



EUROPEAN CENTRAL BANK

EUROSYSTEM



In 2014 all ECB publications feature a motif taken from the €20 banknote.



WORKING PAPER SERIES

NO 1742 / NOVEMBER 2014

WITHIN- AND CROSS-COUNTRY PRICE DISPERSION IN THE EURO AREA

Ádám Reiff and Fabio Rumler

**GROCERY PRICES IN THE EURO AREA:
FINDINGS FROM INFORMAL ESCB
EXPERT GROUP SET-UP TO ANALYSE A
DISAGGREGATED PRICE DATASET**

NOTE: This Working Paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

Grocery prices in the euro area: Findings from informal ESCN expert group to analyse a disaggregated price dataset

This paper was prepared as part of a Eurosystem project group established to analyse a large-scale disaggregated dataset on grocery prices in the euro area. This proprietary dataset was obtained as a follow up to the 2011 Eurosystem Structural Issues Report (SIR) entitled “Structural features of the distributive trades and their impact on prices in the euro area”. The main motivation for obtaining these data was to enable the analysis of a variety of issues that was previously not possible owing to data limitations. More specifically (i) analysis of Single Market issues and quantification of border effects (ii) measuring the impact of competition – both at the producer and retail level – on consumer price levels and (iii) consider potential implications for inflation measurement arising from structural changes in retail sector such as the growing importance of discounters and private label brands.

The data were obtained from Nielsen, an international market information and measurement company. The dataset is multi-dimensional with approximately 3.5 million observations each for the price, value and volume variables across a number of dimensions (13 countries; approximately 45 product categories; approximately 70 regions; approximately 10 store types on average per country; 4 brands per product category and 3 stock-keeping units - skus - per brand). The data are generally collected from barcode scanners. These cross country data are unique in a number of respects, in particular in that (a) there are data on average price levels across regions within countries, (b) there is information on both prices and volumes, and (c) there are data on aggregated private label sales and prices. The data have been cross-checked against HICP and PPP data and found to be highly congruent.

The expert group was chaired by Bob Anderton (ECB) and Aidan Meyler (ECB) acted as Secretary. We are also grateful to Stefanos Dimitrakopoulos (Warwick University) who, whilst at the ECB as a trainee, provided invaluable assistance in compiling and working with the database.

Preliminary results from the project group were initially presented at an informal Eurosystem workshop which took place in Frankfurt on 22 November 2013. Apart from the members of the expert group a small number of external participants were invited to the workshop. The following participants acted as discussants: Mario Crucini (Professor of Economics, Vanderbilt University, and Senior Fellow, Globalization and Monetary Policy Institute, Dallas Fed); Daniel S Hosken, US Federal Trade Commission (Deputy Assistant Director); Jarko Pasanen, Eurostat (Team Leader: Price Statistics, Purchasing Power Parities, Housing Statistics) and Thomas Westermann, European Central Bank (Head of Section: Prices and Costs). The refereeing process for the papers from this project was coordinated by the Secretary of the expert group (Aidan Meyler).

As the dataset is proprietary, it cannot be made available to outside researchers. Thus this paper is released in order to make the working papers and accompanying research carried out by the expert group publicly available. Additional papers from the project group will be published as they are finalised. Any queries regarding the project may be addressed to Aidan Meyler (aidan.meyler@ecb.europa.eu).

Acknowledgements:

We would like to thank Mario Crucini, Attila Ratfai, Aidan Meyler and other group members of the Eurosystem Nielsen Price Analysis Project, the participants of the ECB Workshop “Analysis and Understanding of Grocery Prices in the Euro Area” and the participants of an internal seminar at the Oesterreichische Nationalbank for valuable comments and suggestions. Eniko Gabor provided excellent research assistance. The views expressed in this paper are our own and do not necessarily reflect those of the Magyar Nemzeti Bank, of the Oesterreichische Nationalbank, and of the European Central Bank.

Ádám Reiff

European Central Bank and Magyar Nemzeti Bank; e-mail: adam.reiff@gmail.com

Fabio Rumler

Oesterreichische Nationalbank; e-mail: fabio.rumler@oenb.at

© European Central Bank, 2014

Address	Kaiserstrasse 29, 60311 Frankfurt am Main, Germany
Postal address	Postfach 16 03 19, 60066 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Internet	http://www.ecb.europa.eu

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors. This paper can be downloaded without charge from <http://www.ecb.europa.eu> or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=2528399. Information on all of the papers published in the ECB Working Paper Series can be found on the ECB’s website, <http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html>

ISSN	1725-2806 (online)
ISBN	978-92-899-1150-4
DOI	10.2866/3441
EU Catalogue No	QB-AR-14-116-EN-N (online)

Abstract

Using a comprehensive data set on retail prices across the euro area, we analyze within- and cross-country price dispersion in European countries. First, we study price dispersion over time, by investigating the time-series evolution of the coefficient of variation, calculated from price levels. Second, since we find that cross-sectional price dispersion by far dominates price dispersion over time, we study price dispersion across space and investigate the role of geographical barriers (distance and national borders). We find that (i) prices move together more closely in locations that are closer to each other; (ii) cross-country price dispersion is by an order of magnitude larger than within-country price dispersion, even after controlling for product heterogeneity; (iii) a large part of cross-country price differences can be explained by different tax rates, income levels and consumption intensities. In addition, we find some indication that price dispersion in the euro area has declined since the inception of the Monetary Union.

JEL codes: E31, F41

Keywords: price dispersion, international relative prices, border effect

Non-technical summary

This paper uses a new, disaggregated data set (collected by AC Nielsen) on prices of 45 products between 2008-2011 to evaluate price dispersion within the euro area. One advantage of the data is that it covers approximately 70 regions in 13 euro area countries, which makes it suitable to study both within- and cross-country price dispersions and evaluate the extent to which prices have become harmonized across countries almost a decade after the introduction of the Euro. A further advantage of the data is that although it is international, it is not affected by exchange rate fluctuations.

