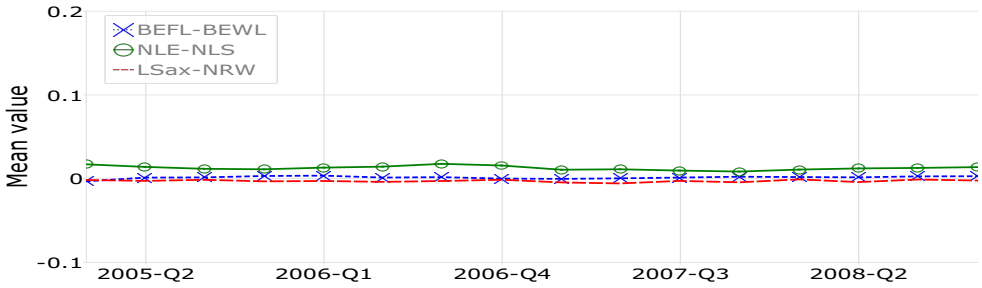
TABLE D.2: Counterfactual evidence: regression discontinuity results, absolute values

	BEFL-BEWL				LSax-NRW				NLE-NLS			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.040	0.088	0.189	16.0	0.003	0.040	0.080	10.1	0.041	0.096	0.143	13.5
Border	0.017	0.034	0.078	11.2	0.003	0.030	0.062	6.5	0.028	0.057	0.077	14.3
Border x Distance	0.057	0.115	0.207	12.7	0.009	0.067	0.129	8.3	0.066	0.132	0.180	11.2
Constant	1.146	1.417	1.085	90.8	1.237	1.394	0.929	58.2	1.420	1.615	1.053	84.2
Observations	2554.000	.	.	.	2012.000	.	.	.	2300.000	.	.	.
	BEFL-BEWL				LSax-NRW				NLE-NLS			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.031	0.087	0.419	13.5	0.014	0.053	0.100	12.8	0.034	0.083	0.125	13.2
Border	0.013	0.035	0.133	11.0	0.010	0.041	0.069	8.8	0.021	0.048	0.069	14.1
Border x Distance	0.045	0.113	0.459	11.1	0.023	0.086	0.146	10.3	0.052	0.116	0.166	12.3
Constant	1.022	1.330	1.125	88.9	1.094	1.310	0.980	70.2	1.229	1.423	0.989	93.4
Observations	4046.000	.	.	.	3414.000	.	.	.	3942.000	.	.	.

Notes: Table D.2 reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each hypothetical country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

FIGURE D.2: Counterfactual evidence: dynamics of estimates for border coef cients across within-country regions

(A) Median values of the border estimates



(B) Mean values of the border estimates

Notes: Figure D.2 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 rst in the legend. Thus, positive values for the BEFL-BEWL border coef cient indicate that prices are higher in Flanders than in Wallonia.

2.E Additional results on border analysis using cross-border shopping information (Section 2.4.3)

2.E.1 Statistics on cross-border purchases

TABLE E.1.1: Cross-border shopping: detailed descriptive statistics for BE-NL

	All eans						National brands						Private labels					
	Eans used for estim.			CB shopping			Eans used for estim.			CB shopping			Eans used for estim.			CB shopping		
	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.
2005q1	680	70	115	187	1	2	393	71	115	101	1	2	287	69	115	86	2	2
2005q2	732	71	121	202	1	2	445	73	124	112	2	3	287	69	118	90	1	2
2005q3	734	72	125	212	2	3	429	73	127	124	2	3	305	70	122	88	1	2
2005q4	836	75	125	215	2	2	481	77	128	136	2	3	355	73	122	79	2	2
2006q1	900	78	128	268	1	2	522	79	129	163	1	2	378	76	127	105	1	2
2006q2	872	79	133	241	1	2	511	81	134	142	1	2	361	76	132	99	1	2
2006q3	871	78	130	275	1	2	500	80	134	165	1	2	371	75	125	110	1	2
2006q4	872	79	130	272	2	2	497	81	132	144	2	3	375	77	128	128	1	2
2007q1	896	80	129	300	1	2	512	79	127	150	1	2	384	80	132	150	2	2
2007q2	992	80	131	318	1	2	564	82	133	165	1	2	428	77	128	153	2	2
2007q3	988	78	130	365	2	2	534	80	133	184	1	2	454	75	126	181	2	2
2007q4	1,038	82	134	391	2	2	580	83	135	225	2	2	458	80	132	166	2	2
2008q1	1,031	86	142	456	2	3	581	86	139	272	2	3	450	87	145	184	2	3
2008q2	1,064	81	136	457	2	3	589	82	136	260	2	3	475	81	136	197	2	3
2008q3	1,051	80	134	445	2	3	562	80	134	259	2	3	489	81	135	186	2	2
2008q4	1,021	86	140	443	2	3	561	87	140	271	2	3	460	85	141	172	2	3
Total	927	79	131	327	2	2	523	80	132	186	2	2	403	78	130	141	2	2

Notes: The table reports descriptive statistics on cross-border purchases for all quarters and compares them with analogous statistics from the full sample. Statistics for all goods are presented in the columns "EANs used for estim." and statistics for those for which we have observed cross-border shopping in the columns "CB shopping". The column denoted by "N" reports the number of goods considered for statistics, "Hhs" reports average number of household buying (the mean value of the number of households buying the goods over all goods), "Purch" reports average number of purchases observed per good (the mean value of the number of purchases per good over all goods) in the corresponding quarter.

TABLE E.1.2: Cross-border shopping: detailed descriptive statistics for GE-NL

	All eans						National brands						Private labels					
	Eans used for estim.			CB shopping			Eans used for estim.			CB shopping			Eans used for estim.			CB shopping		
	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.
2005q1	354	68	109	107	1	2	129	71	116	38	1	2	225	66	105	69	1	2
2005q2	394	68	112	118	1	2	133	73	128	30	1	2	261	66	104	88	1	2
2005q3	413	67	111	101	1	2	145	68	119	30	2	2	268	66	106	71	1	2
2005q4	448	73	117	136	1	2	155	78	128	45	1	2	293	71	111	91	1	2
2006q1	504	79	123	169	2	2	167	79	120	55	1	2	337	79	125	114	2	3
2006q2	543	77	119	172	2	3	186	75	109	52	2	2	357	78	124	120	2	3
2006q3	544	78	120	174	2	2	182	72	106	47	2	2	362	81	127	127	2	2
2006q4	542	87	134	189	2	3	180	88	130	57	2	2	362	87	135	132	2	3
2007q1	536	84	130	156	2	2	180	81	120	39	1	2	356	86	135	117	2	2
2007q2	601	82	127	166	1	2	202	79	117	44	1	2	399	83	132	122	2	2
2007q3	553	84	132	194	2	2	180	76	114	52	1	2	373	87	140	142	2	2
2007q4	579	88	137	191	2	2	210	84	125	52	2	2	369	89	145	139	2	2
2008q1	563	87	135	169	2	2	188	87	129	45	2	2	375	87	138	124	2	2
2008q2	561	83	130	167	2	2	197	77	116	57	1	2	364	86	138	110	2	3
2008q3	548	81	127	139	2	2	196	78	117	43	1	2	352	83	133	96	2	2
2008q4	524	89	139	148	2	2	180	89	132	44	2	2	344	89	143	104	2	2
Total	522	80	126	159	2	2	179	79	120	46	1	2	344	81	129	113	2	2

Notes: The table reports descriptive statistics on cross-border purchases for all quarters and compares them with analogous statistics from the full sample. Statistics for all goods are presented in the columns "EANs used for estim." and statistics for those for which we have observed cross-border shopping in the columns "CB shopping". The column denoted by "N" reports the number of goods considered for statistics, "Hhs" reports average number of household buying (the mean value of the number of households buying the goods over all goods), "Purch" reports average number of purchases observed per good (the mean value of the number of purchases per good over all goods) in the corresponding quarter.

TABLE E.1.3: Cross-border shopping: detailed descriptive statistics for BE-GE

	All eans						National brands						Private labels					
	Eans used for estim.			CB shopping			Eans used for estim.			CB shopping			Eans used for estim.			CB shopping		
	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.	N	Hhs	Purch.
2005q1	68	45	73	4	1	2	14	38	59	1	1	1	54	46	77	3	1	2
2005q2	86	47	79	6	1	1	15	54	89	2	1	2	71	46	77	4	1	1
2005q3	77	43	71	8	1	1	18	33	50	3	1	2	59	46	78	5	1	1
2005q4	104	47	76	8	1	2	22	40	61	6	1	1	82	48	80	2	1	3
2006q1	110	49	82	8	1	2	23	41	61	3	1	1	87	52	87	5	1	2
2006q2	103	52	86	11	1	2	21	51	83	3	1	2	82	52	86	8	1	2
2006q3	98	51	82	11	1	2	16	36	56	1	1	4	82	54	87	10	1	1
2006q4	114	51	80	15	1	2	26	36	51	6	1	2	88	55	88	9	1	2
2007q1	137	53	85	12	1	2	27	47	70	3	1	2	110	55	89	9	1	2
2007q2	128	52	83	17	1	1	28	40	59	9	1	2	100	56	90	8	1	1
2007q3	128	53	84	24	1	1	30	38	56	10	1	2	98	57	93	14	1	1
2007q4	146	52	80	20	1	2	41	42	59	10	1	2	105	56	88	10	1	1
2008q1	131	54	82	10	1	2	32	46	66	3	1	3	99	56	88	7	1	1
2008q2	147	53	84	27	1	1	38	46	65	14	1	2	109	56	90	13	1	1
2008q3	149	50	76	19	1	2	39	38	52	13	1	2	110	54	85	6	1	2
2008q4	159	52	81	23	1	2	44	45	66	14	1	2	115	54	86	9	1	2
Total	124	51	81	15	1	2	29	42	62	7	1	2	95	53	86	8	1	1

Notes: The table reports descriptive statistics on cross-border purchases for all quarters and compares them with analogous statistics from the full sample. Statistics for all goods are presented in the columns "EANs used for estim." and statistics for those for which we have observed cross-border shopping in the columns "CB shopping". The column denoted by "N" reports the number of goods considered for statistics, "Hhs" reports average number of household buying (the mean value of the number of households buying the goods over all goods), "Purch" reports average number of purchases observed per good (the mean value of the number of purchases per good over all goods) in the corresponding quarter.

2.E.2 Border estimates (incl. cross-border shopping observations), grouped by signi cance and the occurrence of cross border shopping

The following tables present summary statistics border estimates (including cross-border purchases) of goods, grouped by signi cance level of the border coef cient and the occurrence of cross-border purchases.

TABLE E.2.1: Border estimates (incl. cross-border shopping observations), grouped by signi cance and cross border shopping, BE-NL

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	680	77	142	0.121	0.097	0.280	379	0.130	0.105	0.271	45	0.007	0.008	0.051	114	0.004	0.016	0.110
2005q2	732	75	144	0.088	0.049	0.269	406	0.142	0.112	0.365	58	0.002	0.011	0.069	124	0.000	0.011	0.072
2005q3	734	73	141	0.104	0.087	0.289	394	0.138	0.095	0.277	71	0.003	0.011	0.052	128	0.003	0.007	0.072
2005q4	836	77	152	0.099	0.070	0.340	488	0.112	0.072	0.283	63	0.005	0.017	0.059	133	0.000	0.004	0.063
2006q1	900	78	198	0.104	0.089	0.233	504	0.118	0.095	0.280	70	0.008	0.011	0.054	128	0.008	0.014	0.068
2006q2	872	79	182	0.102	0.088	0.219	506	0.127	0.101	0.294	59	0.009	0.023	0.062	125	0.003	0.005	0.088
2006q3	871	81	227	0.134	0.119	0.283	476	0.114	0.085	0.270	48	0.013	0.019	0.051	120	0.016	0.033	0.145
2006q4	872	82	233	0.122	0.107	0.260	478	0.126	0.091	0.253	39	0.023	0.039	0.064	122	0.011	0.017	0.072
2007q1	896	81	232	0.112	0.096	0.222	492	0.147	0.115	0.298	68	0.013	0.019	0.064	104	0.008	0.020	0.083
2007q2	992	79	265	0.136	0.112	0.235	515	0.161	0.138	0.280	53	0.009	0.012	0.058	159	0.009	0.016	0.073
2007q3	988	82	302	0.112	0.100	0.226	505	0.129	0.100	0.296	63	0.020	0.024	0.068	118	0.006	0.012	0.065
2007q4	1,038	81	308	0.139	0.128	0.281	535	0.103	0.086	0.296	83	0.004	0.008	0.063	112	0.014	0.024	0.076
2008q1	1,031	81	362	0.159	0.139	0.246	474	0.152	0.129	0.305	94	0.016	0.020	0.059	101	0.014	0.038	0.091
2008q2	1,064	80	358	0.139	0.141	0.233	496	0.148	0.099	0.287	99	0.009	0.019	0.064	111	0.009	0.021	0.078
2008q3	1,051	80	359	0.127	0.117	0.225	486	0.139	0.115	0.295	86	0.010	0.022	0.063	120	0.008	0.008	0.088
2008q4	1,021	79	336	0.137	0.129	0.248	474	0.103	0.062	0.366	107	0.010	0.017	0.060	104	0.005	0.015	0.080
Total	911	79	246	0.121	0.104	0.256	476	0.131	0.100	0.295	69	0.010	0.017	0.060	120	0.007	0.016	0.083

Notes: The table presents summary statistics on border estimates (including cross-border purchases) of goods for BE-NL. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by N contain the number of goods included in the respective regression/group. The column $Sig(\%)$ provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.2.2: Border estimates (incl. cross-border shopping observations) - absolute values, grouped by signi cance and cross border shopping, BE-NL

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	680	77	142	0.161	0.216	0.202	379	0.197	0.239	0.164	45	0.025	0.037	0.036	114	0.036	0.060	0.094
2005q2	732	75	144	0.171	0.212	0.171	406	0.215	0.263	0.277	58	0.019	0.043	0.054	124	0.039	0.055	0.048
2005q3	734	73	141	0.209	0.237	0.187	394	0.214	0.246	0.159	71	0.024	0.038	0.038	128	0.035	0.048	0.055
2005q4	836	77	152	0.154	0.216	0.272	488	0.212	0.244	0.161	63	0.037	0.044	0.043	133	0.034	0.046	0.043
2006q1	900	78	198	0.164	0.200	0.150	504	0.217	0.244	0.167	70	0.039	0.042	0.035	128	0.046	0.052	0.045
2006q2	872	79	182	0.159	0.190	0.139	506	0.217	0.248	0.187	59	0.032	0.047	0.047	125	0.046	0.062	0.062
2006q3	871	81	227	0.194	0.233	0.199	476	0.209	0.236	0.156	48	0.035	0.041	0.036	120	0.041	0.067	0.133
2006q4	872	82	233	0.176	0.218	0.177	478	0.194	0.227	0.143	39	0.046	0.055	0.051	122	0.044	0.055	0.048
2007q1	896	81	232	0.173	0.201	0.133	492	0.223	0.260	0.187	68	0.037	0.048	0.045	104	0.042	0.058	0.062
2007q2	992	79	265	0.184	0.215	0.147	515	0.220	0.255	0.180	53	0.025	0.041	0.043	159	0.044	0.054	0.052
2007q3	988	82	302	0.165	0.201	0.143	505	0.202	0.248	0.189	63	0.034	0.050	0.052	118	0.033	0.048	0.046
2007q4	1,038	81	308	0.188	0.227	0.210	535	0.204	0.240	0.192	83	0.025	0.043	0.046	112	0.048	0.060	0.051
2008q1	1,031	81	362	0.197	0.229	0.166	474	0.224	0.260	0.205	94	0.034	0.044	0.044	101	0.047	0.067	0.072
2008q2	1,064	80	358	0.183	0.215	0.167	496	0.206	0.243	0.181	99	0.026	0.044	0.050	111	0.040	0.057	0.057
2008q3	1,051	80	359	0.179	0.203	0.151	486	0.221	0.252	0.191	86	0.033	0.046	0.048	120	0.041	0.060	0.065
2008q4	1,021	79	336	0.194	0.223	0.169	474	0.203	0.253	0.272	107	0.030	0.044	0.045	104	0.045	0.058	0.056
Total	911	79	246	0.178	0.215	0.174	476	0.211	0.247	0.188	69	0.031	0.044	0.045	120	0.041	0.057	0.062

Notes: The table presents summary statistics on border estimates (including cross-border purchases) of goods for BE-NL. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by *N* contain the number of goods included in the respective regression/group. The column *Sig*(%) provides information about the proportion of signi cant coef cients. *Median*/*Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.2.3: Border estimates (incl. cross-border shopping observations), grouped by signi cance and cross border shopping, GE-NL

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	354	77	80	-0.062	-0.035	0.193	192	-0.117	-0.055	0.263	27	0.000	0.005	0.073	55	0.005	-0.001	0.072
2005q2	394	76	93	-0.072	-0.031	0.185	207	-0.103	-0.056	0.255	25	0.000	-0.004	0.043	69	-0.000	-0.009	0.092
2005q3	413	74	73	-0.077	-0.040	0.186	232	-0.112	-0.091	0.266	28	0.000	-0.001	0.063	80	-0.001	-0.011	0.085
2005q4	448	72	98	-0.069	-0.031	0.212	226	-0.139	-0.100	0.314	38	0.000	-0.011	0.064	86	0.000	-0.010	0.080
2006q1	504	78	131	-0.100	-0.039	0.279	263	-0.105	-0.097	0.254	38	-0.009	-0.021	0.058	72	-0.007	-0.021	0.082
2006q2	543	78	127	-0.089	-0.060	0.210	298	-0.107	-0.060	0.278	45	0.003	0.005	0.058	73	-0.002	-0.004	0.087
2006q3	544	76	129	-0.079	-0.026	0.223	285	-0.109	-0.084	0.275	45	-0.002	-0.007	0.047	85	-0.012	-0.011	0.067
2006q4	542	78	142	-0.071	-0.048	0.197	280	-0.122	-0.071	0.291	47	0.002	-0.000	0.059	73	-0.001	-0.006	0.085
2007q1	536	77	123	-0.067	-0.040	0.206	292	-0.108	-0.066	0.278	33	-0.002	-0.009	0.062	88	0.006	-0.003	0.103
2007q2	601	77	127	-0.085	-0.068	0.194	334	-0.094	-0.054	0.286	39	0.002	0.012	0.074	101	-0.008	-0.016	0.083
2007q3	553	75	145	-0.088	-0.053	0.185	269	-0.088	-0.047	0.304	49	0.002	-0.004	0.055	90	-0.018	-0.033	0.102
2007q4	579	75	141	-0.094	-0.084	0.215	294	-0.072	-0.035	0.337	50	-0.003	-0.015	0.050	94	0.001	-0.008	0.101
2008q1	563	72	126	-0.087	-0.064	0.241	280	-0.075	-0.039	0.266	43	-0.004	-0.014	0.071	114	-0.001	-0.007	0.095
2008q2	561	75	121	-0.069	-0.052	0.181	300	-0.075	-0.061	0.267	46	-0.022	-0.044	0.085	94	-0.004	-0.010	0.076
2008q3	548	75	105	-0.061	-0.053	0.177	307	-0.076	-0.047	0.271	34	0.001	0.002	0.057	102	-0.001	-0.005	0.084
2008q4	524	77	112	-0.116	-0.131	0.224	291	-0.089	-0.065	0.277	36	0.004	-0.002	0.060	85	-0.004	-0.016	0.084
Total	513	76	117	-0.080	-0.053	0.207	272	-0.099	-0.064	0.280	39	-0.002	-0.007	0.061	85	-0.003	-0.011	0.086

Notes: The table presents summary statistics on border estimates (including cross-border purchases) of goods for GE-NL. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by N contain the number of goods included in the respective regression/group. The column $Sig(\%)$ provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.2.4: Border estimates (incl. cross-border shopping observations) - absolute values, grouped by signi cance and cross border shopping, GE-NL

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	354	77	80	0.141	0.165	0.105	192	0.189	0.225	0.147	27	0.020	0.044	0.057	55	0.034	0.050	0.051
2005q2	394	76	93	0.129	0.154	0.107	207	0.179	0.214	0.148	25	0.013	0.026	0.034	69	0.036	0.059	0.071
2005q3	413	74	73	0.143	0.161	0.099	232	0.176	0.225	0.168	28	0.020	0.043	0.046	80	0.042	0.058	0.063
2005q4	448	72	98	0.124	0.172	0.126	226	0.198	0.246	0.218	38	0.024	0.042	0.049	86	0.035	0.050	0.063
2006q1	504	78	131	0.136	0.182	0.214	263	0.184	0.216	0.165	38	0.039	0.046	0.039	72	0.050	0.061	0.058
2006q2	543	78	127	0.142	0.172	0.134	298	0.188	0.225	0.174	45	0.020	0.035	0.046	73	0.039	0.062	0.061
2006q3	544	76	129	0.146	0.176	0.138	285	0.189	0.231	0.171	45	0.022	0.032	0.036	85	0.034	0.049	0.046
2006q4	542	78	142	0.134	0.158	0.127	280	0.194	0.230	0.191	47	0.029	0.039	0.043	73	0.041	0.059	0.061
2007q1	536	77	123	0.117	0.163	0.131	292	0.197	0.229	0.171	33	0.024	0.042	0.046	88	0.039	0.065	0.080
2007q2	601	77	127	0.137	0.163	0.125	334	0.187	0.230	0.177	39	0.035	0.049	0.057	101	0.036	0.057	0.062
2007q3	553	75	145	0.129	0.151	0.118	269	0.184	0.236	0.197	49	0.016	0.035	0.042	90	0.044	0.074	0.077
2007q4	579	75	141	0.142	0.179	0.145	294	0.203	0.240	0.239	50	0.017	0.034	0.039	94	0.033	0.061	0.081
2008q1	563	72	126	0.139	0.175	0.178	280	0.178	0.220	0.153	43	0.020	0.045	0.056	114	0.039	0.064	0.070
2008q2	561	75	121	0.119	0.146	0.119	300	0.179	0.219	0.165	46	0.032	0.063	0.072	94	0.034	0.054	0.053
2008q3	548	75	105	0.117	0.148	0.110	307	0.173	0.214	0.172	34	0.029	0.039	0.040	102	0.040	0.058	0.061
2008q4	524	77	112	0.135	0.182	0.184	291	0.180	0.225	0.174	36	0.038	0.045	0.038	85	0.033	0.058	0.063
Total	513	76	117	0.133	0.165	0.135	272	0.186	0.227	0.177	39	0.025	0.041	0.046	85	0.038	0.059	0.064

Notes: The table presents summary statistics on absolute border estimates (including cross-border purchases) of goods for GE-NL. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by *N* contain the number of goods included in the respective regression/group. The column *Sig*(%) provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.2.5: Border estimates (incl. cross-border shopping observations), grouped by signi cance and cross border shopping, BE-GE

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	68	81	3	0.169	0.190	0.104	52	-0.032	0.009	0.288	1	0.043	0.043	.	12	-0.000	-0.011	0.084
2005q2	86	78	4	0.056	0.073	0.211	63	0.042	0.046	0.315	2	-0.085	-0.085	0.148	17	0.000	0.011	0.099
2005q3	77	74	3	-0.049	-0.039	0.115	54	0.103	0.089	0.302	5	-0.045	0.019	0.118	15	0.001	0.023	0.080
2005q4	104	77	6	0.207	0.184	0.174	74	0.063	0.053	0.301	2	-0.196	-0.196	0.307	22	0.000	-0.032	0.088
2006q1	110	75	6	0.159	0.218	0.242	77	0.038	0.064	0.302	2	-0.011	-0.011	0.074	25	-0.020	-0.020	0.077
2006q2	103	75	5	0.140	0.124	0.155	72	0.010	0.051	0.317	6	0.004	0.027	0.078	20	-0.000	0.020	0.166
2006q3	98	82	9	0.106	0.029	0.221	71	-0.061	0.017	0.330	2	0.034	0.034	0.058	16	-0.003	-0.002	0.062
2006q4	114	70	9	0.116	0.066	0.148	71	0.034	0.045	0.354	6	0.001	0.011	0.115	28	0.006	0.027	0.122
2007q1	137	74	8	-0.047	0.018	0.238	93	0.061	0.052	0.302	4	0.068	0.052	0.120	32	-0.001	0.008	0.058
2007q2	128	78	12	0.075	-0.020	0.293	88	-0.023	0.009	0.268	5	0.070	0.051	0.071	23	0.005	0.044	0.102
2007q3	128	73	16	0.092	0.112	0.265	78	0.002	0.004	0.297	8	0.028	0.100	0.243	26	-0.001	0.023	0.129
2007q4	146	75	11	-0.115	0.032	0.286	99	0.059	0.029	0.260	9	0.026	0.022	0.096	27	-0.012	0.003	0.072
2008q1	131	73	8	-0.040	0.004	0.145	87	0.051	0.020	0.278	2	-0.021	-0.021	0.091	34	0.019	0.024	0.093
2008q2	147	82	17	0.053	0.040	0.207	104	0.047	0.052	0.255	10	0.025	0.022	0.077	16	0.000	0.003	0.076
2008q3	149	72	10	0.060	0.010	0.339	97	0.060	0.041	0.250	9	-0.003	0.009	0.125	33	-0.001	-0.017	0.054
2008q4	159	73	11	0.106	0.110	0.258	105	0.039	0.040	0.241	12	0.036	0.068	0.114	31	0.015	0.011	0.090
Total	118	76	9	0.068	0.072	0.213	80	0.031	0.039	0.291	5	-0.002	0.009	0.122	24	0.000	0.007	0.091

Notes: The table presents summary statistics on border estimates (including cross-border purchases) of goods for BE-GE. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by *N* contain the number of goods included in the respective regression/group. The column *Sig*(%) provides information about the proportion of signi cant coef cients. *Median*/*Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.2.6: Border estimates (incl. cross-border shopping observations) - absolute values, grouped by signi cance and cross border shopping, BE-GE

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	68	81	3	0.169	0.190	0.104	52	0.181	0.231	0.169	1	0.043	0.043	.	12	0.022	0.050	0.067
2005q2	86	78	4	0.158	0.174	0.104	63	0.149	0.237	0.210	2	0.104	0.104	0.120	17	0.059	0.067	0.072
2005q3	77	74	3	0.080	0.093	0.051	54	0.185	0.253	0.184	5	0.083	0.093	0.058	15	0.041	0.056	0.060
2005q4	104	77	6	0.207	0.218	0.119	74	0.182	0.241	0.187	2	0.217	0.217	0.277	22	0.031	0.060	0.072
2006q1	110	75	6	0.159	0.243	0.212	77	0.180	0.239	0.194	2	0.052	0.052	0.016	25	0.048	0.061	0.051
2006q2	103	75	5	0.140	0.160	0.106	72	0.187	0.254	0.193	6	0.008	0.040	0.071	20	0.031	0.081	0.145
2006q3	98	82	9	0.186	0.192	0.093	71	0.158	0.248	0.216	2	0.041	0.041	0.048	16	0.031	0.041	0.045
2006q4	114	70	9	0.116	0.131	0.086	71	0.215	0.267	0.235	6	0.089	0.096	0.048	28	0.044	0.083	0.093
2007q1	137	74	8	0.151	0.182	0.139	93	0.150	0.224	0.208	4	0.116	0.101	0.066	32	0.039	0.043	0.039
2007q2	128	78	12	0.168	0.219	0.184	88	0.162	0.210	0.166	5	0.070	0.076	0.030	23	0.021	0.054	0.097
2007q3	128	73	16	0.196	0.230	0.165	78	0.191	0.236	0.178	8	0.063	0.128	0.227	26	0.035	0.066	0.112
2007q4	146	75	11	0.226	0.242	0.138	99	0.188	0.218	0.144	9	0.029	0.065	0.071	27	0.028	0.048	0.053
2008q1	131	73	8	0.115	0.116	0.074	87	0.181	0.227	0.161	2	0.065	0.065	0.030	34	0.040	0.063	0.071
2008q2	147	82	17	0.196	0.179	0.103	104	0.158	0.205	0.160	10	0.043	0.057	0.053	16	0.031	0.048	0.058
2008q3	149	72	10	0.274	0.276	0.173	97	0.163	0.203	0.150	9	0.043	0.090	0.081	33	0.018	0.033	0.045
2008q4	159	73	11	0.207	0.220	0.162	105	0.147	0.194	0.148	12	0.083	0.101	0.083	31	0.060	0.065	0.062
Total	118	76	9	0.172	0.192	0.126	80	0.174	0.230	0.181	5	0.072	0.086	0.085	24	0.036	0.057	0.071

Notes: The table presents summary statistics on absolute border estimates (including cross-border purchases) of goods for BE-GE. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by N contain the number of goods included in the respective regression/group. The column $Sig(\%)$ provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

2.E.3 Border estimates (excl. cross-border shopping observations), grouped by signi cance and the occurrence of cross-border shopping

TABLE E.3.7: Border estimates (excl. cross-border shopping observations), grouped by signi cance and cross border shopping, BE-NL

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	669	78	140	0.123	0.109	0.272	379	0.130	0.105	0.271	36	0.002	-0.004	0.042	114	0.004	0.016	0.110
2005q2	720	75	134	0.111	0.052	0.252	406	0.142	0.112	0.365	56	0.000	-0.004	0.067	124	0.000	0.011	0.072
2005q3	723	74	138	0.106	0.081	0.275	394	0.138	0.095	0.277	63	-0.000	0.006	0.052	128	0.003	0.007	0.072
2005q4	825	77	149	0.099	0.061	0.389	488	0.112	0.072	0.283	55	0.000	0.014	0.062	133	0.000	0.004	0.063
2006q1	885	79	196	0.099	0.085	0.230	504	0.118	0.095	0.280	57	0.014	0.011	0.055	128	0.008	0.014	0.068
2006q2	864	80	186	0.102	0.087	0.230	505	0.128	0.101	0.294	47	0.001	0.012	0.065	126	0.002	0.003	0.090
2006q3	860	81	221	0.141	0.122	0.291	476	0.114	0.085	0.270	43	0.010	0.017	0.054	120	0.016	0.033	0.145
2006q4	862	82	229	0.129	0.113	0.263	478	0.126	0.091	0.253	33	0.013	0.036	0.071	122	0.011	0.017	0.072
2007q1	882	82	229	0.114	0.096	0.229	492	0.147	0.115	0.298	57	0.015	0.024	0.057	104	0.008	0.020	0.083
2007q2	976	79	257	0.146	0.113	0.244	515	0.161	0.138	0.280	45	0.009	0.015	0.049	159	0.009	0.016	0.073
2007q3	971	82	290	0.120	0.106	0.235	505	0.129	0.100	0.296	58	0.021	0.026	0.055	118	0.006	0.012	0.065
2007q4	1,019	82	305	0.155	0.129	0.282	535	0.103	0.086	0.296	67	0.002	0.007	0.049	112	0.014	0.024	0.076
2008q1	998	81	339	0.174	0.142	0.261	474	0.152	0.129	0.305	84	0.016	0.029	0.061	101	0.014	0.038	0.091
2008q2	1,040	81	346	0.146	0.144	0.240	496	0.148	0.099	0.287	87	0.010	0.026	0.084	111	0.009	0.021	0.078
2008q3	1,022	81	341	0.136	0.124	0.241	486	0.139	0.115	0.295	75	0.007	0.024	0.061	120	0.009	0.009	0.089
2008q4	993	81	331	0.139	0.133	0.254	473	0.103	0.062	0.367	84	0.007	0.014	0.060	105	0.005	0.016	0.081
Total	894	80	239	0.127	0.106	0.262	475	0.131	0.100	0.295	59	0.008	0.016	0.059	120	0.007	0.016	0.083

Notes: The table presents summary statistics on border estimates (excluding cross-border purchases) of goods for GE-NL. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by N contain the number of goods included in the respective regression/group. The column $Sig(\%)$ provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.3.8: Border estimates (excl. cross-border shopping observations) - absolute values, grouped by signi cance and cross border shopping, BE-NL

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	669	78	140	0.175	0.218	0.195	379	0.197	0.239	0.164	36	0.013	0.028	0.031	114	0.036	0.060	0.094
2005q2	720	75	134	0.189	0.217	0.137	406	0.215	0.263	0.277	56	0.023	0.044	0.050	124	0.039	0.055	0.048
2005q3	723	74	138	0.208	0.237	0.159	394	0.214	0.246	0.159	63	0.030	0.037	0.037	128	0.035	0.048	0.055
2005q4	825	77	149	0.180	0.225	0.323	488	0.212	0.244	0.161	55	0.034	0.044	0.046	133	0.034	0.046	0.043
2006q1	885	79	196	0.163	0.198	0.145	504	0.217	0.244	0.167	57	0.041	0.043	0.035	128	0.046	0.052	0.045
2006q2	864	80	186	0.166	0.197	0.147	505	0.217	0.248	0.187	47	0.023	0.040	0.052	126	0.047	0.063	0.064
2006q3	860	81	221	0.196	0.242	0.201	476	0.209	0.236	0.156	43	0.037	0.040	0.040	120	0.041	0.067	0.133
2006q4	862	82	229	0.186	0.225	0.178	478	0.194	0.227	0.143	33	0.035	0.054	0.059	122	0.044	0.055	0.048
2007q1	882	82	229	0.178	0.209	0.134	492	0.223	0.260	0.187	57	0.033	0.044	0.044	104	0.042	0.058	0.062
2007q2	976	79	257	0.195	0.224	0.150	515	0.220	0.255	0.180	45	0.023	0.035	0.037	159	0.044	0.054	0.052
2007q3	971	82	290	0.177	0.212	0.147	505	0.202	0.248	0.189	58	0.036	0.045	0.040	118	0.033	0.048	0.046
2007q4	1,019	82	305	0.196	0.233	0.204	535	0.204	0.240	0.192	67	0.022	0.035	0.035	112	0.048	0.060	0.051
2008q1	998	81	339	0.210	0.241	0.174	474	0.224	0.260	0.205	84	0.029	0.044	0.051	101	0.047	0.067	0.072
2008q2	1,040	81	346	0.196	0.225	0.167	496	0.206	0.243	0.181	87	0.023	0.047	0.074	111	0.040	0.057	0.057
2008q3	1,022	81	341	0.186	0.219	0.160	486	0.221	0.252	0.192	75	0.033	0.045	0.048	120	0.041	0.061	0.065
2008q4	993	81	331	0.201	0.231	0.170	473	0.203	0.253	0.272	84	0.026	0.041	0.046	105	0.046	0.059	0.057
Total	894	80	239	0.188	0.222	0.174	475	0.211	0.247	0.188	59	0.029	0.042	0.045	120	0.041	0.057	0.062

Notes: The table presents summary statistics on absolute border estimates (excluding cross-border purchases) of goods for GE-NL. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by N contain the number of goods included in the respective regression/group. The column $Sig(\%)$ provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.3.9: Border estimates (excl. cross-border shopping observations), grouped by signi cance and cross border shopping, GE-NL

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	348	76	74	-0.067	-0.042	0.216	192	-0.117	-0.055	0.263	27	0.000	0.007	0.065	55	0.005	-0.001	0.072
2005q2	390	77	93	-0.080	-0.037	0.199	207	-0.103	-0.056	0.255	21	0.001	-0.003	0.056	69	-0.000	-0.009	0.092
2005q3	407	74	72	-0.078	-0.039	0.197	231	-0.113	-0.091	0.266	23	0.000	-0.013	0.065	81	-0.002	-0.012	0.085
2005q4	443	73	97	-0.074	-0.042	0.227	227	-0.140	-0.101	0.314	34	0.000	-0.003	0.067	85	0.000	-0.005	0.064
2006q1	493	80	129	-0.100	-0.040	0.293	264	-0.104	-0.097	0.253	29	-0.019	-0.033	0.061	71	-0.007	-0.021	0.083
2006q2	530	79	121	-0.095	-0.061	0.222	298	-0.107	-0.060	0.278	38	0.000	-0.006	0.102	73	-0.002	-0.004	0.087
2006q3	527	78	125	-0.097	-0.038	0.226	285	-0.109	-0.084	0.275	32	-0.001	-0.001	0.054	85	-0.012	-0.011	0.067
2006q4	532	79	140	-0.084	-0.055	0.210	280	-0.122	-0.071	0.291	39	0.003	0.003	0.066	73	-0.001	-0.006	0.085
2007q1	529	79	124	-0.071	-0.045	0.225	292	-0.108	-0.066	0.278	25	-0.002	-0.018	0.057	88	0.006	-0.003	0.103
2007q2	591	78	125	-0.084	-0.072	0.213	336	-0.095	-0.055	0.285	31	0.002	0.009	0.081	99	-0.006	-0.009	0.083
2007q3	541	76	143	-0.095	-0.061	0.201	269	-0.088	-0.047	0.304	39	0.002	-0.012	0.060	90	-0.018	-0.033	0.102
2007q4	570	75	135	-0.114	-0.096	0.229	292	-0.072	-0.035	0.338	47	-0.000	-0.013	0.057	96	0.001	-0.008	0.100
2008q1	554	74	126	-0.094	-0.072	0.250	282	-0.076	-0.041	0.266	34	0.000	0.001	0.065	112	-0.000	-0.001	0.087
2008q2	551	77	124	-0.072	-0.061	0.197	300	-0.075	-0.061	0.267	33	-0.020	-0.036	0.063	94	-0.004	-0.010	0.076
2008q3	541	76	105	-0.065	-0.057	0.190	307	-0.076	-0.047	0.271	27	-0.000	-0.016	0.055	102	-0.001	-0.005	0.084
2008q4	517	77	108	-0.129	-0.141	0.235	291	-0.089	-0.065	0.277	33	0.009	-0.010	0.093	85	-0.004	-0.016	0.084
Total	504	77	115	-0.087	-0.060	0.221	272	-0.100	-0.065	0.280	32	-0.002	-0.009	0.067	85	-0.003	-0.010	0.085

Notes: The table presents summary statistics on border estimates (excluding cross-border purchases) of goods for BE-GE. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by N contain the number of goods included in the respective regression/group. The column $Sig(\%)$ provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.3.10: Border estimates (excl. cross-border shopping observations) - absolute values, grouped by signi cance and cross border shopping, GE-NL

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	348	76	74	0.169	0.188	0.112	192	0.189	0.225	0.147	27	0.031	0.042	0.049	55	0.034	0.050	0.051
2005q2	390	77	93	0.141	0.168	0.112	207	0.179	0.214	0.148	21	0.013	0.031	0.045	69	0.036	0.059	0.071
2005q3	407	74	72	0.150	0.171	0.103	231	0.177	0.225	0.168	23	0.021	0.041	0.051	81	0.042	0.058	0.062
2005q4	443	73	97	0.154	0.186	0.134	227	0.198	0.247	0.218	34	0.018	0.043	0.051	85	0.034	0.046	0.045
2006q1	493	80	129	0.156	0.195	0.222	264	0.183	0.215	0.165	29	0.048	0.054	0.042	71	0.049	0.061	0.059
2006q2	530	79	121	0.151	0.186	0.135	298	0.188	0.225	0.174	38	0.017	0.052	0.088	73	0.039	0.062	0.061
2006q3	527	78	125	0.160	0.188	0.130	285	0.189	0.231	0.171	32	0.026	0.036	0.040	85	0.034	0.049	0.046
2006q4	532	79	140	0.146	0.171	0.133	280	0.194	0.230	0.191	39	0.029	0.043	0.050	73	0.041	0.059	0.061
2007q1	529	79	124	0.130	0.174	0.150	292	0.197	0.229	0.171	25	0.025	0.039	0.045	88	0.039	0.065	0.080
2007q2	591	78	125	0.139	0.174	0.143	336	0.186	0.229	0.177	31	0.041	0.054	0.060	99	0.033	0.056	0.061
2007q3	541	76	143	0.142	0.165	0.129	269	0.184	0.236	0.197	39	0.014	0.035	0.049	90	0.044	0.074	0.077
2007q4	570	75	135	0.161	0.197	0.150	292	0.203	0.241	0.239	47	0.021	0.037	0.044	96	0.034	0.061	0.080
2008q1	554	74	126	0.151	0.190	0.178	282	0.178	0.221	0.153	34	0.018	0.037	0.054	112	0.039	0.060	0.063
2008q2	551	77	124	0.134	0.161	0.128	300	0.179	0.219	0.165	33	0.026	0.049	0.053	94	0.034	0.054	0.053
2008q3	541	76	105	0.125	0.159	0.119	307	0.173	0.214	0.172	27	0.019	0.040	0.040	102	0.040	0.058	0.061
2008q4	517	77	108	0.146	0.196	0.191	291	0.180	0.225	0.174	33	0.039	0.054	0.076	85	0.033	0.058	0.063
Total	504	77	115	0.147	0.179	0.142	272	0.186	0.227	0.177	32	0.025	0.043	0.052	85	0.038	0.058	0.062

Notes: The table presents summary statistics on absolute border estimates (excluding cross-border purchases) of goods for BE-GE. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by N contain the number of goods included in the respective regression/group. The column $Sig(\%)$ provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.3.11: Border estimates (excl. cross-border shopping observations), grouped by signi cance and cross border shopping, BE-GE

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	67	81	2	0.107	0.107	0.039	52	-0.032	0.009	0.288	1	0.037	0.037	.	12	-0.000	-0.011	0.084
2005q2	85	76	3	0.169	0.129	0.164	62	0.048	0.049	0.316	2	-0.055	-0.055	0.085	18	-0.002	-0.000	0.107
2005q3	75	76	3	-0.044	-0.043	0.110	54	0.103	0.089	0.302	3	-0.047	-0.054	0.028	15	0.001	0.023	0.080
2005q4	104	76	6	0.194	0.176	0.173	73	0.052	0.052	0.303	2	-0.202	-0.202	0.317	23	0.000	-0.026	0.092
2006q1	108	76	6	0.156	0.219	0.237	76	0.049	0.068	0.302	1	0.026	0.026	.	25	-0.020	-0.020	0.077
2006q2	103	76	6	0.065	0.097	0.155	72	0.010	0.051	0.317	5	-0.007	0.029	0.089	20	-0.000	0.020	0.166
2006q3	98	82	9	0.072	0.033	0.186	71	-0.061	0.017	0.330	2	0.033	0.033	0.059	16	-0.003	-0.002	0.062
2006q4	114	71	10	0.023	0.036	0.157	71	0.034	0.045	0.354	5	0.054	0.033	0.091	28	0.006	0.027	0.122
2007q1	136	75	9	-0.044	-0.002	0.222	93	0.061	0.052	0.302	2	0.069	0.069	0.120	32	-0.001	0.008	0.058
2007q2	126	79	12	0.075	-0.014	0.298	88	-0.023	0.009	0.268	3	0.104	0.060	0.085	23	0.005	0.044	0.102
2007q3	127	73	15	0.095	0.175	0.304	78	0.002	0.005	0.297	8	0.030	0.032	0.073	26	-0.001	0.023	0.129
2007q4	146	76	12	-0.087	0.020	0.280	99	0.059	0.029	0.260	8	0.021	0.020	0.117	27	-0.012	0.003	0.072
2008q1	130	72	7	-0.033	0.013	0.154	87	0.051	0.020	0.278	2	-0.009	-0.009	0.112	34	0.019	0.024	0.093
2008q2	146	82	15	0.010	0.025	0.231	104	0.047	0.052	0.255	11	0.010	0.025	0.089	16	0.000	0.004	0.075
2008q3	147	74	11	-0.056	-0.011	0.338	98	0.059	0.039	0.249	6	0.073	0.053	0.074	32	-0.001	-0.012	0.045
2008q4	158	76	14	0.153	0.137	0.249	106	0.027	0.038	0.241	8	0.028	0.040	0.124	30	0.016	0.017	0.085
Total	117	76	9	0.053	0.068	0.206	80	0.030	0.039	0.291	4	0.010	0.009	0.105	24	0.000	0.008	0.091

Notes: The table presents summary statistics on border estimates (excluding cross-border purchases) of goods for BE-NL. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by N contain the number of goods included in the respective regression/group. The column $Sig(\%)$ provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

TABLE E.3.12: Border estimates (excl. cross-border shopping observations) - absolute values, grouped by signi cance and cross border shopping, BE-GE

	Number of eans		Signi cant estimates								Not signi cant estimates							
			CB shopping				no CB shopping				CB shopping				no CB shopping			
	N	Sig.(%)	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std	N	Median	Mean	Std
2005q1	67	81	2	0.107	0.107	0.039	52	0.181	0.231	0.169	1	0.037	0.037	.	12	0.022	0.050	0.067
2005q2	85	76	3	0.169	0.163	0.109	62	0.148	0.238	0.212	2	0.060	0.060	0.078	18	0.062	0.074	0.076
2005q3	75	76	3	0.067	0.088	0.058	54	0.185	0.254	0.184	3	0.047	0.054	0.028	15	0.041	0.056	0.060
2005q4	104	76	6	0.194	0.210	0.119	73	0.183	0.242	0.188	2	0.224	0.224	0.286	23	0.038	0.062	0.071
2006q1	108	76	6	0.156	0.239	0.212	76	0.178	0.239	0.195	1	0.026	0.026	.	25	0.048	0.061	0.051
2006q2	103	76	6	0.092	0.127	0.126	72	0.187	0.254	0.193	5	0.046	0.063	0.064	20	0.031	0.081	0.145
2006q3	98	82	9	0.186	0.166	0.071	71	0.158	0.248	0.216	2	0.042	0.042	0.046	16	0.031	0.041	0.045
2006q4	114	71	10	0.112	0.132	0.083	71	0.215	0.267	0.235	5	0.072	0.079	0.044	28	0.044	0.083	0.093
2007q1	136	75	9	0.137	0.171	0.129	93	0.150	0.224	0.208	2	0.085	0.085	0.098	32	0.039	0.042	0.039
2007q2	126	79	12	0.177	0.225	0.184	88	0.162	0.210	0.166	3	0.104	0.085	0.042	23	0.021	0.054	0.097
2007q3	127	73	15	0.196	0.267	0.222	78	0.191	0.236	0.179	8	0.065	0.061	0.048	26	0.035	0.066	0.112
2007q4	146	76	12	0.222	0.232	0.142	99	0.188	0.218	0.144	8	0.037	0.077	0.086	27	0.028	0.048	0.054
2008q1	130	72	7	0.122	0.125	0.075	87	0.181	0.227	0.161	2	0.079	0.079	0.013	34	0.040	0.063	0.071
2008q2	146	82	15	0.205	0.192	0.121	104	0.158	0.205	0.160	11	0.044	0.064	0.065	16	0.031	0.047	0.057
2008q3	147	74	11	0.268	0.282	0.163	98	0.167	0.202	0.149	6	0.073	0.078	0.041	32	0.017	0.029	0.037
2008q4	158	76	14	0.194	0.226	0.165	106	0.147	0.193	0.148	8	0.050	0.088	0.091	30	0.048	0.061	0.060
Total	117	76	9	0.163	0.184	0.126	80	0.174	0.230	0.182	4	0.068	0.075	0.073	24	0.036	0.057	0.071

Notes: The table presents summary statistics on absolute border estimates (excluding cross-border purchases) of goods for BE-NL. Results are grouped by signi cance level of the border coef cient and by the occurrence of cross-border purchases. The rst column reports the time period..Columns denoted by N contain the number of goods included in the respective regression/group. The column $Sig(\%)$ provides information about the proportion of signi cant coef cients. *Median/Mean* refers to the median/mean estimate of the respective group, while *Std* reports their standard deviation.