The paper first analyzes the evolution of price dispersion over time, asking the question of whether intra- or international price dispersion has decreased, as was perhaps expected from the introduction of the common currency. We find some evidence that cross-country price dispersion decreases over time. The decline is statistically significant for about half of the products, but numerically small. In general, we find that approximately 90% of price variation is due to cross-location price differences, and only 10% can be attributed to time-series variation. This might explain why we only find a numerically small decrease in price dispersion over time.

Motivated by the larger importance of cross-location price variation, we then turn to the analysis of cross-location price differences. We have two sets of results: one regards *relative price* variation, when we analyze how closely prices move together across locations. The other regards *price level* dispersion, when we directly compare price levels across locations.

Regarding cross-location relative price variations, we first find that geographic distance between locations matters: prices at different locations move together more strongly if the locations are closer to each other. We also find that price co-movement is much stronger if the different locations are within the same country, which means that national borders continue to matter.

Regarding cross-location price level dispersion, we first find that products are extremely heterogeneous in different countries: we could only identify 13 product varieties (out of 180 available brands in our data set) that have significant market shares in at least six different countries. But even for this narrow subset of exactly identical products, the average cross-country price level difference is around 6 times larger than the average within-country price level difference (20% vs 3.5%). Around 40% of this difference can be explained with factors like different tax rates, per capita GDP, local unemployment rates and consumption

intensities, but even after correction for these factors there remains a significant difference between cross-country and within-country price level dispersion. This remaining difference might be due to omitted variables (like net wages or the share of different store types), measurement errors of explanatory variables, country heterogeneity with respect to domestic price dispersion, or true border effects.

Finally, we use a similar data set covering the years 2000-2003 (also provided by AC Nielsen for the European Commission) to compare our results with a time period close to the introduction of the Euro. Overall, we find some weak evidence of declining price level dispersion over time: both the cross-country coefficient of variation of price levels, and pair-wise average price level differences have decreased somewhat from 2000-2003 to 2008-2011 (although this decline is numerically small). On the other hand, we find that price co-movement between locations was actually stronger between in the earlier than in the later data period.

1 Introduction

In a frictionless world, prices of identical products would be the same across locations and the Law of One Price (LOP) would hold: if prices were different in two locations, excess demand and supply in the cheap and expensive location would eventually equalize them. Empirical evidence on cross-location prices suggests that the world is not frictionless: there are large and persistent differences between prices at different locations.

Isard (1977) shows that for a set of narrowly defined manufactured goods in the U.S., Germany and Japan deviations from the LOP can persist for several years and that exchange rate changes play a key role in distorting relative international prices. According to Haskel and Wolf (2001) deviations from the LOP can be up to 50 % and persist for more than 10 years. Asplund and Friberg (2001) find that even in Scandinavian duty-free outlets where prices are quoted in different currencies and thus the potential for arbitrage would be relatively strong deviations from the LOP can persist. Also when controlling for observed and unobserved product heterogeneity, Lach (2002) still finds a considerable and persistent degree of price dispersion among homogenous consumer goods. Considering over 1,800 consumer goods and services in EU countries in 5-year intervals from 1975 to 1990, Crucini et al. (2005) show that price dispersion within the EU has not declined over the time period investigated.

The reasons for this violation of the LOP are numerous. In reality, products are not exactly identical across locations (Broda and Weinstein, 2008), or even if they are identical, they might be perceived differently by consumers. Ghosh and Wolf (1994) consider the prices of *The Economist* magazine in 12 countries and find pricing to market as well as menu costs to be the main explanatory factors of the failure of the LOP. Furthermore, transportation frictions prevent prices at different regions from being equalized, and imply closer co-movement of prices in closer locations (Engel and Rogers, 1996). Different local conditions (like higher rental rates or wages) might lead to different marginal cost of selling the products, which should lead to price differences. Local market structures and the degree of local competition might also differ, which can give rise to different mark-ups and prices. There are many more location-specific factors that could lead to different prices across space.

In this empirical paper we use a new, highly disaggregated data set (collected by AC Nielsen) on prices and quantities for 45 products between 2008-2011 to evaluate price dispersion within the Euro Area. Our data covers 70 regions in

13 countries, which makes it suitable to study both within- and cross-country price differences. An advantage of the common currency is that although our data set is international, it is not influenced by exchange rate variations.

We first analyze price dispersion measures over time, asking the question of whether intra- or international price dispersion has decreased, as was perhaps expected from the introduction of the common currency. Then we focus on the cross-sectional dispersion of prices, by calculating various relative price dispersion measures, comparing within- and cross-country price dispersion, and estimating the effect of geographic distance and national borders.

We find some evidence that cross-country price dispersion decreases over time. The decline is statistically significant for about half of the products, but numerically small. In addition, based on a similar dataset of AC Nielsen ranging from 2000 to 2003 we also find that price dispersion has on average declined between 2003 and 2008. In terms of cross-location price differences, we have results both for relative price *changes* and price *level* differences. For relative price changes, distance matters: we find evidence that price co-movement is stronger between locations that are closer to each other, and are within the same country, i.e. we find a strong national border effect in relative price changes. For price level differences, we find that for a subset of exactly identical products, the average price difference for two randomly selected regions of two different countries is around 20%, while the same within-country number is around 3.5%; hence the estimated border effect is huge.

We explain around 25% (or 5 percentage points of the 20 percentage points difference) of observed international price differences, and around 43% of the estimated border effect by differences in consumption intensities, per capita income and national tax rates. Quantitatively, value-added tax differences matter the most: they account for around 10% of the international price differences and 20% of the estimated border effects. We conclude that there are large international price differences left within the Eurozone that we are unable to explain.

The remainder of this paper is structured as follows. In the next section we briefly describe our dataset. Section 3 presents and analyses measures of price dispersion over time while section 4 does so for price dispersion across locations. Within section 4 we also present regression results on the determinants of relative price variation and price level differences in the euro area. Section 5 compares our results with those obtained from a similar dataset for the period 2000-2003 and section 6 concludes the paper.