2.E.4 Summary statistics of goods subject to cross-border shopping

TABLE E.4.1: Descriptive statistics on goods, Q4 2008.

Belgium - Germany										
	N	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Cross-Border										
Price	23	0.73	0.49	0.03	0.04	0.33	0.65	1.26	1.47	1.70
Purchases	23	89.91	62.48	39.00	42.00	51.00	66.00	104.00	240.00	290.00
Households	23	60.26	41.96	26.00	34.00	38.00	43.00	69.00	142.00	207.00
Abs. ln. price gap	23	0.17	0.12	0.02	0.02	0.06	0.18	0.26	0.34	0.38
No Cross-Border										
Price	136	1.00	3.12	0.03	0.09	0.25	0.47	0.78	1.86	32.65
Purchases	136	78.07	70.13	26.00	30.00	41.00	55.00	83.50	194.00	401.00
Households	136	50.29	34.97	21.00	23.00	31.00	40.50	56.00	126.00	213.00
Abs. ln. price gap	136	0.16	0.15	0.00	0.00	0.06	0.12	0.22	0.48	0.69
Total										
Price	159	0.96	2.89	0.03	0.09	0.26	0.52	0.86	1.70	32.65
Purchases	159	79.79	69.02	26.00	30.00	42.00	58.00	85.00	240.00	401.00
Households	159	51.74	36.09	21.00	23.00	31.00	41.00	57.00	142.00	213.00
Abs. ln. price gap	159	0.16	0.14	0.00	0.00	0.06	0.12	0.23	0.47	0.69
Observations	159									

Notes: Table E.4.1 provides summary statistics on goods for which cross-border shopping information is available (upper panel) and those for which no such information is available (medium panel). Summary statistics on all goods are provided in the lower panel. The variable "Price" corresponds to the average price of the good, the variable "Purchases" provides information on the number of purchases of the good, similarly the variable "Households" captures the number of households that a given good bought. "Abs. ln. price gap" denotes the percentage difference in the average price between the two countries.

TABLE E.4.2: Descriptive statistics on goods, Q4 2008.

Germany- the Netherlands										
	N	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Cross-Border										
Price	148	2.44	15.05	0.02	0.07	0.21	0.41	0.79	1.89	176.18
Purchases	148	205.13	262.05	28.00	37.00	59.50	113.00	224.00	624.00	2,080.00
Households	148	127.75	128.88	20.00	31.00	45.00	80.00	154.50	394.00	763.00
Abs. ln. price gap	148	0.15	0.16	0.00	0.00	0.06	0.12	0.19	0.39	1.17
No Cross-Border										
Price	376	2.36	20.18	0.03	0.09	0.25	0.49	0.78	2.10	339.91
Purchases	376	110.78	133.17	21.00	29.00	46.00	68.00	125.00	302.00	1,229.00
Households	376	73.70	70.88	20.00	24.00	34.00	50.00	85.50	196.00	566.00
Abs. ln. price gap	376	0.18	0.17	0.00	0.00	0.06	0.14	0.25	0.49	1.41
Total										
Price	524	2.38	18.85	0.02	0.09	0.24	0.49	0.78	2.06	339.91
Purchases	524	137.43	183.91	21.00	30.00	48.00	76.00	149.50	438.00	2,080.00
Households	524	88.97	94.15	20.00	25.00	37.00	56.50	101.00	260.00	763.00
Abs. ln. price gap	524	0.18	0.17	0.00	0.00	0.06	0.13	0.24	0.47	1.41
Observations	524									

Notes: Table E.4.2 provides summary statistics on goods for which cross-border shopping information is available (upper panel) and those for which no such information is available (medium panel). Summary statistics on all goods are provided in the lower panel. The variable "Price" corresponds to the average price of the good, the variable "Purchases" provides information on the number of purchases of the good, similarly the variable "Households" captures the number of households that a given good bought. "Abs. ln. price gap" denotes the percentage difference in the average price between the two countries.

TABLE E.4.3: Number of goods by category, Q4 2008.

Belgium - Germany		
	Cross-Border	No Cross-Border
AlcholfreeCO2	.	1.00
AlcholfreeNoCO2	1.00	5.00
Animalcare	1.00	2.00
Basicfood	3.00	6.00
Candy	1.00	11.00
Cereals	.	3.00
DairyWhite	.	16.00
DairyYellow	7.00	7.00
Delicasees	.	3.00
Fatoils	.	2.00
Frozenproducts	2.00	9.00
HotDrinks	.	5.00
HouseholdCleansers	.	2.00
Hygieneproducts	.	2.00
Liquor	1.00	1.00
Meat	1.00	22.00
PreservedFood	.	2.00
Readymade	.	5.00
Rest	.	4.00
Snacks	4.00	18.00
Spreads	.	5.00
Vegetables	2.00	5.00
Observations	159	

Notes: Table E.4.3 presents the number of goods used in the estimation by category.

TABLE E.4.4: Number of goods by category, Q4 2008.

Germany - the Netherlands		
	Cross-Border	No Cross-Border
AlcholfreeCO2	1.00	1.00
AlcholfreeNoCO2	7.00	14.00
Alcohol	.	1.00
Animalcare	3.00	17.00
Babyproducts	1.00	3.00
Basicfood	2.00	6.00
Bodycare	3.00	20.00
Candy	30.00	71.00
Cereals	.	5.00
DairyWhite	18.00	39.00
DairyYellow	4.00	12.00
Delicasees	4.00	12.00
Fatoils	1.00	1.00
Frische	.	1.00
Frozenproducts	12.00	27.00
HotDrinks	2.00	15.00
HouseholdCleansers	5.00	12.00
Hygieneproducts	6.00	4.00
Laundry	2.00	4.00
Liquor	1.00	.
Meat	22.00	25.00
MouthTooth	.	2.00
PreservedFood	3.00	8.00
Readymade	4.00	15.00
Rest	1.00	3.00
Snacks	10.00	34.00
Spreads	.	6.00
Vegetables	5.00	11.00
Wine	1.00	7.00
Observations	524	

Notes: Table E.4.4 presents the number of goods used in the estimation by category.

2.F Co-movement of prices in border- and non-border regions of different bandwidths (Section 4.4)

TABLE F.1: Co-movement of prices within and across countries: 40 km region width

	All observations											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.63	1.00			0.46	1.00			0.67	1.00		
p^*	0.23	0.06	1.00		0.06	0.07	1.00		0.09	0.11	1.00	
p_{nb}^*	0.30	0.13	0.85	1.00	0.08	0.07	0.57	1.00	0.10	0.09	0.58	1.00
	Signi cant border estimates, cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.52	1.00			0.46	1.00			0.68	1.00		
p^*	0.20	-0.08	1.00		0.07	0.07	1.00		0.09	0.10	1.00	
p_{nb}^*	0.28	0.00	0.83	1.00	0.08	0.07	0.57	1.00	0.09	0.09	0.58	1.00
	Signi cant border estimates, no cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.63	1.00			0.46	1.00			0.69	1.00		
p^*	0.23	0.06	1.00		0.07	0.07	1.00		0.09	0.13	1.00	
p_{nb}^*	0.30	0.13	0.85	1.00	0.08	0.07	0.56	1.00	0.12	0.10	0.62	1.00
	Not signi cant border estimates, cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.16	1.00			0.41	1.00			0.61	1.00		
p^*	0.38	-0.18	1.00		0.04	0.08	1.00		0.08	0.08	1.00	
p_{nb}^*	0.38	-0.14	0.81	1.00	0.05	0.10	0.53	1.00	0.07	0.03	0.49	1.00
	Not signi cant border estimates, no cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.16	1.00			0.40	1.00			0.55	1.00		
p^*	0.38	-0.18	1.00		0.05	0.09	1.00		0.07	0.09	1.00	
p_{nb}^*	0.38	-0.14	0.81	1.00	0.04	0.09	0.50	1.00	0.07	0.04	0.50	1.00

Notes: Table F.1 presents summary statistics of simple correlation coef cients of four price time series. These time series measure the price of a given good measured over the sample period in four regions, namely the border regions of the country pairs indicated on top of each panel and one non-border region of each country. Border regions comprise all observations within a distance of 40 km or less to the border, whereas non-border regions are observations which are more than 40 km but less or equal to 160 km away from the border. The time frequency chosen is monthly.

TABLE F.2: Co-movement of prices within and across countries: 60 km region width

	All observations											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.59	1.00			0.47	1.00			0.62	1.00		
p^*	0.18	0.15	1.00		0.08	0.08	1.00		0.09	0.11	1.00	
p_{nb}^*	0.21	0.20	0.77	1.00	0.08	0.08	0.56	1.00	0.09	0.11	0.57	1.00
	Signi cant border estimates, cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.52	1.00			0.47	1.00			0.65	1.00		
p^*	0.13	0.11	1.00		0.08	0.09	1.00		0.09	0.10	1.00	
p_{nb}^*	0.16	0.15	0.78	1.00	0.09	0.09	0.57	1.00	0.10	0.10	0.55	1.00
	Signi cant border estimates, no cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.61	1.00			0.47	1.00			0.64	1.00		
p^*	0.19	0.16	1.00		0.08	0.08	1.00		0.09	0.12	1.00	
p_{nb}^*	0.23	0.23	0.79	1.00	0.08	0.09	0.56	1.00	0.11	0.11	0.59	1.00
	Not signi cant border estimates, cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.28	1.00			0.43	1.00			0.64	1.00		
p^*	0.21	0.19	1.00		0.06	0.08	1.00		0.11	0.10	1.00	
p_{nb}^*	0.24	0.19	0.69	1.00	0.07	0.08	0.50	1.00	0.10	0.11	0.56	1.00
	Not signi cant border estimates, no cross-border shopping											
	BE-GE				BE-NL				GE-NL			
	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*	p_{nb}	p_b	p_b^*	p_{nb}^*
p_{nb}	1.00				1.00				1.00			
p	0.53	1.00			0.40	1.00			0.57	1.00		
p^*	0.24	0.13	1.00		0.06	0.09	1.00		0.08	0.11	1.00	
p_{nb}^*	0.28	0.24	0.65	1.00	0.06	0.09	0.49	1.00	0.09	0.12	0.53	1.00

Notes: Table F.2 presents summary statistics of simple correlation coef cients of four price time series. These time series measure the price of a given good measured over the sample period in four regions, namely the border regions of the country pairs indicated on top of each panel and one non-border region of each country. Border regions comprise all observations within a distance of 40 km or less to the border, whereas non-border regions are observations which are more than 40 km but less or equal to 160 km away from the border. The time frequency chosen is monthly.

2.G Additional results on the role of retailer variation (Section 5)

2.G.1 Within and between country price dispersion: all retailers vs same retailer

Tables in this section provide statistic on within- and between- country price dispersion using all common goods between countries (only common goods in a corresponding quarter).

TABLE G.1.1: Within- and cross-country price dispersion - All retailers

2005Q1			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	63464	0.26	0.00	10.86	5.07	1.93	9.61
NL-NL	158177	0.22	0.00	11.88	6.03	2.13	10.24
BE-NL	145704	7.11	3.45	28.52	19.85	14.43	21.68
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	862541	0.22	0.00	9.93	5.08	1.11	8.54
NL-NL	39475	0.11	0.00	12.81	6.25	1.54	11.18
GE-NL	278886	-5.90	-4.20	24.68	18.00	13.53	17.88
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	29652	0.42	0.00	11.49	5.14	1.99	10.29
GE-GE	538349	0.16	0.00	10.72	5.30	1.10	9.32
BE-GE	208621	9.79	7.63	26.82	20.35	15.25	20.02
2008Q4			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	99412	-0.07	0.00	13.22	4.40	1.29	12.47
NL-NL	281911	0.25	0.00	10.78	5.40	1.85	9.34
BE-NL	234810	7.36	3.39	28.85	18.75	12.76	23.12
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	1527189	0.03	0.00	10.45	5.88	2.32	8.64
NL-NL	59609	0.29	0.00	11.08	5.07	0.96	9.86
GE-NL	427319	-7.25	-4.10	26.62	19.02	13.45	19.98
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	42637	0.03	0.00	14.05	4.11	0.74	13.44
GE-GE	1052441	-0.02	0.00	10.46	5.83	2.26	8.68
BE-GE	325217	12.86	11.55	27.25	21.81	17.30	20.80

Notes: The table reports descriptive statistics on within and between-country price dispersion. Figures are computed for the case that purchases from all retailers are considered. The table therefore corresponds to Table 2 of the main text. 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. Then, for each good all possible bi-regional price gaps are computed. Finally, for each considered subsample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (1) of the main text. 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 G.1.2: Within- and cross-country price dispersion - Common retailers

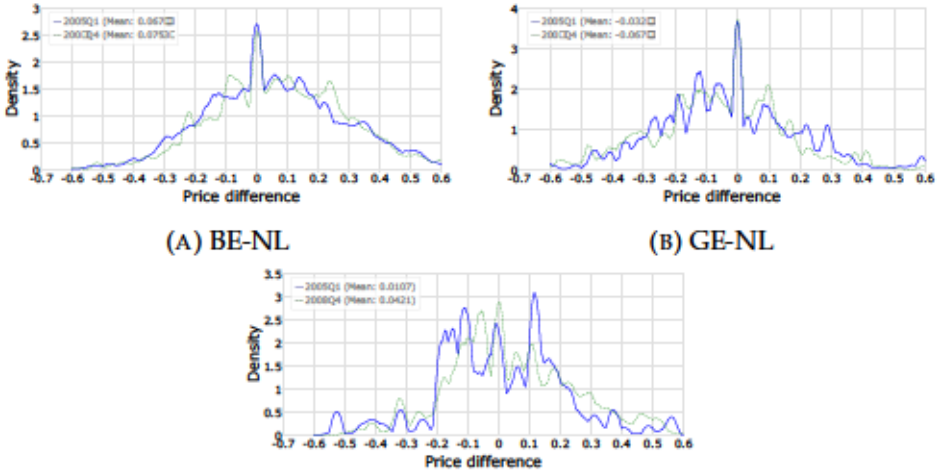
2005Q1			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	18737	0.08	0.00	6.70	2.15	0.01	6.35
NL-NL	29310	-0.08	0.00	6.49	1.73	0.00	6.25
BE-NL	42966	2.06	0.00	26.02	18.52	14.41	18.39
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	267580	-0.01	0.00	4.79	1.42	0.00	4.58
NL-NL	17226	-0.19	0.00	5.77	1.29	0.00	5.62
GE-NL	128734	-6.99	-6.50	21.94	17.26	13.53	15.25
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	11565	0.23	0.00	8.08	2.84	0.00	7.57
GE-GE	225876	-0.10	0.00	5.07	1.52	0.00	4.84
BE-GE	101884	7.76	6.14	24.32	18.20	13.98	17.90
2008Q4			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	28761	-0.04	0.00	8.26	1.31	0.00	8.15
NL-NL	51036	0.05	0.00	6.96	1.92	0.07	6.69
BE-NL	73568	0.32	0.00	25.69	18.03	13.59	18.30
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	360824	-0.14	0.00	6.02	2.41	0.00	5.51
NL-NL	30116	0.09	0.00	4.33	1.25	0.07	4.14
GE-NL	195089	-10.43	-8.71	23.93	19.18	14.08	17.71
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	20849	0.03	0.00	4.78	1.12	0.00	4.65
GE-GE	353426	-0.12	0.00	6.34	2.64	0.00	5.76
BE-GE	166060	8.20	6.95	23.76	18.99	15.03	16.48

Notes: The table reports descriptive statistics on within and between-country price dispersion. Figures are based on the prices of goods purchased from a common retailer across two given markets. 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. Then, for each good all possible bi-regional price gaps are computed. Finally, for each considered subsample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (1) of the main text. 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.

2.G.2 Within- and between-country price dispersion using only goods for which we have a border estimate: all retailers vs. same retailer

Tables in this section provide statistic on within- and between- country price dispersion using only common goods for which we have a border estimate in a corresponding quarter.

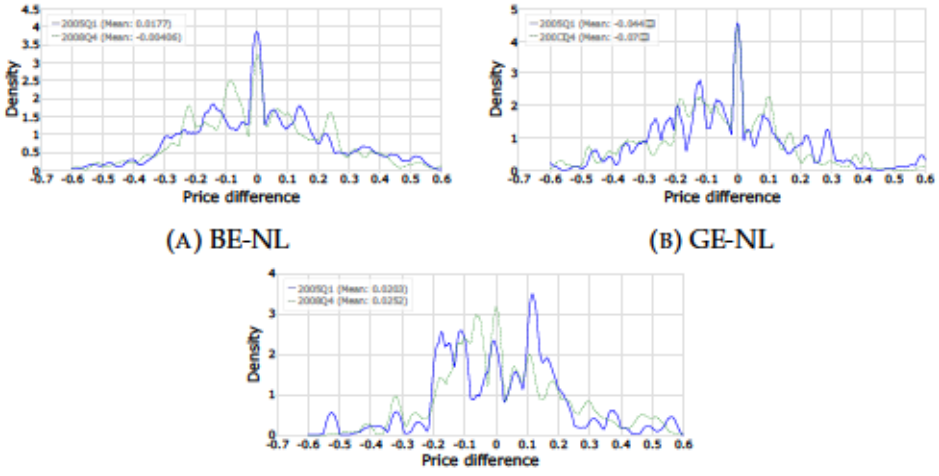
FIGURE G.2.1: Distribution of mean price differences - All retailer



(A) BE-NL (B) GE-NL

(C) BE-GE

FIGURE G.2.2: Distribution of mean price differences - Common retailers



(A) BE-NL (B) GE-NL

(C) BE-GE

TABLE G.2.3: Within- and cross-country price dispersion - All retailers

2005Q1			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	24417	0.21	0.00	7.51	3.95	1.55	6.39
NL-NL	32489	0.30	0.00	9.30	4.87	1.76	7.92
BE-NL	60770	6.78	5.06	24.91	19.37	15.32	17.06
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	196298	0.10	0.00	6.99	3.16	0.10	6.24
NL-NL	17191	0.02	0.00	8.11	3.63	0.26	7.26
GE-NL	121159	-3.29	-4.31	22.28	17.54	14.31	14.13
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	3165	0.26	0.00	4.47	2.13	0.41	3.94
GE-GE	45283	-0.25	0.00	4.67	1.83	0.00	4.30
BE-GE	25110	1.07	-0.04	24.98	18.49	13.76	16.83
2008Q4			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	38357	0.04	0.00	12.35	3.42	0.88	11.86
NL-NL	50619	0.20	0.00	8.48	4.05	1.19	7.45
BE-NL	95447	7.53	6.54	26.46	19.99	16.08	18.89
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	290942	-0.11	0.00	8.02	4.00	0.58	6.95
NL-NL	26453	0.23	0.00	7.71	3.28	0.51	6.98
GE-NL	182126	-6.78	-5.42	24.06	18.32	13.40	17.01
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	7624	0.29	0.00	4.60	2.08	0.37	4.11
GE-GE	104222	0.06	0.00	6.57	3.20	0.61	5.74
BE-GE	59391	4.21	0.66	21.99	17.00	12.79	14.57

Notes: The table reports descriptive statistics on within and between-country price dispersion employing regional goods-level price gaps. Only goods for which we have a border estimate are considered. 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 subsample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (1) of the main text. 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.

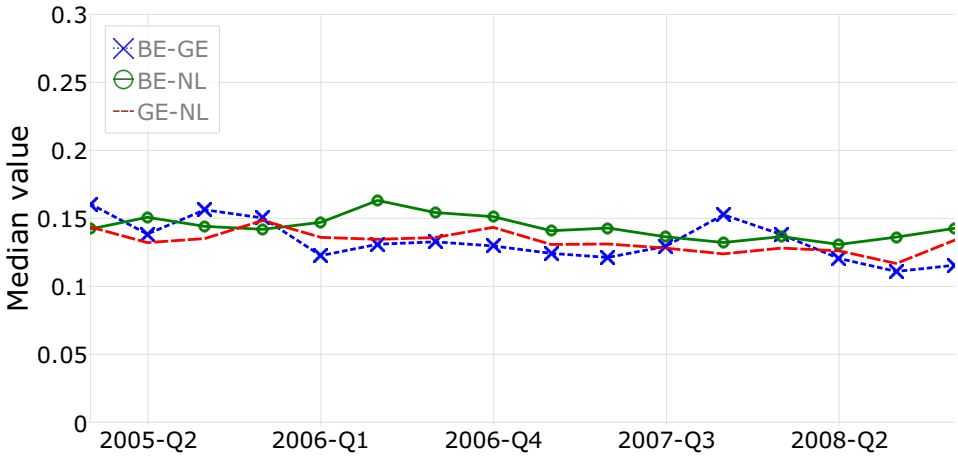
TABLE G.2.4: Within- and cross-country price dispersion - Common retailers

2005Q1			Price gap		Absolute price gap		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	10674	-0.02	0.00	4.47	1.55	0.01	4.20
NL-NL	14657	-0.05	0.00	3.76	1.14	0.00	3.59
BE-NL	26945	1.77	0.00	24.57	18.45	14.79	16.32
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	133910	0.00	0.00	4.57	1.42	0.00	4.34
NL-NL	11562	-0.24	0.00	4.37	1.21	0.00	4.20
GE-NL	82233	-4.49	-5.20	22.47	17.60	14.04	14.67
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	2590	0.47	0.00	4.82	1.86	0.09	4.47
GE-GE	38721	-0.40	0.00	4.60	1.47	0.00	4.37
BE-GE	21007	2.03	0.06	25.75	19.01	14.04	17.49
2008Q4			Price gap		Absolute price gap		
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	19090	0.03	0.00	3.14	0.98	0.00	2.98
NL-NL	25965	0.00	0.00	4.69	1.28	0.11	4.51
BE-NL	48313	-0.41	-0.13	23.97	18.00	14.26	15.84
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	201210	-0.19	0.00	5.47	2.16	0.00	5.03
NL-NL	18753	0.08	0.00	3.68	1.16	0.11	3.49
GE-NL	128072	-7.80	-6.48	22.69	17.49	12.82	16.42
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	5673	-0.01	0.00	2.73	1.10	0.02	2.50
GE-GE	79560	-0.02	0.00	4.61	2.00	0.00	4.15
BE-GE	44740	2.52	0.00	22.47	16.83	11.90	15.11

Notes: The table reports descriptive statistics on within and between-country price dispersion employing regional goods-level price gaps. Only goods for which we have a border estimate are considered. 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 subsample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (1) of the main text. 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.

2.G.3 Border estimates only for goods sold by common retailers

FIGURE G.3.3: Median border effects over time - Common retailers



Notes: plots median and mean values of estimated quarterly border coef cients(absolute values). 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 rst in the legend.

TABLE G.3.5: Regression discontinuity results: baseline speci cation

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.001	0.039	22.2	0.000	-0.005	0.057	25.3	0.000	0.010	0.051	30.4
Border	0.002	0.019	0.244	85.7	-0.060	-0.053	0.217	85.0	0.005	0.023	0.266	85.7
Border x Distance	0.000	0.001	0.065	9.6	0.000	0.001	0.072	22.3	-0.002	-0.027	0.086	8.9
Constant	-1.192	-1.216	1.192	84.3	-1.051	-1.153	1.067	86.7	-0.933	-1.174	1.227	80.4
Observations	293.000	.	.	.	233.000	.	.	.	56.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.002	0.038	10.2	0.000	0.001	0.049	14.7	0.000	0.005	0.051	25.2
Border	-0.002	-0.005	0.242	87.7	-0.066	-0.087	0.220	86.7	0.005	0.024	0.221	84.9
Border x Distance	-0.001	-0.005	0.077	9.2	0.000	-0.008	0.075	14.7	0.000	-0.012	0.069	16.0
Constant	-0.944	-0.884	1.294	95.8	-0.965	-0.962	1.054	95.5	-0.926	-1.003	1.007	87.4
Observations	479.000	.	.	.	354.000	.	.	.	119.000	.	.	.

Notes: The table reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. Positive values of the border coef cient 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 coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

TABLE G.3.6: Regression discontinuity results: baseline speci cation, only private-label goods

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.001	0.039	22.2	0.000	-0.006	0.057	26.6	0.000	0.007	0.043	31.5
Border	0.004	0.021	0.247	85.6	-0.067	-0.059	0.218	85.6	0.005	0.023	0.262	87.0
Border x Distance	0.000	0.002	0.065	9.5	0.000	0.003	0.067	23.4	-0.002	-0.025	0.084	9.3
Constant	-1.206	-1.221	1.191	83.8	-1.059	-1.144	1.062	86.9	-0.933	-1.136	1.188	79.6
Observations	284.000	.	.	.	222.000	.	.	.	54.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.001	0.034	10.8	0.000	0.002	0.047	15.0	0.000	0.005	0.047	25.2
Border	-0.008	-0.012	0.243	87.7	-0.067	-0.089	0.222	87.3	0.005	0.022	0.220	86.1
Border x Distance	0.000	-0.005	0.075	8.8	0.000	-0.010	0.073	14.7	0.000	-0.012	0.068	15.7
Constant	-0.966	-0.897	1.253	96.0	-0.976	-0.959	1.061	95.3	-0.923	-0.977	0.997	87.0
Observations	454.000	.	.	.	339.000	.	.	.	115.000	.	.	.

Notes: The table reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. Positive values of the border coef cient 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 coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

TABLE G.3.7: Regression discontinuity results: absolute values, only private-label goods

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.001	0.015	0.036	22.2	0.000	0.021	0.053	26.6	0.000	0.011	0.042	31.5
Border	0.146	0.184	0.166	85.6	0.144	0.175	0.142	85.6	0.156	0.196	0.174	87.0
Border x Distance	0.009	0.032	0.056	9.5	0.011	0.032	0.059	23.4	0.013	0.040	0.078	9.3
Constant	1.290	1.434	0.922	83.8	1.097	1.276	0.897	86.9	1.043	1.320	0.975	79.6
Observations	284.000	.	.	.	222.000	.	.	.	54.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.003	0.016	0.030	10.8	0.004	0.017	0.044	15.0	0.000	0.020	0.043	25.2
Border	0.142	0.182	0.162	87.7	0.134	0.174	0.164	87.3	0.116	0.164	0.148	86.1
Border x Distance	0.009	0.031	0.068	8.8	0.012	0.036	0.064	14.7	0.012	0.036	0.058	15.7
Constant	1.059	1.262	0.884	96.0	1.016	1.170	0.822	95.3	0.972	1.106	0.849	87.0
Observations	454.000	.	.	.	339.000	.	.	.	115.000	.	.	.

Notes: The table reports absolute vlaues of regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

TABLE G.3.8: Regression discontinuity results: baseline specification, national-brand goods

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.007	0.026	22.2	0.000	0.011	0.058	0.0	0.090	0.090	0.173	0.0
Border	-0.067	-0.062	0.083	88.9	0.037	0.070	0.168	72.7	0.023	0.023	0.491	50.0
Border x Distance	0.000	-0.036	0.068	11.1	0.000	-0.042	0.133	0.0	-0.085	-0.085	0.166	0.0
Constant	-0.504	-1.068	1.308	100.0	-0.716	-1.337	1.206	81.8	-2.206	-2.206	2.411	100.0
Observations	9.000	.	.	.	11.000	.	.	.	2.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.001	0.016	0.083	0.0	0.000	-0.029	0.079	6.7	0.032	0.009	0.129	25.0
Border	0.122	0.118	0.181	88.0	-0.034	-0.041	0.171	73.3	0.007	0.067	0.282	50.0
Border x Distance	-0.007	-0.011	0.112	16.0	0.001	0.043	0.097	13.3	-0.010	0.002	0.121	25.0
Constant	-0.704	-0.649	1.907	92.0	-0.937	-1.013	0.918	100.0	-1.627	-1.758	1.175	100.0
Observations	25.000	.	.	.	15.000	.	.	.	4.000	.	.	.

Notes: The table 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) of the main text. 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 80 km.

TABLE G.3.9: Regression discontinuity results: absolute values, national-brand goods

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	0.011	0.025	22.2	0.000	0.027	0.051	0.0	0.122	0.122	0.127	0.0
Border	0.081	0.089	0.048	88.9	0.103	0.121	0.131	72.7	0.347	0.347	0.033	50.0
Border x Distance	0.001	0.037	0.067	11.1	0.012	0.083	0.109	0.0	0.117	0.117	0.120	0.0
Constant	0.504	1.068	1.308	100.0	0.716	1.337	1.206	81.8	2.206	2.206	2.411	100.0
Observations	9.000	.	.	.	11.000	.	.	.	2.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.006	0.038	0.075	0.0	0.005	0.040	0.074	6.7	0.112	0.103	0.052	25.0
Border	0.159	0.172	0.129	88.0	0.085	0.139	0.102	73.3	0.110	0.171	0.216	50.0
Border x Distance	0.021	0.066	0.090	16.0	0.014	0.061	0.087	13.3	0.090	0.092	0.058	25.0
Constant	1.017	1.463	1.358	92.0	0.937	1.044	0.880	100.0	1.627	1.758	1.175	100.0
Observations	25.000	.	.	.	15.000	.	.	.	4.000	.	.	.

Notes: The table reports absolute vlaues of regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

TABLE G.3.10: Regression discontinuity results: absolute values

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.001	0.015	0.036	22.2	0.000	0.022	0.053	25.3	0.000	0.015	0.049	30.4
Border	0.142	0.181	0.164	85.7	0.144	0.173	0.142	85.0	0.160	0.201	0.173	85.7
Border x Distance	0.009	0.032	0.057	9.6	0.011	0.034	0.063	22.3	0.015	0.042	0.080	8.9
Constant	1.284	1.423	0.935	84.3	1.091	1.279	0.911	86.7	1.043	1.351	1.024	80.4
Observations	293.000	.	.	.	233.000	.	.	.	56.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.003	0.017	0.034	10.2	0.004	0.018	0.046	14.7	0.000	0.023	0.046	25.2
Border	0.143	0.182	0.160	87.7	0.134	0.173	0.162	86.7	0.116	0.164	0.149	84.9
Border x Distance	0.010	0.033	0.070	9.2	0.012	0.037	0.065	14.7	0.013	0.038	0.059	16.0
Constant	1.055	1.273	0.914	95.8	1.006	1.164	0.824	95.5	0.979	1.128	0.863	87.4
Observations	479.000	.	.	.	354.000	.	.	.	119.000	.	.	.

Notes: The table reports absolute vlaues of regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

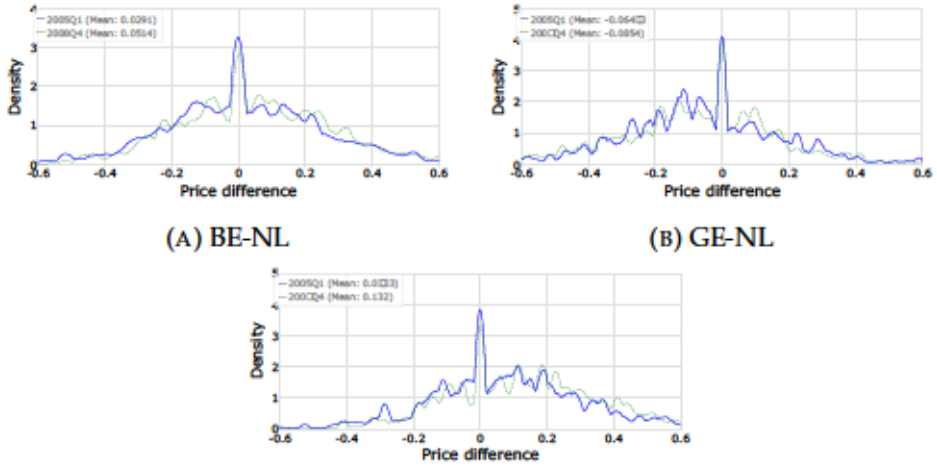
2.H Additional results on the role of sample characteristics for border estimation results (Section 2.6)

2.H.1 Price dispersion and border estimation results for homogeneous samples across country groups

Tables in this subsection provide descriptive statistic on within- and between-country price dispersion using only goods for which we have common purchases in all three sample countries in a given quarter.

Within- and between-country price dispersion for homogeneous samples across countries

FIGURE H.1.1: Distribution of mean price differences for goods with common observations in all countries



(A) BE-NL (B) GE-NL

(C) BE-GE

Notes: Figure H.1.1 plots the kernel density estimates of regional goods-level price gaps 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). Only those goods are considered for which we have common observations in all three sample countries in the respective quarter. To compute regional goods-level price gaps, 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 cross country bi-regional price gaps are computed. 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).

TABLE H.1.1: Price dispersion of goods with common observations in all countries

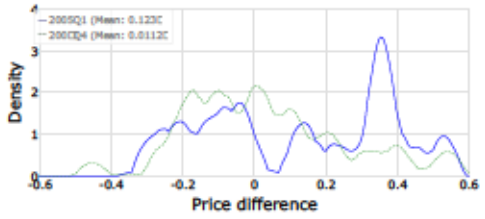
2005Q1			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	19000	0.31	0.00	12.03	4.91	1.69	10.99
NL-NL	25379	0.13	0.00	12.57	6.17	1.58	10.96
BE-NL	41231	2.91	0.00	26.88	19.74	15.41	18.48
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	331389	0.19	0.00	9.37	4.44	0.58	8.26
NL-NL	25379	0.13	0.00	12.57	6.17	1.58	10.96
GE-NL	160772	-6.47	-6.50	23.91	18.82	14.89	16.10
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	19000	0.31	0.00	12.03	4.91	1.69	10.99
GE-GE	331389	0.19	0.00	9.37	4.44	0.58	8.26
BE-GE	141651	8.83	7.41	26.28	19.88	15.23	19.32
2008Q4			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	29011	-0.06	0.00	11.17	3.66	0.66	10.55
NL-NL	41782	0.31	0.00	10.57	4.83	0.89	9.41
BE-NL	65885	5.14	3.82	27.22	20.04	16.21	19.12
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	615571	-0.04	0.00	9.85	5.43	1.90	8.22
NL-NL	41782	0.31	0.00	10.57	4.83	0.89	9.41
GE-NL	265604	-8.54	-6.57	26.10	20.02	14.81	18.80
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	29011	-0.06	0.00	11.17	3.66	0.66	10.55
GE-GE	615571	-0.04	0.00	9.85	5.43	1.90	8.22
BE-GE	221312	13.25	12.00	25.61	21.83	17.99	18.84

Notes: Table H.1.1 reports descriptive statistics on within and between-country price dispersion employing regional goods-level price gaps. Only those goods are considered for which we have common observations in all three sample countries in the respective quarter. 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 subsample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (1) of the main text. 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.

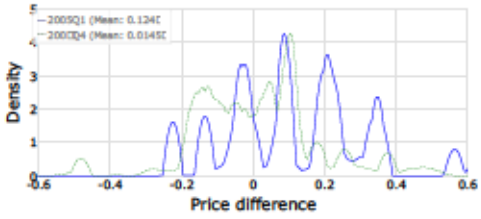
Within- and between- country price dispersion for goods common across all countries and included in our estimation sample

Tables and plots in this section provide statistic on within- and between- country price dispersion using goods common for all countries for which we have a border estimate in a corresponding quarter.

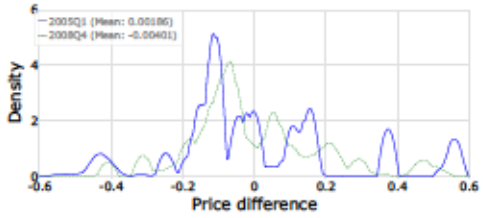
FIGURE H.1.2: Distribution of mean price differences for goods with common observations in all countries that are included in our estimation sample



(A) BE-NL



(B) GE-NL



(C) BE-GE

Notes: Figure H.1.2 plots the kernel density estimates of regional goods-level price gaps 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). Only those goods are considered for which we have common observations in all three sample countries in the respective quarter and for which the number of available observations is large enough to be included in our estimation analysis. To compute regional goods-level price gaps, 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 cross country bi-regional price gaps are computed. 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).

TABLE H.1.2: Within- and cross-country price dispersion for goods common across all countries and used for estimations

2005Q1			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	737	0.04	0.00	3.23	1.54	0.21	2.84
NL-NL	991	-0.23	0.00	4.15	1.75	0.39	3.77
BE-NL	1874	12.28	14.04	25.73	24.55	24.19	14.49
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	11443	-0.07	0.00	2.33	0.83	0.00	2.18
NL-NL	991	-0.23	0.00	4.15	1.75	0.39	3.77
GE-NL	7129	12.42	9.28	20.16	18.34	17.77	14.98
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	737	0.04	0.00	3.23	1.54	0.21	2.84
GE-GE	11443	-0.07	0.00	2.33	0.83	0.00	2.18
BE-GE	6173	0.19	-4.35	23.84	18.05	12.57	15.57
2008Q4			Price gap			Absolute price gap	
BE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	2089	0.18	0.00	4.19	2.01	0.44	3.68
NL-NL	2641	0.37	0.00	4.31	1.73	0.35	3.96
BE-NL	5147	1.12	0.10	25.37	18.77	14.12	17.11
GE-NL	N	Mean	Median	St.Dev	Mean	Median	St.Dev
GE-GE	29415	0.04	0.00	4.16	1.80	0.00	3.75
NL-NL	2641	0.37	0.00	4.31	1.73	0.35	3.96
GE-NL	18643	1.45	1.92	17.26	13.22	10.52	11.19
BE-GE	N	Mean	Median	St.Dev	Mean	Median	St.Dev
BE-BE	2089	0.18	0.00	4.19	2.01	0.44	3.68
GE-GE	29415	0.04	0.00	4.16	1.80	0.00	3.75
BE-GE	16648	-0.40	-5.26	23.42	17.08	10.97	16.03

Notes: Table H.1.2 reports descriptive statistics on within and between-country price dispersion employing regional goods-level price gaps. Only those goods are considered for which we have common observations in all three sample countries in the respective quarter and for which the number of available observations is large enough to be included in our estimation analysis. 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 subsample, summary statistics are computed based on the available, goods-level price gaps. Price gaps are computed according to Equation (1) of the main text. 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.

RD estimates for common goods

The following tables present results from regression discontinuity regressions (excluding cross-border shopping observations) for those goods for which we have common observations across all three countries in a given time period (quarter). The number of goods is smaller for the counterfactual regressions since not all of the common goods have a suf cient number of observations in each of the regions.

TABLE H.1.3: Regression discontinuity results using only goods common across countries: baseline speci cation

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.001	0.006	0.029	37.5	-0.001	0.002	0.038	12.5	0.000	-0.009	0.033	43.8
Border	0.132	0.124	0.263	100.0	0.090	0.123	0.200	93.8	-0.075	-0.010	0.237	87.5
Border x Distance	-0.003	-0.008	0.029	18.8	0.001	-0.002	0.047	6.3	-0.002	-0.008	0.020	6.3
Constant	-1.023	-1.335	0.899	93.8	-1.005	-1.334	0.908	100.0	-0.924	-1.209	0.908	75.0
Observations	16.000	.	.	.	16.000	.	.	.	16.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.002	0.008	0.024	4.9	0.000	-0.007	0.051	14.6	0.000	0.004	0.052	22.0
Border	0.003	0.008	0.265	90.2	0.022	0.009	0.168	87.8	-0.052	-0.016	0.232	87.8
Border x Distance	-0.008	-0.021	0.042	7.3	0.000	0.011	0.071	14.6	0.000	-0.018	0.080	14.6
Constant	-1.035	-1.214	1.181	100.0	-1.046	-1.202	1.179	97.6	-0.988	-1.194	1.201	85.4
Observations	41.000	.	.	.	41.000	.	.	.	41.000	.	.	.

Notes: The table reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Only those goods are considered for which we have common observations in all three sample countries in the respective quarter and for which the number of available observations is large enough to be included in our estimation analysis. Results are based on estimating Equation (2) of the main text. For each country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. Positive values of the border coef cient 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 coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

TABLE H.1.4: Regression discontinuity results using only goods common across countries: absolute values

	BE-NL				GE-NL				BE-GE			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.004	0.015	0.025	37.5	0.008	0.020	0.033	12.5	0.000	0.010	0.032	43.8
Border	0.235	0.242	0.152	100.0	0.156	0.180	0.148	93.8	0.138	0.176	0.152	87.5
Border x Distance	0.006	0.017	0.024	18.8	0.019	0.030	0.036	6.3	0.004	0.013	0.018	6.3
Constant	1.023	1.335	0.899	93.8	1.005	1.334	0.908	100.0	0.924	1.209	0.908	75.0
Observations	16.000	.	.	.	16.000	.	.	.	16.000	.	.	.
	BE-NL				GE-NL				BE-GE			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.005	0.015	0.021	4.9	0.009	0.023	0.045	14.6	0.001	0.020	0.048	22.0
Border	0.169	0.196	0.175	90.2	0.105	0.128	0.107	87.8	0.097	0.166	0.160	87.8
Border x Distance	0.016	0.030	0.037	7.3	0.011	0.035	0.062	14.6	0.014	0.048	0.067	14.6
Constant	1.106	1.404	0.940	100.0	1.106	1.394	0.938	97.6	1.033	1.377	0.980	85.4
Observations	41.000	.	.	.	41.000	.	.	.	41.000	.	.	.

Notes: Table H.1.4 reports absolute values of regression discontinuity results for the first quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimatingEquation (2) of the main text. Only those goods are considered for which we have common observations in all three sample countries in the respective quarter and for which the number of available observations is large enough to be included in our estimation analysis. 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 80 km.

TABLE H.1.5: Counterfactual evidence using only goods common across countries: regression discontinuity results

	BEFL-BEWL				LSax-NRW				NLE-NLS			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	-0.002	0.035	0.0	0.000	-0.002	0.013	18.8	-0.001	-0.039	0.065	12.5
Border	0.000	-0.010	0.028	6.3	0.000	0.003	0.013	0.0	0.000	0.000	0.017	0.0
Border x Distance	0.000	-0.010	0.055	6.3	0.000	0.010	0.024	6.3	0.007	0.050	0.081	0.0
Constant	-0.898	-1.205	0.822	75.0	-0.924	-1.208	0.911	43.8	-1.002	-1.326	0.903	75.0
Observations	16.000	.	.	.	16.000	.	.	.	16.000	.	.	.
	BEFL-BEWL				LSax-NRW				NLE-NLS			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.000	-0.017	0.114	31.7	0.000	-0.003	0.053	2.4	0.000	-0.005	0.067	7.3
Border	-0.002	-0.005	0.057	19.5	0.000	0.005	0.053	7.3	0.002	0.007	0.033	4.9
Border x Distance	0.000	0.019	0.119	12.2	0.000	0.006	0.085	0.0	0.000	0.005	0.081	2.4
Constant	-1.207	-1.196	1.172	87.8	-0.988	-1.191	1.200	63.4	-1.048	-1.209	1.171	82.9
Observations	41.000	.	.	.	41.000	.	.	.	41.000	.	.	.

Notes: The table reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. Only those goods are considered for which we have common observations in all three sample countries in the respective quarter and for which the number of available observations is large enough to be included in our estimation analysis. For each hypothetical country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

TABLE H.1.6: Counterfactual evidence using only goods common across countries: regression discontinuity results, absolute values

	BEFL-BEWL				LSax-NRW				NLE-NLS			
2005Q1	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.002	0.019	0.029	0.0	0.000	0.007	0.011	18.8	0.008	0.041	0.064	12.5
Border	0.001	0.013	0.027	6.3	0.001	0.007	0.011	0.0	0.009	0.012	0.012	0.0
Border x Distance	0.015	0.031	0.045	6.3	0.000	0.014	0.021	6.3	0.024	0.061	0.073	0.0
Constant	0.898	1.214	0.807	75.0	0.924	1.208	0.911	43.8	1.002	1.326	0.903	75.0
Observations	16.000	.	.	.	16.000	.	.	.	16.000	.	.	.
	BEFL-BEWL				LSax-NRW				NLE-NLS			
2008Q4	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)	Median	Mean	Std.	Sign.(%)
Distance	0.010	0.061	0.097	31.7	0.002	0.024	0.047	2.4	0.017	0.034	0.057	7.3
Border	0.013	0.031	0.048	19.5	0.003	0.024	0.048	7.3	0.006	0.016	0.030	4.9
Border x Distance	0.019	0.067	0.099	12.2	0.007	0.044	0.073	0.0	0.022	0.044	0.067	2.4
Constant	1.241	1.385	0.935	87.8	1.033	1.374	0.980	63.4	1.108	1.397	0.932	82.9
Observations	41.000	.	.	.	41.000	.	.	.	41.000	.	.	.

Notes: The table reports regression discontinuity results for the rst quarter of 2005 (2005Q1) and the forth quarter of 2008 (2008Q4). Results are based on estimating Equation (2) of the main text. Only those goods are considered for which we have common observations in all three sample countries in the respective quarter and for which the number of available observations is large enough to be included in our estimation analysis. For each hypothetical country pair (indicated in the rst row of each sample period), the reference country is the respectively rstly mentioned country. The columns denoted "Mean", "Median" and "Std" report the mean, median and standard deviation of the absolute values of estimated coef cients given in the rst column. The column "Sign(%)" contains the fraction of regressions in which the corresponding coef cient is signi cant at the 10% level. The chosen bandwidth in all speci cations is 80 km.

2.H.2 RD estimates grouped by product characteristics and product categories

RD estimates grouped by price and purchase frequency

The following tables present border estimates (after excluding cross-border shopping observations), grouped by price and purchase frequency.

To identify "Low price" and "High price" goods, we first compute the average price of a given good in a given quarter using data from both countries. Goods are then separated into a "Low price" and a "High price" category, depending on whether its average price is below or above the calculated median price.

TABLE H.2.7: Border estimates grouped by price: BE-NL

		Border coeff.			Absolute border coeff.		
				Low-price goods			
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	335.000	0.061	0.077	0.260	0.150	0.198	0.185
2005q2	360.000	0.071	0.075	0.259	0.152	0.206	0.174
2005q3	362.000	0.040	0.071	0.262	0.160	0.203	0.180
2005q4	413.000	0.043	0.061	0.264	0.157	0.203	0.178
2006q1	443.000	0.058	0.066	0.254	0.159	0.198	0.172
2006q2	432.000	0.079	0.073	0.260	0.164	0.203	0.178
2006q3	430.000	0.088	0.102	0.265	0.166	0.213	0.188
2006q4	431.000	0.110	0.099	0.253	0.172	0.210	0.171
2007q1	441.000	0.083	0.092	0.250	0.163	0.207	0.167
2007q2	488.000	0.086	0.103	0.260	0.165	0.211	0.183
2007q3	486.000	0.058	0.076	0.265	0.150	0.205	0.184
2007q4	510.000	0.070	0.092	0.279	0.166	0.209	0.206
2008q1	499.000	0.072	0.104	0.266	0.164	0.208	0.196
2008q2	520.000	0.037	0.080	0.251	0.150	0.195	0.177
2008q3	511.000	0.043	0.078	0.258	0.135	0.193	0.188
2008q4	497.000	0.032	0.051	0.343	0.151	0.214	0.273
				High-price goods			
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	334.000	0.071	0.093	0.232	0.148	0.188	0.164
2005q2	360.000	0.075	0.074	0.337	0.147	0.197	0.283
2005q3	361.000	0.056	0.067	0.221	0.144	0.180	0.145
2005q4	412.000	0.054	0.050	0.288	0.151	0.187	0.225
2006q1	442.000	0.067	0.085	0.228	0.154	0.188	0.154
2006q2	432.000	0.065	0.085	0.249	0.165	0.195	0.176
2006q3	430.000	0.064	0.066	0.248	0.171	0.196	0.165
2006q4	431.000	0.068	0.070	0.217	0.150	0.181	0.138
2007q1	441.000	0.085	0.095	0.261	0.172	0.210	0.182
2007q2	488.000	0.104	0.110	0.230	0.161	0.197	0.162
2007q3	485.000	0.095	0.098	0.239	0.163	0.197	0.167
2007q4	509.000	0.091	0.082	0.257	0.163	0.201	0.180
2008q1	499.000	0.128	0.127	0.262	0.188	0.223	0.187
2008q2	520.000	0.126	0.119	0.244	0.169	0.207	0.175
2008q3	511.000	0.122	0.119	0.245	0.187	0.213	0.170
2008q4	496.000	0.109	0.103	0.240	0.162	0.200	0.167

Notes: The table presents border estimates (after excluding cross-border shopping observations), grouped by price. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given group. To identify "Low price" and "High price" goods, we first compute the average price of a given good in a given quarter using data from both countries. Goods are then separated into a "Low price" and a "High price" category, depending on whether its average price is below or above the calculated median price.

TABLE H.2.8: Border estimates grouped by price: GE-NL

	Border coeff.				Absolute border coeff.		
	Low-price goods						
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	174.000	-0.045	-0.030	0.234	0.141	0.178	0.153
2005q2	195.000	-0.050	-0.027	0.222	0.128	0.165	0.151
2005q3	204.000	-0.065	-0.069	0.237	0.123	0.177	0.173
2005q4	222.000	-0.060	-0.053	0.299	0.145	0.200	0.228
2006q1	247.000	-0.077	-0.085	0.248	0.148	0.195	0.175
2006q2	265.000	-0.062	-0.047	0.259	0.151	0.198	0.174
2006q3	264.000	-0.062	-0.043	0.260	0.144	0.194	0.177
2006q4	266.000	-0.069	-0.042	0.253	0.146	0.190	0.172
2007q1	265.000	-0.069	-0.056	0.263	0.139	0.194	0.185
2007q2	296.000	-0.051	-0.039	0.254	0.134	0.184	0.179
2007q3	271.000	-0.053	-0.052	0.241	0.131	0.176	0.171
2007q4	285.000	-0.033	-0.047	0.237	0.134	0.174	0.168
2008q1	277.000	-0.030	-0.046	0.216	0.129	0.166	0.145
2008q2	276.000	-0.043	-0.057	0.212	0.134	0.164	0.146
2008q3	271.000	-0.041	-0.051	0.218	0.122	0.164	0.152
2008q4	259.000	-0.056	-0.078	0.218	0.132	0.170	0.157
	High-price goods						
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	174.000	-0.038	-0.048	0.212	0.156	0.173	0.131
2005q2	195.000	-0.045	-0.054	0.205	0.142	0.167	0.130
2005q3	203.000	-0.045	-0.053	0.208	0.144	0.167	0.134
2005q4	221.000	-0.063	-0.072	0.199	0.125	0.159	0.139
2006q1	246.000	-0.061	-0.050	0.237	0.116	0.161	0.180
2006q2	265.000	-0.057	-0.050	0.216	0.130	0.164	0.148
2006q3	263.000	-0.058	-0.070	0.205	0.141	0.165	0.140
2006q4	266.000	-0.059	-0.063	0.228	0.127	0.165	0.170
2007q1	264.000	-0.023	-0.042	0.212	0.121	0.165	0.139
2007q2	295.000	-0.048	-0.056	0.226	0.141	0.175	0.153
2007q3	270.000	-0.038	-0.040	0.243	0.130	0.175	0.173
2007q4	285.000	-0.046	-0.040	0.302	0.143	0.193	0.236
2008q1	277.000	-0.017	-0.029	0.242	0.128	0.173	0.170
2008q2	275.000	-0.021	-0.044	0.231	0.129	0.172	0.161
2008q3	270.000	-0.006	-0.027	0.231	0.125	0.167	0.162
2008q4	258.000	-0.053	-0.061	0.262	0.137	0.191	0.189

Notes: The table presents border estimates (after excluding cross-border shopping observations), grouped by price. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given group. To identify "Low price" and "High price" goods, we first compute the average price of a given good in a given quarter using data from both countries. Goods are then separated into a "Low price" and a "High price" category, depending on whether its average price is below or above the calculated median price.

TABLE H.2.9: Border estimates grouped by price: BE-GE

	Border coeff.			Absolute border coeff.		
			Low-price goods			
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	34.000	-0.110	-0.028	0.321	0.185	0.251	0.198
2005q2	43.000	-0.037	-0.015	0.309	0.138	0.213	0.223
2005q3	38.000	0.034	0.086	0.320	0.170	0.251	0.213
2005q4	52.000	0.000	0.040	0.321	0.160	0.234	0.221
2006q1	54.000	0.029	0.079	0.320	0.150	0.238	0.226
2006q2	52.000	-0.019	0.039	0.319	0.159	0.236	0.216
2006q3	49.000	-0.062	0.016	0.342	0.151	0.242	0.239
2006q4	57.000	-0.047	0.005	0.348	0.132	0.233	0.257
2007q1	68.000	-0.027	0.020	0.318	0.138	0.212	0.236
2007q2	63.000	0.000	0.016	0.275	0.141	0.200	0.187
2007q3	64.000	-0.001	0.018	0.303	0.153	0.222	0.205
2007q4	73.000	-0.052	0.000	0.265	0.167	0.209	0.161
2008q1	65.000	-0.041	-0.007	0.271	0.141	0.204	0.176
2008q2	73.000	-0.048	0.019	0.261	0.150	0.195	0.173
2008q3	74.000	-0.015	0.002	0.242	0.136	0.177	0.164
2008q4	79.000	-0.043	0.007	0.236	0.132	0.172	0.161
			High-price goods			
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	33.000	0.037	0.047	0.161	0.124	0.131	0.103
2005q2	42.000	0.056	0.094	0.229	0.151	0.179	0.169
2005q3	37.000	0.000	0.043	0.188	0.147	0.147	0.123
2005q4	52.000	0.029	0.035	0.205	0.137	0.167	0.122
2006q1	54.000	-0.008	0.032	0.198	0.115	0.153	0.128
2006q2	51.000	0.004	0.055	0.229	0.122	0.172	0.160
2006q3	49.000	-0.001	0.016	0.221	0.128	0.162	0.149
2006q4	57.000	0.055	0.073	0.212	0.132	0.170	0.144
2007q1	68.000	0.058	0.056	0.180	0.104	0.139	0.126
2007q2	63.000	0.021	0.013	0.214	0.118	0.160	0.142
2007q3	63.000	0.013	0.043	0.225	0.106	0.165	0.157
2007q4	73.000	0.051	0.045	0.191	0.117	0.150	0.125
2008q1	65.000	0.060	0.048	0.190	0.105	0.148	0.127
2008q2	73.000	0.060	0.065	0.192	0.131	0.156	0.129
2008q3	73.000	0.028	0.048	0.204	0.114	0.154	0.141
2008q4	79.000	0.074	0.078	0.191	0.134	0.160	0.128

Notes: The table presents border estimates (after excluding cross-border shopping observations), grouped by price. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given group. To identify "Low price" and "High price" goods, we first compute the average price of a given good in a given quarter using data from both countries. Goods are then separated into a "Low price" and a "High price" category, depending on whether its average price is below or above the calculated median price.

TABLE H.2.10: Border estimates grouped by purchase frequency: BE-NL

		Border coeff.			Absolute border coeff.		
			Goods purchased at low frequency				
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	335.000	0.055	0.083	0.252	0.140	0.191	0.184
2005q2	363.000	0.069	0.069	0.355	0.155	0.214	0.291
2005q3	366.000	0.053	0.067	0.252	0.157	0.195	0.174
2005q4	417.000	0.047	0.047	0.305	0.157	0.200	0.234
2006q1	444.000	0.057	0.074	0.245	0.158	0.196	0.165
2006q2	435.000	0.056	0.071	0.267	0.163	0.202	0.187
2006q3	432.000	0.067	0.071	0.279	0.172	0.213	0.195
2006q4	431.000	0.087	0.077	0.232	0.158	0.194	0.148
2007q1	444.000	0.092	0.086	0.269	0.179	0.214	0.184
2007q2	488.000	0.076	0.094	0.259	0.162	0.206	0.182
2007q3	488.000	0.071	0.068	0.268	0.152	0.205	0.185
2007q4	515.000	0.070	0.068	0.280	0.167	0.210	0.196
2008q1	501.000	0.104	0.109	0.279	0.182	0.224	0.199
2008q2	523.000	0.091	0.084	0.266	0.167	0.206	0.187
2008q3	513.000	0.091	0.083	0.273	0.172	0.212	0.191
2008q4	498.000	0.060	0.070	0.274	0.164	0.209	0.191
			Goods purchased at high frequency				
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	334.000	0.070	0.087	0.240	0.156	0.194	0.165
2005q2	357.000	0.075	0.080	0.231	0.151	0.188	0.157
2005q3	357.000	0.043	0.071	0.232	0.150	0.188	0.154
2005q4	408.000	0.050	0.064	0.243	0.153	0.190	0.165
2006q1	441.000	0.076	0.077	0.237	0.154	0.190	0.161
2006q2	429.000	0.087	0.087	0.241	0.165	0.196	0.165
2006q3	428.000	0.082	0.097	0.232	0.162	0.196	0.157
2006q4	431.000	0.081	0.092	0.240	0.162	0.197	0.164
2007q1	438.000	0.080	0.101	0.242	0.158	0.204	0.165
2007q2	488.000	0.109	0.119	0.231	0.162	0.202	0.163
2007q3	483.000	0.089	0.106	0.234	0.162	0.197	0.165
2007q4	504.000	0.094	0.107	0.254	0.164	0.200	0.190
2008q1	497.000	0.107	0.122	0.249	0.174	0.208	0.183
2008q2	517.000	0.105	0.115	0.228	0.157	0.196	0.164
2008q3	509.000	0.087	0.114	0.228	0.153	0.194	0.165
2008q4	495.000	0.095	0.084	0.319	0.152	0.205	0.258

Notes:The table presents border estimates (after excluding cross-border shopping observations), grouped by purchase frequency. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given group. To identify "Low freq." and "High freq." goods, we first compute the total number of purchases of a given good in a given quarter using data from both countries. Goods are then separated into a "Low freq." and a "High freq." category, depending on whether its total number of purchases is below or above the calculated median price.

TABLE H.2.11: Border estimates grouped by purchase frequency: GE-NL

	Border coeff.			Absolute border coeff.		
	Goods purchased at low frequency					
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	179.000	-0.106	-0.097	0.203	0.144	0.174	0.143
2005q2	195.000	-0.076	-0.084	0.193	0.118	0.157	0.140
2005q3	204.000	-0.072	-0.102	0.221	0.117	0.171	0.173
2005q4	223.000	-0.082	-0.101	0.237	0.143	0.177	0.187
2006q1	248.000	-0.090	-0.115	0.192	0.125	0.165	0.151
2006q2	269.000	-0.089	-0.095	0.226	0.137	0.181	0.165
2006q3	267.000	-0.085	-0.090	0.231	0.137	0.178	0.172
2006q4	269.000	-0.089	-0.081	0.237	0.132	0.168	0.186
2007q1	267.000	-0.071	-0.087	0.251	0.121	0.184	0.192
2007q2	299.000	-0.090	-0.097	0.241	0.131	0.184	0.184
2007q3	276.000	-0.073	-0.086	0.240	0.120	0.172	0.188
2007q4	286.000	-0.068	-0.085	0.260	0.134	0.186	0.200
2008q1	279.000	-0.074	-0.078	0.230	0.131	0.173	0.170
2008q2	276.000	-0.080	-0.099	0.232	0.136	0.181	0.176
2008q3	274.000	-0.071	-0.083	0.233	0.124	0.174	0.176
2008q4	260.000	-0.076	-0.095	0.247	0.154	0.193	0.180
	Goods purchased at high frequency						
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	169.000	0.013	0.022	0.227	0.153	0.177	0.143
2005q2	195.000	-0.017	0.004	0.225	0.144	0.175	0.141
2005q3	203.000	-0.038	-0.020	0.218	0.144	0.173	0.134
2005q4	220.000	-0.020	-0.023	0.265	0.128	0.182	0.194
2006q1	245.000	-0.023	-0.020	0.277	0.148	0.192	0.201
2006q2	261.000	-0.010	-0.001	0.242	0.146	0.181	0.160
2006q3	260.000	-0.029	-0.022	0.232	0.151	0.181	0.147
2006q4	263.000	-0.044	-0.024	0.241	0.149	0.187	0.154
2007q1	262.000	-0.014	-0.010	0.219	0.151	0.175	0.131
2007q2	292.000	-0.003	0.003	0.228	0.144	0.175	0.146
2007q3	265.000	-0.007	-0.005	0.238	0.149	0.180	0.154
2007q4	284.000	-0.001	-0.002	0.277	0.138	0.181	0.210
2008q1	275.000	0.003	0.004	0.221	0.122	0.167	0.145
2008q2	275.000	0.001	-0.002	0.200	0.126	0.154	0.126
2008q3	267.000	0.009	0.006	0.206	0.125	0.156	0.135
2008q4	257.000	-0.024	-0.044	0.233	0.118	0.167	0.167

Notes: The table presents border estimates (after excluding cross-border shopping observations), grouped by purchase frequency. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given group. To identify "Low freq." and "High freq." goods, we first compute the total number of purchases of a given good in a given quarter using data from both countries. Goods are then separated into a "Low freq." and a "High freq." category, depending on whether its total number of purchases is below or above the calculated median price.

TABLE H.2.12: Border estimates grouped by purchase frequency: BE-GE

		Border coeff.			Absolute border coeff.		
			Goods purchased at low frequency				
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	34.000	-0.000	-0.023	0.253	0.131	0.176	0.181
2005q2	43.000	0.000	0.002	0.295	0.138	0.194	0.221
2005q3	39.000	0.013	0.030	0.213	0.147	0.164	0.137
2005q4	54.000	0.019	0.003	0.250	0.151	0.188	0.163
2006q1	56.000	0.015	0.024	0.263	0.126	0.190	0.183
2006q2	52.000	-0.001	-0.010	0.262	0.122	0.183	0.186
2006q3	50.000	-0.009	-0.007	0.313	0.129	0.202	0.237
2006q4	57.000	0.025	0.026	0.323	0.137	0.215	0.241
2007q1	68.000	-0.000	0.004	0.212	0.111	0.148	0.150
2007q2	64.000	0.000	-0.014	0.244	0.136	0.178	0.166
2007q3	65.000	0.000	0.017	0.293	0.118	0.205	0.209
2007q4	73.000	0.009	0.012	0.233	0.146	0.176	0.152
2008q1	66.000	0.019	0.002	0.232	0.114	0.165	0.163
2008q2	74.000	0.016	0.048	0.207	0.127	0.161	0.137
2008q3	75.000	0.009	0.037	0.244	0.129	0.178	0.169
2008q4	81.000	0.015	0.028	0.222	0.134	0.167	0.148
			Goods purchased at high frequency				
	N	Median	Mean	Std.	Median	Mean	Std.
2005q1	33.000	0.052	0.041	0.259	0.172	0.209	0.155
2005q2	42.000	0.057	0.077	0.254	0.139	0.199	0.174
2005q3	36.000	-0.005	0.102	0.306	0.162	0.238	0.214
2005q4	50.000	-0.004	0.075	0.284	0.151	0.213	0.200
2006q1	52.000	0.014	0.089	0.267	0.136	0.202	0.195
2006q2	51.000	0.063	0.104	0.283	0.171	0.226	0.197
2006q3	48.000	-0.007	0.040	0.257	0.152	0.203	0.160
2006q4	57.000	0.000	0.052	0.252	0.119	0.188	0.174
2007q1	68.000	0.049	0.072	0.295	0.133	0.203	0.224
2007q2	62.000	0.030	0.045	0.245	0.112	0.182	0.168
2007q3	62.000	0.012	0.044	0.236	0.150	0.182	0.154
2007q4	73.000	0.029	0.034	0.230	0.157	0.183	0.142
2008q1	64.000	0.017	0.039	0.238	0.143	0.188	0.148
2008q2	72.000	0.001	0.037	0.252	0.150	0.190	0.168
2008q3	72.000	-0.002	0.012	0.203	0.119	0.152	0.134
2008q4	77.000	0.039	0.058	0.211	0.127	0.165	0.143

Notes: The table presents border estimates (after excluding cross-border shopping observations), grouped by purchase frequency. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given group. To identify "Low freq." and "High freq." goods, we first compute the total number of purchases of a given good in a given quarter using data from both countries. Goods are then separated into a "Low freq." and a "High freq." category, depending on whether its total number of purchases is below or above the calculated median price.

RD estimates by different groups: product category

The following tables present border estimates (after excluding cross-border shopping observations), grouped by product categories. For each country group, results are presented for 2005Q1 and 2008Q4. Product categories correspond to those speci ed by the data supplier for Germany. Goods from the Netherlands and Belgium were manually assigned to the German categories based on a detailed description of goods characteristics.

TABLE H.2.13: Border estimates grouped by product category, BE-NL, 2005 Q1

2005 Q1		Border coeff.			Abs. border coeff.		
	N	Median	Mean	Std.	Median	Mean	Std.
AlcholfreeCO2	2	0.269	0.269	0.374	0.269	0.269	0.374
AlcholfreeNoCO2	8	0.143	0.180	0.164	0.143	0.180	0.164
Animalcare	17	0.088	0.054	0.145	0.144	0.132	0.075
Babyproducts	3	-0.027	0.260	0.517	0.051	0.311	0.472
Basicfood	12	0.067	-0.022	0.390	0.221	0.287	0.250
Beer	4	-0.113	-0.080	0.272	0.261	0.229	0.112
Bodycare	27	0.088	0.101	0.224	0.141	0.197	0.142
Candy	57	0.043	0.046	0.181	0.123	0.149	0.111
Cereals	9	-0.043	-0.029	0.132	0.069	0.104	0.078
DairyWhite	52	0.100	0.114	0.169	0.137	0.154	0.133
DairyYellow	2	-0.251	-0.251	0.013	0.251	0.251	0.013
Delicasees	18	0.153	0.201	0.407	0.203	0.277	0.356
Fatoils	10	0.274	0.246	0.277	0.349	0.316	0.182
Frische	10	0.160	0.059	0.220	0.208	0.200	0.087
Frozenproducts	36	0.023	0.078	0.197	0.138	0.160	0.138
HotDrinks	5	0.205	0.118	0.212	0.224	0.207	0.092
HouseholdCleansers	35	0.021	0.061	0.289	0.146	0.214	0.201
Hygieneproducts	9	0.038	0.112	0.438	0.213	0.294	0.331
Laundry	15	0.095	0.080	0.213	0.165	0.174	0.141
Liquor	1	0.174	0.174	.	0.174	0.174	.
Meat	11	0.193	0.250	0.270	0.193	0.274	0.243
MouthTooth	6	0.190	0.214	0.337	0.225	0.307	0.236
PreservedFood	42	0.055	0.044	0.237	0.147	0.181	0.157
Readymade	2	0.072	0.072	0.248	0.175	0.175	0.101
Rest	166	0.066	0.091	0.242	0.142	0.190	0.175
Snacks	81	0.004	0.048	0.271	0.167	0.220	0.163
Spreads	14	0.046	0.059	0.081	0.052	0.077	0.064
Vegetables	8	0.216	0.190	0.203	0.216	0.230	0.149
Wine	7	0.127	0.196	0.197	0.127	0.196	0.197

Notes: The table presents border estimates for the country pair BE-NL (after excluding cross-border shopping observations), grouped by product categories. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given category. The considered sample period is 2005Q1.

TABLE H.2.14: Border estimates grouped by product category, BE-NL, 2008 Q4

2008 Q4		Border coeff.			Abs. border coeff.		
	N	Median	Mean	Std.	Median	Mean	Std.
AlcholfreeCO2	2	0.228	0.228	0.474	0.335	0.335	0.323
AlcholfreeNoCO2	18	0.116	0.146	0.232	0.161	0.209	0.173
Animalcare	37	0.072	0.057	0.173	0.114	0.146	0.106
Babyproducts	2	-0.188	-0.188	0.057	0.188	0.188	0.057
Basicfood	17	0.052	0.183	0.445	0.340	0.371	0.295
Beer	6	-0.065	-0.057	0.039	0.065	0.057	0.039
Bodycare	52	0.117	0.088	0.219	0.154	0.191	0.137
Candy	91	0.133	0.134	0.260	0.162	0.207	0.207
Cereals	7	0.174	0.170	0.185	0.174	0.182	0.171
DairyWhite	72	0.048	0.072	0.173	0.124	0.145	0.117
DairyYellow	4	-0.095	-0.089	0.152	0.138	0.131	0.103
Delicasees	26	0.205	0.196	0.572	0.346	0.462	0.381
Fatoils	9	0.138	0.087	0.205	0.152	0.193	0.093
Frische	11	0.057	0.061	0.153	0.065	0.120	0.108
Frozenproducts	55	0.113	0.071	0.524	0.181	0.298	0.434
HotDrinks	26	0.202	0.195	0.214	0.244	0.244	0.152
HouseholdCleansers	36	-0.065	-0.100	0.336	0.185	0.265	0.226
Hygieneproducts	7	-0.090	-0.083	0.306	0.208	0.256	0.159
Laundry	12	-0.027	-0.005	0.147	0.067	0.110	0.091
Liquor	4	-0.028	-0.030	0.051	0.041	0.044	0.035
Meat	39	0.067	0.075	0.204	0.152	0.176	0.126
MouthTooth	10	-0.070	-0.086	0.342	0.322	0.305	0.148
PreservedFood	43	-0.015	-0.022	0.205	0.180	0.168	0.117
Readymade	10	0.156	0.057	0.294	0.225	0.246	0.151
Rest	253	0.074	0.077	0.325	0.155	0.206	0.263
Snacks	106	0.109	0.115	0.212	0.157	0.190	0.148
Spreads	13	0.005	-0.001	0.163	0.121	0.123	0.101
Vegetables	12	0.208	0.159	0.142	0.208	0.180	0.112
Wine	13	-0.044	-0.030	0.146	0.112	0.118	0.084

Notes: The table presents border estimates for the country pair BE-NL (after excluding cross-border shopping observations), grouped by product categories. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given category. The considered sample period is 2008Q4.

TABLE H.2.15: Border estimates grouped by product category, GE-NL, 2005 Q1

2005 Q1		Border coeff.			Abs. border coeff.		
	N	Median	Mean	Std.	Median	Mean	Std.
AlcholfreeCO2	6	0.110	0.114	0.120	0.110	0.129	0.102
AlcholfreeNoCO2	23	-0.060	-0.029	0.185	0.129	0.147	0.112
Animalcare	2	-0.323	-0.323	0.156	0.323	0.323	0.156
Babyproducts	2	0.025	0.025	0.133	0.094	0.094	0.035
Basicfood	6	-0.165	-0.023	0.415	0.249	0.331	0.202
Bodycare	11	0.034	-0.037	0.299	0.181	0.232	0.177
Candy	83	-0.072	-0.075	0.187	0.134	0.160	0.121
Cereals	4	-0.201	-0.183	0.257	0.239	0.221	0.214
DairyWhite	23	0.000	0.010	0.182	0.119	0.142	0.111
DairyYellow	13	-0.187	-0.089	0.216	0.216	0.208	0.090
Delicasees	16	-0.010	-0.013	0.154	0.116	0.123	0.089
Fatoils	2	0.153	0.153	0.161	0.153	0.153	0.161
Frozenproducts	30	-0.101	-0.068	0.185	0.135	0.152	0.123
HotDrinks	4	0.235	0.202	0.147	0.235	0.202	0.147
HouseholdCleansers	11	-0.243	-0.246	0.211	0.243	0.246	0.211
Hygieneproducts	3	-0.209	-0.093	0.263	0.209	0.231	0.039
Laundry	6	-0.159	-0.156	0.052	0.159	0.156	0.052
Liquor	1	0.061	0.061	.	0.061	0.061	.
Meat	27	-0.004	-0.009	0.201	0.107	0.137	0.145
MouthTooth	1	-0.344	-0.344	.	0.344	0.344	.
PreservedFood	7	-0.052	-0.074	0.156	0.163	0.137	0.094
Readymade	8	-0.052	-0.001	0.370	0.294	0.306	0.173
Snacks	38	0.000	0.022	0.275	0.203	0.209	0.178
Spreads	8	0.077	0.016	0.160	0.089	0.118	0.100
Vegetables	13	0.012	0.048	0.313	0.186	0.233	0.204

Notes: The table presents border estimates for the country pair GE-NL (after excluding cross-border shopping observations), grouped by product categories. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given category. The considered sample period is 2005Q1.

TABLE H.2.16: Border estimates grouped by product category, GE-NL, 2008 Q4

2008 Q4		Border coeff.			Abs. border coeff.		
	N	Median	Mean	Std.	Median	Mean	Std.
AlcholfreeCO2	2	-0.080	-0.080	0.016	0.080	0.080	0.016
AlcholfreeNoCO2	21	-0.071	-0.060	0.090	0.093	0.087	0.063
Alcohol	1	-0.703	-0.703	.	0.703	0.703	.
Animalcare	19	-0.124	-0.147	0.115	0.124	0.147	0.115
Babyproducts	4	-0.050	-0.029	0.078	0.071	0.070	0.026
Basicfood	8	-0.143	-0.207	0.215	0.143	0.210	0.212
Bodycare	23	-0.170	-0.220	0.241	0.181	0.258	0.198
Candy	101	-0.066	-0.079	0.258	0.130	0.185	0.195
Cereals	5	0.151	0.106	0.132	0.151	0.133	0.097
DairyWhite	56	-0.036	-0.049	0.230	0.154	0.176	0.154
DairyYellow	16	-0.071	0.014	0.238	0.156	0.190	0.136
Delicasees	16	-0.012	-0.107	0.286	0.239	0.244	0.176
Fatoils	2	0.033	0.033	0.046	0.033	0.033	0.046
Frische	1	-0.233	-0.233	.	0.233	0.233	.
Frozenproducts	39	0.040	0.015	0.170	0.126	0.132	0.105
HotDrinks	17	0.070	0.112	0.238	0.162	0.197	0.170
HouseholdCleansers	16	-0.085	-0.130	0.265	0.206	0.237	0.168
Hygieneproducts	7	-0.269	-0.225	0.185	0.269	0.239	0.163
Laundry	6	-0.287	-0.212	0.286	0.321	0.287	0.192
Liquor	1	-0.107	-0.107	.	0.107	0.107	.
Meat	46	-0.019	-0.030	0.136	0.095	0.109	0.086
MouthTooth	2	-0.553	-0.553	0.006	0.553	0.553	0.006
PreservedFood	11	-0.056	-0.027	0.187	0.069	0.132	0.129
Readymade	19	0.022	-0.043	0.233	0.116	0.159	0.172
Rest	4	-0.019	-0.030	0.137	0.107	0.118	0.040
Snacks	44	-0.042	-0.067	0.324	0.148	0.227	0.239
Spreads	6	-0.007	-0.045	0.075	0.007	0.045	0.075
Vegetables	16	0.052	0.017	0.305	0.089	0.210	0.216
Wine	8	-0.316	-0.373	0.182	0.316	0.373	0.182

Notes: The table presents border estimates for the country pair GE-NL (after excluding cross-border shopping observations), grouped by product categories. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given category. The considered sample period is 2008Q4.

TABLE H.2.17: Border estimates grouped by product category, BE-GE, 2005 Q1

2005 Q1		Border coeff.			Abs. border coeff.		
	N	Median	Mean	Std.	Median	Mean	Std.
AlcholfreeNoCO2	4	0.176	0.138	0.371	0.275	0.313	0.180
Basicfood	5	-0.180	-0.186	0.363	0.191	0.284	0.273
Bodycare	2	-0.345	-0.345	0.254	0.345	0.345	0.254
Candy	2	0.071	0.071	0.251	0.177	0.177	0.100
DairyWhite	7	-0.164	-0.166	0.142	0.164	0.166	0.142
DairyYellow	11	-0.027	-0.014	0.202	0.088	0.146	0.132
Delicasees	1	0.478	0.478	.	0.478	0.478	.
Fatoils	1	0.230	0.230	.	0.230	0.230	.
Frozenproducts	5	0.015	-0.001	0.125	0.056	0.083	0.083
HotDrinks	1	0.112	0.112	.	0.112	0.112	.
HouseholdCleansers	1	0.000	0.000	.	0.000	0.000	.
Hygieneproducts	2	0.017	0.017	0.234	0.166	0.166	0.024
Meat	12	0.128	0.193	0.277	0.163	0.242	0.231
Readymade	3	-0.113	-0.049	0.217	0.192	0.177	0.059
Snacks	7	0.018	0.013	0.138	0.123	0.116	0.059
Spreads	1	-0.107	-0.107	.	0.107	0.107	.
Vegetables	2	-0.095	-0.095	0.324	0.229	0.229	0.134

Notes: The table presents border estimates for the country pair BE-GE (after excluding cross-border shopping observations), grouped by product categories. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given category. The considered sample period is 2008Q4.

TABLE H.2.18: Border estimates grouped by product category, BE-GE, 2008 Q4

2008 Q4		Border coeff.			Abs. border coeff.		
	N	Median	Mean	Std.	Median	Mean	Std.
AlcholfreeCO2	1	0.626	0.626	.	0.626	0.626	.
AlcholfreeNoCO2	6	0.166	0.197	0.084	0.166	0.197	0.084
Animalcare	3	0.001	0.065	0.175	0.069	0.111	0.136
Basicfood	9	0.015	0.032	0.367	0.112	0.245	0.261
Candy	12	0.170	0.187	0.193	0.194	0.226	0.141
Cereals	3	0.015	-0.005	0.042	0.023	0.031	0.020
DairyWhite	16	0.045	0.054	0.200	0.164	0.165	0.118
DairyYellow	14	0.003	0.105	0.269	0.147	0.202	0.200
Delicasees	3	0.116	0.083	0.057	0.116	0.083	0.057
Fatoils	2	-0.174	-0.174	0.039	0.174	0.174	0.039
Frozenproducts	11	0.000	-0.009	0.130	0.069	0.095	0.084
HotDrinks	5	-0.086	-0.140	0.182	0.086	0.174	0.140
HouseholdCleansers	2	-0.109	-0.109	0.036	0.109	0.109	0.036
Hygieneproducts	2	0.037	0.037	0.031	0.037	0.037	0.031
Liquor	2	-0.023	-0.023	0.090	0.064	0.064	0.032
Meat	23	0.088	0.059	0.205	0.140	0.175	0.117
PreservedFood	2	0.179	0.179	0.198	0.179	0.179	0.198
Readymade	5	-0.071	-0.035	0.071	0.071	0.067	0.032
Rest	4	0.111	0.079	0.212	0.200	0.173	0.116
Snacks	22	-0.068	-0.024	0.239	0.126	0.179	0.156
Spreads	5	-0.096	-0.023	0.179	0.132	0.139	0.092
Vegetables	6	-0.018	-0.037	0.134	0.099	0.103	0.083

Notes: The table presents border estimates for the country pair BE-GE (after excluding cross-border shopping observations), grouped by product categories. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given category. The considered sample period is 2008Q4.

TABLE H.2.19: Border estimates grouped by product category, all country pairs, 2005Q1

2005 Q1		Border coeff.			Abs. border coeff.		
	N	Median	Mean	Std.	Median	Mean	Std.
AlcholfreeCO2	8	0.110	0.153	0.188	0.110	0.164	0.178
AlcholfreeNoCO2	35	0.006	0.038	0.221	0.147	0.174	0.139
Animalcare	19	0.081	0.014	0.185	0.145	0.152	0.100
Babyproducts	5	-0.027	0.166	0.393	0.069	0.224	0.355
Basicfood	23	-0.069	-0.058	0.379	0.215	0.298	0.233
Beer	4	-0.113	-0.080	0.272	0.261	0.229	0.112
Bodycare	40	0.067	0.041	0.264	0.173	0.214	0.156
Candy	150	-0.025	-0.022	0.192	0.126	0.153	0.118
Cereals	13	-0.043	-0.076	0.183	0.075	0.140	0.136
DairyWhite	95	0.027	0.058	0.205	0.142	0.164	0.134
DairyYellow	26	-0.070	-0.069	0.207	0.202	0.185	0.110
Delicasees	35	0.082	0.111	0.330	0.186	0.212	0.274
Fatoils	13	0.230	0.230	0.247	0.267	0.284	0.176
Frische	17	-0.005	0.007	0.183	0.167	0.145	0.106
Frozenproducts	80	-0.007	0.008	0.202	0.140	0.157	0.126
HotDrinks	11	0.205	0.177	0.184	0.224	0.218	0.127
HouseholdCleansers	54	0.000	-0.017	0.297	0.147	0.217	0.201
Hygieneproducts	14	-0.041	0.055	0.375	0.209	0.262	0.265
Laundry	22	-0.026	0.010	0.206	0.154	0.163	0.121
Liquor	2	0.117	0.117	0.080	0.117	0.117	0.080
Meat	51	0.034	0.096	0.257	0.117	0.191	0.196
MouthTooth	7	0.087	0.134	0.373	0.294	0.312	0.215
PreservedFood	51	0.043	0.021	0.236	0.150	0.181	0.152
Readymade	14	-0.064	-0.003	0.295	0.212	0.240	0.159
Rest	305	0.037	0.059	0.255	0.145	0.189	0.181
Snacks	129	0.004	0.034	0.265	0.166	0.210	0.164
Spreads	23	0.046	0.037	0.116	0.068	0.092	0.077
Vegetables	24	0.104	0.078	0.277	0.189	0.225	0.174
Wine	8	0.113	0.179	0.189	0.113	0.179	0.189

Notes: The table presents border estimates for all three country pairs (BE-NL, GE-NL and BE-GE) (after excluding cross-border shopping observations), grouped by product categories. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given category. The considered sample period is 2005Q1.