2 Data

For the analysis, we use a disaggregated price data set provided by AC Nielsen to the Eurosystem.¹ The data set itself is three dimensional, with several sub-dimensions. The three main dimensions are countries, product categories and time. In terms of countries, we have 13 euro area countries, which represent the entire euro area (as of 2013) minus Cyprus, Luxemburg, Malta and Finland. In terms of product categories, data is available for 45 product categories, mostly food, consumer durables or personal care items. In terms of time, we have monthly data² for a 3-year period for each country, beginning in late 2008. The exact data periods are varying across countries, but there is at least a 30-months overlap for each country pair.

Out of the several sub-dimensions, we mention two which we heavily use. The first of such sub-dimensions is regions: there are around 70 regions (their number varies somewhat across products) within the 13 countries. These regions are defined by the national Nielsen affiliates, and generally do not correspond to official NUTS-2 or NUTS-3 regions. Nevertheless, with the information from Nielsen, we matched Nielsen-regions and official NUTS-2 and NUTS-3 regions, so that we can obtain Nielsen region-specific data from official Eurostat regional macro data. The unit of observation in this paper is regions, and the regional sub-dimension allows us to study price dispersions both within and across countries.

The second important sub-dimension is within product categories: for each product, we have 4 brands, and within brands, we have 3 different stock keeping units (SKU) or equivalently denoted pack sizes. A driving principle of brand selection in the dataset was to have two pan-European and two (most popular) local brands.³ Although in principle we would like to focus on the most popular brands within each country, whenever a pan-European brand (i.e. a brand that is available in most countries) could be identified, it was selected irrespective of local market shares. For example, in case of the product category beer, Heineken is a pan-European brand, so we have data about Heineken from all countries, even if the local market share is small in some countries.

¹For a detailed description of the data set, see Meyler (2013). In this paper we just summarize briefly the information that is most relevant for us. In section 5 we will also use a similar dataset compiled by AC Nielsen for the period 2000-2003.

²Initially, data frequency was different in different countries. With a series of initial transformations, we created a monthly data set. Details of these are in Meyler (2013).

³For a full list of pan-European brands see Table 1; the 45 product categories are listed in Table 14 in the Appendix.

Within brands, we have information about the three most popular varieties (or SKUs). An example of the variety is Milka Alpenmilch 100g regular chocolate. The basis of selection of particular varieties was local market share, which means that these varieties are generally not comparable internationally. In this paper, whenever needed, we will treat this product heterogeneity problem by selecting a narrow sub-set of products which are identical across countries.

Data is generally available on sales value, sales volume and price, the latter one usually in terms of both pack price and equivalised unit price (e.g. price per kg or price per liter). We also note that price data are derived from information on sales values and sales volumes, i.e. prices are not individual transaction prices but average unit prices over the month observed.

3 Price dispersion over time

In this section we present the results on descriptive statistics of price dispersion calculated over time.

3.1 Measurement

A frequently used measure of price dispersion which can be calculated over time is the coefficient of variation (CV) defined as the standard deviation of prices over their mean.⁴ In order to give a first indication of a possible border effect we calculate both the across- and within-country CV, for all pan-European brands in our dataset.

A general problem in our dataset is that at all levels of aggregation (the product-, brand- and SKU-level) homogeneity of product specifications cannot be guaranteed. In particular for the local brands, we have to accept that there can be large differences in quality across countries which hamper the comparability of prices. This also translates into heterogeneity at the product level given that products are composed of quite different brands across countries. However, there are 47 pan-European brands in our dataset which are equal in most countries (see Table 1 for a list of the pan-European brands),⁵ but even for

⁴Other measures of price dispersion are the standard deviation, the mean or median absolute deviation and distribution measures such as the interquartile range. However, they are all dependent on the scale of the variable while the CV is dimensionless. Apart from these measures, β -convergence (the β -coefficient of a regression of prices on their initial values) is sometimes also used in the literature to describe price dispersion over time.

⁵Different products have different numbers of pan-European brands. For example, the product “Beer” and “Chewing gum” both have only one pan-European brand (Heineken and

them SKUs could still be different across countries, so that there will always be differences in product specifications at the very disaggregate level of the data. Therefore, our strategy is to calculate within- and across-country price dispersion only for the pan-European brands.⁶ By doing so, we hope to “aggregate out” cross-country quality differences in SKUs to some degree and also that these differences remain constant over time. Further, to rule out the effect of differences in package sizes across countries we calculate the CV for prices per unit.

Across countries the CV is directly calculated from the average prices of each pan-European brand at the total country level, i.e. it measures the variation of prices between at most 13 countries. The within-country CV is calculated as the variation of prices of pan-European brands at the total product level across different locations within each country and then averaged over countries. Due to the unbalanced nature of our data at the beginning and at the end of the sample period we calculated all statistics for a common sample ranging from Dec 2008 to Sep 2011.

In addition, to provide a comparable measure of price dispersion within and across countries, we decompose the total variance of prices across different regions into the variance of prices between regions within the same country and between regions in different countries. To do so, we fit a multi-level model for prices per unit with random intercepts at the country level. The so-called intra-class correlation (ICC, sometimes also referred to as intra-cluster correlation), which is a standard diagnostic of random intercepts models, gives the proportion of the total variance that is explained by the higher level unit, in our case the country level, compared to the variance at the lower level after controlling for the higher level, i.e. between locations in the same country.⁷ We calculate the ICC for all pan-European brands to get an idea by how much more prices vary across countries than within countries.

Mentos, respectively). On the other hand, the product “Shampoo” has four pan-European brands (Elseve/Elvital, Head & Shoulders, Fructis and Pantene). This explains why the number of pan-European brands is larger than the number of products.