TABLE H.2.20: Border estimates grouped by product category, all country pairs, 2008Q4

2008 Q4		Border coeff.			Abs. border coeff.		
	N	Median	Mean	Std.	Median	Mean	Std.
AlcholfreeCO2	5	-0.069	0.184	0.376	0.107	0.292	0.278
AlcholfreeNoCO2	48	0.010	0.044	0.195	0.127	0.150	0.129
Alcohol	1	-0.703	-0.703	.	0.703	0.703	.
Animalcare	59	-0.026	-0.008	0.182	0.119	0.144	0.109
Babyproducts	7	-0.060	-0.021	0.189	0.099	0.144	0.110
Basicfood	35	-0.048	0.036	0.408	0.177	0.305	0.269
Beer	6	-0.065	-0.057	0.039	0.065	0.057	0.039
Bodycare	75	-0.003	-0.007	0.266	0.171	0.212	0.160
Candy	211	0.006	0.031	0.275	0.159	0.196	0.195
Cereals	15	0.039	0.113	0.156	0.060	0.135	0.137
DairyWhite	168	0.002	0.011	0.197	0.124	0.150	0.128
DairyYellow	35	-0.025	0.051	0.252	0.164	0.195	0.163
Delicasees	46	0.063	0.076	0.480	0.300	0.354	0.329
Fatoils	13	0.065	0.039	0.194	0.147	0.166	0.098
Frische	20	0.011	0.005	0.176	0.064	0.131	0.113
Frozenproducts	109	0.070	0.039	0.388	0.143	0.213	0.327
HotDrinks	49	0.114	0.127	0.237	0.183	0.217	0.157
HouseholdCleansers	62	-0.065	-0.094	0.294	0.176	0.236	0.198
Hygieneproducts	16	-0.093	-0.130	0.245	0.171	0.221	0.161
Laundry	19	-0.058	-0.092	0.226	0.115	0.182	0.159
Liquor	7	-0.068	-0.039	0.060	0.068	0.058	0.037
Meat	109	0.023	0.028	0.183	0.119	0.147	0.112
MouthTooth	12	-0.295	-0.164	0.359	0.343	0.346	0.165
PreservedFood	58	-0.026	-0.014	0.198	0.134	0.157	0.120
Readymade	41	0.025	0.014	0.267	0.144	0.198	0.177
Rest	459	0.044	0.055	0.343	0.152	0.212	0.275
Snacks	177	0.042	0.049	0.258	0.152	0.195	0.175
Spreads	25	-0.010	-0.014	0.143	0.096	0.105	0.095
Vegetables	35	0.066	0.075	0.259	0.137	0.194	0.184
Wine	22	-0.168	-0.185	0.254	0.196	0.237	0.203

Notes: The table presents border estimates for all three country pairs (BE-NL, GE-NL and BE-GE) (after excluding cross-border shopping observations), grouped by product categories. Reported statistics include the median (Median), the mean (Mean) and the standard deviation (Std.) of estimates in a given category. The considered sample period is 2008Q4.

2.H.3 Exploring the role of geographic heterogeneity

The tables in this section report raw price gaps (column "Price Gap") and median values of obtained border estimates (after excluding cross-border shopping observations) (column "Border") for different bandwidths for the 4th quarter of 2008.

TABLE H.3.1: Raw price gaps and border estimates for different bandwidths, BE-NL, 2008Q4

	All goods						
	N Eans	N	Price Gap	Border	Abs Price Gap	Abs Border	Sign
20	310	41	0.057	0.057	0.148	0.140	76
40	571	51	0.089	0.078	0.173	0.173	80
60	831	55	0.079	0.077	0.165	0.164	82
80	993	60	0.071	0.071	0.165	0.158	81
100	1,136	63	0.077	0.069	0.163	0.159	81
120	1,187	65	0.076	0.067	0.164	0.160	81
140	1,228	66	0.075	0.067	0.164	0.158	81
160	1,239	67	0.075	0.070	0.164	0.159	81
Total	1,136	60	0.076	0.070	0.164	0.159	81
	Goods present at all bandwidths						
	N Eans	N	Price Gap	Border	Abs Price Gap	Abs Border	Sign
20	310	41	0.057	0.057	0.148	0.140	76
40	310	72	0.064	0.055	0.151	0.138	81
60	310	100	0.068	0.059	0.153	0.142	85
80	310	126	0.063	0.060	0.152	0.143	86
100	310	148	0.067	0.061	0.152	0.152	88
120	310	158	0.067	0.061	0.152	0.152	88
140	310	165	0.067	0.063	0.152	0.154	89
160	310	170	0.066	0.065	0.152	0.156	89

Notes: The table reports raw price gaps (column "Price Gap") and median values of obtained border estimates for BE-NL (after excluding cross-border shopping observations) (column "Border") for different bandwidths for the 4th quarter of 2008. Raw price gaps are computed as the log difference between means of the prices observed at each side of the border. The upper panel ("All goods") presents values based on all available goods given the respective bandwidth, while the lower panel ("Goods present at all bandwidths") only considers goods present at all considered bandwidths. Column "N Eans" reports the total number of goods included in a given sample, whereas "N" indicates the median number of observations per good included, "Sign" reports the proportion (in percentage terms) of signi cant estimates.

TABLE H.3.2: Raw price gaps and border estimates for different bandwidths, GE-NL, 2008Q4

	All goods						
	N Eans	N	Price Gap	Border	Abs Price Gap	Abs Border	Sign
20	74	46	-0.009	-0.024	0.115	0.110	82
40	187	46	-0.028	-0.026	0.112	0.118	77
60	360	53	-0.038	-0.035	0.123	0.120	74
80	517	57	-0.054	-0.055	0.134	0.134	77
100	623	60	-0.056	-0.053	0.136	0.137	78
120	738	64	-0.066	-0.065	0.143	0.143	76
140	778	66	-0.064	-0.071	0.143	0.139	77
160	804	68	-0.063	-0.070	0.140	0.143	78
Total	738	61	-0.057	-0.057	0.136	0.136	77
	Goods present at all bandwidths						
	N Eans	N	Price Gap	Border	Abs Price Gap	Abs Border	Sign
20	74	46	-0.009	-0.024	0.115	0.110	82
40	74	76	-0.006	-0.013	0.114	0.116	84
60	74	123	-0.007	-0.007	0.113	0.113	85
80	74	175	-0.012	-0.017	0.114	0.112	85
100	74	214	-0.007	-0.007	0.116	0.111	86
120	74	255	-0.010	-0.009	0.114	0.112	86
140	74	277	-0.011	-0.009	0.115	0.109	88
160	74	289	-0.013	-0.006	0.113	0.111	89

Notes: The table reports raw price gaps (column "Price Gap") and median values of obtained border estimates for GE-NL (after excluding cross-border shopping observations) (column "Border") for different bandwidths for the 4th quarter of 2008. Raw price gaps are computed as the log difference between means of the prices observed at each side of the border. The upper panel ("All goods") presents values based on all available goods given the respective bandwidth, while the lower panel ("Goods present at all bandwidths") only considers goods present at all considered bandwidths. Column "N Eans" reports the total number of goods included in a given sample, whereas "N" indicates the median number of observations per good included, "Sign" reports the proportion (in percentage terms) of signi cant estimates.

TABLE H.3.3: Raw price gaps and border estimates for different bandwidths, BE-GE, 2008Q4

	All goods						
	N Eans	N	Price Gap	Border	Abs Price Gap	Abs Border	Sign
20	1	22	-0.131	-0.139	0.131	0.139	100
40	20	29	-0.048	0.017	0.090	0.094	80
60	72	34	0.003	0.029	0.111	0.114	75
80	158	41	0.006	0.021	0.116	0.132	76
100	222	49	0.020	0.020	0.127	0.133	77
120	390	51	0.087	0.083	0.148	0.154	79
140	500	53	0.106	0.104	0.164	0.165	79
160	572	55	0.108	0.112	0.169	0.170	80
Total	500	50	0.081	0.076	0.151	0.154	79
	Goods present at all bandwidths						
	N Eans	N	Price Gap	Border	Abs Price Gap	Abs Border	Sign
60	72	34	0.003	0.029	0.111	0.114	75
80	72	56	0.002	0.016	0.114	0.115	79
100	72	72	0.005	0.001	0.113	0.114	82
120	72	113	0.005	0.006	0.115	0.113	86
140	72	137	0.001	0.002	0.114	0.115	86
160	72	163	0.003	0.006	0.114	0.115	85

Notes: The table reports raw price gaps (column "Price Gap") and median values of obtained border estimates for BE-GE (after excluding cross-border shopping observations) (column "Border") for different bandwidths for the 4th quarter of 2008. Raw price gaps are computed as the log difference between means of the prices observed at each side of the border. The upper panel ("All goods") presents values based on all available goods given the respective bandwidth, while the lower panel ("Goods present at all bandwidths") only considers goods present at all considered bandwidths. Column "N Eans" reports the total number of goods included in a given sample, whereas "N" indicates the median number of observations per good included, "Sign" reports the proportion (in percentage terms) of signi cant estimates.

Chapter 3

Regional in ation rates: measurement and biases

Natalia Zabelina

3.1 Introduction

The dynamics of aggregate prices play a key role in many macro-oriented economic elds such as macroeconomics, monetary economics or international economics, and in related policy areas. Since price level data are generally not available, aggregate price dynamics are usually measured employing corresponding indices.[1] Two widely used indices are the producer price index (PPI) which focuses on producer prices and the consumer price index (CPI) which is - as its name already indicates - based on prices paid by consumers. The CPI generally underlies the calculation of the in ation rate of a country (or any other regional entity) which "measures the rate at which the prices of consumption goods and services are changing from one period to another" (ILO, IMF, OECD, Eurostat, UNECE, The World Bank, 2020, p. 14). The in ation rate thus provides information about the intertemporal change in the purchasing power of a given nominal amount of currency. It plays an important role in the decision-making process of many political institutions in very different areas and at different geographical levels.

[1] Since aggregate price-level rather than price-index data are crucial for being able to properly address a number of important questions related, e.g., to the international comparison of real incomes or the construction of absolute poverty measures, there have been made several efforts to provide such absolute price level measures. The most noteworthy one is conducted in the context of the International Comparison Program of the World Bank Group. Whilst the price-level statistics provided as part of this initiative certainly provide considerable value added, their computation is based on fairly few goods only and data are available at a very low (annual) frequency only.

In terms of geographical coverage, CPIs are usually computed only at a national level and used at this level of aggregation for various economic policy purposes such as a major target variable of monetary policy or the computation of real values for GDP and wages. At sub-national levels, CPI measures are available only in relatively few selected countries (such as the United States, Germany, Italy, or Spain). However, even for these countries, the level of geographic disaggregation at which these indices are calculated is not very granular though. This lack represents a serious shortcoming since (correct) measures of regional price developments also play an important role for decision making at this level and for understanding the interaction between local prices and other economic variables. Such measures are, e.g., important to assess the degree of regional in ation inequality which plays an important role in the context of competitiveness and social welfare assessments.[2] Moreover, such regional statistics would allow to assess the role that price adjustments play for the response of local economies both to local and nationwide/global shocks. Existing empirical evidence on the topic shows that there seems to exist substantial variation in price developments across regions. As a consequence, the common practice to employ national in ation rates to de ate regional nominal values can lead to substantial mis-calculations of the underlying real values. The - rare - evidence on this topic is either based on - not bias-corrected - of cial price statistics (see e.g. Beck, Hubrich, and Marcellino, 2009) or only available for the U.S. (see, e.g., Handbury and David E. Weinstein, 2015).

The primary measure of national in ation in the EU is the Harmonised Index of Consumer Prices (HICP), which is compiled using the harmonised methodology across the EU countries. Since its introduction some of the countries have fully adopted the HICP index as the national CPI, but most of the countries compute the national CPI separately. Both price indexes are based on the same data sources, but computed with slight differences in concept or methodology.[3] For intra-countries regions there is no systematic

[2]Within the European Union, e.g., an important policy application could be the use of regional price statistics to properly compute regional real GDP values which are highly important for European Union regional policy, known as "Cohesion Policy", which aims at the diminishing economic activity disparities across regions and has budget equal to - almost a third of the total European Union (EU) budget, see European Comission (2018b).

[3] European Comission (2018a) provides an overview of differences between HICP and the national CPI for each country.

or harmonised approach of measuring price indexes across EU countries. Moreover, regional price indexes are only computed as sub-products of the national CPI computation and published only in a few countries. The of cial price indexes (whether HICP/CPI or existing regional price indexes) computed in the EU countries on a basis of the Laspeyres price index for a xed (although regularly updated) basket of selected goods and services. These indexes are also called " xed basket" price indexes since they are designed to re ect the average price change of a representative xed basket of goods and services. The main concern about the usage of " xed basket" price indexes in computation of in- ation rates is the potential overestimation of in ation due to the measurement biases arising from how " xed basket" price indexes handle substitution between goods, quality changes and introduction of new goods (see e.g. ECB, 2014, Broda and D. Weinstein, 2010, Stephen J. Redding and David E. Weinstein, 2020).

First, there is a potential "quality bias", that arises when sales of a good from the basket become insigni cant and the good is resampled with a replacement, more representative good, which might be of a different quality. In this case, the change in prices between the two periods shows the difference in prices for goods with different quality features. HICP methodology (see European Comission (2017)) requires the prices of replacement goods to undergo a quality adjustment, which is implemented by the national statistical of ces using a number of methods including "hedonic methods", expert judgments, and "overlap methods". These quality adjustments methods are, however, can still lead to an upward bias of the price index if the quality improvement occurs faster than they are measured or if they are not fully measured.[4] Another kind of "quality bias" arises when completely new goods appear at the market and they are not directly introduced in the basket. New goods are frequently introduced at a higher price, which declines very rapidly as the product starts to penetrate the market and this decline is missed in a price index that does not introduce new goods immediately[5] (ECB, 2014).

Secondly, there exists a "substitution bias", which arises from the fact that consumers can substitute away from the goods that become more expensive to a cheaper alternative.

[4] For a comprehensive discussion about potential biases from the most frequently applied quality adjustment methods see Ahnert and Kenny (2004).

[5] In HICP a new good is only introduced into the basket within 12 months after it reached 0.1% of households nal expenditures (ECB (2014)), which is a quite signi cant value for individual good.

This bias is partly accounted for in the CPIs by regularly updating the weights[6]. However, in the case of changes in relative prices consumers might rapidly substitute away to cheaper goods, so the old weights do not re ect the true current preferences leading to less expensive goods not being represented enough and more expensive goods being overrepresented.[7] This might result in an upward bias in the price index because the statistical of ces only collect information about a sample of good s varieties so it is not possible for the CPI to accurately re ect the substitution effects.

Thirdly, conventionally used in ation measures are subject to the "consumer valuation bias", which was recently discovered by Stephen J. Redding and David E. Weinstein (2020). The bias arises from the fact that conventional measures of in ation rates assume that consumers do not change their valuations of goods over time. In a sense the "consumer valuation bias" is similar to the "substitution bias" with the difference that the "substitution bias" arises in the case of relative price changes, while the "consumer valuation bias" arises due to relative demand changes.

In this study, I employ European household panel scanner price data to compute regional price indices that are not prone to the aforementioned biases and thus provide reliable measures to gauge the size of regional in ation rates and their differences within a given country. Simultaneously, I evaluate the magnitude of each of the three biases at the regional level. This exercise allows to draw insights into questions related to the extent of heterogeneities in consumer and retailer behavior across space that have started to gain attention in the literature recently.[8] To compute price indices, I make use of standard approaches such as the Laspeyres, Paasche, Fischer, and Törnquvist indices. Moreover, I employ the recently developed methodology by Stephen J. Redding and David E. Weinstein (2020) and thus to the best of my knowledge the current study provides the rst quantitative assessment of the size and heterogeneity of consumer valuation biases at the regional level.

[6]For HICP weights are updated annually by EU countries starting from 2012. For some national CPIs the weights are updated even less frequently with periods between the updates lasting up to 5 years.

[7]Recent literature (see e.g. Argente and M. Lee (2017), Nevo and Wong (2019), and Coibion, Gorodnichenko, and Hong (2015)) provides support for changing behaviour of households during the Great Recession, such that households started to buy more goods of worse quality, at less expensive stores and increased purchasing on sales.

[8]See, e.g., the works by C. Hottman (2014), C. J. Hottman, Stephen J Redding, and David E Weinstein (2016), Handbury and David E. Weinstein (2015), and Handbury (2019).

To conduct the study, data for Spain (for the period 2008 – 2016) and Germany (for the period 2009 – 2016) are used. These countries are well suited for the purpose of this study given that both their historical development and their current economic and political structure would suggest that there might exist signi cant heterogeneities in retail markets across regions. Moreover, the availability of regional in ation rates in Spain allows for a direct comparison of the price dynamics obtained using scanner data with that provided by the Spanish national statistical of ce.

The scanner price data have the following advantages compared to the data collected by statistical of ces: 1) the data contain information about the whole universe of goods, which the households buy including the new goods, 2) the data provide the exact prices paid and quantities bought that allow to additionally control for substitution between goods and changes in consumer valuation of unique goods. The results show that in a-tion rates computed using the Laysperes price index tend to overestimate in ation rates given by the UPI on average by 0.88% in Germany and by 0.46% in Spain. The "substitution bias" is found to be equal on average to 0.36% in Germany and 0.24% in Spain. The "variety bias" equals to 0.26% in Germany and 0.18% in Spain, and the "consumer valuation bias" amounts to 0.36% in Germany and 0.04% in Spain. The biases exhibit considerable heterogeneity across regions and some substantial volatility over time.

This study is the most closely related to two major strands of the literature. First, it complements the literature that uses scanner price data and employs the CES demand system to incorporate quality adjustment and substitution patterns into the cost of living price index and estimates biases in conventional in ation measures. The literature builds on the following contributions: Feenstra (1994), Broda and D. Weinstein (2010), and Stephen J. Redding and David E. Weinstein (2020) that modify and re ne an economic approach to measure the cost of living price index using CES preferences. Using the Nielsen Homescan Database, Broda and D. Weinstein (2010) nds that the "variety bias" in the US is around 0.8 percentage points per year, Stephen J. Redding and David E. Weinstein (2020) con rms this result with a longer period of data and additionally nds that the "consumer valuation bias" is around 0.4 percentage points per year.

Secondly, this study is related to the literature that uses scanner price data to construct regional in ation rates. Coibion, Gorodnichenko, and Hong (2015), Stroebel and Vavra

(2019), and Beraja, Hurst, and Ospina (2019) all use US data provided either by Nielsen Retail Scanner Database, by Nielsen Homescan Database, or by IRI Marketing Data Set to compute local (state or metropolitan) price in ation rates and relate them to the local economic condition. Closely related to this study are Lecznar and Smith (2018) and Gilbert (2018) who use the economic approach to construct regional cost of living measures. Gilbert (2018) studies the relationship between local welfare gains from product introduction and local business cycles and Lecznar and Smith (2018) studies the implications of the heterogeneity in regional consumption patters for the aggregate measure of in ation.

The study is organized as follows. Section 3.2 introduces different concepts underlying the computation of in ation. Section 4.3 describes the employed data and provides descriptive statistics. Section 3.4 presents the regional in ation rates computed using the different measures and quanti es the sizes of the "substitution", "variety", and "consumer valuation" biases in in ation computation. Section 4.6 summaries the results and concludes.

3.2 Measures of in ation

Section 3.2 presents various approaches for computing price indexes. Section 3.2.1 describes methodologies used by of cial statistical of ces and in most of the academic literature on the topic. Section 3.2.2 introduces the uni ed price index (UPI) developed in Stephen J. Redding and David E. Weinstein (2020), which allows to estimate in ation rate by controlling for substitution patterns, product turnover, and intertemporal demand shifts. Section 3.2.3 provides with the details on elasticity estimation.

3.2.1 Conventional (non-UPI) price indexes

Statistical of ces of European (and also non-European) countries primarily use a Laspeyres index to compute of cial price index statistics. The formula in Equation (3.1) shows that the Laspeyres price index for a given category c in region r is computed by taking a weighted average over the changes in prices of a set of goods $\Omega_{b,t}^{cr}$ common between base b and current t periods using the initial period expenditure shares as weights.

$$\text{Laspeyres price index formula: } \Phi_{b,t}^{Lcr} = \sum_{k \in \Omega_{b,t}^{cr}} s_{kb}^{*cr} \frac{p_{kt}^{cr}}{p_{kb}^{cr}}, \tag{3.1}$$

where p_{kt}^{cr} and p_{kb}^{cr} are current and base period prices for good $k \in \Omega_{b,t}^{cr}$ and s_{kb}^{*cr}[9] is expenditure share of good $k \in \Omega_{b,t}^{cr}$ in base period.

Closely related to the Laysperes price index is the Paasche price index. Equation (3.2) reveals that the Paasche price index also computes the aggregate measure as a weighted arithmetic average of price changes. However, in contrast to the Laspeyres index, it uses current expenditure weights to aggregate the price relatives.

$$\text{Paasche price index formula: } \Phi_{b,t}^{Pcr} = \left[\sum_{k \in \Omega_{b,t}^{cr}} s_{kt}^{*cr} \left(\frac{p_{kt}^{cr}}{p_{kb}^{cr}} \right)^{-1} \right]^{-1}, \tag{3.2}$$

where s_{kt}^{*cr} is expenditure share of good $k \in \Omega_{b,t}^{cr}$ in current period.

Independently of which weights are used, the assumption of constant expenditure shares makes both Laspeyres and Paasche price indexes prone to the "substitution bias" because neither of the two approaches takes into account that consumers tend to substitute the goods that become more expensive with cheaper alternatives. As a consequence, the Laspeyres price index using base period expenditure shares tends to overstate the in ation rate, while the Paasche price index using nal period expenditure shares tends to underestimate the in ation rate.

The rst step towards a more accurate measure of price changes – taking potential substitution patterns into account – can be done by using geometric rather than arithmetic averages of the price relatives. The Laspeyres and Paasche price indexes implicitly assume the elasticity of substitution between goods being equal to zero, while the geometric Laspeyres and Paasche price indexes (given by Equations (3.3) and (3.4) respectively) allow for substitution between goods and implicitly assume the elasticity of substitution being equal to one.

[9] In the following text I differentiate between expenditure shares computed using the data on all goods observed in current or base period and expenditure shares computed using common goods dataset only. Hence, symbol $*$ here indicates that the dataset underlying the computation of expenditure shares is composed from the common goods only (the goods that are observed in both current and base periods).

$$\text{Laspeyres price index formula, geometric averaging: } \Phi_{b,t}^{GLcr} = \prod_{k \in \Omega_{b,t}^{cr}} \left(\frac{p_{kt}^{cr}}{p_{kb}^{cr}} \right)^{s_{kb}^{*cr}}. \quad (3.3)$$

$$\text{Paasche price index formula, geometric averaging: } \Phi_{b,t}^{GPcr} = \prod_{k \in \Omega_{b,t}^{cr}} \left(\frac{p_{kt}^{cr}}{p_{kb}^{cr}} \right)^{s_{kt}^{*cr}}. \quad (3.4)$$

While allowing for (the more realistic scenario of a) unitary elasticity of substitution between goods, geometric averaging in the Laspeyres and Paasche price indexes still uses expenditure information from a single period only and, therefore, can be improved by combining the information of both periods. Such a combination is realized in the Fisher (given by Equation (3.5)) and the Törnqvist price index (given by Equation (3.6)). Both the Fisher index, which is the geometric average of the arithmetically averaged Laspeyres and Paasche indexes, and the Törnqvist index, which is the geometric average of the geometric Laspeyres and Paasche indexes, correct the upward substitution bias implied by the Laspeyres price index and downward bias implied by the Paasche index. Additionally, both the Fisher and Törnqvist price index are superlative price indexes implying that they represent an index number function form "that is exact for (consistent with) a homothetic preference function that can approximate arbitrary homothetic preferences" (Diewert, 1998). The Fisher price index is an exact approximation to the cost of living if consumers preferences are described by a homogeneous quadratic utility function. The Törnqvist price index is an exact approximation to the cost of living if consumers preferences are given by translog indirect utility function. Both provide a measure of the in ation rate that is free from the "substitution bias".

$$\text{Fisher price index formula: } \Phi_{b,t}^{Fcr} = \sqrt{\Phi_{b,t}^{Lcr} \Phi_{b,t}^{Pcr}}. \quad (3.5)$$

$$\text{Törnqvist price index formula: } \Phi_{b,t}^{Tcr} = \prod_{k \in \Omega_{b,t}^{cr}} \left(\frac{p_{kt}^{cr}}{p_{kb}^{cr}} \right)^{\frac{1}{2}(s_{kb}^{*cr} + s_{kt}^{*cr})} \quad (3.6)$$

The price indexes presented in this subsection differ in how they handle substitution between goods, but all of them are computed for a basket of continued goods and, therefore, do not take product turnover into account. Moreover, all of them additionally assume that consumers value goods equally over time and, consequently, they do not respond to the demand shocks of the unique goods. A more accurate measure of in ation rate that incorporates product substitution, product turnover, and intertemporal demand shocks can be obtained employing the Redding-Weinstein CES uni ed price index, which is presented in the next subsection.

3.2.2 The Redding-Weinstein CES uni ed price index (UPI)

The CES uni ed price index developed in Stephen J. Redding and David E. Weinstein (2020) is a cost-of-living price index, which is based on CES preferences and which takes into account the exit/entry of goods and taste shocks in addition to substitution behavior. For brevity, I only list the major equations here, while the extended derivations are available in Stephen J. Redding and David E. Weinstein (2020).

I assume that in each region r consumers have homothetic CES preferences and derive utility in time period t from consuming goods $k \in \Omega_t^{cr}$ of category c , where Ω_t^{cr} indicates the set of goods available in the category in region r and period t and its cardinality is denoted by $N_t^{cr} = |\Omega_t^{cr}|$. The corresponding utility function is given by the following equation:

$$U_t^{cr} = \left[\sum_{k \in \Omega_t^{cr}} (\varphi_{kt}^{cr} q_{kt}^{cr})^{\frac{\sigma^{cr}-1}{\sigma^{cr}}} \right]^{\frac{\sigma^{cr}}{\sigma^{cr}-1}}, \tag{3.7}$$

where q_{kt}^{cr} denotes consumption of good k, $\varphi_{kt}^{cr} > 0$ is an unobserved demand parameter associated with good k and σ^{cr} is the elasticity of substitution between goods in category c in region r, which does not vary over time and is assumed to be larger than 1. The last assumption assures that consumers exhibit "love-for-variety" preferences implying that utility is increasing when the number of varieties available in a category raises.

The maximisation of the utility function given by Equation (3.7) over N_t^{cr} goods consumption within the category yields the following unit expenditure function:

$$P_t^{cr} = \left[\sum_{k \in \Omega_t^{cr}} \left(\frac{p_{kt}^{cr}}{\varphi_{kt}^{cr}} \right)^{1-\sigma^{cr}} \right]^{\frac{1}{1-\sigma^{cr}}}, \tag{3.8}$$

where p_{kt}^{cr} is the price of good k in region r at time period t. Applying Shepard s lemma to Equation (3.8), the following demand system can be derived:

$$s_{kt}^{cr} = \frac{p_{kt}^{cr} q_{kt}^{cr}}{\sum_{k \in \Omega_t^{cr}} p_{kt}^{cr} q_{kt}^{cr}} = \frac{(p_{kt}^{cr} / \varphi_{kt}^{cr})^{1-\sigma^{cr}}}{\sum_{k \in \Omega_t^{cr}} (p_{kt}^{cr} / \varphi_{kt}^{cr})^{1-\sigma^{cr}}} = \frac{(p_{kt}^{cr} / \varphi_{kt}^{cr})^{1-\sigma^{cr}}}{(P_t^{cr})^{1-\sigma^{cr}}} \tag{3.9}$$

The unit expenditure function in Equation (3.8) presents the price of obtaining one unit of utility from consumption of goods Ω_t^{cr} at time period t. Therefore, the price index capturing how the price of one unit of utility changes from base period b to period t can be derived as follows:

$$\Phi_{b,t}^{UPIcr} = \frac{P_t^{cr}}{P_b^{cr}} = \left[\frac{\sum_{k \in \Omega_t^{cr}} (p_{kt}^{cr} / \varphi_{kt}^{cr})^{1-\sigma^{cr}}}{\sum_{k \in \Omega_b^{cr}} (p_{kb}^{cr} / \varphi_{kb}^{cr})^{1-\sigma^{cr}}} \right]^{\frac{1}{1-\sigma^{cr}}}. \tag{3.10}$$

The nominator and the denominator of Equation (3.10) include different sets of varieties. To deal with changes in the product sets available in the category in periods t and b Stephen J. Redding and David E. Weinstein (2020) separate Equation (3.10) on price index for the set of common goods (available in both periods) denoted by $\Omega_{b,t}^{cr}$ and the variety adjustment term, which controls for changes in the goods set. The variety adjustment is done by introducing the expenditure share of common goods in total expenditures in time period t:

$$\lambda_{t,b}^{cr} = \frac{\sum_{k \in \Omega_{b,t}^{cr}} (p_{kt}^{cr} / \varphi_{kt}^{cr})^{1-\sigma^{cr}}}{\sum_{k \in \Omega_t^{cr}} (p_{kt}^{cr} / \varphi_{kt}^{cr})^{1-\sigma^{cr}}} = \frac{\sum_{k \in \Omega_{b,t}^{cr}} p_{kt}^{cr} q_{kt}^{cr}}{\sum_{k \in \Omega_t^{cr}} p_{kt}^{cr} q_{kt}^{cr}} \tag{3.11}$$

and in base period b:

$$\lambda_{b,t}^{cr} = \frac{\sum_{k \in \Omega_{b,t}^{cr}} (p_{kb}^{cr} / \varphi_{kb}^{cr})^{1-\sigma^{cr}}}{\sum_{k \in \Omega_b^{cr}} (p_{kb}^{cr} / \varphi_{kb}^{cr})^{1-\sigma^{cr}}} = \frac{\sum_{k \in \Omega_{b,t}^{cr}} p_{kb}^{cr} q_{kb}^{cr}}{\sum_{k \in \Omega_b^{cr}} p_{kb}^{cr} q_{kb}^{cr}} \tag{3.12}$$

in the equation Equation (3.10):

$$\Phi_{b,t}^{UPIcr} = \frac{P_t^{cr}}{P_b^{cr}} = \left(\frac{\lambda_{b,t}^{cr}}{\lambda_{t,b}^{cr}} \right)^{\frac{1}{1-\sigma^{cr}}} \left[\frac{\sum_{k \in \Omega_{b,t}^{cr}} (p_{kt}^{cr} / \varphi_{kt}^{cr})^{1-\sigma^{cr}}}{\sum_{k \in \Omega_{b,t}^{cr}} (p_{kb}^{cr} / \varphi_{kb}^{cr})^{1-\sigma^{cr}}} \right]^{\frac{1}{1-\sigma^{cr}}}. \tag{3.13}$$

Using the demand system in Equation (3.9) and additionally assuming that geometric mean of log demand changes is equal to zero $\frac{1}{N^{cr}_{t,b}} \sum_{k \in \Omega^{cr}_{b,t}} \Delta ln(\varphi^{cr}_{kt}) = 0$,[10] the uni ed price index can be expressed just in terms of observed prices and expenditure shares and the elasticity of substitution:

$$\Phi^{UPIcr}_{b,t} = \frac{P^{cr}_t}{P^{cr}_b} = \underbrace{\left(\frac{\lambda^{cr}_{t,b}}{\lambda^{cr}_{b,t}}\right)^{\frac{1}{\sigma^{cr}-1}}}_{\text{Variety adjustment}} \underbrace{\left[\frac{\tilde{p}^{*cr}_t}{\tilde{p}^{*cr}_b}\left(\frac{\tilde{s}^{*cr}_t}{\tilde{s}^{*cr}_b}\right)^{\frac{1}{\sigma^{cr}-1}}\right]}_{\Phi^{UPI\_CGcr}_{b,t}}, \tag{3.14}$$

where $\Phi^{UPI\_CGcr}_{b,t}$ is the common-goods UPI and $\tilde{x}^*$ denotes the geometric average (indicated by a tilde) of the corresponding variable computed using the common goods dataset (indicated by an asterisk).

The rst term of Equation (3.14) represents the variety adjustment component documenting how much consumers value new varieties relatively to exiting ones. A ratio of less than one implies that new varieties have lower demand adjusted prices and therefore higher market share than exiting varieties. In this case, the price index will decrease because new varieties are preferred by consumers and allow them to receive a level of utility at a given unit expenditure cost. The effect is more pronounced if the demand in the category is less elastic.

The second term of Equation (3.14), which is the common-goods UPI ($\Phi^{UPI\_CGcr}_{b,t}$), presents the effect from the changes in demand adjusted prices of the continuing goods on the cost-of-living. It is composed of two terms. The rst term, given by $\frac{\tilde{p}^{*cr}_t}{\tilde{p}^{*cr}_b}$, represents the geometric average of price relatives for the common goods and corresponds to the Jevons price index, which increases when prices rise. The second term, $\frac{\tilde{s}^{*cr}_t}{\tilde{s}^{*cr}_b}$, is the ratio of geometric averages of expenditures shares for the common goods and captures the heterogeneity in common goods shares. A ratio of smaller than one implies that expenditure shares become more dispersed across goods, which happens when consumers react to demand shocks and rebalance their consumption in favour of the currently more appealing goods. Consumers value that they can redistribute its consumption so their utility additionally grows and the cost of living decreases.

[10] This assumption implies that consumers do not prefer all of the goods more in the current period than in the last period, but rather shifts their preferences away from some goods in favour of others. Additionally, this assumption implies that the geometric mean of demand shocks is constant over time such that $\tilde{\varphi}^{cr}_t = \tilde{\varphi}^{cr}$ holds for any time period.

The UPI is derived from the CES preferences assuming that both product turnover and intertemporal demand shocks occur. If one assumes no product turnover, the UPI collapses to the common-goods UPI, and, therefore, the difference between the UPI and the common-goods UPI can be treated as the "variety bias". Furthermore, if one additionally assumes that the intertemporal demand shocks are zero, the common-goods UPI collapses to the Sato-Vartia price index[11] given by Equation (3.15), which is exact for CES preferences. Therefore, the difference between the common-goods UPI and the Sato-Vartia price index can be treated as the "consumer valuation bias".

$$\Phi_{b,t}^{SVcr} = \prod_{k \in \Omega_{b,t}^{cr}} \left(\frac{p_{kt}^{cr}}{p_{kb}^{cr}} \right)^{w_{kt}^{cr}}, \tag{3.15}$$

where

$$w_{kt}^{cr} \equiv \frac{\frac{s_{kt}^{*cr} - s_{kb}^{*cr}}{\ln s_{kt}^{*cr} - \ln s_{kb}^{*cr}}}{\sum_{k \in \Omega_{b,t}^{cr}} \frac{s_{kt}^{*cr} - s_{kb}^{*cr}}{\ln s_{kt}^{*cr} - \ln s_{kb}^{*cr}}}. \tag{3.16}$$

Computation of the UPI requires the estimation of an elasticity of substitution σ^{cr} for each product category in each region. Given that some categories in the regional data do not exhibit a suf cient number of observations to obtain a reliable parameter estimate for this variable, I assume that for each category, the parameter σ^{cr} is the same across all regions and equal to σ^{c}, which is then estimated at the country level. To estimate the elasticities of substitution I employ the estimator developed by Feenstra (1994) and introduced in the next subsection.

3.2.3 Elasticity estimation

In this section I present the formal derivation of the estimator for the parameter σ^{c}, closely following Feenstra (1994). The demand equation for each good can be obtained directly from Equation (3.9). Taking logarithms and differencing over time yields:

$$\Delta \ln s_{kt}^{*c} = \phi_t^c - (\sigma^c - 1)\Delta \ln p_{kt}^c + \varepsilon_{kt}^c, \tag{3.17}$$

where $\phi_t^c = (\sigma^c - 1)\Delta \ln P_t^{*c}$ is the intercept and $\varepsilon_{kt}^c = (\sigma - 1)\Delta \ln \varphi_{kt}^c$ captures good-speci c tastes shocks to demand. The supply curve of the product category c is assumed

[11] See Stephen J. Redding and David E. Weinstein (2020) for the derivation.

to be:

$$\Delta \ln p_{kt}^c = w^c \Delta \ln q_{kt}^c + \xi_{kt}^c, \tag{3.18}$$

where $w^c \geq 0$ is the inverse supply elasticity and ξ_{kt}^c is the random supply error. Using the de nition of the expenditure share in Equation (3.9), the quantity term in Equation (3.18) can be replaced with the expenditure share:

$$\Delta \ln p_{kt}^c = w^c (\Delta \ln s_{kt}^{*c} - \Delta \ln p_{kt}^c + \Delta \ln E_t^c) + \xi_{kt}^c, \tag{3.19}$$

where $\ln E_t^{*c} = \ln \sum_{k \in \Omega_{b,t}^{cr}} p_{kt}^{cr} q_{kt}^{cr}$ is the total expenditures for common goods. Additionally, the log-change in the expenditure share is substituted with Equation (3.17):

$$\Delta \ln p_{kt}^c = w^c (\phi_t^c - (\sigma^c - 1)\Delta \ln p_{kt}^c + \varepsilon_{kt}^c - \Delta \ln p_{kt}^c + \Delta \ln E_t) + \xi_{kt}^c. \tag{3.20}$$

Rearranging the terms in Equation (3.20) the "reduced form" supply curve can be presented as follows:

$$\Delta \ln p_{kt}^c = \psi_{kt}^c + \frac{\rho^c \varepsilon_{kt}^c}{\sigma^c - 1} + \delta_{kt}^c, \tag{3.21}$$

where $\psi_{kt}^c = \frac{w^c(\phi_t^c + \Delta \ln E_t^c)}{1 + w^c \sigma^c}$, $\rho^c = \frac{w^c(\sigma^c - 1)}{1 + w^c \sigma^c}$, and $\delta_{kt}^c = \frac{\xi_{kt}^c}{1 + w^c \sigma^c}$. From Equation (3.21) it can be seen that the demand error ε_{kt}^c is correlated with the price creating the simultaneity problem in Equation (3.17): in the case of an exogenous positive shock to the consumer tastes for a particular good its price rises along the supply curve. Feenstra (1994) suggests that the endogeneity problem can be avoided by exploiting the panel nature of the data. The identifying assumption is that the errors terms ε_{kt}^c and δ_{kt}^c are independent with zero mean and variances $\nu_{\varepsilon_k^c}$ and $\nu_{\delta_k^c}$. According to Leamer (1981) this assumption assures that the possible set of maximum likelihood estimates of supply and demand elasticities lie on a hyperbola de ned by the second moments of the data. Feenstra (1994) argues that using this insight in the panel data provides an opportunity to identify the parameters. The main idea is that as long as idiosyncratic supply and demand shocks do not come from the same distribution there is a unique hyperbola for each good and the intersection of these hyperbolas provides with the estimates of supply and demand elasticities[12].

To be able to use the identi cation assumption the log-changes in expenditure shares

[12] See Feenstra (1994) and Broda and D. Weinstein (2010) for a more comprehensive discussion.

given by Equation (3.17) and the log-changes in prices given by Equation (3.21) must be purged of time-speci c category shocks. Following Feenstra (1994) log-changes in expenditure shares and prices are differenced with respect to the selected reference good R to remove ϕ_t^c and ψ_{kt}^c:

$$\tilde{\varepsilon}_{kt}^c = \varepsilon_{kt}^c - \varepsilon_{kt}^{cR} = (\Delta \ln s_{kt}^{*c} - \Delta \ln s_{kt}^{*cR}) + (\sigma^c - 1)(\Delta \ln p_{kt}^c - \Delta \ln p_{kt}^{cR}), \tag{3.22}$$

$$\tilde{\delta}_{kt}^c = \delta_{kt}^c - \delta_{kt}^{cR} = (\Delta \ln p_{kt}^c - \Delta \ln p_{kt}^{cR}) - \frac{\rho^c \tilde{\varepsilon}_{kt}^c}{\sigma^c - 1} \tag{3.23}$$

$$= (1 - \rho^c)(\Delta \ln p_{kt}^c - \Delta \ln p_{kt}^{cR}) - \left(\frac{\rho^c}{\sigma^c - 1}\right)(\Delta \ln s_{kt}^{*c} - \Delta \ln s_{kt}^{*cR}). \tag{3.24}$$

Multiplying Equation (3.22) with Equation (3.23) one arrives at the following relationship:

$$Y_{kt}^c = \theta_1^c X_{1kt}^c + \theta_2^c X_{2kt}^c + u_{kt}^c, \tag{3.25}$$

where

$$Y_{kt}^c = (\Delta \ln p_{kt}^c - \Delta \ln p_{kt}^{cR})^2, \tag{3.26}$$

$$X_{1kt}^c = (\Delta \ln s_{kt}^{*c} - \Delta \ln s_{kt}^{*cR})^2, \tag{3.27}$$

$$X_{2kt}^c = (\Delta \ln p_{kt}^c - \Delta \ln p_{kt}^{cR})(\Delta \ln s_{kt}^{*c} - \Delta \ln s_{kt}^{*cR}), \tag{3.28}$$

$$\theta_1^c = \frac{\rho^c}{(\sigma^c - 1)^2(1 - \rho^c)}, \theta_2^c = \frac{2\rho^c - 1}{(\sigma^c - 1)(1 - \rho^c)}, \tag{3.29}$$

and

$$u_{kt}^c = \frac{\tilde{\varepsilon}_{kt}^c \tilde{\delta}_{kt}^c}{(\sigma^c - 1)(1 - \rho^c)}. \tag{3.30}$$

Averaging over time Equations (3.25) to (3.28):

$$\bar{Y}_k^c = \theta_1^c \bar{X}_{1k}^c + \theta_2^c \bar{X}_{2k}^c + \bar{u}_k^c. \tag{3.31}$$

Equation (3.31) presents second moments of the data and $\bar{u}_k^c$ vanishes in its probability limit as $t \to \infty$. Hence, OLS (Ordinary Least Squares) is a consistent estimator of θ_1^c and θ_2^c when there is some heteroskedasticity in supply and demand shocks across goods assuring that $\bar{X}_{1k}^c$ and $\bar{X}_{2k}^c$ are not co-linear. Given the heteroskedasticity, OLS provides with consistent but not ef cient estimates. To adjust for heteroskedasticity and obtain an

ef cient estimate I employ WLS (Weighted Least Squares) using the weighting scheme developed in Broda and D. Weinstein (2006) to estimate Equation (3.31). Using the relationship given in Equation (3.29) the estimates of θ_1^c and θ_2^c allow to solve for the demand elasticity. In the case when $\hat{\theta}_1^c > 0$ and $\hat{\theta}_2^c > 0$:

$$\hat{\rho}^c = \frac{1}{2} + \left(\frac{1}{4} - \frac{1}{4 + (\hat{\theta}_1^c)^2 / \hat{\theta}_1^c} \right)^{\frac{1}{2}}. \tag{3.32}$$

In the case when $\hat{\theta}_1^c > 0$ and $\hat{\theta}_2^c < 0$:

$$\hat{\rho}^c = \frac{1}{2} - \left(\frac{1}{4} - \frac{1}{4 + (\hat{\theta}_1^c)^2 / \hat{\theta}_1^c} \right)^{\frac{1}{2}}. \tag{3.33}$$

The estimate of the demand elasticity in both cases is given by:

$$\hat{\sigma}^c = 1 + \left(\frac{2\hat{\rho}^c - 1}{1 - \hat{\rho}^c} \right) \frac{1}{\hat{\theta}_2^c}. \tag{3.34}$$

3.3 Data and descriptive statistics

The computation of the price indexes in this study is undertaken by using household panel data for Germany over the period 2009 – 2016 and for Spain over the period 2008 – 2016. The data are provided by Aimark.[13] in Germany, the household panel is operated by GfK and in Spain, it is operated by Kantar. The data contain detailed information about purchases of FMCG (Fast Moving Consumer Goods) made by a representative group of households in each country. More speci cally, households report the date of every purchase they make, the name of the retailer, quantity, volume, and price paid for every product bought. A product is identi ed by its GTIN (Global Trade Item Number) or by a product identi cation number assigned to it according to the organisations standards[14]. Additionally, the data provide information on product characteristics, such as

[13] Aimark is a non-pro t foundation that supports academic research by providing access to various price data on behalf of its data partners (GfK, Kantar, and others). More information about the organisation can be obtained on the Aimark s website: https://aimark.net/.

[14] Both GfK and Kantar develop vocabularies with detailed product characteristics that allow households to precisely identify goods that do not have a GTIN.

brand, private label identi er, product volume, product category, manufacturer, households socio-economic characteristics, and households areas of residence identi ed by postal codes.

TABLE 3.1: Summary statistics on the dataset

Germany				
	Mean	SD	Min	Max
Purchases	34,099,573	973,709	32,943,516	35,881,104
Households	36,264	1,357	33,772	38,010
Products	242,634	6,926	232,982	252,594
Observations	8			
Spain				
	Mean	SD	Min	Max
Purchases	10,661,336	1,858,554	7,434,131	12,717,956
Households	12,671	1,746	9,682	14,310
Products	158,855	11,166	140,719	174,530
Observations	9			

Notes: Table 3.1 reports descriptive statistics on the number of purchases, the number of unique households and products observed yearly in the dataset. All values of the descriptive statistics are rounded to the closest integer.

Table 3.1 presents descriptive statistics on the raw dataset. In Germany, the panel tracks on average around 36 thousands households yearly who make more than 34 millions of purchases buying on average around 242 thousands unique products. In Spain, the yearly panel is on average composed from around 12.5 thousands of households who buy around 159 thousands unique products and make in total 10.5 million of purchases.

To identify regions in the data I use the TERCET NUTS-postal codes matching tables provided by Eurostat containing the matching between postal codes and NUTS regions and apply this matching to the postal codes reported by households. Before starting with the analysis I restrict the sample to the observations where product category, region, price, and volume bought are available and the last two variables are larger than zero.

In choosing the frequency of the calculated price indexes and consequently product price aggregation, I follow Stephen J. Redding and David E. Weinstein (2020) and perform the analysis for year-on-year changes in quarterly prices. Stephen J. Redding and David E. Weinstein (2020) argue that higher frequency of aggregation such as daily or weekly might disconnect the events of purchase and consumption of the good, while

quarterly aggregation should mitigate this concern. Before aggregating the prices I again follow Stephen J. Redding and David E. Weinstein (2020) and clean the data from potential misreporting by removing observations where the reported price per unit of a good is 3 times higher or lower than the average unit price paid of the corresponding good in the corresponding quarter and observations where the purchased volume is 25 times higher than the average volume purchased.

Tables A.1 and A.2 of Section 4.A present quarterly mean values of the number of purchases, the number of households observed, the number of categories and the number of unique products per region. It is noticeable that samples are somewhat heterogeneous in size across regions, but the differences in samples re ect the differences in population across the regions presented in the column "Popul.". I further restrict the sample to the categories that are available in all regions in all quarters to prevent the results being driven by differences in product categories composition across time and space. Tables A.1 and A.2 of Section 4.A show that the sample restricted to the common categories captures on average 86% of regional expenditures in Spain and 97% of regional expenditures in Germany, so the selected categories remain representative of the sample.

Next, to obtain the regional quarterly price per good I average all observed prices of the good over all households living in the region in the quarter using total expenditures as weights. At last, I compute 4-quarter changes in logarithms of quarterly prices and shares and drop the observations where these changes are in 1% or 99% percentiles of their distributions. Figure 3.1 shows that the remaining data can closely replicate of cial country food in ation rate with the coef cient of correlation between of cial in ation rate and scanner data in ation rate being equal to 0.79 for Germany and 0.88 for Spain[15]. Therefore, the data can provide useful insights applicable to the of cial statistics.

Table 3.2 reports descriptive statistics on all region-category combinations observed in the common categories quarterly sample. The median number of observations per category in a region is 3,404 in Germany and 2,831 in Spain with categories in 5th percentile having only 589 observations in Germany and 627 in Spain and categories in 95th

[15]In Spain food in ation rate is also provided at the NUTS-2 regional level by the Spanish Statistical Office. Figure 3.B.1 of Section 4.B plots 4-quarter percentage changes in "Food and non-alcoholic beverages" component of the regional of cial price index and in scanner data price index. It can be seen from the gure that scanner data in ation rate tracks the changes in the of cial regional food in ation rate very closely.

FIGURE 3.1: HICP food in ation rate and scanner price data in ation rate

(A) Germany (B) Spain

Notes: Figure 3.1 plots 4-quarter percentage changes in the of cial food price index and scanner data price index. Scanner data in ation rates (denoted as in _L) are computed using 4-quarter percentage changes in the Laspeyres price index. National food in ation rates (denoted as in _food) are computed using 12-months percentage changes in "Food including alcohol and tobacco" component of the HICP and aggregating them to the quarterly level by taking a mean over months in a quarter.

percentile having 178,022 observations in Germany and 57,196 in Spain [16]. The median number of unique goods observed in a category in a region in Germany is 411 (with 59 goods being observed in the categories in 5th and 15,646 in the 95th percentile) and 307 in Spain (with 75 goods being observed in the categories in 5th and 6,042 in the 95th percentile).

Common goods share in current and past expenditures (denoted by $\lambda_{t,t-4}$ and $\lambda_{t-4,t}$ respectively) have a similar distribution in both countries being on average 0.65 and 0.67 in Germany and 0.66 and 0.68 in Spain, which indicates quite substantial product turnover in the data. More speci cally, average product entry and exit rates are close to 48% in both countries, which implies that 48% of the products were not available in a category four quarters ago. Interestingly, given that the average value for $\frac{\lambda_{t,t-4}}{\lambda_{t-4,t}}$ is equal to 1 in Germany and 1.01 in Spain on average new products are as attractive to the consumers as the old ones.

The last four rows of Table 3.2 show mean and standard deviation of 4-quarter log

[16] These numbers are much smaller than those reported in Broda and D. Weinstein (2010), which is due to the fact that their dataset is covering a longer period (the authors observe data for 10 years) and the categories in the dataset are more broadly de ned.

TABLE 3.2: Descriptive statistics on product categories

Germany	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Reg. Cat. Sample Size	7,040	12,560	149	589	1,512	3,405	7,637	22,381	178,022
Reg. Cat. Number of Goods	845	1,431	13	59	180	411	951	2,746	15,646
Reg. Cat. Mean $\lambda_{t,t-4}$	0.65	0.19	0.04	0.29	0.55	0.69	0.80	0.89	0.99
Reg. Cat. Mean $\lambda_{t-4,t}$	0.67	0.19	0.05	0.29	0.56	0.71	0.81	0.90	0.99
Reg. Cat. Mean $\frac{\lambda_{t,t-1}}{\lambda_{t-4,t}}$	1.00	0.08	0.70	0.93	0.97	0.99	1.01	1.12	2.36
Reg. Cat. Mean Entry Rate	48.17	14.09	10.86	27.76	38.58	46.41	55.94	74.49	94.33
Reg. Cat. Mean Exit Rate	48.10	14.15	11.72	27.42	38.34	46.53	55.90	74.50	94.14
Reg. Cat. Mean $\Delta_4 \ln p_{kt}$	0.00	0.02	-0.18	-0.03	-0.01	0.00	0.01	0.03	0.15
Reg. Cat. Mean $\Delta_4 \ln s_{kt}$	-0.02	0.05	-0.49	-0.09	-0.04	-0.02	0.00	0.05	0.36
Reg. Cat. SD $\Delta_4 \ln p_{kt}$	0.12	0.03	0.01	0.08	0.10	0.11	0.13	0.17	0.26
Reg. Cat. SD $\Delta_4 \ln s_{kt}$	0.81	0.09	0.45	0.68	0.76	0.81	0.86	0.96	1.33
Observations	9804								
Spain	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Reg. Cat. Sample Size	4,128	4,347	206	627	1,556	2,831	5,146	12,028	57,196
Reg. Cat. Number of Goods	437	443	21	75	175	307	548	1,179	6,042
Reg. Cat. Mean $\lambda_{t,t-4}$	0.66	0.15	0.13	0.37	0.57	0.69	0.78	0.86	0.93
Reg. Cat. Mean $\lambda_{t-4,t}$	0.68	0.15	0.13	0.38	0.58	0.71	0.80	0.88	0.96
Reg. Cat. Mean $\frac{\lambda_{t,t-4}}{\lambda_{t-4,t}}$	1.01	0.08	0.83	0.94	0.97	1.00	1.03	1.15	2.00
Reg. Cat. Mean Entry Rate	48.42	10.80	21.29	32.25	40.62	47.35	55.23	68.39	84.17
Reg. Cat. Mean Exit Rate	47.49	11.05	21.15	31.27	39.51	46.41	54.43	67.88	84.93
Reg. Cat. Mean $\Delta_4 \ln p_{kt}$	0.01	0.01	-0.04	-0.01	-0.00	0.00	0.01	0.02	0.05
Reg. Cat. Mean $\Delta_4 \ln s_{kt}$	-0.03	0.04	-0.33	-0.09	-0.05	-0.03	-0.00	0.04	0.37
Reg. Cat. SD $\Delta_4 \ln p_{kt}$	0.11	0.02	0.04	0.09	0.10	0.11	0.12	0.14	0.19
Reg. Cat. SD $\Delta_4 \ln s_{kt}$	0.83	0.07	0.60	0.72	0.79	0.83	0.88	0.95	1.22
Observations	3383								

Notes: Table 3.2 reports descriptive statistics across all region-category combinations of the quarterly aggregated dataset. "Reg. Cat. Sample Size" presents the number of total observations (here, one observation presents variables for a good in a quarter) available for a category in a region. "Reg. Cat. Unique Goods" presents the number of unique goods observed in a category in a region over the whole period. All values of the descriptive statistics for the rst two variables are rounded to the closest integer. The next ve variables denoted by "Reg. Cat. Mean" presents mean values of respective variables aggregated over all quarters for each region-category combination. Last four variables "Reg. Cat. Mean" and "Reg. Cat. SD" of $\Delta_4 \ln p_{kt}$ and $\Delta_4 \ln s_{kt}$ present mean values and standard deviations of 4-quarter price and share changes observed in a category in a region computed by pooling all quarters and goods together and estimating one parameter per a region-category combination.

price changes ($\Delta_4 \ln p_{kt}$, where k denotes good and t quarter) and 4-quarter log expenditure share changes ($\Delta_4 \ln s_{kt}$) in a category in a region, which indicate that although changes in expenditure share are small on average (being equal to -0.02 in Germany and -0.03 in Spain) they are much more volatile than changes in prices.

3.4 Results

This section presents results of the study including the ndings from employing all approaches discussed in Section 3.2 and an assessment of the substitution, variety and consumer valuation biases. Section 3.4.1 reports the results from the elasticity estimation while Section 3.4.2 presents regional in ation rate gures. Section 3.4.2 contains the ndings for the sizes of the "substitution", "variety", and "consumer valuation" biases.

3.4.1 Elasticities of substitution

The computation of regional in ation rates using the Redding-Weinstein uni ed price index requires an estimate of the elasticity of substitution for each product category presented in the data. As outlined in Section 3.2 I am using the Feenstra elasticity estimator to obtain values for these parameters. Table 3.3 provides summary statistics of the estimated elasticities obtained for the different goods categories.

TABLE 3.3: Estimated elasticities

Germany	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Elast. FN	10.71	5.07	2.86	5.34	8.05	9.77	11.72	18.06	51.27
Observations	258								
Spain	Mean	SD	Min	p5	p25	p50	p75	p95	Max
Elast. FN	9.05	2.29	4.67	6.18	7.50	8.56	10.09	13.57	18.27
Observations	199								

Notes: Table 3.3 reports summary statistics for elasticities estimated for each category using the Feenstra estimator.

Elasticities are on average larger in Germany, with median of 9.77, than in Spain where the median amounts to 8.56. The resulting values are somewhat higher than usually reported but very similar to the values found in the literature using the most comparable data: Stephen J. Redding and David E. Weinstein (2020) reports a median value of estimates being equal to 6.48 and Broda and D. Weinstein (2010) nds that the median within brand module elasticity is 11.5.

Table 3.3 also reveal that there exists considerable heterogeneity in substitutions elasticities across categories with differences being distinctly more pronounced in Germany than in Spain. For German categories values between the 5th and 95th percentile range from around 5 to 18, whereas for Spain the comparable numbers are between 6 and 13.5.

3.4.2 Regional in ation rates

I compute price indexes for each category in a region and aggregate the price indexes over the categories to a regional price index using initial period expenditure shares as weights.[17] Regional in ation rates are then computed by taking logarithms of the price indexes, computing year-on-year changes, and multiplying the obtained numbers by 100. The resulting in ation rates for Germany and Spain are presented at Figures 3.B.2 and 3.B.3 of Section 4.B. It can be seen from the gures that in both countries conventional in ation rates lie close to each other, while the UPI in ation rates exhibit somewhat different dynamics. In Spain, it falls much more than conventional price indexes during the Great Recession and rises above the conventional price indexes from Q1 2016. In Germany, the UPI in ation rate lies generally below conventional rates, except for the period from Q1 2012 to Q1 2013, where the UPI in ation rate lies above.

Tables A.3 and A.4 of Section 4.A present summary statistics on regional in ation rates in each quarter of the sample time period. In both countries, there exist quite substantial regional in ation differentials as shown by the standard deviation and the spread between maximum and minimum of the observed regional in ation rates. In Germany, the standard deviation computed across regions ranges from 0.15 to 0.47 depending on the measure in ation and the time period. The standard deviation of the UPI in ation rates takes the highest values across all measures of in ation in almost every period (except 2016 (Q2)) and ranges from 0.24 to 0.47. The standard deviation of the conventional in ation rates indicates a somewhat smaller dispersion across regions and ranges from 0.15 to 0.38. The spread between maximum and minimum of regional in ation rates is around 1% on average for the conventional in ation rate with the observed minimum at 0.66% (in _L) and maximum at 1.98% (in _GP). The spread for the UPI in ation rates reaches 1.59% on average, 0.85% at the minimum, and 2.09% at the maximum. In Spain, the standard deviation indicates a higher dispersion across regions in comparison with Germany ranging from 0.18 - 0.22 (at the minimum) to 0.72 - 0.88 (at the maximum) for the conventional in ation rates and from 0.35 to 0.98 for the UPI in ation rates. The spread in Spain is also higher comparing to Germany and takes values between 1.45% to

[17] Aggregation using current expenditure share and average between initial and current expenditure share lead to very similar results

1.72% on average for the conventional in ation rates and 2.27% on average for the UPI in ation rates. The range of observed spreads is quite similar across all measures of in- ation and lies between 0.71% - 0.87% (at minimum) and 3.38% - 4.30% (at maximum) depending on the measure of in ation. In Spain, both measures of dispersion indicate an increase in the regional dispersion of the conventional in ation rates at the beginning of 2009, during the period from 2012 (Q3) to 2013 (Q3), and in 2014 (Q4). The regional dispersion in the UPI in ation rates increases in the aforementioned periods as well but also develops a few additional spikes (2010 (Q2), 2012(Q1), and 2015 (Q3), which are not observed in the conventional in ation rates.

Table 3.4 present a numerical comparison between the Laysperes and the UPI in ation rates. On average the Laypseres in ation rate overestimates the in ation rate given by UPI by 0.88% in Germany and 0.46% in Spain. The between regions standard deviation is two times smaller in Germany than in Spain (0.12% versus 0.24%), while the within regions standard deviation is practically the same in both countries (0.74% in Germany and 0.75% in Spain).

TABLE 3.4: Differences between the Laysperes and the uni ed price index in ation rates

Germany	Mean	SD	SD_b	SD_w	Min	p5	p25	p50	p75	p95	Max
Diff.	0.88	0.75	0.12	0.74	-1.40	-0.31	0.43	0.85	1.28	2.21	3.80
Observations	1064		38	28							
Spain	Mean	SD	SD_b	SD_w	Min	p5	p25	p50	p75	p95	Max
Diff.	0.46	0.78	0.24	0.75	-1.84	-0.83	-0.02	0.43	0.96	1.85	3.04
Observations	544		17	32							

Notes: Table 3.4 reports summary statistics for differences between the Laysperes and the UPI in ation rates computed over regions and time periods. SD shows the overall standard deviation in the sample from the overall mean, SD_b shows the standard deviation between regions (SD_b is computed over regional mean values), SD_w shows the standard deviation within regions while excluding the between variance from the computation.

In general, the differences between in ation rates computed using different price indexes can arise from three reasons: differences in capturing substitution patterns, differences in assumptions about variety adjustment and differences in assumption about demand shocks. The conventional price indexes computed for a basket of common goods do not account for product turnover and assume no intertemporal shifts in demand for

unique goods. Therefore, these conventional price indexes only differ by their assumption about substitution between goods, while the UPI additionally differs from the conventional measures by allowing for product turnover and intertemporal demand shifts.

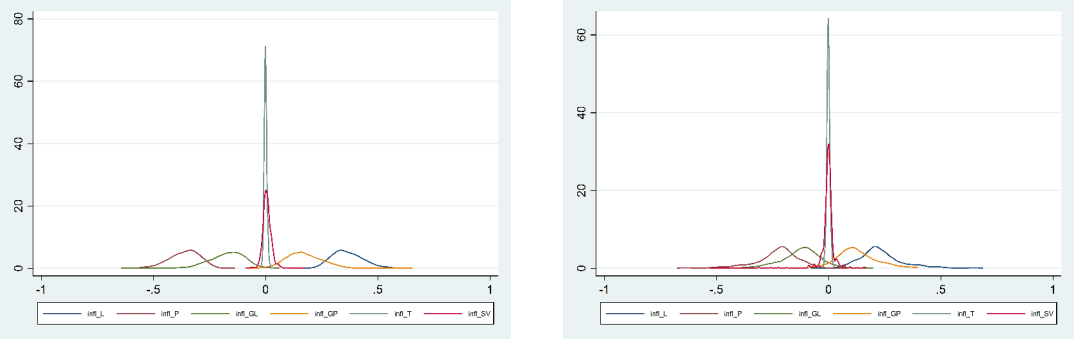
To asses how different assumptions about substitution affect the in ation measures I compare the differences between Laysperes, Paasche, geometric Laysperes, geometric Paasche, Törnqvist and Fisher in ation rates. To be able to understand how the assumption of CES preferences affects the difference between conventional price indexes and the UPI I use the Sato-Vartia price index, which can be derived from the UPI under the assumptions of no product turnover and zero intertemporal demand shifts.

Figure 3.3 presents kernel densities of differences between aforementioned in ation rates and the Fisher in ation rates over all regions and time periods. In both countries conventional in ation rates exhibit the expected behaviour with the Laysperes in- ation rate differences distribution shifted to the right of the origin, con rming that the Laysperes price index tends to overestimate price changes, and with the Paasche in ation rate differences distribution shifted to the left of the origin, con rming that the Paasche price index tends to underestimate price changes. The geometric Laysperis and Paasche in ation rates differences lie closer to the origin but as was mentioned before they still inherit some errors by using expenditure share of one period only. Finally, the Törnqvist and Sato-Vartia in ation rate differences is distributed very close to the origin, indicating that both measure the in ation rate very close to the Fisher price index. Therefore, the usage of CES preferences is not responsible for the differences I observe between the UPI and conventional in ation rates. Hence, the difference must come from assumption about product turnover and intertemporal demand shifts. In the next sections I quantify the biases arising from ignoring these aspects in the price index computation.

3.4.3 Biases in regional in ation rates

As was discussed in Section 3.2.2 the UPI collapses to the common-goods UPI when one assumes no product turnover and the common-goods UPI collapses to the Sato-Vartia price index if one assumes no intertemporal shifts in demand. Therefore, the differences between these price indexes measure the "variety bias" and the "consumer valuation

FIGURE 3.3: Kernel densities of differences between in ation rates and Fisher.



(A) Germany (B) Spain

Notes: Figure 3.3 depicts kernel densities of differences between Laysperes, Paasche, geometric Laysperes, geometric Paasche, Törnqvist, Sato-Vartia in ation rates and Fisher in ation rate. The values of differences close to zero indicate that in ation rate is similar to the Fisher in ation rate.

bias" correspondingly. To present all biases affecting the Laysperes price index I additionally compute the "substitution bias", which, as suggested by Boskin et al. (1996), can be computed as the difference between the Laysperes price index and one of the superlative price indexes.

Table 3.5 presents the summary statistics computed over all regions and time periods and Figure 3.B.4 of Section 4.B shows the evolution of biases in regions over time. From Table 3.5 it can be seen that the "substitution biases" computed with respect to the Törnqvist is positive on average in both countries and higher in Germany (0.36%) than in Spain (0.23%), what also in line with the result obtained in Section 3.4.2 showing that the households in Germany exhibit more elastic demand than in Spain and, therefore, tend to substitute more between goods triggered by a change in relative prices. The absolute values of the "substitution bias" provide with almost exact number so that the bias remains positive over the whole period.

The "variety bias" and the "consumer valuation bias" are on average positive in both countries, but both exhibit much higher dispersion than the "substitution bias". The "variety bias" is 0.26% on average in Germany and 0.18% in Spain and the "consumer valuation bias" is 0.26% on average in Germany and 0.04% in Spain. However, looking at the absolute values, which are 0.29% and 0.58% in Germany and 0.46% and 0.37% in Spain, it is obvious that these biases are more substantial than the average values suggest

TABLE 3.5: In ation biases

	Germany					Spain				
	Mean	SD	SD_b	SD_w	p50	Mean	SD	SD_b	SD_w	p50
Subst. T.	0.36	0.07	0.03	0.06	0.36	0.24	0.10	0.06	0.09	0.22
Variety	0.26	0.22	0.07	0.21	0.26	0.18	0.59	0.13	0.57	0.21
Cons. val.	0.26	0.71	0.08	0.71	0.23	0.04	0.45	0.10	0.44	0.08
Subst. T. Abs.	0.36	0.07	0.03	0.06	0.36	0.24	0.10	0.06	0.09	0.22
Variety Abs.	0.29	0.18	0.04	0.17	0.27	0.46	0.40	0.11	0.39	0.37
Cons. val. Abs.	0.58	0.50	0.11	0.48	0.43	0.37	0.26	0.08	0.25	0.31
Observations	1064		38	28		544		17	32	

Notes: Table 3.5 reports summary statistics for in ation biases computed over regions and time periods."Abs." indicates that statistics is computed over absolute values of the corresponding bias. "Subst. T." presents "substitution bias" with respect to Törnqvist price index. SD shows the overall standard deviation in the sample from the overall mean, SD_b shows the standard deviation between regions (SD_b is computed over regional mean values), SD_w shows the standard deviation within regions while excluding the between variance from the computation.

since they also take negative values.

Figure 3.B.4 of Section 4.B shows that in Germany the "substitution bias" and the "variety bias" remain relatively stable over time with a slight increase in the "variety bias" by the end of the observation period, while the "consumer valuation bias" varies quite substantially over time. Therefore, the substitution due to the changes in consumer valuation have a signi cant effect on the UPI and common-goods UPI in ation rates dynamics. In Spain the gures indicate an increase in the "substitution bias" during 2009 and 2010, which con rms the ndings in the literature that the consumers change their behaviour during the recession and switch from more expensive product to less expensive products. The "variety bias" and the "consumer valuation bias" are both increasing during the period of the Great Recession. The observed dynamics in the "variety" and "consumer valuation" biases suggests that the UPI might react more to the movements in business cycles than the conventional measures of in ation rate.

Figure 3.B.4 of Section 4.B illustrates that the biases exhibit considerable dispersion across regions. To undestand the magnitude of the regional dispersion, Table 3.6 presents summary statistics computed across values of quarterly standard deviation (estimated across regions in each quarter) and quarterly spread (difference between maximum and minimum value of the bias observed in the corresponding quarter). Both standard deviation and spread indicate that the "variety bias" and the "consumer valuation bias" exhibit signi cantly higher dispersion across regions in comparison to the "substitution bias" in

both countries. In Germany, the spread observed in the "substitution bias" is 0.25% on average and does not exceed 0.4%. In Spain, the spread in the "substitution bias" exhibit the similar magnitude with mean value of 0.36% and the maximum value being 0.54%. Regional dispersion observed in the "variety bias" and the "consumer valuation bias" is quite different in Spain and Germany. In Germany, the spread observed in the "variety bias" is 0.77% on average and reaches 1.61% at the maximum, while the "consumer valuation bias" exhibits the highest dispersion across regions with the mean and the maximum value of the observed spread being 1.47% and 2.61% correspondingly. In Spain, the "consumer valuation bias" is less dispersed than the "variety bias" and takes values of 1.09% on average and 1.67% at the maximum, while the "variety bias" exhibit the highest dispersion reaching 1.37% on average and 2.18% at the maximum. Observed regional heterogeneity indicates that the differences in the regional access to variety and regional differences in preferences of households have a signi cant impact on consumer utility and provide an important insight on the signi cance of capturing these differences in regional in ation rates.

TABLE 3.6: Dispersion in in ation biases

	Germany					Spain				
	Mean	SD	p50	Min	Max	Mean	SD	p50	Min	Max
Subst. T. Std.Dev.	0.05	0.01	0.05	0.04	0.07	0.09	0.02	0.08	0.05	0.13
Variety Std.Dev.	0.16	0.03	0.16	0.11	0.24	0.34	0.10	0.32	0.18	0.55
Cons. val. Std.Dev.	0.32	0.08	0.32	0.17	0.51	0.30	0.07	0.30	0.16	0.44
Subst. T. Spread	0.25	0.07	0.23	0.16	0.40	0.36	0.11	0.36	0.18	0.54
Variety Spread	0.77	0.27	0.71	0.43	1.61	1.37	0.39	1.37	0.85	2.18
Cons. val. Spread	1.47	0.46	1.38	0.77	2.61	1.09	0.29	1.09	0.45	1.67
Observations	28					32				

Notes: Table 3.6 reports summary statistics for two measures of within-regional dispersion across in ation biases computed over time periods. "Std.Dev." indicates that statistics is computed over quarterly standard deviation of the corresponding bias. "Spread" indicates that statistics is computed over quarterly spreads in the corresponding bias. Quarterly spread is computed as the difference between the maximum and the minimum of the observed regional biases. "Subst. T." stands for "substitution bias" computed with respect to Törnqvist price index.

3.5 Conclusions

Regional in ation rates are important for the computation of real values that are important for economic policy and sub-national economic analysis, which are recently frequently employed in the literature[18]. Traditional statistical price indexes used for in ation computation are usually biased due to how they handle introduction of new goods into basket and consumer substitution between goods triggered by changes in relative prices or demand.

In this study I employ scanner price data to compute regional in ation rates using the uni ed price index developed by Stephen J. Redding and David E. Weinstein (2020), which takes into account changes in consumer demand, allows for substitution between goods and adjusts for changes in available varieties. Additionally, I compare conventional measures with the uni ed price index and quantify the "substitution", "variety" and "consumer valuation" biases in the computation of regional in ation rates.

Results show that the average difference per annum between the Laysperes and the UPI in ation rates is 0.88% in Germany and 0.46% in Spain implying that the Laysperes index overestimates the cost-of-living in ation rate given by the UPI. Interestingly, UPI seems to exhibit somewhat different dynamic than the conventional in ation rates and the graphical evidence suggests that the UPI falls more signi cantly than the Laysperes price index in Spain during the Great Recession implying that there might be some implication for the analysis of the business cycles. Looking at the contribution of the aforementioned biases separately I nd that the "substitution bias" is on average 0.36% per year in Germany and 0.24% in Spain, the "variety bias" is on average 0.26% in Germany and 0.18% in Spain and the "consumer valuation bias" is 0.26% in Germany and 0.04% in Spain. The biases exhibit some substantial volatility over time and across regions.

[18] see Beraja, Hurst, and Ospina (2019) for a review of the recent economic literature that uses the variation in regional data to study the national uctuations.

Appendices

3.A Additional tables

TABLE A.1: Summary statistics on the regional datasets

Markets		All categories				Common categories				
	Popul.	Purch.	Hh.	Cat.	Prod.	Purch.	Hh.	Cat.	Prod.	Exp. Sh.
ES11	2,756	107,133	641	339	21,712	97,648	641	199	17,918	86
ES12	1,066	45,065	256	331	13,206	40,963	256	199	11,132	85
ES13	587	20,381	120	319	8,356	18,569	120	199	7,235	86
ES21	2,173	102,202	604	339	20,680	93,676	604	199	17,215	86
ES22	634	20,516	122	320	8,556	18,708	122	199	7,409	86
ES23	318	15,807	92	311	6,870	14,462	92	199	6,011	86
ES24	1,335	61,159	321	337	16,458	55,789	321	199	13,947	86
ES30	6,372	246,913	1,323	346	33,230	222,235	1,323	199	26,902	85
ES41	2,519	123,960	751	341	23,659	112,609	751	199	19,571	86
ES42	2,074	77,230	411	337	17,929	69,923	411	199	15,000	86
ES43	1,096	42,039	240	328	11,830	38,074	240	199	9,976	86
ES51	7,435	294,628	1,443	346	35,643	266,380	1,443	199	28,783	85
ES52	4,969	208,294	1,025	345	26,827	188,395	1,024	199	21,741	86
ES53	1,097	41,201	199	334	11,857	37,122	199	199	9,989	85
ES61	8,323	316,219	1,587	345	33,717	285,690	1,587	199	27,109	86
ES62	1,454	53,057	265	336	13,897	48,064	265	199	11,646	86
ES70	2,077	61,265	334	331	14,033	54,822	334	199	11,705	86
Total	2,723	108,063	573	334	18,733	97,831	572	199	15,488	86

Notes: Table A.1 reports mean values of the number of purchases ("Purch."), the number of unique households ("Hh.""), the number of categories ("Cat.""), and the number of unique products ("Prod.") observed quarterly in the Spanish dataset. Additionally, column "Popul." reports yearly mean of the regional population. All values of the descriptive statistics are rounded to the closest integer.

TABLE A.2: Summary statistics on the regional datasets

Markets		All categories				Common categories				
	Popul.	Purch.	Hh.	Cat.	Prod.	Purch.	Hh.	Cat.	Prod.	Exp. Sh.
DE11	3,980	324,770	1,141	387	48,816	319,189	1,141	258	46,810	97
DE12	2,711	234,882	822	383	42,422	230,732	822	258	40,768	97
DE13	2,180	153,612	547	376	33,013	151,044	547	258	31,795	97
DE14	1,789	135,971	480	374	32,721	133,547	480	258	31,534	97
DE21	4,419	345,804	1,256	390	52,486	339,176	1,256	258	50,218	97
DE22	1,189	106,040	341	372	27,616	104,174	341	258	26,615	97
DE23	1,080	113,581	378	371	28,655	111,438	378	258	27,574	97
DE24	1,065	150,990	486	375	33,997	148,348	486	258	32,759	98
DE25	1,706	295,731	1,046	387	47,915	290,294	1,045	258	45,865	97
DE26	1,306	146,912	502	375	32,766	144,379	502	258	31,584	98
DE27	1,800	178,000	578	380	37,023	174,819	578	258	35,598	97
DE30	3,408	405,928	1,462	392	56,344	398,473	1,462	258	53,965	97
DE40	2,474	364,422	1,161	387	52,107	357,711	1,161	258	49,957	97
DE50	659	57,323	210	354	19,345	56,348	210	258	18,779	97
DE60	1,750	153,778	579	377	33,427	151,135	579	258	32,222	97
DE71	3,809	354,042	1,270	391	53,074	347,468	1,270	258	50,848	97
DE72	1,033	109,709	381	370	29,546	107,883	381	258	28,551	98
DE73	1,211	145,727	497	375	32,995	143,273	497	258	31,828	98
DE80	1,618	177,612	609	377	37,414	174,412	609	258	36,053	97
DE91	1,590	177,833	588	379	37,587	174,755	588	258	36,205	97
DE92	2,114	206,359	703	380	38,853	202,829	703	258	37,398	97
DE93	1,680	174,233	558	378	35,150	171,237	558	258	33,876	97
DE94	2,459	225,458	756	380	37,518	221,584	756	258	36,079	97
DEA1	5,122	456,600	1,602	392	57,732	448,738	1,602	258	55,364	98
DEA2	4,348	381,382	1,323	390	52,451	374,988	1,323	258	50,343	98
DEA3	2,586	200,797	691	380	38,851	197,334	691	258	37,444	98
DEA4	2,036	160,139	542	377	34,090	157,395	542	258	32,870	98
DEA5	3,599	329,753	1,114	388	48,545	324,255	1,114	258	46,655	98
DEB1	1,484	134,356	465	375	32,406	132,114	465	258	31,305	98
DEB2	520	39,521	144	347	14,729	38,869	144	258	14,312	98
DEB3	2,005	202,264	714	383	39,838	198,683	714	258	38,323	98
DEC0	1,003	174,337	584	379	34,691	171,536	584	258	33,446	98
DED2	1,604	299,404	973	385	45,174	294,238	973	258	43,334	98
DED4	1,479	269,307	848	382	43,021	264,883	848	258	41,350	98
DED5	986	169,212	565	377	36,063	166,293	565	258	34,739	98
DEE0	2,287	283,009	955	383	45,873	277,971	955	258	44,071	97
DEF0	2,823	290,443	1,000	386	44,030	285,429	1,000	258	42,293	97
DEG0	2,194	363,817	1,188	386	52,251	357,531	1,188	258	50,121	98
Total	2,146	223,501	765	380	39,488	219,592	765	258	37,970	97

Notes: Table A.2 reports mean values of the number of purchases ("Purch."), the number of unique households ("Hh.""), the number of categories ("Cat.""), and the number of unique products ("Prod.") observed quarterly in the German dataset. Additionally, column "Popul." reports yearly mean of the regional population. All values of the descriptive statistics are rounded to the closest integer.

TABLE A.3: Summary statistics on regionally in ation rates, Germany

In ation	Metric	2010				2011				2012				2013				2014				2015				2016			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
in _L	mean	-1.33	-0.44	0.78	1.01	1.50	2.13	2.39	3.07	2.68	2.25	1.96	1.89	2.02	2.28	2.18	2.94	1.33	0.46	0.57	1.11	-0.48	0.09	0.84	1.82	1.61	1.28	-0.04	0.58
in _L	sd	0.24	0.24	0.31	0.30	0.22	0.28	0.24	0.23	0.20	0.21	0.19	0.20	0.26	0.25	0.19	0.31	0.17	0.24	0.16	0.24	0.21	0.24	0.30	0.30	0.29	0.32	0.19	0.21
in _L	min	-1.76	-0.99	0.22	0.58	0.90	1.41	1.79	2.67	2.29	1.79	1.60	1.49	1.51	1.97	1.84	2.28	0.89	-0.14	0.19	0.66	-1.07	-0.62	0.36	1.38	1.14	0.73	-0.52	0.27
in _L	max	-0.69	0.07	1.60	1.63	1.94	2.56	2.94	3.79	3.27	2.63	2.36	2.49	2.58	2.89	2.64	3.62	1.78	0.98	0.85	1.61	-0.14	0.54	1.43	2.50	2.59	2.01	0.39	1.36
in _L	spread	1.07	1.05	1.38	1.05	1.03	1.15	1.15	1.12	0.99	0.84	0.76	0.99	1.07	0.92	0.80	1.34	0.90	1.13	0.66	0.95	0.93	1.17	1.07	1.12	1.45	1.28	0.92	1.09
in _P	mean	-2.08	-1.19	0.10	0.28	0.67	1.41	1.68	2.36	1.96	1.53	1.24	1.31	1.40	1.57	1.51	2.12	0.70	-0.17	-0.11	0.20	-1.09	-0.50	0.05	0.99	0.76	0.34	-0.73	-0.12
in _P	sd	0.22	0.22	0.26	0.26	0.20	0.26	0.23	0.24	0.17	0.20	0.19	0.17	0.23	0.24	0.20	0.26	0.18	0.24	0.19	0.27	0.21	0.25	0.32	0.31	0.36	0.34	0.22	0.22
in _P	min	-2.51	-1.67	-0.39	-0.25	0.13	0.67	1.14	1.89	1.60	1.06	0.82	1.00	1.01	1.28	1.20	1.71	0.32	-0.70	-0.52	-0.36	-1.51	-1.24	-0.57	0.57	0.23	-0.24	-1.20	-0.52
in _P	max	-1.67	-0.78	0.67	0.74	0.96	1.78	2.16	2.98	2.45	1.95	1.60	1.81	1.84	2.12	1.88	2.64	1.15	0.35	0.31	0.76	-0.64	-0.08	0.65	1.96	2.10	1.11	-0.20	0.51
in _P	spread	0.84	0.88	1.06	0.99	0.83	1.11	1.02	1.09	0.85	0.89	0.78	0.81	0.84	0.85	0.68	0.92	0.83	1.05	0.83	1.12	0.87	1.17	1.22	1.39	1.87	1.35	1.00	1.03
in _GL	mean	-1.80	-0.94	0.32	0.52	0.96	1.60	1.84	2.55	2.16	1.73	1.44	1.42	1.56	1.79	1.70	2.37	0.86	-0.04	0.06	0.46	-0.99	-0.42	0.20	1.19	0.97	0.65	-0.63	-0.01
in _GL	sd	0.24	0.25	0.31	0.30	0.23	0.29	0.24	0.23	0.20	0.22	0.21	0.22	0.27	0.26	0.19	0.29	0.18	0.22	0.16	0.23	0.22	0.26	0.31	0.27	0.27	0.29	0.21	0.23
in _GL	min	-2.24	-1.50	-0.29	0.06	0.32	0.84	1.22	2.15	1.80	1.22	1.06	0.99	0.95	1.44	1.40	1.71	0.42	-0.62	-0.34	0.08	-1.67	-1.23	-0.38	0.74	0.50	0.11	-1.14	-0.38
in _GL	max	-1.17	-0.44	1.12	1.15	1.38	2.04	2.35	3.21	2.70	2.06	1.89	2.04	2.14	2.42	2.19	2.99	1.25	0.46	0.36	0.96	-0.64	-0.05	0.77	1.75	1.84	1.25	-0.14	0.74
in _GL	spread	1.07	1.06	1.41	1.10	1.06	1.21	1.13	1.07	0.90	0.85	0.83	1.05	1.19	0.98	0.79	1.28	0.82	1.08	0.70	0.89	1.03	1.19	1.15	1.01	1.34	1.14	1.00	1.12
in _GP	mean	-1.61	-0.69	0.57	0.78	1.21	1.94	2.22	2.87	2.47	2.04	1.74	1.78	1.87	2.06	1.99	2.70	1.17	0.33	0.41	0.85	-0.58	0.00	0.69	1.63	1.40	0.98	-0.14	0.47
in _GP	sd	0.23	0.22	0.26	0.25	0.20	0.26	0.24	0.25	0.18	0.20	0.19	0.15	0.22	0.24	0.20	0.29	0.19	0.26	0.20	0.30	0.22	0.25	0.34	0.35	0.38	0.37	0.23	0.22
in _GP	min	-2.02	-1.15	0.12	0.26	0.65	1.28	1.73	2.44	2.08	1.66	1.36	1.47	1.51	1.72	1.66	2.27	0.77	-0.24	0.00	0.21	-1.05	-0.62	0.03	1.14	0.87	0.36	-0.58	0.01
in _GP	max	-1.11	-0.22	1.15	1.21	1.52	2.33	2.72	3.54	3.01	2.49	2.06	2.25	2.28	2.58	2.44	3.29	1.73	0.95	0.84	1.44	-0.04	0.50	1.33	2.72	2.85	1.89	0.47	1.12
in _GP	spread	0.91	0.93	1.03	0.95	0.87	1.05	0.99	1.10	0.93	0.83	0.71	0.78	0.76	0.86	0.79	1.03	0.96	1.20	0.84	1.23	1.01	1.12	1.31	1.58	1.98	1.53	1.05	1.11
in _F	mean	-1.71	-0.82	0.44	0.64	1.08	1.77	2.03	2.71	2.32	1.89	1.60	1.60	1.71	1.92	1.85	2.53	1.02	0.14	0.23	0.65	-0.79	-0.21	0.44	1.40	1.18	0.81	-0.39	0.23
in _F	sd	0.22	0.22	0.28	0.27	0.20	0.27	0.23	0.23	0.18	0.20	0.19	0.18	0.24	0.24	0.19	0.28	0.16	0.23	0.17	0.25	0.21	0.24	0.30	0.30	0.32	0.33	0.20	0.20
in _F	min	-2.10	-1.33	-0.06	0.19	0.52	1.10	1.47	2.30	1.95	1.42	1.21	1.25	1.27	1.64	1.54	2.00	0.60	-0.42	-0.17	0.16	-1.29	-0.94	-0.11	0.98	0.68	0.26	-0.87	-0.10
in _F	max	-1.19	-0.42	1.13	1.18	1.38	2.16	2.55	3.38	2.86	2.22	1.98	2.15	2.21	2.50	2.26	3.10	1.42	0.66	0.53	1.19	-0.39	0.23	1.04	2.23	2.34	1.56	0.10	0.93
in _F	spread	0.92	0.91	1.19	1.00	0.86	1.06	1.08	1.08	0.91	0.80	0.77	0.90	0.94	0.87	0.72	1.10	0.82	1.09	0.70	1.03	0.90	1.17	1.15	1.25	1.66	1.30	0.96	1.03
in _T	mean	-1.71	-0.82	0.44	0.65	1.08	1.77	2.03	2.71	2.31	1.88	1.59	1.60	1.71	1.92	1.84	2.53	1.02	0.14	0.23	0.65	-0.79	-0.21	0.44	1.41	1.18	0.82	-0.39	0.22
in _T	sd	0.22	0.22	0.28	0.27	0.20	0.26	0.23	0.23	0.18	0.20	0.19	0.18	0.24	0.24	0.19	0.28	0.16	0.23	0.17	0.25	0.21	0.24	0.31	0.30	0.32	0.33	0.20	0.20
in _T	min	-2.10	-1.33	-0.05	0.19	0.51	1.12	1.47	2.30	1.94	1.44	1.21	1.25	1.29	1.63	1.54	2.01	0.61	-0.43	-0.17	0.15	-1.31	-0.93	-0.11	0.98	0.69	0.26	-0.86	-0.10
in _T	max	-1.19	-0.42	1.14	1.18	1.38	2.17	2.54	3.38	2.85	2.22	1.98	2.14	2.21	2.50	2.26	3.11	1.42	0.66	0.53	1.19	-0.39	0.22	1.04	2.23	2.34	1.57	0.09	0.93
in _T	spread	0.91	0.91	1.19	0.99	0.87	1.05	1.06	1.08	0.91	0.79	0.77	0.89	0.92	0.87	0.72	1.10	0.82	1.09	0.70	1.04	0.92	1.15	1.15	1.25	1.65	1.31	0.95	1.02
in _UPI	mean	-2.28	-1.21	-0.26	0.69	0.82	1.12	1.42	2.30	2.07	2.16	2.18	2.22	2.14	2.07	1.57	0.84	0.69	-0.65	-0.35	-0.74	-1.15	-0.51	-0.78	-0.29	-0.46	-0.21	-0.98	-0.54
in _UPI	sd	0.33	0.32	0.37	0.36	0.37	0.31	0.35	0.30	0.36	0.47	0.40	0.41	0.40	0.31	0.31	0.32	0.34	0.38	0.25	0.34	0.35	0.34	0.36	0.40	0.39	0.33	0.24	0.24
in _UPI	min	-2.97	-1.74	-0.98	0.10	0.23	0.66	0.70	1.75	1.31	1.31	1.42	1.47	1.25	1.49	1.01	0.31	-0.06	-1.15	-1.09	-1.84	-2.10	-1.35	-2.20	-1.03	-1.52	-1.19	-1.50	-1.02
in _UPI	max	-1.67	-0.42	0.65	2.08	2.10	2.20	2.16	3.23	2.88	3.16	3.17	3.26	3.22	2.66	2.49	1.57	1.51	0.94	0.16	-0.23	-0.50	0.25	-0.29	0.70	0.54	0.63	-0.50	-0.17
in _UPI	spread	1.30	1.33	1.63	1.98	1.87	1.54	1.46	1.48	1.58	1.85	1.76	1.78	1.96	1.17	1.47	1.26	1.57	2.09	1.25	1.60	1.60	1.60	1.91	1.73	2.06	1.82	0.99	0.85