⁶By scrutinizing the data we identified 13 SKUs which are exactly identical in a sufficiently large number of countries. To check whether non-homogeneity of products has a large effect on price dispersion we calculate the cross- and within-country CV also for these SKUs.

⁷Specifically, we estimate the following random intercepts model for each pan-European brand: $P_{ijt} = \mu + \alpha_j + \varepsilon_{ijt}$, where P is unit prices of the specific brand, i is the region identifier and j is the county identifier. The ICC is then defined as the ratio of the variance of the country-specific random intercepts ($Var(\alpha)$) and the total variance ($Var(\alpha) + Var(\varepsilon)$).

3.2 Results

The results for the average CV across and within countries along with the intra-class correlation for all 47 pan-European brands are given in Table 1. The time series of the cross- and within-country CVs are depicted in the 47 graphs of Figures 1-5 in the Appendix.

As expected from our earlier discussion, we find that price dispersion across countries as measured by the CV is by an order of magnitude larger than price dispersion within countries. For all products – with the notable exception of bouillon and cigarettes – the CV across countries is between 3 times (Uncle Ben’s rice) and 22 times (Sensodyne toothpaste) larger than within countries. However, these large differences have to be interpreted in the light of possible quality differences across countries, as mentioned before. Different taxation, competition levels and local cost structures may additionally contribute to the large dispersion across countries. Thus, the descriptive results on price dispersion already indicate that there might be a remaining effect relating to national borders, but this can only be assessed properly when controlling for other factors contributing to international price differences in a regression analysis, which will be done in later sections.

As already mentioned, bouillon and cigarettes stand out as the products for which price dispersion is extremely larger across than within countries (up to 60 times for Marlboro cigarettes). In the case of cigarettes this is most likely due to price administration at the national level which drives down within-country price dispersion to almost 0. For bouillon we observe an exceptionally high price dispersion across countries. This could be due to the fact the units can be different across varieties (sometimes milliliters sometimes grams) which would distort price variation measures.

The intra-class correlation reported in the last column of the Table 1 reveals that for many brands the variance of prices across countries is more than 90% of the total variance of prices across all locations. It is highest for bouillon and dishwasher tablets (0.99) and lowest for Marlboro cigarettes (0.71) and Nestle cereals (0.76).⁸ Overall, these figures confirm that national borders are

⁸Note that our specification for calculating the intra-class correlation implies that the estimated country effects are constant over time. So any time variation in the price level will increase the variance of the idiosyncratic error term, ε_{ijt} , hence it will reduce the reported ICC. As we discuss in subsection 3.3, this time variation is usually much smaller than cross-location variation, so it should not distort our results in any significant way. One exception is cigarettes, where national price regulation implies that there is effectively no cross-location variation within the countries, so the only reason why the variance of ε_{ijt} can be positive is

Table 1: Average coefficient of variation (CV) across and within countries and intra-class correlation

Product	Brand	CV across	CV within	ICC
Beer	Heineken	0.25	0.03	0.93
Bouillon	Knorr	1.33	0.04	0.99
Bouillon	Maggi	1.38	0.04	0.99
Cat food	Whiskas	0.18	0.04	0.89
Cat food	Gourmet	0.26	0.03	0.95
Cereals	Kellogg's Cornflakes	0.16	0.03	0.91
Cereals	Kellogg's Special K	0.13	0.02	0.89
Cereals	Nestle Fitness	0.13	0.02	0.76
Cereals	Weetabix	0.14	0.03	0.80
Chewing gum	Mentos	0.78	0.04	0.94
Cigarettes	Marlboro	0.18	0.003	0.71
Cigarettes	Camel	0.17	0.004	0.79
Cleaners (all purpose)	Ajax	0.44	0.04	0.98
Cleaners (all purpose)	Mr Proper/Mr Muscle	0.55	0.03	0.98
Coffee, ground	Lavazza	0.25	0.03	0.94
Coffee, instant	Nescafe	0.22	0.05	0.92
Condoms	Durex	0.14	0.03	0.89
Carbonated soft drink	Coca Cola	0.24	0.03	0.95
Carbonated soft drink	Fanta	0.21	0.03	0.91
Deodorant	Nivea	0.30	0.04	0.93
Deodorant	Dove	0.42	0.04	0.96
Diapers	Pampers	0.21	0.04	0.84
Diapers	Huggies	0.59	0.04	0.87
Dishwasher tablet	Finish	0.79	0.04	0.99
Fabric softener	Lenor	0.38	0.03	0.95
Fabric softener	Silan	0.26	0.03	0.89
Laundry detergent	Ariel	0.30	0.03	0.92
Laundry detergent	Persil	0.23	0.04	0.80
Margarine	Becel	0.35	0.03	0.96
Panty liners	Carefree	0.20	0.02	0.83
Panty liners	Always	0.33	0.03	0.93
Pasta, dry	Barilla	0.20	0.04	0.82
Peas, frozen	Iglo	0.17	0.03	0.82
Peas, tinned	Bonduelle	0.42	0.05	0.94
Rice	Uncle Ben's	0.14	0.04	0.79
Shampoo	Elseve/Elvital	0.24	0.03	0.89
Shampoo	Head & Shoulders	0.26	0.02	0.91
Shampoo	Fructis	0.23	0.03	0.88
Shampoo	Pantene	0.23	0.03	0.82
Shave preps	Gillette	0.59	0.03	0.97
Shave preps	Nivea	0.55	0.04	0.98
Toothpaste	Colgate	0.44	0.03	0.98
Toothpaste	Sensodyne	0.38	0.02	0.97
Vodka	Absolut	0.20	0.02	0.86
Vodka	Smirnoff	0.26	0.02	0.90
Whiskey	Ballentines	0.28	0.02	0.94
Whiskey	Jack Daniel's	0.18	0.02	0.89

The table shows the average CV calculated across and within the 13 countries along with the intra-class correlation of a random intercepts model at the country level. It indicates that price dispersion across countries is of an order of magnitude higher than within countries.

price variation over time. This is the reason for the relatively small ICC-estimates for cigarette brands.

very important for explaining price dispersion in the euro area and that the variation of prices is much larger across countries than within countries.

Looking at the results for the individual products over time in Figures 1-5, we cannot see much time variation in price dispersion across and even less so within countries. When formally testing for a trend in the time series of dispersion, we do find a significantly negative trend of cross-country price dispersion over time for 24 out of 47 brands.⁹ However, for most of these brands the trend is numerically small. In the graphs it is visible only for Gourmet cat food, Camel cigarettes, Marlboro cigarettes, Ajax all-purpose cleaner, Lavazza ground coffee, Huggies diapers, Becel margarine, Carefree pantyliners, Absolut vodka, Smirnoff vodka, Ballentines whiskey and Jack Daniel's whiskey. For 12 out of 47 brands we find a significantly positive trend of cross-country price dispersion over time. Surprisingly, for within-country price dispersion we have the opposite result. The test reveals a mild positive time trend in the time series of within-country price dispersion for 20 out of 47 brands whereas the trend is negative only for 7 brands. These trends are masked in the Figures by the scale effect of displaying across- and within-country price dispersion on the same axis.

In order to check to what extent heterogeneity of (SKUs within) brands across countries affects cross-country price dispersion, we identified 13 exactly identical products at the SKU-level which are available in at least 6 countries. They are listed in Table 8 in section 4.4. where more explanation on the selection of the homogenous products can be found. Table 2 shows the results of the CV across and within countries calculated at the SKU-level for these products. Comparing the figures with those in Table 1, we generally find a lower level of price dispersion, both across and within countries, for the set of only homogenous products. However, the relation between cross-country and within-country price dispersion is similarly large also for the homogenous products: cross-country price dispersion is between 4 times (Becel) and 24 times (Gillette gel) higher than within-country dispersion. This indicates that heterogeneity explains only a small part of the difference between price dispersion across and within countries.

Summing up, for most products in our sample we observe vast differences between price dispersion across and within countries with the former exceeding the latter by a factor of more than 8. This does not change even when considering only products which are homogenous across countries. For about half of the

⁹By including country-dummies in the test regression we account for changes in the country-composition over time.

Table 2: Average coefficient of variation (CV) across and within countries for homogenous products

Brands	CV across	CV within
Milka	0.19	0.02
Snickers	n.a.	0.03
Coca Cola	0.22	0.02
Fanta small	0.26	0.03
Fanta big	0.21	0.02
Kellogg's Cornflakes small	0.18	0.03
Kellogg's Cornflakes big	0.14	0.03
Becel	0.06	0.02
Nivea deodorant	0.23	0.03
Gillette gel	0.66	0.03
Nivea gel	0.59	0.04
Elseve shampoo	0.12	0.03
Fructis shampoo	0.14	0.03

The table shows the average coefficient of variation calculated across and within countries for the 13 homogenous products. Price dispersion turns out to be somewhat smaller for the homogenous products than for the broader pan-European brands shown in the previous table.

products in our sample there is a mild downward trend in cross-country price dispersion over time.

3.3 Variance decomposition of price dispersion

To assess the relative importance of price dispersion over time versus across space, we perform a decomposition of the total variance in regional (pairwise) price differences into the cross-sectional and the times-series component according to Crucini and Telmer (2012). Let P_{irt} denote the nominal price of product i in region r at time t .¹⁰ Then the relative price of product i at time t between regions r and r' is the log difference between the respective nominal prices:

$$p_{irr't} = \log P_{irt} - \log P_{ir't}. \quad (1)$$

Note that the total variation in the relative prices of product i can be decomposed into a cross-sectional and a time-series component:

¹⁰All prices are in Euros.

$$Var_{rr',t}(p_{irr't}|i) = Var_{rr'}(E_t[p_{irr't}|i, rr']) + E_{rr'}[Var_t(p_{irr't}|i, rr')] = T_i + F_i. \quad (2)$$

where $E_x(\cdot|y)$ and $Var_x(\cdot|y)$ denote the conditional mean and variance calculated by integrating over variable x while conditioning on y . So, for instance, $E_t[p_{irr't}|i, rr']$ is the mean of the time series of relative prices for product i between regions r and r' and $Var_{rr'}(E_t[p_{irr't}|i, rr'])$ is the variance across region-pairs in these time-series means. This term is denoted T_i and is commonly associated with trade costs and trade barriers which do not vary over time. The other term F_i captures the time-series variation around the long-term means of relative prices of product i averaged across region-pairs. It is associated with transitory fluctuations in international relative prices which die out over time (see Crucini and Telmer, 2012).

Table 3 presents the results of the variance decomposition. The numbers in the first column are averages of $T_i + F_i$ across the 13 homogenous goods, the second and third columns show the averages of T_i and F_i , respectively. The last column is the ratio of the cross-sectional variance to the total variance. The results are shown for all region-pairs (combined) in the first line and separately for only international region-pairs in the second and only intranational region pairs in the third line. The results indicate that for both, the international as well as only intranational region-pairs, the cross-sectional variance accounts for almost 90% of the total variance while the time-series variation accounts only for slightly more than 10%. This result is consistent with the ICC-results of Table 1, which also imply large cross-country price dispersion. It is also in line with – though even more extreme than – Crucini and Telmer (2012), who find a proportion of about 60% cross-sectional to 40% time-series variance. Reasons for the relatively low time-series variation in our data could be the short observation period of only 3 years whereas the sample of Crucini and Telmer (2012) spans 15 years and that – contrary to our data – in Crucini and Telmer (2012) exchange rate volatility additionally affects the variation of prices over time. Another interesting observation from Table 3 is that in absolute terms the international cross-sectional variance of 0.057 is about 5 times as large as the intranational cross-sectional variance of 0.012 which is a further indication of a possibly large border effect.

Given our finding that cross-sectional price dispersion exceeds by far price dispersion over time, we now concentrate on the price dispersion across space.