Notes: Table A.3 reports summary statistics on the regional in ation rates for every quarter from 2010 (Q1) to 2016 (4). Rows "Mean", "SD", "Minimum" and "Maximum" present the mean, standard deviation, minimum and maximum computed across the regional in ation rates in the indicated quarter. Row "Spread" shows the difference between the maximum and the minimum of the observed regional in ation rates in the indicated quarter. in _L denotes the Laspeyres in ation rate, in _P denotes the Paasche in ation rate, in _F denotes the Fisher in ation rate, in _T denotes the Törnqvist in ation rate, in _GL denotes the geometric Laysperes in ation rate, in _GP denotes the geometric Paasche in ation rate and in _UPI denotes the UPI in ation rate

TABLE A.4: Summary statistics on regionally in ation rates, Spain

In ation	Metric	2009				2010				2011				2012				2013				2014				2015				2016			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
in _L	mean	-0.12	-2.36	-3.08	-2.73	-1.49	-0.70	-0.22	-0.02	0.79	1.83	2.23	2.45	2.64	2.34	2.24	2.61	2.37	2.18	2.06	1.62	0.68	-0.72	-1.28	-1.18	-0.87	0.33	0.67	1.06	0.96	0.67	0.15	-0.26
in _L	sd	0.70	0.57	0.57	0.38	0.24	0.39	0.29	0.38	0.29	0.29	0.32	0.36	0.34	0.32	0.49	0.71	0.73	0.71	0.51	0.34	0.21	0.22	0.18	0.36	0.27	0.19	0.28	0.25	0.30	0.32	0.28	0.25
in _L	min	-1.04	-3.25	-3.98	-3.27	-1.88	-1.31	-0.73	-0.79	0.36	1.30	1.56	1.85	2.14	1.55	0.85	0.72	0.40	0.45	0.44	0.87	0.19	-1.22	-1.62	-1.88	-1.39	-0.16	0.26	0.50	0.49	0.27	-0.15	-0.98
in _L	max	1.28	-1.44	-1.90	-2.03	-0.87	-0.04	0.26	0.72	1.44	2.37	2.89	3.09	3.32	2.90	2.97	3.85	3.78	3.47	2.77	2.05	1.01	-0.49	-0.91	-0.19	-0.38	0.59	1.22	1.56	1.50	1.33	0.80	0.18
in _L	spread	2.32	1.81	2.08	1.24	1.00	1.27	0.99	1.51	1.08	1.08	1.33	1.24	1.19	1.36	2.12	3.13	3.38	3.02	2.33	1.18	0.82	0.73	0.72	1.69	1.01	0.75	0.96	1.06	1.01	1.06	0.95	1.16
in _P	mean	-0.86	-3.13	-3.79	-3.39	-2.04	-1.19	-0.71	-0.40	0.33	1.33	1.76	2.04	2.16	1.94	1.90	2.23	1.89	1.72	1.68	1.19	0.15	-1.24	-1.76	-1.63	-1.33	-0.07	0.34	0.72	0.60	0.34	-0.26	-0.56
in _P	sd	0.76	0.66	0.55	0.41	0.22	0.40	0.29	0.36	0.29	0.34	0.28	0.33	0.30	0.35	0.55	0.71	0.88	0.79	0.63	0.39	0.33	0.30	0.31	0.52	0.43	0.30	0.40	0.28	0.31	0.29	0.32	0.25
in _P	min	-1.76	-3.99	-4.48	-3.90	-2.44	-1.91	-1.32	-1.12	-0.21	0.67	1.21	1.66	1.69	0.95	0.19	0.05	-0.86	-0.63	-0.47	0.02	-0.63	-2.13	-2.62	-3.08	-2.69	-0.91	-0.49	0.06	0.11	-0.13	-0.63	-1.20
in _P	max	0.81	-2.08	-2.70	-2.41	-1.71	-0.37	-0.21	0.25	0.82	2.00	2.44	2.64	2.74	2.45	2.79	3.48	3.42	2.94	2.58	1.66	0.62	-0.97	-1.18	-0.46	-0.78	0.27	1.38	1.22	1.25	0.97	0.57	-0.20
in _P	spread	2.57	1.90	1.78	1.49	0.74	1.54	1.11	1.37	1.03	1.34	1.24	0.99	1.05	1.50	2.60	3.43	4.28	3.57	3.05	1.63	1.26	1.16	1.44	2.62	1.91	1.17	1.88	1.16	1.15	1.10	1.20	1.00
in _GL	mean	-0.56	-2.83	-3.55	-3.17	-1.90	-1.08	-0.59	-0.39	0.42	1.47	1.87	2.10	2.31	2.02	1.91	2.26	2.04	1.85	1.73	1.27	0.35	-1.08	-1.63	-1.53	-1.22	0.00	0.35	0.73	0.65	0.37	-0.17	-0.59
in _GL	sd	0.71	0.56	0.56	0.37	0.24	0.37	0.26	0.37	0.32	0.32	0.31	0.35	0.32	0.31	0.49	0.70	0.72	0.70	0.52	0.37	0.23	0.24	0.20	0.38	0.30	0.22	0.30	0.26	0.29	0.29	0.26	0.26
in _GL	min	-1.48	-3.68	-4.43	-3.68	-2.29	-1.77	-1.13	-1.18	-0.08	0.86	1.11	1.52	1.83	1.13	0.43	0.30	-0.04	0.02	0.02	0.48	-0.19	-1.60	-1.97	-2.41	-1.94	-0.52	-0.18	0.22	0.21	-0.03	-0.49	-1.31
in _GL	max	0.83	-1.93	-2.39	-2.51	-1.31	-0.47	-0.14	0.25	1.00	1.95	2.45	2.72	2.97	2.52	2.58	3.44	3.37	3.07	2.42	1.73	0.68	-0.80	-1.27	-0.56	-0.74	0.29	0.91	1.22	1.16	0.98	0.44	-0.16
in _GL	spread	2.31	1.76	2.04	1.17	0.98	1.29	0.99	1.43	1.07	1.10	1.35	1.20	1.14	1.38	2.15	3.14	3.41	3.06	2.40	1.25	0.87	0.80	0.70	1.85	1.20	0.81	1.08	0.99	0.95	1.01	0.92	1.15
in _GP	mean	-0.42	-2.65	-3.33	-2.95	-1.65	-0.82	-0.34	-0.03	0.69	1.69	2.11	2.38	2.49	2.25	2.23	2.58	2.23	2.04	2.01	1.54	0.49	-0.87	-1.39	-1.26	-0.99	0.26	0.65	1.05	0.90	0.64	0.06	-0.24
in _GP	sd	0.77	0.66	0.58	0.42	0.22	0.42	0.32	0.37	0.26	0.33	0.30	0.35	0.32	0.36	0.55	0.72	0.88	0.79	0.61	0.36	0.29	0.27	0.27	0.49	0.40	0.26	0.37	0.29	0.32	0.31	0.33	0.26
in _GP	min	-1.33	-3.46	-4.03	-3.44	-2.01	-1.46	-0.94	-0.72	0.23	1.09	1.61	1.96	1.93	1.35	0.62	0.49	-0.38	-0.18	-0.03	0.49	-0.11	-1.56	-2.05	-2.53	-2.18	-0.42	-0.05	0.34	0.35	0.13	-0.28	-0.91
in _GP	max	1.29	-1.58	-2.16	-1.96	-1.32	0.08	0.25	0.70	1.25	2.43	2.87	3.01	3.08	2.78	3.20	3.91	3.84	3.32	2.91	1.99	0.98	-0.59	-0.80	-0.07	-0.43	0.61	1.67	1.56	1.59	1.30	0.91	0.13
in _GP	spread	2.62	1.88	1.86	1.48	0.69	1.54	1.19	1.41	1.01	1.34	1.27	1.05	1.15	1.44	2.58	3.43	4.22	3.50	2.94	1.50	1.09	0.97	1.25	2.45	1.75	1.03	1.72	1.23	1.24	1.18	1.19	1.04
in _F	mean	-0.50	-2.75	-3.44	-3.06	-1.77	-0.95	-0.46	-0.21	0.56	1.58	1.99	2.24	2.40	2.14	2.07	2.42	2.13	1.95	1.87	1.40	0.41	-0.98	-1.52	-1.40	-1.10	0.13	0.50	0.89	0.78	0.51	-0.06	-0.41
in _F	sd	0.73	0.60	0.55	0.39	0.21	0.38	0.28	0.37	0.28	0.31	0.29	0.34	0.31	0.33	0.51	0.70	0.80	0.75	0.56	0.35	0.25	0.24	0.22	0.43	0.34	0.22	0.33	0.26	0.29	0.29	0.29	0.25
in _F	min	-1.39	-3.53	-4.24	-3.56	-2.14	-1.51	-1.03	-0.96	0.18	0.99	1.38	1.77	1.92	1.25	0.52	0.39	-0.23	-0.09	-0.01	0.53	-0.11	-1.45	-1.99	-2.48	-2.05	-0.36	-0.12	0.28	0.33	0.07	-0.37	-1.09
in _F	max	1.04	-1.84	-2.31	-2.23	-1.35	-0.24	0.02	0.49	1.13	2.19	2.66	2.86	3.03	2.65	2.85	3.66	3.59	3.20	2.67	1.85	0.79	-0.74	-1.05	-0.33	-0.58	0.39	1.30	1.39	1.38	1.15	0.66	-0.01
in _F	spread	2.43	1.69	1.93	1.33	0.80	1.27	1.05	1.44	0.94	1.20	1.28	1.09	1.11	1.40	2.33	3.28	3.83	3.30	2.69	1.32	0.90	0.72	0.94	2.16	1.46	0.75	1.42	1.11	1.05	1.08	1.02	1.08
in _T	mean	-0.49	-2.74	-3.44	-3.07	-1.78	-0.95	-0.47	-0.21	0.56	1.58	1.99	2.24	2.39	2.14	2.07	2.42	2.13	1.94	1.87	1.40	0.42	-0.98	-1.52	-1.40	-1.11	0.13	0.50	0.89	0.78	0.51	-0.06	-0.41
in _T	sd	0.73	0.60	0.55	0.39	0.21	0.38	0.28	0.36	0.28	0.31	0.30	0.34	0.31	0.33	0.51	0.70	0.80	0.74	0.56	0.35	0.25	0.24	0.22	0.43	0.34	0.23	0.32	0.26	0.29	0.29	0.28	0.25
in _T	min	-1.39	-3.52	-4.23	-3.56	-2.15	-1.52	-1.04	-0.95	0.18	0.99	1.35	1.77	1.91	1.24	0.52	0.39	-0.21	-0.08	-0.01	0.54	-0.10	-1.45	-1.99	-2.47	-2.06	-0.38	-0.11	0.28	0.33	0.08	-0.37	-1.11
in _T	max	1.04	-1.85	-2.31	-2.25	-1.34	-0.23	0.01	0.47	1.12	2.19	2.66	2.86	3.02	2.63	2.84	3.67	3.60	3.20	2.67	1.86	0.78	-0.74	-1.04	-0.32	-0.58	0.38	1.29	1.39	1.37	1.14	0.65	-0.01
in _T	spread	2.43	1.67	1.92	1.32	0.81	1.29	1.05	1.42	0.94	1.20	1.31	1.10	1.12	1.40	2.32	3.28	3.81	3.28	2.67	1.32	0.88	0.71	0.96	2.15	1.48	0.76	1.40	1.11	1.05	1.07	1.02	1.10
in _UPI	mean	-0.56	-2.37	-2.92	-2.97	-2.91	-2.20	-2.30	-1.84	-0.12	0.83	1.22	1.44	1.85	1.80	1.84	2.32	2.45	2.06	1.80	1.24	0.26	-0.73	-1.48	-1.35	-0.96	-0.17	0.59	0.51	0.53	0.70	0.67	0.84
in _UPI	sd	0.75	0.49	0.64	0.51	0.58	0.66	0.46	0.35	0.45	0.57	0.64	0.50	0.64	0.61	0.84	0.98	0.96	0.72	0.65	0.65	0.36	0.44	0.40	0.53	0.57	0.47	0.80	0.54	0.44	0.46	0.35	0.41
in _UPI	min	-2.27	-3.20	-4.06	-3.73	-4.20	-4.00	-3.20	-2.52	-1.41	0.25	0.38	0.70	0.79	0.61	-0.04	-0.26	-0.29	-0.19	0.02	0.43	-0.38	-1.48	-2.11	-2.32	-2.14	-1.03	-0.59	-0.40	0.03	0.02	0.23	0.32
in _UPI	max	0.76	-1.46	-1.68	-1.99	-1.79	-1.08	-1.56	-1.19	0.67	2.36	2.72	2.85	3.79	2.74	3.13	3.98	4.01	2.97	2.59	2.79	1.13	0.37	-0.64	0.22	0.31	0.87	2.31	1.34	1.49	1.72	1.33	1.59
in _UPI	spread	3.03	1.74	2.38	1.74	2.40	2.92	1.64	1.33	2.08	2.10	2.34	2.15	3.00	2.14	3.17	4.24	4.30	3.15	2.56	2.36	1.51	1.85	1.46	2.53	2.45	1.90	2.90	1.75	1.46	1.70	1.10	1.27

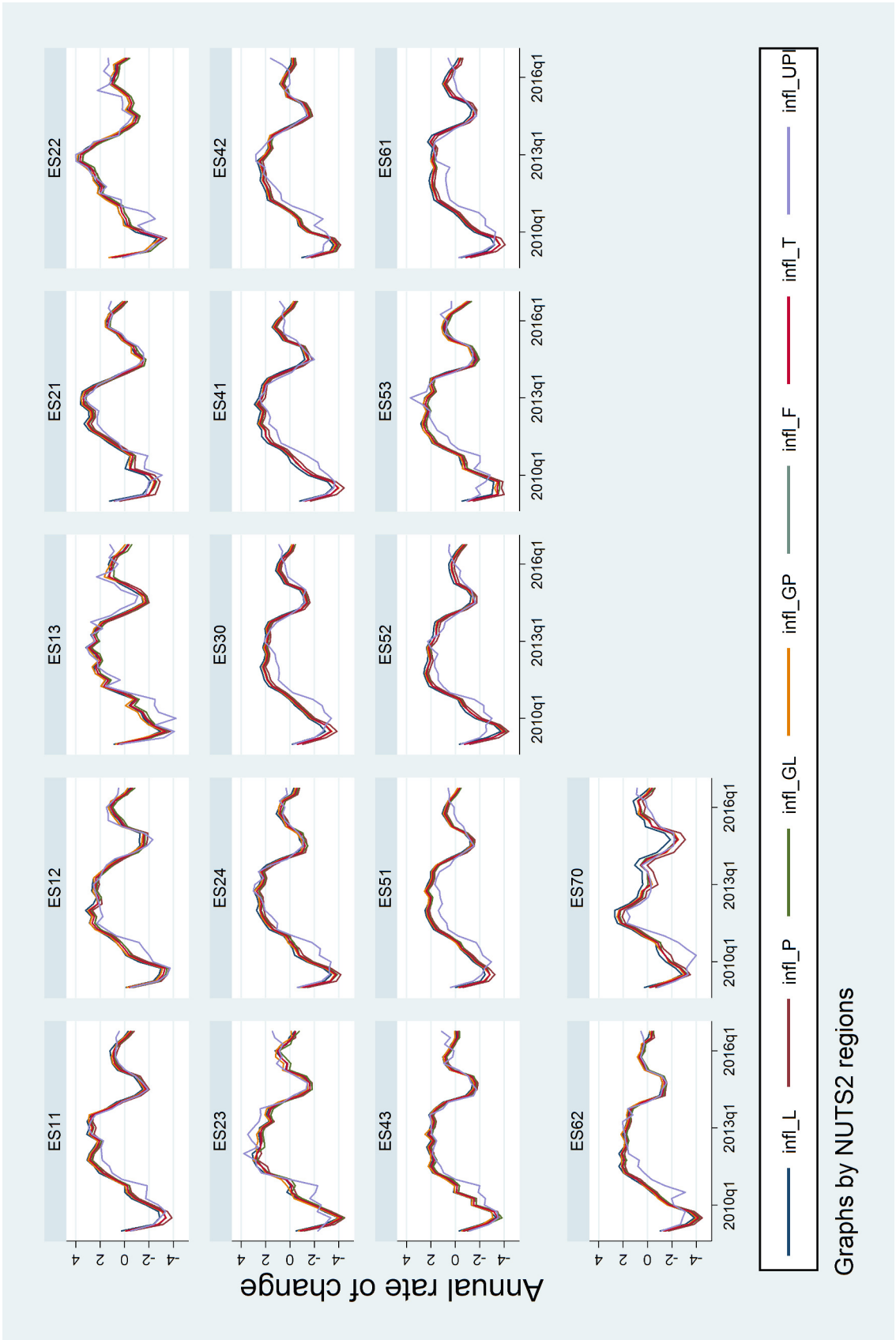
Notes: Table A.4 reports summary statistics on the regional in ation rates for every quarter from 2010 (Q1) to 2016 (4). Rows "Mean", "SD", "Minimum" and "Maximum" present the mean, standard deviation, minimum and maximum computed across the regional in ation rates in the indicated quarter. Row "Spread" shows the difference between the maximum and the minimum of the observed regional in ation rates in the indicated quarter. in _L denotes the Laspeyres in ation rate, in _P denotes the Paasche in ation rate, in _F denotes the Fisher in ation rate, in _T denotes the Törnqvist in ation rate, in _GL denotes the geometric Laysperes in ation rate, in _GP denotes the geometric Paasche in ation rate and in _UPI denotes the UPI in ation rate.

3.B Additional gures

FIGURE 3.B.1: Of cial regional food in ation rates and scanner data in ation rates

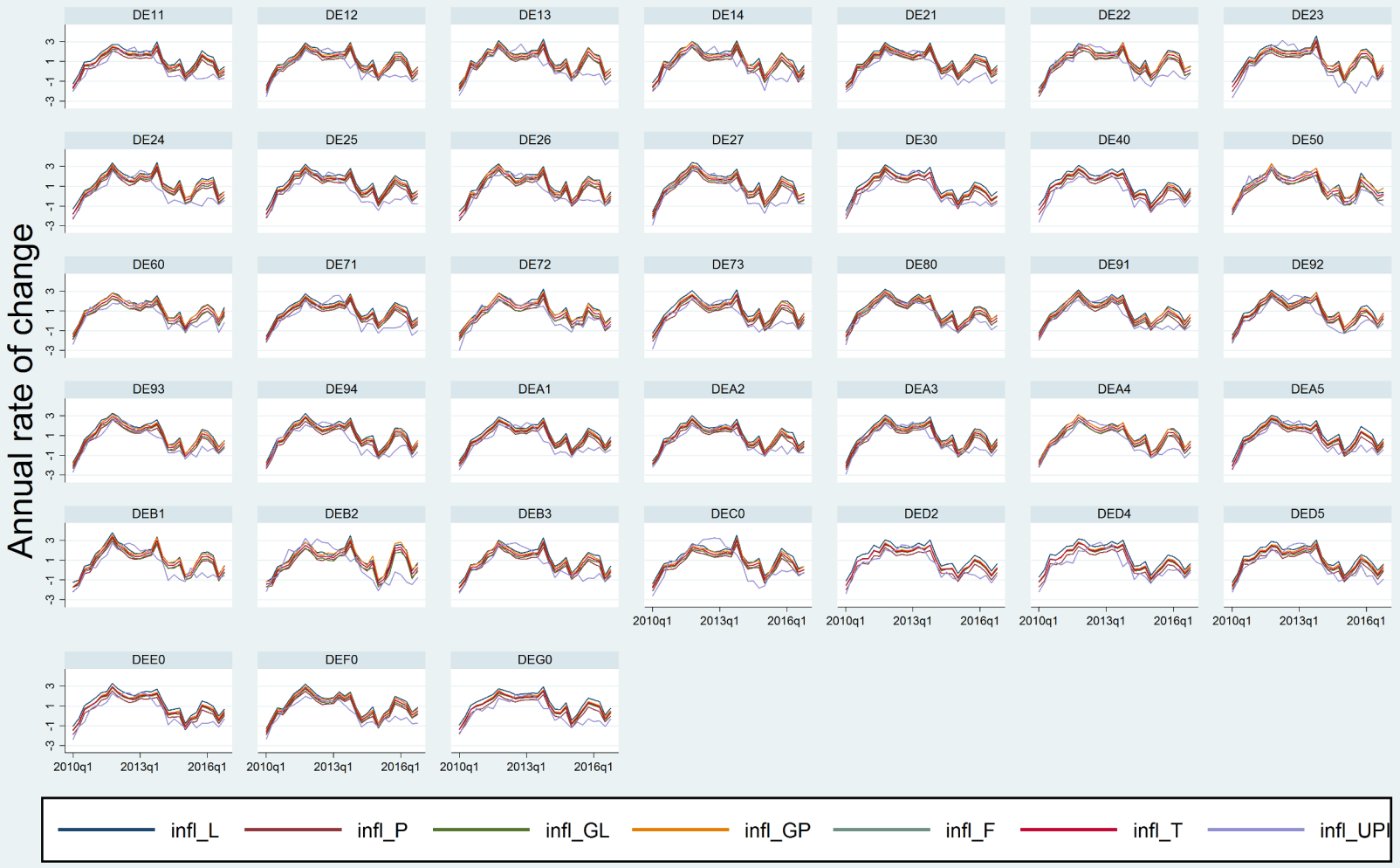
Notes: Figure 3.B.1 plots 4-quarter percentage changes in of cial food price index and the household scanner data price index. The scanner data in ation rates (denoted as in _L) are computed using 4-quarter percentage changes in the Laspeyres price index. Regional food in ation rates (denoted as in _food) are computed using 12-months percentage changes in "Food and non-alcoholic beverages" component of the regional price index and aggregating them to the quarterly level by taking a mean over months in a quarter.

FIGURE 3.B.2: Regional in ation rates computed from the scanner price data, Spain



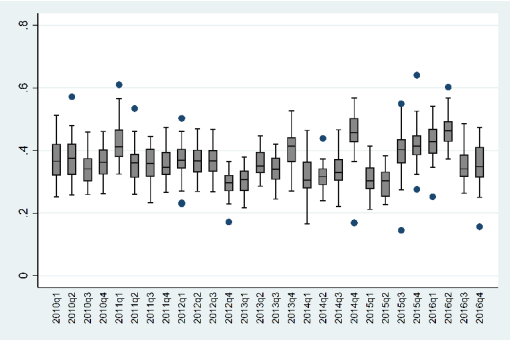
Notes: Figure 3.B.2 plots 4-quarter percentage changes of price indexes computed using the household scanner data. in _L denotes the Laspeyres in ation rate, in _P denotes the Paasche in ation rate, in _F denotes the Fisher in ation rate, in _T denotes the Törnqvist in ation rate, in _GL denotes the geometric Laysperes in ation rate, in _GP denotes the geometric Paasche in ation rate and in _UPI denotes the UPI in ation rate. Price index was rst computed for each category in each region and then aggregated to the regional price index by using initial period expenditure shares as weights.

FIGURE 3.B.3: Regional in ation rates computed from the scanner price data, Germany

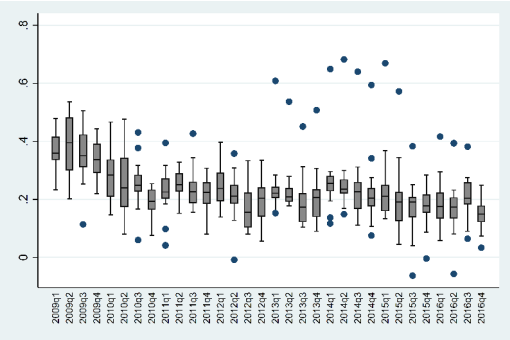


Notes: Figure 3.B.3 plots 4-quarter percentage changes of price indexes computed using the household scanner data. in _L denotes the Laspeyres in ation rate, in _P denotes the Paasche in ation rate, in _F denotes the Fisher in ation rate, in _T denotes the Törnqvist in ation rate, in _GL denotes the geometric Laysperes in ation rate, in _GP denotes the geometric Paasche in ation rate and in _UPI denotes the UPI in ation rate. Price index was rst computed for each category in each region and then aggregated to the regional price index by using initial period expenditure shares as weights.

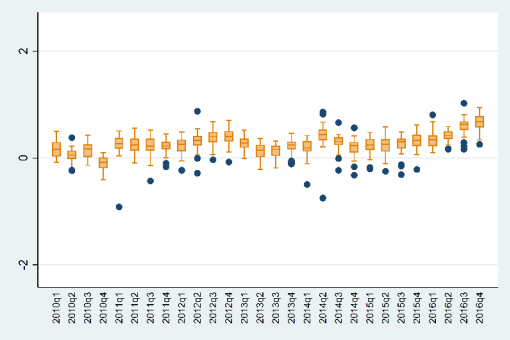
FIGURE 3.B.4: In ation biases



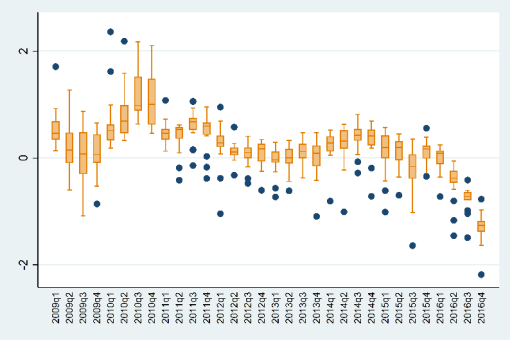
(A) Substitution bias (Törnqvist), Germany



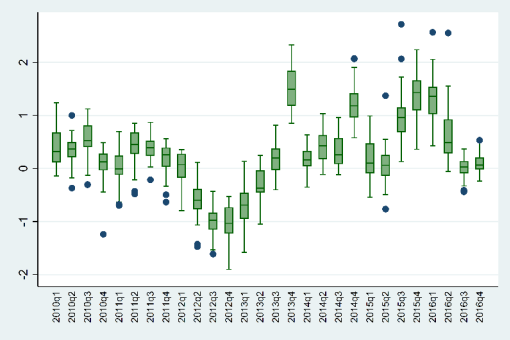
(B) Substitution bias (Törnqvist), Spain



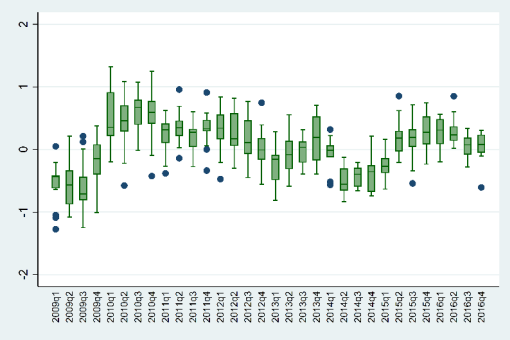
(C) Variety bias, Germany



(D) Variety bias, Spain



(E) Consumer valuation bias, Germany



(F) Consumer valuation bias, Spain

Notes: Figure 3.B.4 depicts box plots of biases in in ation computation plotted using regional values for each time period. The "variety bias" is computed as the difference between the common-goods UPI and the UPI in ation rates. The "consumer valuation bias" is computed as the difference between the the Sato-Vartia and the common-goods UPI in ation rates. The "substitution bias (Törnqvist)" is computed as the difference between the Laspeyres and the Törnqvist in ation rate. Positive values of biases imply that the in ation rate is being overestimated.

Chapter 4

Regional in ation rates and local economic conditions

Natalia Zabelina

4.1 Introduction

The relationship between in ation and economic activity is at the core of many macroeconomics models that are used to understand and predict in ation, estimate welfare effects of business cycles and guide policy makers in their decisions. The measure of in ation usually used to study this relationship is computed with the use of " xed basket" price indexes that only capture the pure price changes of the same basket of goods over time. While these price indexes are mostly suf cient to study the effects of economic uctuations on the supply side, they are not suf cient to measure the consumer welfare effects of business cycles, since these price indexes do not take into account consumer preferences and other processes that affect consumer utility. Broda and D. Weinstein (2010) nd that rms introduce less new products during economic downturns and more new products during economic upturns. Given that consumers derive positive utility from being able to access the larger variety of goods, the procyclicality of product creation could amplify the effects of business cycles on consumer welfare. Indeed, the authors demonstrate that " xed basket" price indexes, which fail to account for the changes in product variety, are signi cantly less volatile than the cost of living price index taking the changes in product variety into account and, therefore, do not adequately measure the welfare effects of business cycles.

Additionally, a vast amount of literature documents changes in the shopping behaviour of households during the different phases of a business cycle. It has been found

by Argente and M. Lee (2017), Nevo and Wong (2019), Coibion, Gorodnichenko, and Hong (2015), and Stroebel and Vavra (2019) that in times of an economic downturn consumers increase their purchases of goods of a worse quality, purchase more often on sales and shop at less expensive stores. The abilities of consumers to change their shopping habits or to switch to a cheaper alternative have important implications for consumer welfare, but are not re ected in " xed basket" price indexes that do not take into account substitution of goods due to the changes in the relative prices of goods or due to the changes in consumers tastes. Cost of living price indexes in turn use this information and allow to quantify it in terms of in ation rate: price indexes such as Törnqvist or Fisher account for the substitution due to the relative price changes and newly developed by S. Redding and D. Weinstein (2019) price index incorporates the substitution due to the changes in consumers tastes. Therefore, different price indexes could react differently to economic conditions depending if they only use the supply side data like pure price changes or if they combine supply side data with demand side data like consumers substitution behaviour.

Access to new varieties and substitution of goods driven by the price or taste changes could differ not only across time but also across space. Lecznar and Smith (2018) document differences in consumption of products across different regions in the USA. First, the authors document that consumers buy different products in different regions, so it is to expect that regional in ation rates cover the different set of products, which might create some in ation rate differential across regions. Second, the authors observe variation in the market shares of products conditional on relative price, which indicates that consumers might value the same goods differently in different markets. Third, the authors nd that variety growth and its cyclicality vary across regions, this suggests that some rms might only operate locally or only introduce new goods in some regions and potentially react to local business cycles. These regional differences in consumption have important implications for in ation computation in different regions and indicate the importance of regional analysis. In fact, literature devoted to the investigation of the Phillips curve relationship, which postulates a negative relationship between changes in prices and unemployment gap, have recently also switched to a regional dimension of the data due to incapability to observe the relationship in the national data (see T. Fitzgerald

and Nicolini (2014), Kiley (2015), Kumar and Orrenius (2016), Hooper, Mishkin, and Su (2019), and McLeay and Tenreyro (2020)).

In this study, I employ scanner price data to compute regional in ation rates using different price indexes and investigate their sensitivity to local economic conditions. To the best of my knowledge, there exists no study investigating this question using scanner price data of the Euro area countries and estimating regional in ation rates using both " xed basket" and cost of living price indexes. Moreover, regional analysis of the relationship between " xed basket" price index and economic activity, described by the means of the Phillis curve, is also not present for the Euro area countries, which is mostly caused by the regional data availability: the regional in ation rates are scarcely available in the Euro area countries and the regional data for unemployment or GDP are available yearly but rarely at a higher frequency.

The results support the existence of the negative relationship between regional in ation rates and the unemployment gap in the polled data model and the model with region xed effects. Additionally, the difference in the sensitivities indicates that in ation measures accounting for changes in consumers tastes are more sensitive to the business cycles. However, when the time xed effects are included in the model the relationship disappears for all measures of in ation rate in Spain and all measures that do not account for changes in consumers tastes in Germany. These results indicate potentially signi - cant co-movement of the regional variables and motivate the second contribution of this study that is to assess the importance of national and regional factors affecting the variance in the data using a dynamic factor model. The results of the dynamic factor model s estimations reveal a signi cant contribution of the national factor to the regional in ation variation in both Spain and Germany (explaining on average 84% – 92% in Germany and 79% – 91% in Spain) and to the local business cycles (the national component explains on average 73% of the variation in Germany and 69% in Spain). The numbers suggest that the local business cycles have a stronger regional component than the in ation rates. Together with the unresponsiveness of the regional in ation rates to local economic conditions, this result points to the possibility that the majority of retailers in Germany and Spain follow a uniform pricing strategy adjusting prices in all regions simultaneously[1].

[1]See recent funding of DellaVigna and Gentzkow (2019) who document that considerable number of retailers in the US follow a uniform pricing strategy and set price nationally rather than regionally.

Additionally, the study provides a rst insight on using regional data in Phillips curve analysis for Spain. The robustness analysis performed in the study using the of cial regional CPI in ation rate indicates no relationship between regional in ation and the unemployment gap.

The rest of the study is organized as follows. Section 4.2 summaries the related literature. Section 4.3 presents the data and descriptive statistics. Section 4.4 discusses the empirical strategy of the estimation and presents the results. Section 4.5 introduces the dynamic factor model and presents the results of its estimation. Section 4.6 summarizes and concludes.

4.2 Related literature

The study compliments the following literature. First, the study enriches the literature that uses scanner price data to make inferences about regional in ation rates and relates them to local demand conditions. Coibion, Gorodnichenko, and Hong (2015) use the IRI Marketing Data Set and aggregate the data to a city level in ation. The authors nd that the "posted" price in ation — computed from the prices set by retailers — reacts mildly to the changes of the local unemployment rate, while the "effective" price in ation — computed from the prices paid by consumers — is much more sensitive to the business cycle. Additionally, the authors emphasize that the difference in the cyclicality of posted and effective prices is due to the expenditure reallocation by the consumers from more to less expensive stores. Using the same data Stroebel and Vavra (2019) investigate the cyclicality of local retailer prices with respect to the changes in the local house prices and document a positive relationship. The authors provide further evidence that this relationship exists due to the decreasing price sensitivity of the homeowners with a rise in house price to which retailers react by increasing the markups. Beraja, Hurst, and Ospina (2019) use the price data from the Nielsen Retail Scanner Database to construct price indexes at the state level and use them to de ate nominal wages. In the earlier version of the paper (Beraja, Hurst, and Ospina (2014)) the authors additionally relate state price indexes to local unemployment and document a negative relationship with 1% increase in unemployment translating into a 0.4% decline in price growth.

While the contributions discussed above use conventional price indexes for their analysis, Gilbert (2018) uses the Nielsen Retail Scanner Database and constructs city-level cost of living price indexes derived from the CES utility function adjusted by the product turnover as in Broda and D. Weinstein (2010). Decomposing these price indexes on welfare gains from the product turnover and relative price changes, the author investigates the transmission of local business cycles through the local product creation and destruction. The author documents that regional variety growth does not respond to local economic condition, but is instead driven by the multi-regional retailers that introduce new varieties in all cities simultaneously.

Secondly, the study is related to the literature investigating the Phillips curve at the regional level. The literature can be summarised by the following contributions: T. Fitzgerald and Nicolini (2014) propose a theoretical model following the Old Keynesian framework to show that the aggregate relationship between in ation and unemployment can be broken or weaken if a central bank is successful in keeping the in ation rate close to its target. The authors estimate the proposed model using the data on MSAs (metropolitan statistical areas) and nd the robust and stable relationship with a 1% increase in the unemployment rate leading to a 0.3% decline in the in ation rate. Kiley (2015) exploit the MSAs data to study the effect of short-term and long-term unemployment on the in ation rate and nd that both in uence the in ation negatively. Kumar and Orrenius (2016) nd that the wage-price Phillips curve estimated using the state-level panel data is non-linear and strongly convex and that the shape and the slope of the Phillips curve are heterogeneous across the US states. McLeay and Tenreyro (2020) reformulate T. Fitzgerald and Nicolini (2014) model using the New Keynesian framework and discusses the endogeneity problem associated with the identi cation of the Phillips curve using the national data. The authors update the analysis of Kiley (2015) using MSAs data and con rm their results. Hooper, Mishkin, and Su (2019) using data on both MSAs and states present a comprehensive analysis of price and wage Phillips curves in the US con rming the previous ndings.

Thirdly, this study contributes to the stream of literature studying the importance of the global, regional or national, and local components in explaining more disaggregated in ation rates and business cycles. Forni and Reichlin (2001) employ a dynamic

latent factor model to regional real GDP growth in the European countries and identify European, national, and regional idiosyncratic components. European and national components together explain from 60% and up to 90% of in ation variability depending on the country and the sample time span. In Germany and Spain regional idiosyncratic components explain 10% and 27.3% correspondingly suggesting a relatively high synchronisation of regional business cycles in Germany but not in Spain. Kose, Otrok, and Whiteman (2003) employ a Bayesian dynamic latent factor model to quantify the importance of global, regional, and national cycles using international macroeconomic data from 60 countries. The authors nd that a signi cant part of business cycles in developed countries is explained by the global factor, while business cycles in developing countries are more nationally driven.

Monacelli and Sala (2009) study product-level monthly price series of US, Germany, France and UK in the framework of approximate dynamic factor model and nd that one global factor explains on average 15 – 30% depending on the transformation of the data. Additionally, the authors nd that the global factor explains more of the variation for the series in the sectors that are more open to trade. Ciccarelli and Mojon (2010) employ an approximate dynamic factor model to investigate the global nature of 22 OECD countries for the period 1961 – 2008. The authors nd that on average the global factor explains around 70% of the national in ation in the considered countries, 74% of Spanish national in ation and 57% of German national in ation. Parker (2017) extends this study to a large set of countries and con rms these results. Neely and Rapach (2011) analyse the co-movement of in ation rates from 64 countries over the period from 1951 to 2009 applying a dynamic factor model and retaining one global and 7 regional factors. The results indicate that the global factor explains 35% of the variation in the national in ation rates, the regional factors explain additional 16% and the remaining half of the variance is attributed to the national idiosyncratic component.

Most closely related to the analysis performed in this study is the contribution of Beck, Hubrich, and Marcellino (2009) who use the regional in ation rates from Austria, Finland, Germany, Italy, Norway, and Spain to investigate the importance of the common Euro area factor and national factors in explaining regional in ation rates variance. The authors employ an approximate dynamic factor model and nd that around 50% of

the variance in the regional in ation rates can be explained by the global factor, while the national factor explains at least an additional 26.9% (in Spain) and at most 48.8% (in Finland). The regional idiosyncratic component of the variance is found to be quite substantial (except for Finland) ranging from 14.4% in Germany to 24.7% in Spain.

4.3 Data and descriptive statistics

Section 4.3 describes the data used in empirical estimations and presents the summary statistics. The data are collected or computed at the NUTS-2 regional level. Section 4.3.1 presents the data on regional unemployment and Section 4.3.2 describes the regional in ation rates computed using consumer panel scanner price data.

4.3.1 Regional unemployment

To conduct the study I employ regional unemployment rates to make an inference about local economic conditions[2]. The data for Spanish NUTS-2 regions are available quarterly and collected from the Spanish Statistical Of ce for the period Q1 2002 – Q1 2018. In Germany regional unemployment rates are provided by the Federal Employment Agency monthly for NUTS-3 regions only. Hence, to obtain quarterly NUTS-2 regional unemployment rates I rst average NUTS-3 values to NUTS-2 level using the labor force s values at NUTS-3 regional level as weights and then I additionally average the resulted values over months in a quarter to receive quarterly numbers for the period Q1 2007 – Q4 2017. To mitigate the impact of seasonal patterns in the data all time-series are seasonally adjusted using X-11 seasonal adjustment method.

To separate the trend and cyclical component of the data I apply Christiano-Fitzgerald band pass lter[3] to each regional unemployment rate separately using the complete available sample. Table 4.1 reports the descriptive statistics on quarterly regional unemployment rates (denoted by un) and estimated cyclical components that are interchangeably

[2]This choice is guided by the availability of the regional data. As has been mentioned before the regional data at frequencies higher than yearly are scarcely available in the Euro area countries and regional unemployment rates are the only demand-related time series available in both countries.

[3]Issever Grochová and Rozmahel (2015) show that the Christiano-Fitzgerald band pass lter is better suited for estimation of business cycles in the European economies. The lter estimates the business cycle component as being composed of stochastic shocks between 6 and 32 quarters. For further details on the ltering procedure see Christiano and T. J. Fitzgerald (2003).

referred to as unemployment gaps (denoted by un_gap_cf) in the following text. The period under study is conditioned by the availability of the scanner price data used for in ation measurement and therefore reduced to Q2 2009 – Q4 2016 for Germany and Q2 2008 – Q4 2016 for Spain.

The gures reported in Table 4.1 show that in Germany regional unemployment on average is around 6.8%. The standard deviations indicate that unemployment rates are much more dispersed across regions rather than within the regions across time with the corresponding values of the standard deviations being 2.58% and 0.79%. The heterogeneity in unemployment rates across regions is also quite noticeable in Figure 4.B.1 of Section 4.B.1 that depicts mean region unemployment rate and the change in unemployment rates over the period under study. Figure 4.B.1a shows that the regions in the East Germany have on average the highest unemployment rates ranging between 8.48% and 12%, while the regions in the West Germany have in general smaller unemployment rates (between 3.6% and 9.28%) with the majority of the lowest rates concentrating in the South of Germany in the states Bavaria (3.66% to 4.79%) and Baden-Württemberg (3.6% to 4.62%). Additionally, Figure 4.B.1b shows that unemployment rates in all regions have fallen over the sample period with larger reductions in between -6.65% and -4.53% in the East Germany compared to in between -2.68% and -0.72% in the West Germany, re ecting the lasting catching up process of the eastern regions after the German uni cation.

TABLE 4.1: Descriptive statistics on regional unemployment

Germany	Mean	SD	SD_b	SD_w	Min	p5	p25	p50	p75	p95	Max
un	6.84	2.67	2.58	0.79	3.11	3.41	4.50	6.50	8.52	11.95	14.44
un_gap_cf	0.05	0.22	0.02	0.22	-0.59	-0.29	-0.06	0.04	0.14	0.48	0.92
Observations	1178										
Spain	Mean	SD	SD_b	SD_w	Min	p5	p25	p50	p75	p95	Max
un	19.80	6.48	5.22	4.03	5.93	10.20	15.21	19.08	23.64	32.29	36.52
un_gap_cf	0.13	1.15	0.10	1.14	-3.32	-1.61	-0.62	0.09	0.73	2.33	3.59
Observations	595										

Notes: Table 4.1 reports descriptive statistics for regional unemployment rates and estimated unemployment gaps. SD shows the overall standard deviation in the sample from the overall mean, SD_b shows the standard deviation between the regions (SD_b is computed over regional mean values), SD_w shows the standard deviation within regions while excluding the between variance from the computation.

Over the last decade, the unemployment rate in Spain has been one of the highest across European countries what is well re ected in the regional data presented in Table 4.1. Mean unemployment in Spain over all regions and time periods is almost 19.8%,

ranging from around 10,2% in its 5th percentile to more than 32% in its 95th percentile. Its between and within standard deviations show that there are considerable differences in unemployment rates between regions with a standard deviation of around 5% as well as over time with the within regions standard deviation being equal to 4.03%. Figure 4.B.2 of Section 4.B.1 shows that the mean regional unemployment rates presented in Figure 4.B.2a differ quite signi cantly across the regions being equal to 30.05% in Andalusia and 13.2% in Basque Community. The unemployment rates are generally higher in the south of Spain compared to the northern regions. In contrast to Germany, the unemployment values in Spain have been increasing over the sample period as shown in Figure 4.B.2b with the larger increases in the southern regions.