Table 3: Variance in relative prices

	Total variance	Cross-sectional variance	Time-series variance	cross-sectional total
Combined	0.059	0.053	0.007	0.889
International	0.065	0.057	0.007	0.886
Intranational	0.014	0.012	0.001	0.896

The table reports averages of the cross-sectional (T_i) and time-series variances (F_i) according to the equation:

$$\text{Var}_{rr',t}(p_{irr't}|i) = \text{Var}_{rr'}(E_t[p_{irr't}|i, rr']) + E_{rr'}[\text{Var}_t(p_{irr't}|i, rr')] = T_i + F_i.$$

The definition of the variables and a discussion are in the text.

4 Price dispersion and relative price variation across space

The goal of this section is to obtain a comprehensive view on within and cross-country (absolute and relative) price variation in the 13 Eurozone countries in our data set. As we saw previously, price dispersion in the spatial dimension is more pronounced than over time. So in this section we concentrate only on the spatial dimension, by aggregating out the time-series dimension, as in Engel and Rogers (1996).¹¹

We first define and discuss various measures of relative price variation. Then we present simple descriptive statistics of them, calculated both within- and across countries. Next, we run regressions to evaluate the exact role that distance and border plays in relative price variation. Finally, we will repeat the whole exercise on a narrow subset of products that are homogenous internationally; controlling for product heterogeneity enables us to compare *absolute price levels*.

4.1 Measures of relative price variation

In this section – motivated by the strand of literature that was started by the influential paper of Engel and Rogers (1996) – we will work with relative prices between *region pairs*, calculated from around 70 “Nielsen-regions” available in our data set. Therefore our unit of observation is region pairs, and our main

¹¹Note that this is exactly the opposite of what we have done in the previous section. There, we have calculated the coefficient of variation (a cross-sectional measure) first, and then reported the average value (over time) of this. In this section we do some kind of time-averaging first, and then analyze the determinants of the cross-sectional variation in the data.

interest is to study the impact of region pair-specific characteristics (like physical distance, separation by a border, relative income and tax rates etc) in relative price variation and price dispersion.

Recall the definition of the relative price of product i between regions r and r' at time t ($p_{irr't}$) from equation (1).¹² One can calculate various statistics of these product- and region pair-specific time series. For example, the means of the relative price time series, $E_t[p_{irr't}|i, rr']$, or in short, $\bar{p}_{i,rr'}$ inform us whether there are any systematic price level differences between the respective regions (i.e. whether a specific product is systematically more expensive in one region than in the other):¹³

$$\bar{p}_{i,rr'} = \frac{1}{T} \sum_{t=1}^T p_{irr't}. \quad (3)$$

In practice, we would not expect these mean relative prices to be zero, for various reasons. First, products might be different across regions (Broda and Weinstein, 2007), and these product differences can lead to systematic price level differences. But even if we have exactly the same products in the respective regions, local market conditions (different tax rates, differences in market concentration, differences in local wage and/or rental rates) or geographic barriers to arbitrage (like transportation costs or times) might prevent prices from being equalized, and lead to systematic price level differences.

In the presence of systematic price level differences between regions, the mean absolute difference informs us about the size of these:

$$MAD(p)_{i,rr'} = \frac{1}{T} \sum_{t=1}^T |p_{irr't}|. \quad (4)$$

In a stylized model with cross-location arbitrage and transportation costs, the width of no-arbitrage bands might depend on physical distance between locations. For regions close to each other, a relatively small price differential might induce arbitrage; while for more distant regions, a larger price dispersion would be needed for the cross-location arbitrage. Therefore in such an environ-

¹²Note that this is a different strategy to the one applied in the previous section. There we investigated the cross section of prices, while now we concentrate on the cross-section of region pairs. The reason of this is that in this setup, we can naturally control for some physical barriers – like distance and national borders – that would be difficult to be controlled for in the “simple” cross-section.

¹³Previously, we denoted the cross-sectional variance of $\bar{p}_{i,rr'}$ by T_i ($T_i = Var_{rr'}(\bar{p}_{i,rr'})$), and called this the “cross-sectional variance”.

ment mean absolute price differences can increase with distance. Note, however, that the measure of mean absolute difference is also influenced by systematic differences in product attributes and local market conditions; so in order to study the role of physical distance with this measure, we have to control for these other characteristics.

There are several ways to filter out the impact of product heterogeneity and other time-invariant local factors from price dispersion measures. One possibility is to calculate the time-series standard deviation of the relative price series (Engel and Rogers, 1996). Assuming that differences in product attributes and local market conditions are time invariant, and thus these lead to a price difference that is constant over time, the standard deviation will reflect the relative price variation around this constant price differential:

$$SD(p)_{i,rr'} = \sqrt{\frac{1}{T} \sum_{t=1}^T (p_{irr't} - \bar{p}_{i,rr'})^2}. \quad (5)$$

Further possibilities to control for systematic price level differences between regions is first differencing the relative price series, by which the systematic price level differences are eliminated. This is in fact what one has to do when only CPI-based relative price series are available (see, for example, Engle and Rogers 1996), without any information on price *levels*. The differenced relative price (or relative price *change*) series is

$$\Delta p_{i,rr't} = (\log P_{irt} - \log P_{ir't}) - (\log P_{ir,t-1} - \log P_{ir',t-1}), \quad (6)$$

and the mean absolute difference and the standard deviation of the relative price changes – $MAD(\Delta p)_{i,rr'}$ and $SD(\Delta p)_{i,rr'}$, respectively – are defined similarly as before for the relative price time series.

Note that the measures of $SD(p)$, $SD(\Delta p)$ and $MAD(\Delta p)$ all describe the extent of relative price variation. Their small values reflect that relative prices change little over time,¹⁴ while large time-variation in relative prices will lead to larger values of these measures. In contrast, the mean absolute deviation of relative prices, $MAD(p)$ is informative about the respective price levels in the two regions: its small values mean that price *levels* are similar. Therefore when investigating relative price variation, we will concentrate on $SD(p)$, $SD(\Delta p)$ and $MAD(\Delta p)$; and we will only use $MAD(p)$ when studying cross- and within-

¹⁴In articular: if they are zero, then relative prices are constant over time.

country price levels.

4.2 Descriptive statistics of relative price variation

In this subsection we report simple averages of the above-defined measures of relative price variation. The averages are calculated for within- and cross-country region pairs. For example, the average relative price standard deviation within Austria (AT) for a specific product i is calculated as the simple unweighted average of relative price standard deviations for this product between all possible Austrian region pairs. The cross-country averages are based on region pairs in which the two regions are in different countries.

We calculate these relative price measures for the 47 pan-European brands of 27 products introduced in section 3.¹⁵ Note, however, that our focus on pan-European brands does not eliminate all systematic cross-border differences between price levels. First, “packsize 1” in France is generally not the same pack size as “packsize 1” in Spain: the local Nielsen offices collect price and quantity data about the most popular pack sizes and varieties in the respective countries, and nothing ensures that these are identical.¹⁶ So this product heterogeneity can itself lead to systematic differences in price levels. But even if we happen to have exactly the same pack sizes and varieties (of the same brand of the same product), many other factors might create systematic price level differences:

- Differences in value-added tax rates in the different countries.
- Differences in the local distribution costs. The rental rate of a similar store, and the wage of the sellers might be different in Graz (Austria) and in Maribor (Slovenia), despite the fact that the distance between these two locations is less than 100 kilometers. Even if the products themselves are tradable, they include a set of non-tradable inputs whose prices might differ.
- Differences in local market conditions. Concentration on Italian pasta market might be different from that in Greece.

¹⁵We focus on pan-European brands as we would like to ensure a minimum level of similarity between countries: while we might expect that the price of Colgate toothpaste moves similarly in different countries, that might not be the case for toothpastes belonging to different brands (Colgate and Sensodyne, for example). By selecting pan-European brands we ensure that we are comparing the same brands of the same products across the countries (but not necessarily the same varieties or pack sizes).

¹⁶In fact, they are very rarely identical.

- Income levels might be different, even within the countries. (Like the Paris region in France vs all the other French regions.)
- The very same product might have a different customer perception in one country than in other. For example, a specific type of pasta which is considered to be quite regular in Italy might be perceived as something of high quality in Slovakia.

In this subsection we do not control for these potential sources of price differentials, therefore we only report statistics about those relative price variation measures that are not sensitive to systematic price level differences: $SD(p)$, $SD(\Delta p)$ and $MAD(\Delta p)$. Instead of reporting these statistics for all 47 brands in our sample, we just report the (across-product) median values.¹⁷

Table 4: Average relative price variations

Packsizes 1	$SD(p)$	$SD(\Delta p)$	$MAD(\Delta p)$
within BE	0.017	0.017	0.012
within DE	0.021	0.021	0.016
within EE	0.043	0.051	0.036
within IE	0.038	0.038	0.033
within GR	0.032	0.030	0.024
within ES	0.024	0.019	0.020
within FR	0.012	0.009	0.007
within IT	0.036	0.040	0.030
within NL	0.014	0.013	0.010
within AT	0.026	0.029	0.021
within PT	0.024	0.020	0.015
within SI	0.012	0.007	0.006
within SK	0.027	0.017	0.013
Cross-country	0.078	0.060	0.044

The table presents average relative price variations, as measured by three alternative statistics, within 13 Eurozone countries and across them, for the median product. It suggests that relative price variation is bigger across than within the countries.

Table 4 shows the results for the most popular pack sizes (“packsize 1”) in the respective countries.¹⁸ The numbers in the table are medians across the 47 brands, and should be understood as log points. To help interpretation, take, for example, the case of Belgium. For all 47 brands, we calculate the $SD(p)_{i,rr'}$ measures for all possible Belgian region pairs, and then take the simple unweighted average of these. Then we take the median of the resulting

¹⁷Brand-specific results are available from the authors upon request.

¹⁸We did the same calculations for the other pack sizes and the whole brand averages as well. We report “packsize 1” (instead of packsize 2 or 3) as these are the most popular pack sizes within the countries, with highest market shares.

47 brand-specific averages, 0.017, and report it in the table: it means that the average standard deviation of within-Belgian relative price time series, for the median brand in Belgium, is 0.017 log points, or approximately 1.7%.

It is apparent from Table 4 that the three different relative price measures give different numerical values for relative price variations, but these measures are strongly correlated, and lead to similar ranking orders of the countries.

When comparing within- and cross-country relative price variations, all three measures indicate that average cross-country relative price variation is much higher than within-country relative price variations. In Table 4, all the 39 within-country measures are smaller than the respective cross-country measures. If we also look at similar tables about other pack sizes (packsize 2-3) and about the brand averages (all pack sizes taken together), 154 of the 156 within-country measures are smaller than the respective cross-country measures. Looking at all individual products (instead of the median products in each country), the proportion of within-country measures that are smaller than the corresponding cross-country measure is between 91.7% and 97.0%, depending on the relative price measure and pack size.

Taken together, this is strong indication that relative price variation is smaller within the countries than across. In the next subsection we ask the question of whether (or to what extent) this is because international region pairs are more remote than intranational ones.

4.3 Cross-sectional regressions on relative price variation

Part of the difference between international and intranational relative price variation might be explained by larger geographic distance (and possibly other types of geographic separation, like national borders) between international locations. To account for these, in this section we estimate various cross-sectional regressions on the relative price variation measures, using geographic variables as controls.

To control for product heterogeneity, we start by considering only those observations that correspond to intra-national region pairs. This ensures that the products in the respective regions are exactly the same.¹⁹ Besides controlling for product heterogeneity, with this approach we also get rid of country-specific shocks, like changes in the value-added tax rates or in local market conditions

¹⁹Within the countries, it is always exactly the same product (brand, pack size and variety) that Nielsen price collectors observe. Across the countries, in most cases pack sizes and brand characteristics are different.