The unemployment gap is positive on average for both countries and is around 0.05 for Germany and 0.13 for Spain and is much more dispersed in Spain with the standard deviation of 1.15% compared to Germany where the standard deviation is 0.22%. Looking at the between and within standard deviations it can be seen that most of the variation in both countries are coming from over time variation. The standard deviation within regions being 1.14% in Spain and 0.22% in Germany indicates that the business cycles movements are much more volatile in Spanish than in German regions. Figure 4.B.2 of Section 4.B.1 shows that there is always some heterogeneity in the level and the sign of unemployment gaps across the regions in both countries. However, almost all regions in both Spain and Germany have been in a recession simultaneously during the Great Recession period.

4.3.2 Scanner price in ation rates

To compute regional in ation rates I employ consumer panel scanner price data for Spain and Germany. The datasets are provided by AiMark[4] for the period Q1 2008 – Q4 2016 for Spain and Q1 2009 – Q4 2016 for Germany. The data contain information on purchases

[4] Aimark is a non-pro t foundation that supports academic research by providing access to various price data on behalf of its data partners (GfK, Kantar, and others). More information about the organisation can be obtained on the Aimark s website: https://aimark.net/.

of a representative sample of households including price, quantities, goods characteristics, retailer names, and socio-demographical characteristics of households.[5] The region is identi ed in the data via postal codes of the households residence. Using the data on prices and quantities I compute quarter-on-quarter price indexes: Laspeyres ($\Phi^{Lr}_{t-1,t}$), Paasche ($\Phi^{Pr}_{t-1,t}$), geometrical Laspeyres ($\Phi^{GLr}_{t-1,t}$), geometrical Paasche ($\Phi^{GPr}_{t-1,t}$), Fisher ($\Phi^{Fr}_{t-1,t}$), Törnqvist ($\Phi^{Tr}_{t-1,t}$), Sato-Vartia ($\Phi^{SVr}_{t-1,t}$), the common-goods uni ed price index ($\Phi^{UPI\_CGr}_{t-1,t}$), and the uni ed price index ($\Phi^{UPIr}_{t-1,t}$)[6]. The price indexes differentiate with respect to how the substitution patterns are taken into account, whether the new varieties are accounted for, and whether the price index takes into account changes in the consumers valuation of goods.

Laspeyres and Paasche price indexes assume zero elasticity of substitution and, hence, tend to over- and underestimate in ation rates since they are using the expenditure shares of one time period (either the initial in the case of Laspeyres or the nal period in the case of Paasche). Geometrical Laspeyres and Paasche price indexes reduce the effect of the "substitution bias" by incorporating the assumption about the elasticity of substitution being equal to 1 and taking geometric averages of the price relatives instead of arithmetical. Superlative Fisher and Törnqvist price indexes are presenting the second-order approximations to the exact cost of living price indexes and are free from the "substitution bias". These measures are, however, only take into account goods that are available in both periods and assume that consumers do not change their valuation of goods. The UPI in turn is a cost of living price index derived from the CES utility function that allows the consumers to change their valuation of goods over time and accounts for the product turnover and, therefore, provides an unbiased measure of in ation. Additionally, to examine the effects of each of the biases separately I consider two simpli cations of the UPI: the UPI_CG which is the UPI without product turnover and the Sato-Vartia price index which is a price index derived from the CES utility function without allowing for a product turnover and changes in the consumers valuation.

[5]The dataset is described more thoroughly in Chapter 3 of this thesis. Here, I follow the same steps in cleaning and aggregating the data.

[6]The methodology for computation of the price indexes and elasticity estimation is provided in Chapter 3 of this thesis.

Before computation of in ation rates, the quarterly price indexes are seasonally adjusted using X-11 seasonal adjustment method. The annualised quarter on quarter regional in ation rates are then computed using the following formula:

$$\pi_t^r = 100 \ln \left[(\Phi_{t-1,t}^r)^4 - 1 \right] \tag{4.1}$$

where π_t^r is annualised quarterly change in price index $\Phi_{t-1,t}^r$ in period t in region r.

Table 4.2 presents the summary statistics on the computed annualised quarterly regional in ation rates. In both countries different measures of in ation rate provide with a quite different mean and median values, while the standard deviation is similar across all measures with the standard deviation between regions being relatively small compared to the within regions standard deviation, which indicates the presence of a significant overtime variation. The last row of Table 4.2 presents the correlation between the corresponding in ation rate and regional unemployment gap. In both countries the correlation is always negative. In Germany it is between -0.16 and -0.27 for the conventional in ation measures and -0.55 and - 0.52 for the UPI_CG and the UPI correspondingly. In Spain the correlation is similar across all measures of in ation and lies between -0.48 and -0.4. These values indicate the negative relationship between the in ation rates and the unemployment gap, which I formally evaluate in the next section.

4.4 Sensitivity of regional in ation to local economic conditions

This section presents the econometric framework and the results of estimation the relationship between regional in ation rates and local economic conditions. Section 4.4.1 introduces the methodology used for the estimation and discusses the estimation procedure and Section 4.4.2 presents the results.

4.4.1 Methodology

To estimate the sensitivity of different regional in ation rates to local economic conditions I apply the following econometric framework. In each country I use the panel of data across regions and relate each type of in ation rate to the unemployment gap. First, I

TABLE 4.2: Descriptive statistics on regional in ation rates

Germany	Mean	SD	SD_b	SD_w	Min	p5	p25	p50	p75	p95	Max	Corr
in _L	2.26	2.05	0.19	2.04	-3.49	-1.06	0.84	1.99	3.56	6.16	7.80	-0.16
in _P	-0.80	1.74	0.18	1.73	-6.19	-3.76	-1.98	-0.92	0.36	2.26	4.97	-0.27
in _GL	0.51	1.91	0.24	1.90	-5.88	-2.66	-0.74	0.40	1.80	4.02	5.79	-0.19
in _GP	0.93	1.87	0.26	1.85	-3.92	-2.20	-0.33	0.71	2.11	4.39	7.17	-0.22
in _F	0.71	1.87	0.13	1.86	-5.05	-2.45	-0.56	0.54	1.93	4.14	6.45	-0.21
in _T	0.71	1.87	0.13	1.87	-4.97	-2.45	-0.56	0.54	1.94	4.16	6.44	-0.21
in _SV	0.75	1.89	0.13	1.89	-4.98	-2.42	-0.53	0.56	1.96	4.31	6.52	-0.20
in _UPI_CG	-0.44	1.68	0.13	1.68	-5.54	-3.60	-1.35	-0.47	0.76	2.11	4.38	-0.55
in _UPI	-0.59	1.74	0.26	1.72	-6.54	-3.94	-1.49	-0.58	0.61	2.04	4.90	-0.52
Observations	1178											
Spain	Mean	SD	SD_b	SD_w	Min	p5	p25	p50	p75	p95	Max	Corr
in _L	0.87	2.09	0.28	2.07	-6.10	-2.82	-0.32	0.95	2.30	4.20	8.02	-0.45
in _P	-0.26	2.19	0.26	2.17	-8.61	-4.47	-1.35	-0.00	1.21	2.78	8.04	-0.48
in _GL	-0.09	2.10	0.23	2.09	-6.96	-3.94	-1.22	0.03	1.31	3.10	6.84	-0.46
in _GP	0.70	2.16	0.33	2.13	-7.56	-3.31	-0.42	0.85	2.18	3.83	8.66	-0.48
in _F	0.30	2.12	0.25	2.11	-7.30	-3.56	-0.82	0.49	1.75	3.40	8.30	-0.47
in _T	0.30	2.13	0.25	2.11	-7.31	-3.65	-0.83	0.47	1.75	3.39	8.29	-0.47
in _SV	0.30	2.14	0.25	2.13	-7.66	-3.63	-0.78	0.45	1.72	3.44	8.35	-0.47
in _UPI_CG	0.30	1.99	0.29	1.97	-6.57	-3.18	-0.80	0.26	1.71	3.22	6.79	-0.46
in _UPI	0.53	2.08	0.59	2.00	-6.11	-3.36	-0.73	0.64	1.81	3.92	7.22	-0.40
Observations	595											

Notes: Table 4.2 reports descriptive statistics for annualised quarterly regional in ation rates computed from the household scanner price data. SD shows the overall standard deviation in the sample from the overall mean, SD_b shows the standard deviation between the regions (SD_b is computed over regional mean values), SD_w shows the standard deviation within regions while excluding the between variance from the computation. Corr presents the correlation coef cient between the corresponding in ation rate and the unemployment gap.

assume homogeneous parameters across regions, pool the regions together, and estimate the cross-regional sensitivity using the following model:

$$\pi_t^r = \alpha + \gamma \pi_{t-1}^r + \beta x_t^r + \epsilon_t^r, \tag{4.2}$$

where π_t^r is the annualised quarter on quarter in ation rate in period t in region r. Demand pressure in a region is captured by the unemployment gap $x_t^r = (u_t^r - \overline{u_t^r})$ – the difference between the regional rate of unemployment u_t^r and the regional natural rate of unemployment $\overline{u_t^r}$. A positive deviation of the unemployment rate from its natural level indicates the presence of a slack in the labor market that creates a downwards pressure on the in ation rate. A negative deviation implies that the labor market is tight and pushes the in ation up. Coef cient β presents the sensitivity of the in ation rate to the unemployment gap. With the higher values of β a change in the unemployment gap translates

into a bigger change in the in ation rate. ϵ_t^r captures random errors. Taking into consideration the literature on the Phillips curve I augment the model with a lag of in ation rate as an additional explanatory variable. This variable serves as a proxy of price setters in ation expectations, which according to the literature on the Phillips curve affects the setting of prices by rms. Since all of the considered in ation rates include relative pure price changes of goods, it is to expect that the price-setting behaviour underlying the Phillips curve will affect all in ation rates to some extent. Hence, π_{t-1}^r presents adaptive in ation expectations, which let the in ation rate of the previous period affect the current in ation with the coef cient γ. Higher values of γ signal that a change in in ation rate is more persistent and an increase in in ation in the previous period translates into a bigger change in the current in ation.

Second, to capture time-invariant distinctive regional features I introduce region xed effects α^r to the model:

$$\pi_t^r = \alpha^r + \gamma \pi_{t-1}^r + \beta x_t^r + \epsilon_t^r, \tag{4.3}$$

Third, to control for the common macroeconomic shocks I additionally include time dummies α_t in the model:

$$\pi_t^r = \alpha^r + \alpha_t + \gamma \pi_{t-1}^r + \beta x_t^r + \epsilon_t^r. \tag{4.4}$$

To estimate the model given by Equation (4.2) I employ the OLS (Ordinary Least Squares) estimator and to estimate the models given Equations (4.3) and (4.4) I employ the LSDV (Least-Squares Dummy Variables) estimator. One concern that arises when estimating dynamic panel models with xed effects is that the inclusion of the lagged dependent variable as explaining variable introduces Nickell bias[7] to the coef cients estimated with the LSDV estimator. While there exist a number of alternative IV and GMM estimators developed by Anderson and Hsiao (1982), Arellano and Bond (1991), and Blundell and Bond (1998) that circumvent Nickell bias, the properties of these estimators hold only if the number of units in the panel is large, which is not the case here with the number of regions being equal to 17 in Spain and 38 in Germany.

According to the Monte Carlo evidence provided by Judson and Owen (1999) and

[7] Nickell (1981) shows that during the time demeaning of dependent and explaining variables creates correlation between lagged dependent variable used as a regressor and the error term that causes inconsistent estimation of coef cients.

Bruno (2005b) for the panels with a small number of entities and time dimension equal to 30[8] IV and GMM estimators do not perform better than the LSDV, while the best performance is achieved by the biased-corrected LSDV (developed by Kiviet (1995) and extended by Bruno (2005a)). Given that the bias-corrected LSDV does not provide analytical standard errors I present the results of both LSDV estimation in the main text and bias-corrected LSDV estimation with bootstrapped standard errors in Section 4.A.1.

4.4.2 Results

Tables 4.3 and 4.4 present the results of the models estimation given by Equations (4.2) to (4.4) for Germany and Spain correspondingly. In both countries Panel A and Panel B indicate signi cant estimates of all coef cients. Given that the estimated coef cients of these two models are very close, in the following I refer to the coef cients estimated with regional xed effects. In Spain the coef cient of interest β lies in between -0.31 and -0.39, while in Germany the coef cient is much higher and lies between -0.84 and -2.17 depending on the measure of in ation rates. Additionally, in both countries the measures of in ation that control for in ation biases exhibit higher sensitivity to the unemployment gap.

The largest change in the sensitivity occurs with the adjustment by the "consumer valuation bias", which can be observed by comparing coef cients of Sato-Vartia in ation rate, which present an unadjusted CES price index, with UPI_CG in ation rate, which adjusts Sato-Vartia in ation rate for the changes in consumers tastes. After adjustment of the explanatory variable by the "consumer valuation bias" its sensitivity falls by 0.05 in Spain and by 0.93 in Germany. The effects of adjustment for the "substitution bias" and the "variety bias" are smaller. Fisher and Törnqvist in ation rates, which are adjusted for the "substitution bias", provide with the slopes that 0.01 smaller than the slope given by the Laspeyres in Spain and 0.26 and 0.25 in Germany. UPI in ation rate that is different from the UPI_CG by the "variety bias" provides with the slope smaller by 0.1 in Spain and 0.18 in Germany.

[8] I concentrate on the value of 30 because the time dimension of the panel in Spain is 34 and in Germany is 30.

TABLE 4.3: Regional in ation and unemployment gap, Germany

Panel A: pooled OLS

	in _L	in _P	in _GL	in _GP	in _F	in _T	in _SV	in _UPI_CG	in _UPI
un_gap_cf	-0.81***	-1.27***	-0.95***	-1.14***	-1.07***	-1.06***	-1.04***	-1.93***	-2.03***
	(0.28)	(0.19)	(0.24)	(0.22)	(0.23)	(0.23)	(0.23)	(0.18)	(0.21)
L_in	0.16***	0.24***	0.22***	0.20***	0.20***	0.20***	0.19***	0.45***	0.40***
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
_cons	2.06***	-0.44***	0.56***	0.91***	0.73***	0.74***	0.77***	-0.03	-0.13***
	(0.09)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.04)	(0.04)
N	1140	1140	1140	1140	1140	1140	1140	1140	1140
R2	0.04	0.11	0.07	0.07	0.07	0.07	0.06	0.45	0.39
F-stat	17.25	91.70	37.84	46.57	41.70	40.30	36.87	517.69	281.01
Quarter FE	No	No	No	No	No	No	No	No	No
Region FE	No	No	No	No	No	No	No	No	No
VCE	robust	robust	robust	robust	robust	robust	robust	robust	robust

Panel B: region xed effects

	in _L	in _P	in _GL	in _GP	in _F	in _T	in _SV	in _UPI_CG	in _UPI
un_gap_cf	-0.84***	-1.31***	-0.99***	-1.20***	-1.10***	-1.09***	-1.06***	-1.99***	-2.17***
	(0.28)	(0.19)	(0.24)	(0.23)	(0.23)	(0.23)	(0.24)	(0.18)	(0.21)
L_in	0.15***	0.23***	0.20***	0.18***	0.20***	0.19***	0.19***	0.44***	0.38***
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)
_cons	2.08***	-0.44***	0.57***	0.93***	0.74***	0.74***	0.77***	-0.03	-0.14***
	(0.09)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)
N	1140	1140	1140	1140	1140	1140	1140	1140	1140
R2	0.05	0.12	0.08	0.09	0.07	0.07	0.07	0.45	0.40
R2 w	0.04	0.11	0.07	0.07	0.07	0.07	0.06	0.45	0.39
F-stat	16.35	89.06	35.14	44.28	40.76	39.36	36.03	505.40	272.62
Quarter FE	No	No	No	No	No	No	No	No	No
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VCE	robust	robust	robust	robust	robust	robust	robust	robust	robust

Panel C: region and time xed effects

	in _L	in _P	in _GL	in _GP	in _F	in _T	in _SV	in _UPI_CG	in _UPI
un_gap_cf	-0.22	-0.10	-0.22	-0.10	-0.16	-0.16	-0.16	-0.30*	-0.34**
	(0.16)	(0.14)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.16)	(0.17)
L_in	-0.05	-0.05	-0.07*	-0.01	-0.05	-0.05	-0.04	-0.12***	-0.11***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
_cons	2.48***	-0.75***	0.66***	1.04***	0.85***	0.85***	0.89***	-0.38***	-0.53***
	(0.08)	(0.04)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
N	1140	1140	1140	1140	1140	1140	1140	1140	1140
R2	0.91	0.88	0.90	0.90	0.90	0.90	0.90	0.82	0.82
R2 w	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.02
F-stat	1.74	1.33	2.74	0.25	1.48	1.33	1.13	8.12	6.72
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VCE	robust	robust	robust	robust	robust	robust	robust	robust	robust

Notes: Table 4.3 reports the results of model estimation given by Equations (4.2) to (4.4) using German data. The in ation rates are estimated using the household scanner price data. Panel A presents results from the estimation of Equation (4.2), Panel B presents results from the estimation of Equation (4.3), and Panel C presents results from the estimation of Equation (4.4). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 4.4: Regional in ation and unemployment gap, Spain

Panel A: pooled OLS

	in _L	in _P	in _GL	in _GP	in _F	in _T	in _SV	in _UPI_CG	in _UPI
un_gap_cf	-0.31***	-0.34***	-0.31***	-0.33***	-0.32***	-0.32***	-0.33***	-0.38***	-0.39***
	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.06)
L_in	0.58***	0.58***	0.59***	0.59***	0.58***	0.58***	0.57***	0.52***	0.53***
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
_cons	0.28***	-0.18***	-0.12**	0.21***	0.05	0.05	0.05	0.08	0.24***
	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
N	578	578	578	578	578	578	578	578	578
R2	0.49	0.50	0.50	0.51	0.50	0.50	0.49	0.45	0.42
F-stat	213.18	170.72	218.39	197.77	188.38	184.92	177.72	130.13	185.91
Quarter FE	No	No	No	No	No	No	No	No	No
Region FE	No	No	No	No	No	No	No	No	No
VCE	robust	robust	robust	robust	robust	robust	robust	robust	robust

Panel B: region xed effects

	in _L	in _P	in _GL	in _GP	in _F	in _T	in _SV	in _UPI_CG	in _UPI
un_gap_cf	-0.31***	-0.34***	-0.31***	-0.33***	-0.32***	-0.32***	-0.33***	-0.38***	-0.39***
	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
L_in	0.57***	0.57***	0.58***	0.58***	0.58***	0.58***	0.57***	0.52***	0.49***
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
_cons	0.29***	-0.18***	-0.11**	0.22***	0.05	0.05	0.05	0.08	0.26***
	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
N	578	578	578	578	578	578	578	578	578
R2	0.50	0.50	0.50	0.52	0.50	0.50	0.49	0.45	0.44
R2 w	0.49	0.50	0.50	0.50	0.50	0.49	0.49	0.44	0.38
F-stat	204.35	164.61	210.70	188.10	181.96	179.00	171.97	121.50	165.10
Quarter FE	No	No	No	No	No	No	No	No	No
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VCE	robust	robust	robust	robust	robust	robust	robust	robust	robust

Panel C: region and time xed effects

	in _L	in _P	in _GL	in _GP	in _F	in _T	in _SV	in _UPI_CG	in _UPI
un_gap_cf	0.01	0.02	0.02	0.01	0.03	0.02	0.02	0.00	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)
L_in	0.12**	0.09	0.12**	0.12**	0.09	0.09	0.07	-0.03	0.11**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
_cons	0.65***	-0.36***	-0.20***	0.49***	0.15***	0.15***	0.16***	0.20***	0.39***
	(0.07)	(0.03)	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
N	578	578	578	578	578	578	578	578	578
R2	0.88	0.90	0.88	0.90	0.90	0.89	0.89	0.83	0.80
R2 w	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01
F-stat	2.00	1.27	1.99	2.26	1.32	1.17	0.86	0.10	2.43
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VCE	robust	robust	robust	robust	robust	robust	robust	robust	robust

Notes: Table 4.4 reports the results of model estimation given by Equations (4.2) to (4.4) using Spanish data. The in ation rates are estimated using the household scanner price data. Panel A presents results from the estimation of Equation (4.2), Panel B presents results from the estimation of Equation (4.3), and Panel C presents results from the estimation of Equation (4.4). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The results are, however, not robust to the inclusion of time dummies as can be seen from Panel C of Tables 4.3 and 4.4. In Spain neither measure of in ation exhibit a signi cant estimate of the sensitivity, while in Germany only the measures of in ation that control for the "consumer valuation bias" exhibit a negative relationship with the unemployment gap: the slope for UPI_CG in ation is -0.3 and for UPI is -0.34. Tables A.1 and A.2 of Section 4.A.1 present the results of the estimation using the bias-corrected LSDV estimator that con rm the results described above.

Given that the consumer panel datasets do not supply the full scope of the categories taken into consideration in the of cial consumer price index (CPI) computation, some of the obtained results might be driven by this fact. Therefore, to examine if the scanner price in ation rate performs differently from the consumer price in ation I additionally collect of cial data for regional (NUTS-2) consumer price indexes for Spain[9] (period Q1 2002 – Q1 2018) provided by the Spanish Statistical Of ce and repeat the analysis. Of - cial in ation rates are then computed as annualised quarterly changes in the seasonally adjusted CPI.

First, I reduce the sample period to the period under consideration to compare the magnitude of the coef cients. Panel A of Table 4.5 summarises the corresponding results which exhibit very similar behaviour to the results obtained employing the scanner price in ation rates: the relationship between the in ation rate and the unemployment gap is signi cantly negative in the polled model and the model with regional xed effect with the coef cient being equal to -0.23, but the relationship disappears once the time xed effects are included. Secondly, to examine if this result is a feature of the particular data period under the analyses I additionally consider the full available sample. The results presented in Panel B of Table 4.5 are numerically very similar to the reduced sample and con rm the previously observed behaviour.

Comparing the results obtained using of cial CPI in ation rate with the results obtained using the Laspeyres price index in ation rates computed from the scanner price data it can be noted that the magnitude of the coef cients are very close and, therefore, scanner price data provide a good representation of the price dynamics observed in the of cial CPI data. Additionally, this funding provides the rst insight into using regional

[9] Unfortunately, in Germany data at the NUTS-2 regional level is not available.

data for the estimation of the Phillips curve in Spain, which shows no signi cant relationship between in ation and the unemployment gap when time dummies are included in the estimation.

TABLE 4.5: Of cial regional CPI in ation and unemployment gap, Spain

Panel A: Q2 2008-Q4 2016					
	1	2	3	4	5
	in _CPI	in _CPI	in _CPI	in _CPI	in _CPI
un_gap_cf	-0.23***	-0.23***	-0.02	-0.22***	-0.02
	(0.07)	(0.07)	(0.03)	(0.07)	(0.03)
L_in	0.45***	0.45***	0.09	0.49***	0.12***
	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)
_cons	0.59***	0.59***	0.92***		
	(0.08)	(0.08)	(0.06)		
N	578	578	578	578	578
R2	0.25	0.25	0.96		
R2 w		0.25	0.01		
F-stat	101.09	97.16	1.34		
Quarter FE	No	No	Yes	No	Yes
Region FE	No	Yes	Yes	Yes	Yes
VCE	robust	robust	robust	bootstrap	bootstrap
Panel B: Q2 2002-Q1 2018					
	1	2	3	4	5
	in _CPI	in _CPI	in _CPI	in _CPI	in _CPI
un_gap_cf	-0.24***	-0.24***	-0.02	-0.23***	-0.02
	(0.06)	(0.06)	(0.02)	(0.06)	(0.02)
L_in	0.52***	0.52***	0.05	0.54***	0.06**
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
_cons	0.95***	0.95***	1.86***		
	(0.08)	(0.08)	(0.08)		
N	1071	1071	1071	1071	1071
R2	0.32	0.32	0.96		
R2 w		0.31	0.00		
F-stat	285.51	277.91	1.09		
Quarter FE	No	No	Yes	No	Yes
Region FE	No	Yes	Yes	Yes	Yes
VCE	robust	robust	robust	bootstrap	bootstrap

Notes: Table 4.5 reports the results of model estimation given by Equations (4.2) to (4.4) using of cial Spanish regional CPI in ation rates. Panel A presents the results from the estimation of the models using the restricted sample, while Panel B presents the results from the estimation of the model using the full sample. Column 1 presents the results from the estimation of Equation (4.2), column 2 presents the results from the estimation of Equation (4.3) using the LSDV, column 3 presents the results from the estimation of Equation (4.4) using the LSDV. Columns 4 and 5 present the biased-corrected LSDV estimation of the last two models correspondingly. Bias correction is initialised by Arellano-Bond estimator. Standard errors are obtained using 200 bootstrap replications and are reported in parentheses. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.5 Co-movement of regional variables

The results of the previous section show that the negative relationship disappears in Spain and in many cases in Germany with the inclusion of the time dummies, which points out to the potentially signi cant co-movement in either regional in ation rates or regional unemployment gaps. To empirically investigate how much variation in the in- ation rates and unemployment gaps is due to the common (national) component and how much is due to a regional idiosyncratic component, in this section I employ an approximate dynamic factor model. Dynamic factor models are frequently used in macroeconomic studies to understand the co-movement of economic indicators and to identify common global or/and national factors using data from less-aggregated units (e.g., countries, regions, states, or cities)[10].

Let x_t^r denote a variable under consideration (regional in ation rates computed as annualised quarterly change in corresponding price index or regional unemployment gap), which is driven by national and local shocks as in:

$$x_t^r = \lambda^r f_t + \varepsilon_t^r, \tag{4.5}$$

where x_t^r is also normalised to have zero mean and unit variance, f_t are static common factors that capture national developments common to all regions, λ^r are the regional-speci c factors loadings associated with the national common factors, and ε_t^r captures region-speci c idiosyncratic component. To achieve the identi cation of model, the national factors are assumed to be uncorrelated with the idiosyncratic component and idiosyncratic components are allowed to develop mild temporal and cross-sectional correlation. Under these assumptions the national factors can be estimated as the principal components of the data[11].

Before proceeding with the estimation, the optimal number of factors to be retrieved

[10] See Barhoumi, Darné, and Ferrara (2014) and Stock and Watson (2016) for an overview of speci cations, methods, and applications of dynamic factor modelling in macroeconomics.

[11] This estimation procedure is used by Beck, Hubrich, and Marcellino (2009) to identify three layers of components: common, national, and region-speci c using regional in ation data from 6 EMU member countries. Additionally, see Stock and Watson (2002) who pioneered the usage of the principal component method for common factor identi cation in approximate static factor models.

needs to be determined. Stock and Watson (2016) suggest several approaches to determine the number of static factors: scree plots, information criteria developed in Bai and Ng (2002), and other less frequently used statistical procedures[12]. Given that the most frequently used information criteria of Bai and Ng (2002) is not adequate when $\min\{N, T\} < 40$, which is the case for both Spain and Germany, to determine the optimal number of factors to be retrieved I use scree plots of eigenvalues for all variables under the analysis.

A scree plot presents the eigenvalues of factors arranged in the descending order and plotted against the sequential number of factor. The number of factors is selected by visually inspecting the plot for changes in the slope: when the size of the eigenvalues drops signi cantly including an additional factor adds relatively little to the already extracted information. Additionally to the scree plots, Kaiser s rule (see Kaiser (1960)) suggests retrieving all factors that have eigenvalues higher than one. This rule, however, can be biased in the case of small samples. Therefore, the eigenvalues should be adjusted by a potential sample bias before the Kaiser s rule can be applied. To estimate this bias Horn (1965) develops "parallel analysis" procedure that simulates arti cial uncorrelated data of the observed sample s size, performs the principal component analysis, and uses estimated eigenvalues to compute a sample bias that is then subtracted from the observed data.

Figures 4.B.3 and 4.B.4 of Section 4.B.2 show the scree plots of eigenvalues estimated from the observed data as well as the eigenvalues adjusted by the bias using the "parallel analysis" procedure. From the scree plots the dominance of the rst factor is obvious for all measures of in ation rates in both Spain and Germany. Moreover, Kaiser s rule applied to the sample-bias adjusted eigenvalues suggests retaining only one common component. In the case of the unemployment gap in both countries the rst factor again clearly dominates the other factors and the "parallel analysis" suggests retaining one common component in Spain and two in Germany (to preserve the consistency of the analysis and given the clear domination of the rst factor in Germany, I will retain only the rst factor in the analysis of the German unemployment gap).

[12] These procedures include methods developed by Onatski (2009), Onatski (2010), and Ahn and Horenstein (2013).

Equation (4.5) is estimated by extracting the rst principal component of the correlation matrix. Figure 4.B.5 of Section 4.B.2 plots the estimated national factors from all measures of in ation and the unemployment gap. In both Spain and Germany, it is visible that there is a negative relationship between the common component of the in ation rate and the unemployment gap. Additionally, conventional measures of in ation are moving very close to each other, while the UPI_CG and the UPI deviate from the common trend.

The variance decomposition of regional in ation rates and the regional unemployment gap is presented in Tables A.3 and A.4 of Section 4.A.2. It shows how much of the variance in a variable is explained by the national factor and how much of the variation is due to the idiosyncratic regional component. In Germany the national factor explains on average 91% – 92% of the variability in regional in ation rates computed using conventional measures of in ation, 84% and 85% of in _UPI_CG and in _UPI respectively, and 73% of the regional unemployment gap. In Spain the results are very similar: 89% – to 91% of the variance in conventional in ation rates, 84% and 79% of in _UPI_CG and in _UPI respectively, and 69% of the regional unemployment gap are explained by the national factor.

Given that the national component in the performed analysis accumulates the effects of global and the Euro area developments (e.g., commodity price changes, global nancial crisis spillovers, the monetary policy of the ECB) with the developments at the national level (e.g., national scal policy) the results are similar to the previous research contributions, which identify the global and national components separately. However, comparing the results for the Laysperes price in ation with the previous nding for Germany and Spain it should be noted that the regional idiosyncratic components explain a much smaller share of the total variance than previously found for these countries: in Spain the corresponding share is 11% on average and in Germany it is 8% on average. The importance of the regional component in explaining the variance of unemployment rates is still high in Spain (31% on average) and is higher than previously found in Germany (27% on average). The big difference in the share of the variation explained by the regional component between in ation rates and the unemployment gap suggests that

prices are not set to react to local economic conditions but rather follow a national pattern.

4.6 Conclusions

This study presents the rst attempt to investigate the sensitivity of in ation rates measured with different price indexes to local economic conditions. Using the unique scanner price datasets for Spain and Germany I computed various measures of regional in ation rates, that take into account biases present in the of cially used " xed basket" price indexes. Empirical results of the estimation show a negative relation between in ation rates and the unemployment gap in the polled model and the model with xed effect. In these models in ation measures that take into account "substitution, variety, and consumer valuation biases" respond stronger to a deterioration in economic conditions. These results are, however, not robust to the inclusion of the time dummies. With the inclusion of time dummies in Spain the sensitivities of all in ation measures turn out to be insigni cant, while in Germany the relationship continues to hold for the uni ed price indexes with and without variety adjustment pointing out the importance of capturing the changes in consumer tastes.

The additional co-movement analysis shows a very strong in uence of the national component on the variance of the regional in ation rates in both countries. The variation in the regional unemployment gaps in both countries is affected less by the national component and has a stronger regional component. This nding is in line with the emerging empirical evidence (see DellaVigna and Gentzkow (2019) for a comprehensive review) documenting that majority of nationally active retailers follow a uniform pricing strategy and set their prices nationally rather than regionally, what in turn causes the unresponsiveness of regional in ation rates to local economic conditions.

Another insight emerged in this study from analysing of cial regional CPI in ation rates in Spain and is relevant for the literature employing regional variation for the Phillips curve estimation. In contrast to the studies using the USA data, the analysis performed in this study using Spanish data show that as soon as the macroeconomic factors are accounted for the relationship between the unemployment gap and in ation cease to exist. This funding is however is a byproduct of the main analysis and requires further investigation, which is outside the scope of this study.

Appendices

4.A Additional tables

4.A.1 Additional tables for Section "Sensitivity of regional in ation to local economic conditions"

TABLE A.1: Regional in ation and unemployment gap: bias-corrected estimation, Germany

	in _L	in _P	in _GL	in _GP	in _F	in _T	in _SV	in _UPI_CG	in _UPI
Panel A: region xed effects									
un_gap_cf	-0.75**	-1.20***	-0.90***	-1.10***	-1.00***	-0.99***	-0.97***	-1.81***	-2.00***
	(0.30)	(0.25)	(0.27)	(0.27)	(0.27)	(0.27)	(0.27)	(0.21)	(0.22)
L.in	0.19***	0.27***	0.24***	0.22***	0.23***	0.23***	0.23***	0.48***	0.42***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
N	1140	1140	1140	1140	1140	1140	1140	1140	1140
Quarter FE	No	No	No	No	No	No	No	No	No
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: region and time xed effects									
un_gap_cf	-0.21	-0.09	-0.21	-0.09	-0.14	-0.14	-0.15	-0.29*	-0.33**
	(0.14)	(0.14)	(0.14)	(0.14)	(0.13)	(0.13)	(0.13)	(0.16)	(0.16)
L.in _L	-0.01	-0.02	-0.04	0.02	-0.02	-0.01	-0.01	-0.09***	-0.08***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
N	1140	1140	1140	1140	1140	1140	1140	1140	1140
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table A.1 reports the results of models estimation given by Equations (4.3) and (4.4) using German data. Panel A presents results from the estimation of Equation (4.3), Panel B presents results from the estimation of Equation (4.4). Both models are estimated using the bias-corrected LSDV estimator. Bias correction is initialised by Arellano-Bond estimator. Standard errors are obtained using 200 bootstrap replications and are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.2: Regional in ation and unemployment gap: bias-corrected estimation, Spain

Panel A: region xed effects									
	in _L	in _P	in _GL	in _GP	in _F	in _T	in _SV	in _UPI_CG	in _UPI
un_gap_cf	-0.30***	-0.33***	-0.30***	-0.32***	-0.31***	-0.31***	-0.32***	-0.36***	-0.38***
	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)
L.in	0.61***	0.61***	0.62***	0.62***	0.62***	0.61***	0.61***	0.55***	0.53***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
N	578	578	578	578	578	578	578	578	578
Quarter FE	No	No	No	No	No	No	No	No	No
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: region and time xed effects									
	in _L	in _P	in _GL	in _GP	in _F	in _T	in _SV	in _UPI_CG	in _UPI
un_gap_cf	0.01	0.02	0.02	0.01	0.03	0.02	0.02	0.00	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)
L.in	0.15***	0.12***	0.15***	0.15***	0.12***	0.12***	0.10**	0.00	0.14***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
N	578	578	578	578	578	578	578	578	578
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table A.2 reports the results of models estimation given by Equations (4.3) and (4.4) using Spanish data. Panel A presents results from the estimation of Equation (4.3), Panel B presents results from the estimation of Equation (4.4). Both models are estimated using the bias-corrected LSDV estimator. Bias correction is initialised by Arellano-Bond estimator. Standard errors are obtained using 200 bootstrap replications and are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.A.2 Additional tables for Section "Co-movement of regional variables"

TABLE A.3: Variance decomposition, Germany

	in _L		in _P		in _GL		in _GP		in _F		in _T		in _SV		in _UPI_CG		in _UPI		un_gap_cf	
	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.
DE11	0.94	0.06	0.90	0.10	0.94	0.06	0.92	0.08	0.94	0.06	0.94	0.06	0.94	0.06	0.88	0.12	0.85	0.15	0.92	0.08
DE12	0.94	0.06	0.94	0.06	0.93	0.07	0.95	0.05	0.94	0.06	0.94	0.06	0.94	0.06	0.92	0.08	0.92	0.08	0.92	0.08
DE13	0.88	0.12	0.85	0.15	0.86	0.14	0.85	0.15	0.87	0.13	0.87	0.13	0.87	0.13	0.92	0.08	0.92	0.08	0.96	0.04
DE14	0.88	0.12	0.91	0.09	0.94	0.06	0.92	0.08	0.93	0.07	0.93	0.07	0.94	0.06	0.83	0.17	0.83	0.17	0.93	0.07
DE21	0.89	0.11	0.84	0.16	0.88	0.12	0.86	0.14	0.87	0.13	0.87	0.13	0.87	0.13	0.91	0.09	0.89	0.11	0.99	0.01
DE22	0.94	0.06	0.92	0.08	0.94	0.06	0.92	0.08	0.93	0.07	0.93	0.07	0.94	0.06	0.76	0.24	0.72	0.28	0.86	0.14
DE23	0.88	0.12	0.76	0.24	0.86	0.14	0.86	0.14	0.87	0.13	0.87	0.13	0.86	0.14	0.78	0.22	0.82	0.18	0.93	0.07
DE24	0.90	0.10	0.89	0.11	0.88	0.12	0.91	0.09	0.90	0.10	0.90	0.10	0.91	0.09	0.86	0.14	0.88	0.12	0.95	0.05
DE25	0.95	0.05	0.94	0.06	0.94	0.06	0.95	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.86	0.14	0.87	0.13	0.98	0.02
DE26	0.95	0.05	0.88	0.12	0.91	0.09	0.90	0.10	0.91	0.09	0.91	0.09	0.92	0.08	0.77	0.23	0.78	0.22	0.93	0.07
DE27	0.95	0.05	0.93	0.07	0.94	0.06	0.94	0.06	0.94	0.06	0.94	0.06	0.95	0.05	0.89	0.11	0.89	0.11	0.93	0.07
DE30	0.90	0.10	0.85	0.15	0.88	0.12	0.86	0.14	0.87	0.13	0.87	0.13	0.88	0.12	0.87	0.13	0.87	0.13	0.06	0.94
DE40	0.91	0.09	0.92	0.08	0.91	0.09	0.91	0.09	0.91	0.09	0.91	0.09	0.91	0.09	0.90	0.10	0.90	0.10	0.00	1.00
DE50	0.86	0.14	0.83	0.17	0.83	0.17	0.85	0.15	0.84	0.16	0.84	0.16	0.85	0.15	0.73	0.27	0.62	0.38	0.15	0.85
DE60	0.93	0.07	0.88	0.12	0.93	0.07	0.91	0.09	0.92	0.08	0.92	0.08	0.92	0.08	0.86	0.14	0.79	0.21	0.67	0.33
DE71	0.95	0.05	0.91	0.09	0.94	0.06	0.92	0.08	0.93	0.07	0.94	0.06	0.94	0.06	0.82	0.18	0.85	0.15	0.85	0.15
DE72	0.94	0.06	0.90	0.10	0.93	0.07	0.92	0.08	0.94	0.06	0.94	0.06	0.94	0.06	0.79	0.21	0.75	0.25	0.89	0.11
DE73	0.91	0.09	0.90	0.10	0.90	0.10	0.92	0.08	0.91	0.09	0.92	0.08	0.91	0.09	0.81	0.19	0.80	0.20	0.69	0.31
DE80	0.87	0.13	0.87	0.13	0.86	0.14	0.88	0.12	0.87	0.13	0.87	0.13	0.87	0.13	0.79	0.21	0.83	0.17	0.21	0.79
DE91	0.91	0.09	0.90	0.10	0.91	0.09	0.90	0.10	0.92	0.08	0.92	0.08	0.90	0.10	0.88	0.12	0.84	0.16	0.56	0.44
DE92	0.93	0.07	0.92	0.08	0.93	0.07	0.91	0.09	0.93	0.07	0.93	0.07	0.93	0.07	0.90	0.10	0.88	0.12	0.73	0.27
DE93	0.90	0.10	0.91	0.09	0.89	0.11	0.91	0.09	0.91	0.09	0.91	0.09	0.91	0.09	0.89	0.11	0.89	0.11	0.50	0.50
DE94	0.95	0.05	0.96	0.04	0.95	0.05	0.96	0.04	0.95	0.05	0.95	0.05	0.95	0.05	0.85	0.15	0.83	0.17	0.89	0.11
DEA1	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.88	0.12	0.89	0.11	0.88	0.12
DEA2	0.96	0.04	0.96	0.04	0.95	0.05	0.96	0.04	0.96	0.04	0.95	0.05	0.96	0.04	0.90	0.10	0.91	0.09	0.75	0.25
DEA3	0.97	0.03	0.92	0.08	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.86	0.14	0.86	0.14	0.93	0.07
DEA4	0.96	0.04	0.95	0.05	0.95	0.05	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.86	0.14	0.86	0.14	0.95	0.05
DEA5	0.95	0.05	0.94	0.06	0.93	0.07	0.93	0.07	0.93	0.07	0.93	0.07	0.93	0.07	0.86	0.14	0.88	0.12	0.95	0.05
DEB1	0.94	0.06	0.91	0.09	0.93	0.07	0.91	0.09	0.92	0.08	0.92	0.08	0.93	0.07	0.78	0.22	0.81	0.19	0.80	0.20
DEB2	0.88	0.12	0.81	0.19	0.86	0.14	0.84	0.16	0.86	0.14	0.86	0.14	0.86	0.14	0.73	0.27	0.74	0.26	0.65	0.35
DEB3	0.93	0.07	0.89	0.11	0.93	0.07	0.89	0.11	0.92	0.08	0.92	0.08	0.93	0.07	0.88	0.12	0.89	0.11	0.88	0.12
DEC0	0.94	0.06	0.89	0.11	0.93	0.07	0.90	0.10	0.92	0.08	0.92	0.08	0.93	0.07	0.87	0.13	0.84	0.16	0.83	0.17
DED2	0.91	0.09	0.88	0.12	0.90	0.10	0.89	0.11	0.89	0.11	0.89	0.11	0.89	0.11	0.88	0.12	0.88	0.12	0.84	0.16
DED4	0.90	0.10	0.86	0.14	0.89	0.11	0.88	0.12	0.88	0.12	0.88	0.12	0.87	0.13	0.82	0.18	0.80	0.20	0.95	0.05
DED5	0.90	0.10	0.87	0.13	0.88	0.12	0.88	0.12	0.89	0.11	0.88	0.12	0.90	0.10	0.81	0.19	0.80	0.20	0.43	0.57
DEE0	0.91	0.09	0.91	0.09	0.91	0.09	0.92	0.08	0.91	0.09	0.91	0.09	0.92	0.08	0.81	0.19	0.88	0.12	0.16	0.84
DEF0	0.95	0.05	0.92	0.08	0.93	0.07	0.93	0.07	0.94	0.06	0.93	0.07	0.94	0.06	0.89	0.11	0.89	0.11	0.74	0.26
DEG0	0.95	0.05	0.94	0.06	0.95	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.83	0.17	0.84	0.16	0.54	0.46
Total	0.92	0.08	0.90	0.10	0.91	0.09	0.91	0.09	0.92	0.08	0.92	0.08	0.92	0.08	0.85	0.15	0.84	0.16	0.73	0.27

Notes: Table A.3 presents variance decomposition of different measures of regional in ation rates and the unemployment gap. "Nat." indicates the proportion of the variance of the relevant variable explained by the common national component, while "Reg." reports the proportion of the variance attributed to the idiosyncratic regional component.

TABLE A.4: Variance decomposition, Spain

	in _L		in _P		in _GL		in _GP		in _F		in _T		in _SV		in _UPI_CG		in _UPI		un_gap_cf	
	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.	Nat.	Reg.
ES11	0.91	0.09	0.91	0.09	0.91	0.09	0.90	0.10	0.91	0.09	0.91	0.09	0.91	0.09	0.89	0.11	0.88	0.12	0.69	0.31
ES12	0.90	0.10	0.91	0.09	0.90	0.10	0.91	0.09	0.91	0.09	0.92	0.08	0.90	0.10	0.86	0.14	0.78	0.22	0.43	0.57
ES13	0.85	0.15	0.80	0.20	0.84	0.16	0.83	0.17	0.82	0.18	0.82	0.18	0.83	0.17	0.87	0.13	0.67	0.33	0.74	0.26
ES21	0.62	0.38	0.74	0.26	0.63	0.37	0.74	0.26	0.70	0.30	0.69	0.31	0.70	0.30	0.71	0.29	0.61	0.39	0.82	0.18
ES22	0.89	0.11	0.92	0.08	0.90	0.10	0.91	0.09	0.91	0.09	0.91	0.09	0.91	0.09	0.81	0.19	0.76	0.24	0.72	0.28
ES23	0.91	0.09	0.88	0.12	0.91	0.09	0.89	0.11	0.91	0.09	0.91	0.09	0.91	0.09	0.81	0.19	0.64	0.36	0.24	0.76
ES24	0.94	0.06	0.95	0.05	0.95	0.05	0.94	0.06	0.95	0.05	0.95	0.05	0.95	0.05	0.91	0.09	0.92	0.08	0.63	0.37
ES30	0.97	0.03	0.97	0.03	0.97	0.03	0.97	0.03	0.97	0.03	0.97	0.03	0.97	0.03	0.91	0.09	0.92	0.08	0.70	0.30
ES41	0.97	0.03	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.96	0.04	0.97	0.03	0.96	0.04	0.85	0.15
ES42	0.92	0.08	0.95	0.05	0.92	0.08	0.95	0.05	0.93	0.07	0.93	0.07	0.94	0.06	0.86	0.14	0.87	0.13	0.92	0.08
ES43	0.87	0.13	0.92	0.08	0.88	0.12	0.92	0.08	0.91	0.09	0.91	0.09	0.92	0.08	0.82	0.18	0.83	0.17	0.28	0.72
ES51	0.97	0.03	0.97	0.03	0.97	0.03	0.97	0.03	0.97	0.03	0.97	0.03	0.97	0.03	0.85	0.15	0.83	0.17	0.87	0.13
ES52	0.95	0.05	0.96	0.04	0.96	0.04	0.96	0.04	0.94	0.06	0.96	0.04	0.94	0.06	0.88	0.12	0.86	0.14	0.89	0.11
ES53	0.91	0.09	0.92	0.08	0.91	0.09	0.93	0.07	0.92	0.08	0.92	0.08	0.92	0.08	0.73	0.27	0.58	0.42	0.65	0.35
ES61	0.97	0.03	0.98	0.02	0.97	0.03	0.98	0.02	0.96	0.04	0.96	0.04	0.97	0.03	0.88	0.12	0.89	0.11	0.87	0.13
ES62	0.89	0.11	0.90	0.10	0.90	0.10	0.89	0.11	0.90	0.10	0.90	0.10	0.90	0.10	0.86	0.14	0.81	0.19	0.74	0.26
ES70	0.72	0.28	0.74	0.26	0.74	0.26	0.75	0.25	0.76	0.24	0.76	0.24	0.77	0.23	0.68	0.32	0.68	0.32	0.75	0.25
Total	0.89	0.11	0.90	0.10	0.90	0.10	0.91	0.09	0.90	0.10	0.90	0.10	0.90	0.10	0.84	0.16	0.79	0.21	0.69	0.31

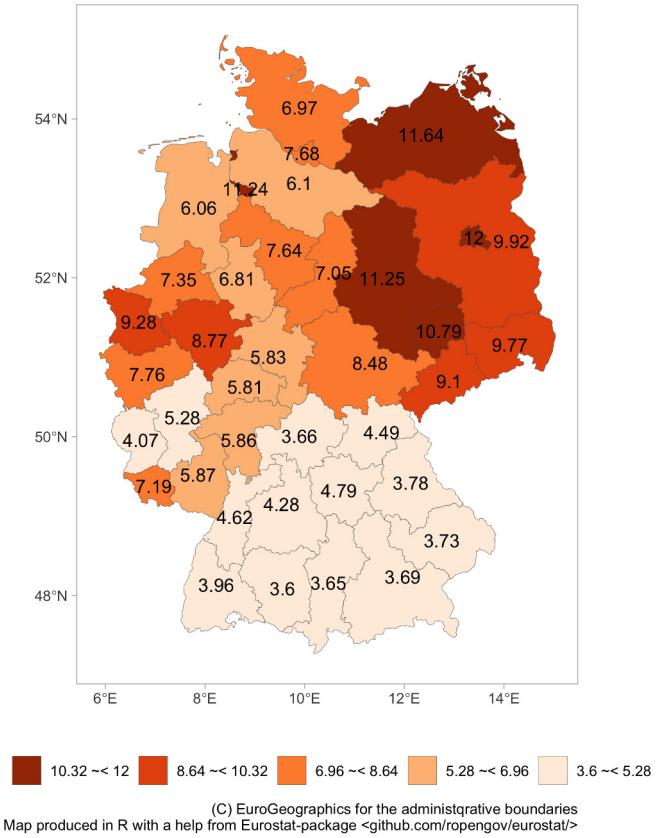
Notes: Table A.4 presents variance decomposition of different measures of regional in ation rates and the unemployment gap. "Nat." indicates the proportion of the variance of the relevant variable explained by the common national component, while "Reg." reports the proportion of the variance attributed to the idiosyncratic regional component.

4.B Additional gures

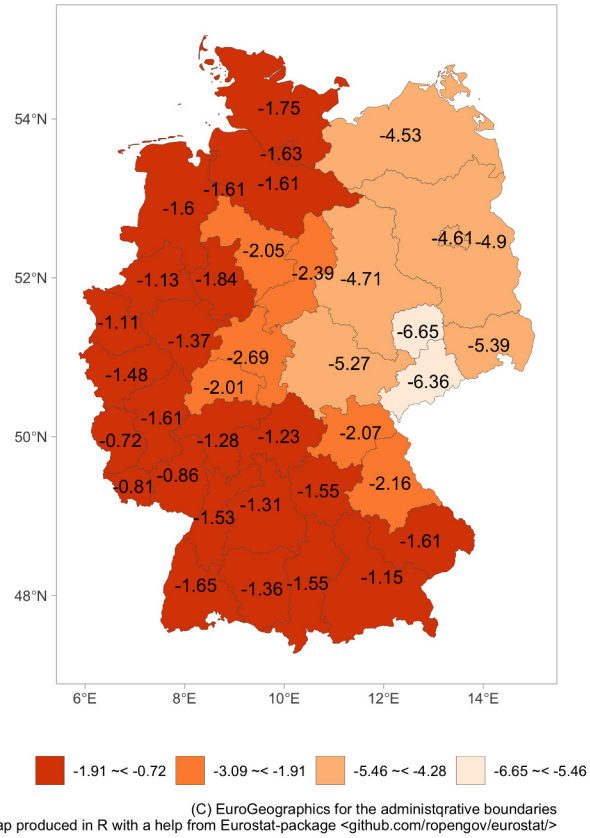
4.B.1 Additional gures for Section "Data and descriptive statistics"

FIGURE 4.B.1: Regional unemployment in Germany

(A) Average quarterly unemployment

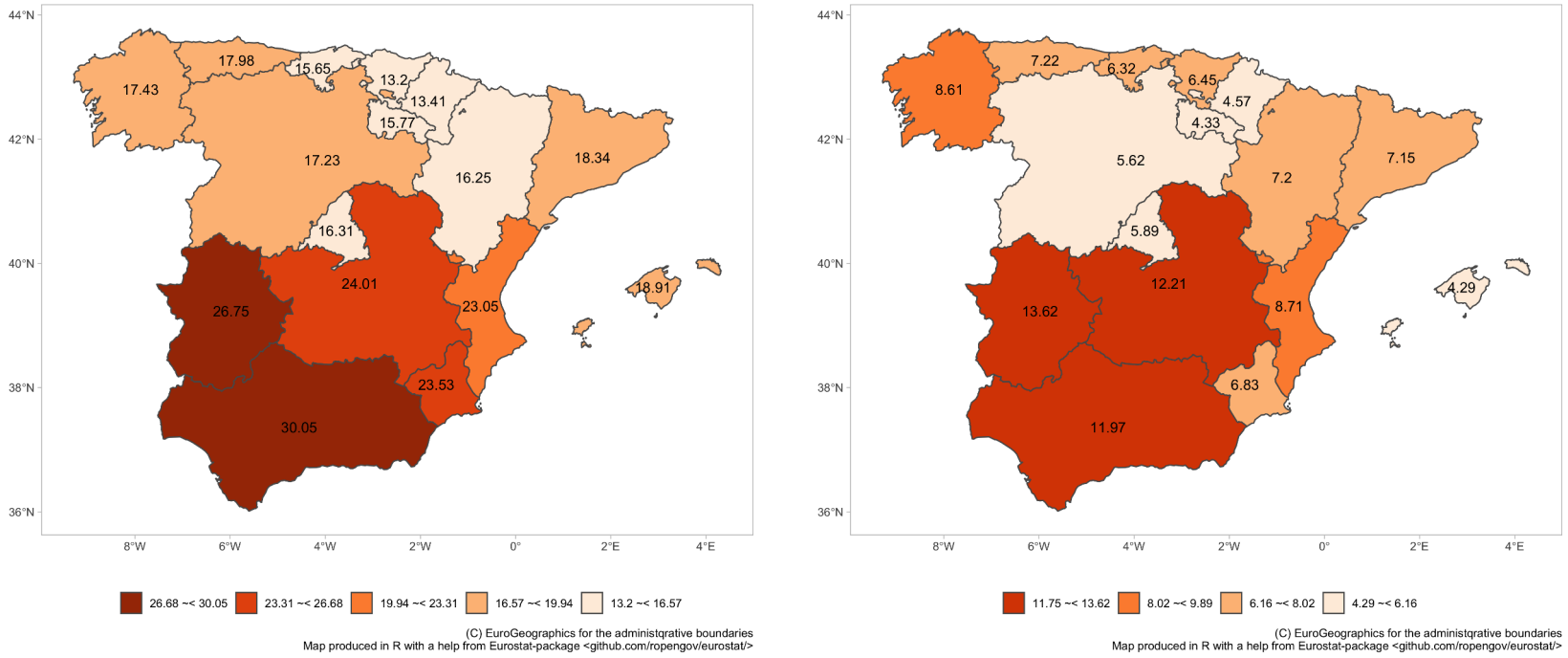


(B) Change in unemployment rate



Notes: Figure 4.B.1a plots average values of quarterly unemployment in German NUTS2 regions. The average is taken over the period Q2 2009 - Q4 2016. Figure 4.B.1b plots changes in regional unemployment rate from the Q2 2009 to Q4 2016.

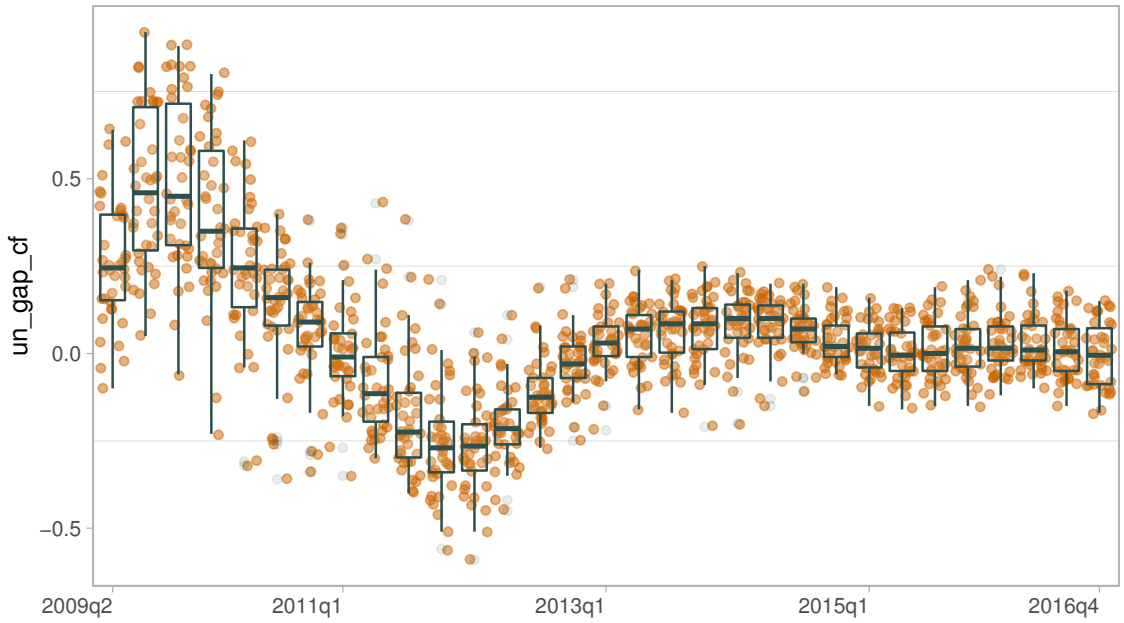
FIGURE 4.B.2: Regional unemployment in Spain



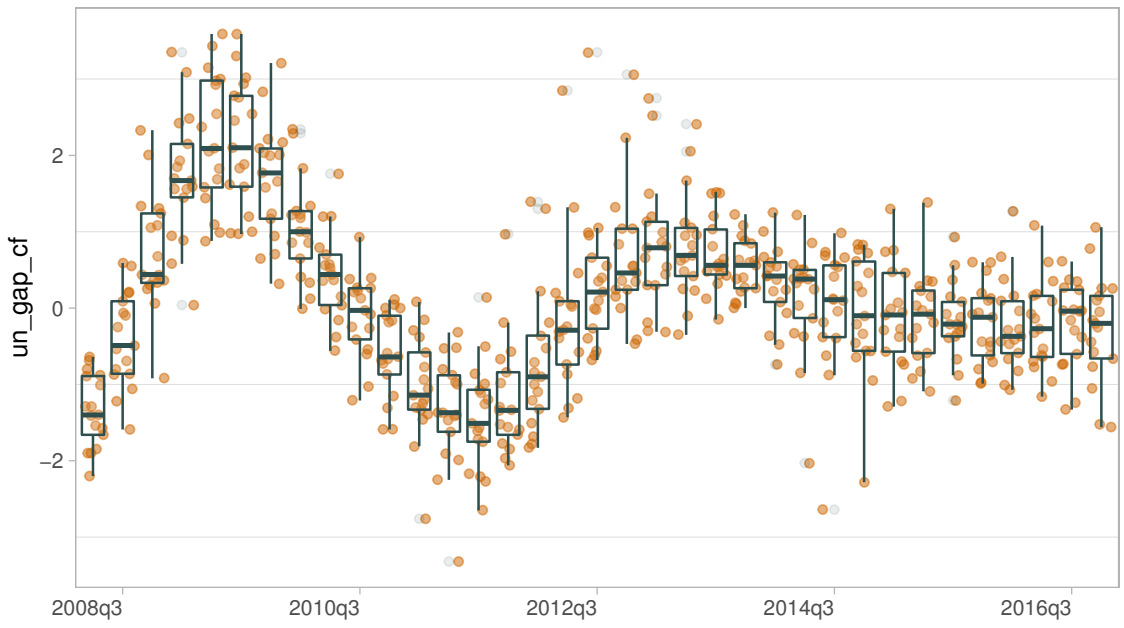
(A) Average quarterly unemployment (B) Change in unemployment rate

Notes: Figure 4.B.2a plots average values of quarterly unemployment in Spanish NUTS2 regions. The average is taken over the period Q2 2008 - Q4 2016. Figure 4.B.2b plots changes in regional unemployment rate from the Q2 2008 to Q4 2016. The Canary Islands have not been plotted on the map for the sake of a more convenient presentation. The corresponding values for the region are: 28.73 and 9.5.

FIGURE 4.B.2: Regional unemployment gaps



(A) Germany

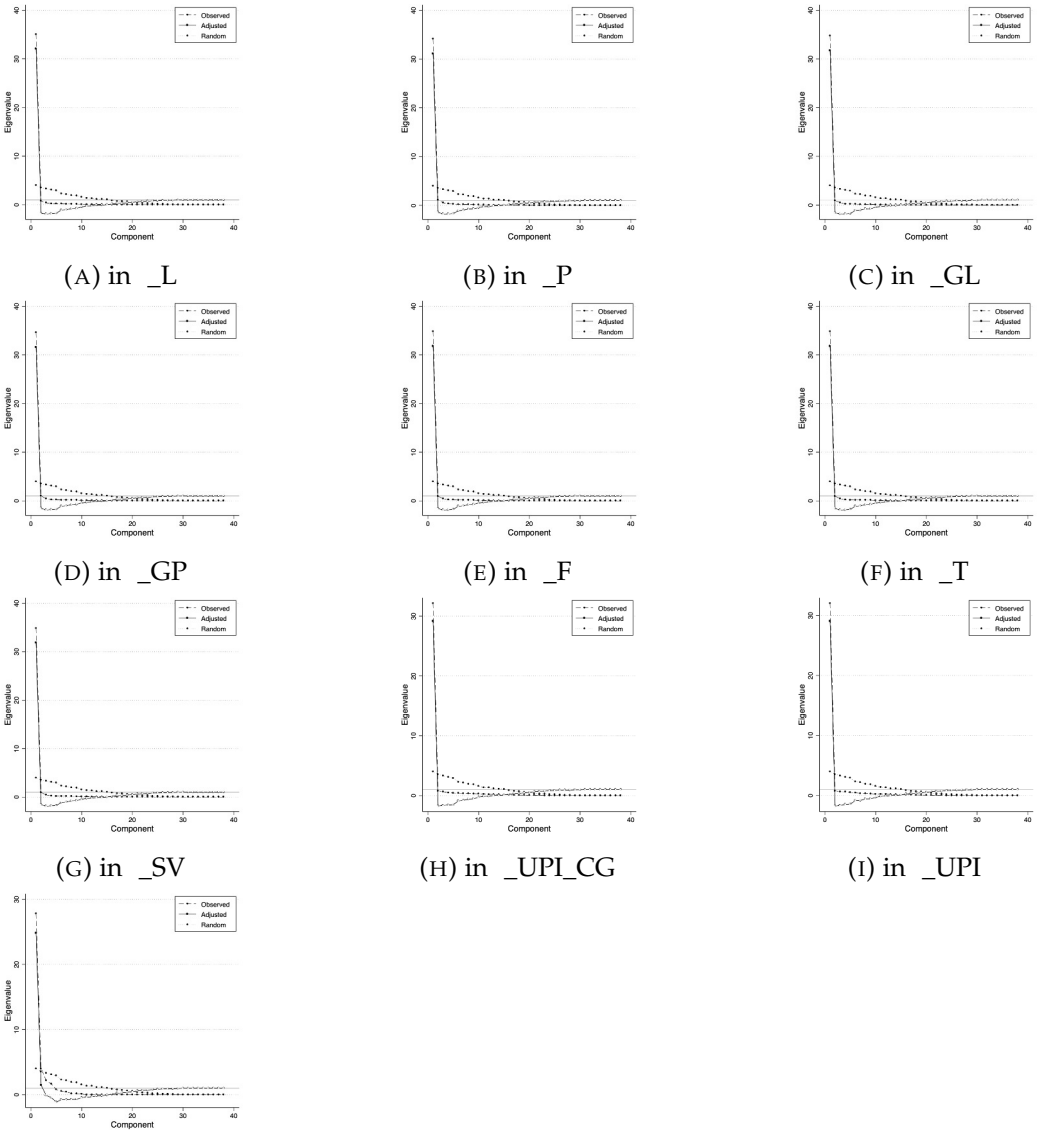


(B) Spain

Notes: Figure 4.B.2 presents box plots of regional unemployment gaps for Germany and Spain, data underlying the box plot is plotted as dots. Regional unemployment gaps are computed using the Christiano-Fitzgerald lter.

4.B.2 Additional gures for Section " Co-movement of regional variables"

FIGURE 4.B.3: Scree plots of eigenvalues, Germany



(A) in _L (B) in _P (C) in _GL

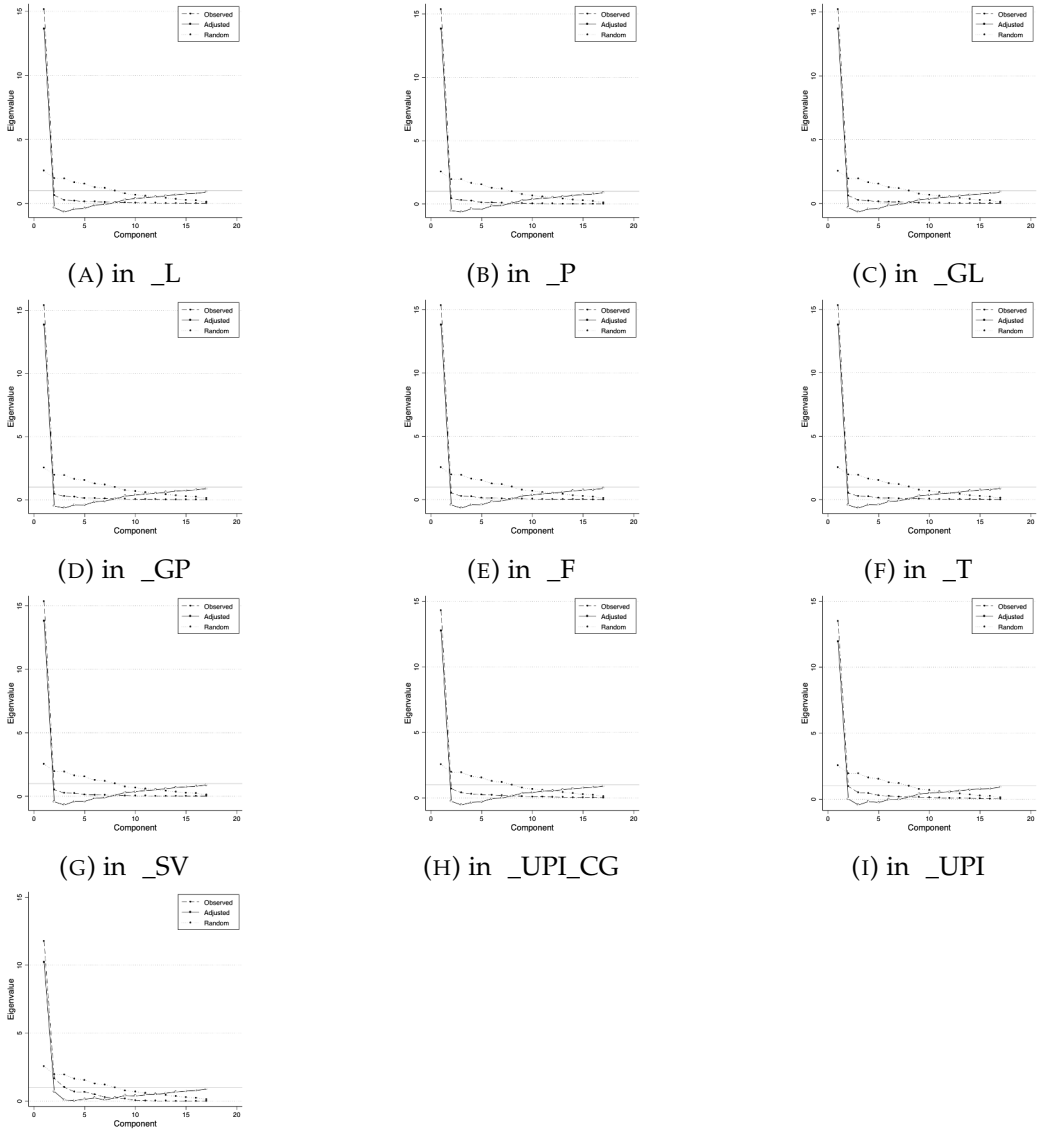
(D) in _GP (E) in _F (F) in _T

(G) in _SV (H) in _UPI_CG (I) in _UPI

(J) un_gap_cf

Notes: Figure 4.B.3 presents scree plots of eigenvalues for different measures of regional in ation rates and regional unemployment gap in Spain. Each panel of Figures 4.B.3a to 4.B.3j plots the eigenvalues resulted from the implementation of the Horn s parallel analysis. "Observed" presents the eigenvalues estimated from the observed data, "Random" presents the eigenvalues estimated from simulated data and "Adjusted" presents the bias-adjusted eigenvalues.

FIGURE 4.B.4: Scree plots of eigenvalues, Spain



(A) in _L (B) in _P (C) in _GL

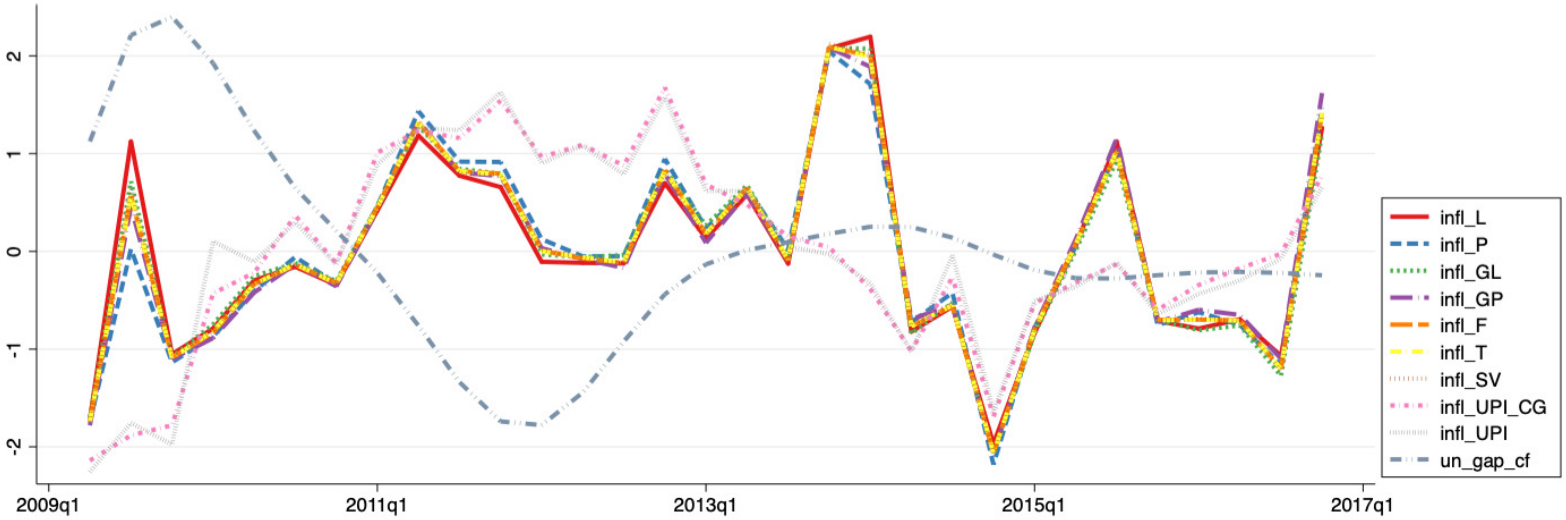
(D) in _GP (E) in _F (F) in _T

(G) in _SV (H) in _UPI_CG (I) in _UPI

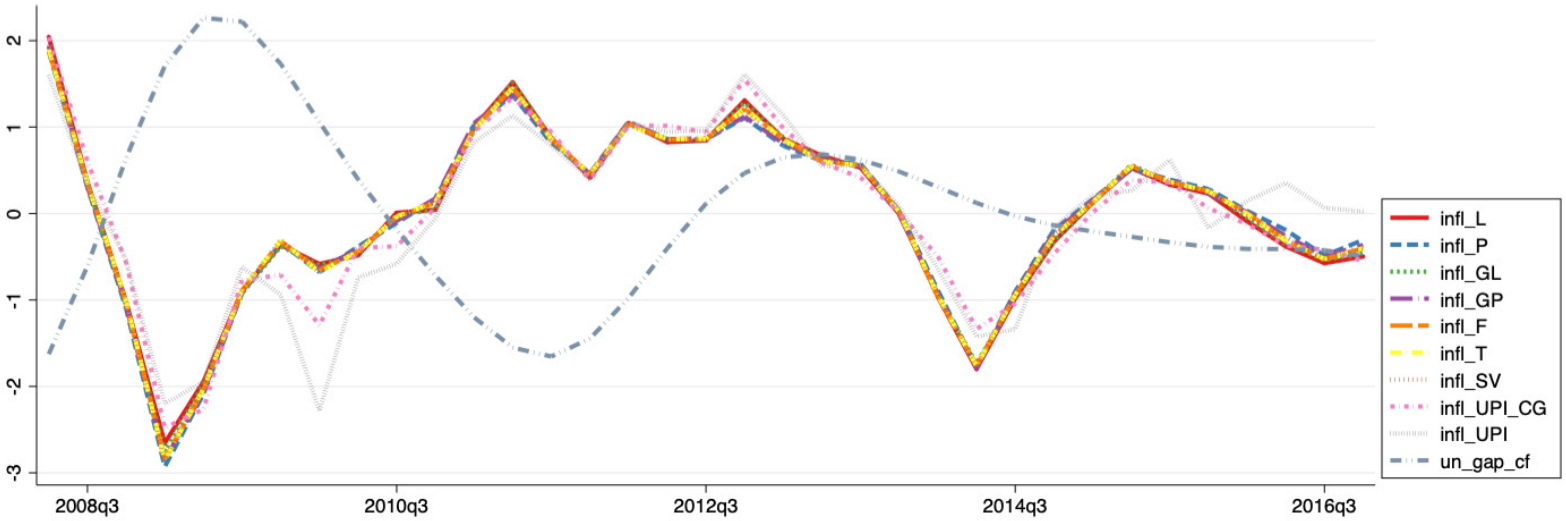
(J) un_gap_cf

Notes: Figure 4.B.4 presents scree plots of eigenvalues for different measures of regional in ation rates and regional unemployment gap in Spain. Each panel of Figures 4.B.4a to 4.B.4j plots the eigenvalues resulted from the implementation of the Horn s parallel analysis. "Observed" presents the eigenvalues estimated from the observed data, "Random" presents the eigenvalues estimated from simulated data and "Adjusted" presents the bias-adjusted eigenvalues.

FIGURE 4.B.5: National factors



(A) Germany



(B) Spain

Notes: Figure 4.B.5 plots estimated national factors for different measures of regional in ation rates and regional unemployment gap. Each panel of regional data is modelled as an approximate dynamic factor model given by Equation (4.5). The national factor for each panel of regional data is estimated as a rst principal component of the data.

References

Ahn, S.C. and A.R. Horenstein (2013). "Eigenvalue Ratio Test for the Number of Factors". In: *Econometrica* 81.3, pp. 1203–1227.

Ahnert, H. and G. Kenny (2004). "Quality Adjustment of European Price Statistics and the Role for Hedonics". In: *ECB Occasional Paper Series*.

Anderson, T.W. and C. Hsiao (1982). "Formulation and Estimation of Dynamic Models Using Panel Data". In: *Journal of Econometrics* 18.1, pp. 47–82.

Arellano, M. and S. Bond (1991). "Some Tests of Speci cation for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". In: *The Review of Economic Studies* 58.2, pp. 277–297.

Argente, D. and M. Lee (2017). "Cost of Living Inequality during the Great Recession". Working Paper.

Bai, J. and S. Ng (2002). "Determining the Number of Factors in Approximate Factor Models". In: *Econometrica* 70.1, pp. 191–221.

Barhoumi, K., O. Darné, and L. Ferrara (2014). "Dynamic Factor Models: A Review of the Literature". In: *OECD Journal: Journal of Business Cycle Measurement and Analysis* 2013.2, pp. 73–107.

Beck, G.W., K. Hubrich, and M. Marcellino (2009). "Regional In ation Dynamics within and across Euro Area Countries and a Comparison with the United States". In: *Economic Policy* 24.57, pp. 142–184.

Beraja, M., E. Hurst, and J. Ospina (2014). "The Regional Evolution of Prices and Wages during the Great Recession". Working Paper.

— (2019). "The Aggregate Implications of Regional Business Cycles". In: *University of Chicago, Becker Friedman Institute for Economics Working Paper* 2019-04.

Blundell, R. and S. Bond (1998). "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models". In: *Journal of Econometrics* 87.1, pp. 115–143.

Boskin, M.J. et al. (1996). *Toward a More Accurate Measurement of In ation*. Tech. rep. Final Report to the Senate Finance Committee, December 4.

Broda, C., E. Leibtag, and D.E. Weinstein (2009). "The Role of Prices in Measuring the Poor s Living Standards". In: *Journal of Economic Perspectives* 23.2, pp. 77–97.

Broda, C. and D.E. Weinstein (2006). "Globalization and the Gains from Variety". In: *The Quarterly Journal of Economics* 121.2, pp. 541–585.

— (2008). "Understanding International Price Differences Using Barcode Data". NBER Working Paper Series No. 14017.

— (2010). "Product Creation and Destruction: Evidence and Price Implications". In: *American Economic Review* 100, pp. 691–723.

Bruno, G.S.F. (2005a). "Approximating the Bias of the LSDV Estimator for Dynamic Unbalanced Panel Data Models". In: *Economics Letters* 87.3, pp. 361–366.

— (2005b). "Estimation and Inference in Dynamic Unbalanced Panel-Data Models with a Small Number of Individuals". In: *The Stata Journal* 5.4, pp. 473–500.

Calonico, S. et al. (2017). "rdrobust: Software for Regression-Discontinuity Designs". In: *The Stata Journal* 17.2, pp. 372–404.

Campbell, J.R. and B. Lapham (2004). "Real Exchange Rate Fluctuations and the Dynamics of Retail Trade Industries on the US-Canada Border". In: *American Economic Review* 94.4, pp. 1194–1206.

Cavallo, A., B. Neiman, and R. Rigobon (2014). "Currency Unions, Product Introductions, and the Real Exchange Rate". In: *The Quarterly Journal of Economics* 129.2, pp. 529–595.

Christiano, L.J. and T. J. Fitzgerald (2003). "The Band Pass Filter". In: *International Economic Review* 44.2, pp. 435–465.

Ciccarelli, M. and B. Mojon (2010). "Global In ation". In: *The Review of Economics and Statistics* 92.3, pp. 524–535.

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

DellaVigna, S. and M. Gentzkow (2019). "Uniform Pricing in U.S. Retail Chainss". In: *The Quarterly Journal of Economics* 134.4, pp. 2011–2084.

Diewert, W. Erwin (1998). "Index number issues in the consumer price index". In: *The Journal of Economic Perspectives* 12.1, pp. 47–58.

Dvir, E. and G. Strasser (2018). "Does Marketing Widen Borders? Cross-Country Price Dispersion in the European Car Market". In: *Journal of International Economics* 112, pp. 134–149.

ECB (2014). "Potential Measurement Issues in Consumer Price Indeces". In: *Monthly Bulletin, April 2014*, pp. 40–42.

Engel, C. and J.H. Rogers (1996). "How Wide Is the Border?" In: *The American Economic Review*, pp. 1112–1125.

— (2004). "European Product Market Integration after the Euro". In: *Economic Policy* 19.39, pp. 348–384.

European Comission (2017). *HICP methodology: Statistics Explained*. Tech. rep.

— (2018a). *Harmonised Index of Consumer Prices (HICP)*. URL: http://ec.europa.eu/eurostat/cache/metadata/en/prc_hicp_esms.htm.

— (2018b). *The EU s Main Investment Policy*. URL: http://ec.europa.eu/regional_policy/en/policy/what/investment-policy/.

Feenstra, R.C. (1994). "New Product Varieties and the Measurement of International Prices". In: *The American Economic Review*, pp. 157–177.

Fischer, C. (2012). "Price Convergence in the EMU? Evidence from Micro Data". In: *European Economic Review* 56.4, pp. 757–776.

Fitzgerald, T. and J.P. Nicolini (2014). "Is There a Stable Relationship between Unemployment and Future In ation? Evidence from U.S. Cities". Federal Reserve Bank of Minneapolis, Working Paper.

Forni, M. and L. Reichlin (2001). "Federal Policies and Local Economies: Europe and the US". In: *European Economic Review* 45.1, pp. 109–134.

Gilbert, L. (2018). "Gains from Product Variety and the Local Business Cycle". Working Paper.

Gopinath, G. et al. (2011). "International Prices, Costs, and Markup Differences". In: *American Economic Review* 101.6, pp. 2450–86.

Gorodnichenko, Y. and Linda L 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. 219–41.

Handbury, Jessie (2019). "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities".

Handbury, Jessie and David E. Weinstein (2015). "Goods Prices and Availability in Cities". In: *The Review of Economic Studies* 82.1, pp. 258–296.

Hooper, P., F.S. Mishkin, and A. Su (2019). "Prospects for In ation in a High Pressure Economy: Is the Phillips Curve Dead or is It Just Hibernating?" National Bureau of Economic Research.

Horn, J.L. (1965). "A Rationale and Test for the Number of Factors in Factor Analysis". In: *Psychometrika* 30.2, pp. 179–185.

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

Hottman, Colin J, Stephen J Redding, and David E Weinstein (2016). "Quantifying the sources of rm heterogeneity". In: *The Quarterly Journal of Economics* 131.3, pp. 1291–1364.

ILO, IMF, OECD, Eurostat, UNECE, The World Bank (2020). *Consumer price index manual: concepts and methods*. Washington: IMF.

Imbens, G. and K. Kalyanaraman (2012). "Optimal Bandwidth Choice for the Regression Discontinuity Estimator". In: *The Review of Economic Studies* 79.3, pp. 933–959.

Imbens, G. and T. Lemieux (2008). "Regression Discontinuity Designs: A Guide to Practice". In: *Journal of Econometrics* 142.2, pp. 615–635.

Imbs, J. et al. (2005). "PPP Strikes Back: Aggregation and the Real Exchange Rate". In: *The Quarterly Journal of Economics* 120.1, pp. 1–43.

— (2010). "One TV, One Price?" In: *Scandinavian Journal of Economics* 112.4, pp. 753–781.

Isard, P. (1977). "How Far Can We Push the Law of One Price?" In: *The American Economic Review* 67.5, pp. 942–948.

Issever Grochová, L. and P. Rozmahel (2015). "On the Ideality of Filtering Techniques in the Business Cycle Analysis Under Conditions of European Economy". In: *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* 63.3, pp. 915–926.

Judson, R.A. and A.L. Owen (1999). "Estimating Dynamic Panel Data Models: a Guide for Macroeconomists". In: *Economics Letters* 65.1, pp. 9–15.

Kaiser, H.F. (1960). "The Application of Electronic Computers to Factor Analysis". In: *Educational and Psychological Measurement* 20.1, pp. 141–151.

Kiley, M.T. (2015). "An Evaluation of the In ationary Pressure Associated with Short-and Long-Term Unemployment". In: *Economics Letters* 137, pp. 5–9.

Kiviet, J.F. (1995). "On Bias, Inconsistency, and Ef ciency of Various Estimators in Dynamic Panel Data Models". In: *Journal of Econometrics* 68.1, pp. 53–78.

Kose, A., C. Otrok, and C. Whiteman (2003). "International Business Cycles: World, Region, and Country-Speci c Factors". In: *American Economic Review* 93.4, pp. 1216–1239.

Kulikov, D. (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.

Kumar, A. and P.M. Orrenius (2016). "A Closer Look at the Phillips Curve Using State-Level Data". In: *Journal of Macroeconomics* 47, pp. 84–102.

Leamer, E.E. (1981). "Is It a Demand Curve, or is It a Supply Curve? Partial Identi cation through Inequality Constraints". In: *The Review of Economics and Statistics*, pp. 319–327.

Lecznar, J. and A.V. Smith (2018). "Geographic Aggregation and the Measurement of Real Consumption Growth and Volatility". Working Paper.

Lee, D.S. and T. Lemieux (2010). "Regression Discontinuity Designs in Economics". In: *Journal of Economic Literature* 48.2, pp. 281–355.

Li, Y., B. Gordon, and O. Netzer (2018). "An Empirical Study of National vs. Local Pricing by Chain Stores under Competition". In: *Marketing Science* 37.5, pp. 812–837.

McAfee, R.P. (2008). "Price Discrimination". In: *Issues in Competition Law and Policy* 1, pp. 465–484.

McLeay, Mi. and S. Tenreyro (2020). "Optimal in ation and the identi cation of the Phillips curve". In: *NBER Macroeconomics Annual* 34.1, pp. 199–255.

Mendoza, R.U. (2011). "Why Do the Poor Pay More? Exploring the Poverty Penalty Concept". In: *Journal of International Development* 23.1, pp. 1–28.

Monacelli, T. and L. Sala (2009). "The International Dimension of In ation: Evidence from Disaggregated Consumer Price Data". In: *Journal of Money, Credit and Banking* 41, pp. 101–120.

Mussa, M. (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.

Neely, C.J. and D.E. Rapach (2011). "International Comovements in In ation Rates and Country Characteristics". In: *Journal of International Money and Finance* 30.7, pp. 1471–1490.

Nevo, A. and A. Wong (2019). "The Elasticity of Substitution between Time and Market Goods: Evidence from the Great Recession". In: *International Economic Review* 60.1, pp. 25–51.

Nickell, S. (1981). "Biases in Dynamic Models with Fixed Effects". In: *Econometrica: Journal of the Econometric Society*, pp. 1417–1426.

Of cer, L.H. (1976). "The purchasing-power-parity theory of exchange rates: A review article". In: *IMF Staff Papers* 23.1, pp. 1–60.

Onatski, A. (2009). "Testing Hypotheses about the Number of Factors in Large Factor Models". In: *Econometrica* 77.5, pp. 1447–1479.

— (2010). "Determining the Number of Factors from Empirical Distribution of Eigenvalues". In: *The Review of Economics and Statistics* 92.4, pp. 1004–1016.

Parker, M. (2017). "Global In ation: The Role of Food, Housing and Energy Prices". ECB Working Paper.

Redding, S.J. and D.E. Weinstein (2019). "Measuring Aggregate Price Indexes with Taste Shocks: Theory and Evidence for CES Preference". In: *Quarterly Journal of Economics, forthcoming*.

Redding, Stephen J. and David E. Weinstein (2020). "Measuring aggregate price indices with taste shocks: Theory and evidence for CES preferences". In: *The Quarterly Journal of Economics* 135.1, pp. 503–560.

Reiff, A. and F. Rumler (2014). "Within-and cross-country price dispersion in the euro area". ECB Working Paper Series No. 1742.

Rogoff, K. (1996). "The Purchasing Power Parity Puzzle". In: *Journal of Economic Literature* 34.2, pp. 647–668.

Skovron, C. and R. Titiunik (2015). "A Practical Guide to Regression Discontinuity Designs in Political Science". In: *American Journal of Political Science* 2015, pp. 1–36.

Stock, J.H. and M.W. Watson (2002). "Forecasting Using Principal Components from a Large Number of Predictors". In: *Journal of the American Statistical Association* 97.460, pp. 1167–1179.

— (2016). "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics". In: *Handbook of Macroeconomics*. Vol. 2. Elsevier, pp. 415–525.

Stroebel, J. and J. Vavra (2019). "House Prices, Local Demand, and Retail Prices". In: *Journal of Political Economy* 127.3, pp. 1391–1436.