(e.g. changes in country-wide minimum wages or any change affecting local distribution costs); all of these would show up in international price differences, but not in intranational ones. By controlling for product heterogeneity, this approach also has the advantage that the mean absolute difference of prices, $MAD(p)$ is a meaningful measure to consider. The drawback, however, is that we lose many observations and the international dimension of the data set, and variation in terms of the distance between region pairs gets much smaller.

In this first step, we estimate the following cross-sectional regression (on region pairs rr' which belong to the same country), separately for each brand and pack size (i):

$$RPV_{i,rr'} = \alpha + \beta \log D_{rr'} + \sum_{j=1}^R \delta_j I_j + \varepsilon_{rr'}, \quad (7)$$

where $RPV_{i,rr'}$ is a relative price variation measure for a specific brand and pack size i ($RPV \in \{SD(p), MAD(p), SD(\Delta p), MAD(\Delta p)\}$), $D_{rr'}$ is the distance between regions r and r' , I_j , $j = 1, \dots, R$ are region-dummies. We have two types of measures for the distance variable: it is either driving minutes or driving kilometers between region centers, as obtained from www.viamichelin.com in May 2013.²⁰

Table 5 summarizes the results on the estimated β parameters for pack size 1 (the most popular pack size in each country). Regression (7) is run for each relative price variation measure 47 times (for the 47 different pan-European brands), and the table reports the median (across products) of the estimated β -s followed by the numbers of the regressions in which the t -value of the estimated β parameter was larger than 2 (significantly positive), between 0 and 2 (positive insignificant), between -2 and 0 (negative insignificant), or below -2 (negative significant). The left/right panels of the table report results when the distance measure is driving minutes or driving kilometers, respectively.

The median values of the estimated parameters are broadly in line with similar estimates on alternative data sources. In Engel-Rogers (1996) and Parsley-Wei (2001), the baseline estimates for the distance parameter were 0.0011 and 0.0022, respectively. Our estimates are somewhat larger than these, but similar

²⁰Our choice of driving minutes and kilometers, instead of the great circle distance, was motivated by Braconier and Pisu (2013). They conclude that it is important to use distance measures related not just to pure geographic distance, but also to road connectivity between locations. Taking into account road connectivity, their estimated border effect on trade flows declines by about 15%.

Table 5: t -values of estimated β parameters

Packsize 1	Median	driving minutes; $t_\beta \in$				Median	driving kilometers; $t_\beta \in$			
		> 2	$[0; 2]$	$[-2; 0]$	< -2		> 2	$[0; 2]$	$[-2; 0]$	< -2
$SD(p)$	0.0036	32	12	3	0	0.0031	33	11	3	0
$MAD(p)$	0.0040	19	25	3	0	0.0035	19	25	3	0
$SD(\Delta p)$	0.0030	32	12	3	0	0.0027	34	10	3	0
$MAD(\Delta p)$	0.0022	33	12	2	0	0.0021	34	11	2	0

The table reports the median – among the 47 different regressions run separately on the most important pack size (“packsize 1”) of different brands – of the estimated β coefficient of regression (7), and the number of regressions in which it is positive significant, positive insignificant, negative insignificant and negative significant, respectively. In the left panel of the table, the distance variable between the region pairs is driving minutes; in the right panel, it is driving kilometers. The rows $SD(p)$, $MAD(p)$, $SD(\Delta p)$ and $MAD(\Delta p)$ refer to different measures of relative price variation, and the integer numbers in each row of each panel add up to 47. The table shows that distance matters: for the majority of regressions it is positive and significant, while there is not a single combination of brand, relative price variation and distance measure for which it would be negative and significant.

in magnitudes.²¹

It is also apparent from Table 5 that distance matters for price dispersion, even within the countries: for the majority of the brands, the estimated coefficient on distance is positive and significant, and there is not a single brand for which it would be negative significant. Even insignificant parameter estimates are mostly positive.²² In words, one can expect that the larger is the distance between two regions, the larger is the relative price variation. This result is robust to alternative measures of price dispersion, and to products, brands and pack sizes.²³

Note that in Table 5, results are very similar for driving minutes or driving kilometers. Given the strong correlation between these two measures, this is hardly surprising. Therefore in subsequent regressions we will only report results that are based on the driving minutes distance measure.

As a next step, we extend our sample to international region pairs. With this we increase the number of cross-sectional observations, at the cost of introducing

²¹One reason for larger estimates can be that pairwise distances in our data are on average smaller than in the US-Canadian data of Engel-Rogers (1996) or in the US-Japanese data of Parsley-Wei (2001).

²²Recall that by dropping international region pairs, the variation in distance decreases a lot, and the number of observations gets smaller. Both of these can increase the estimated standard errors, which might be the reason of some parameter estimates being insignificant.

²³Table 15 in the Appendix contains results for other pack sizes – packsize 2, 3, and all pack sizes taken together. Out of 752 regressions on different brands and packsizes, we have 439 or 446 (depending on distance measure), about 59%, positive significant β estimates, and not a single one which is negative significant.

possible product heterogeneity. Along the lines of Engel and Rogers (1996), we also add a border dummy to the regression equation, which takes a value of one region pairs in which the two regions are in different countries. The new specification is therefore

$$RPV_{i,rr'} = \alpha + \beta \log D_{rr'} + \gamma B_{rr'} + \sum_{j=1}^R \delta_j I_j + \varepsilon_{rr'}, \quad (8)$$

where $B_{rr'}$ is the border dummy. This equation is again estimated separately for the 47 pan-European brands, for certain pack sizes (the baseline is the most popular “packsize 1”).

Table 6 shows the (cross-product) median of the estimated distance and border parameters, β and γ , and the number of regressions in which they are significant in the 47 regressions estimated on 47 brands (always using the most popular pack sizes “packsize 1”). The left panel of the table shows the distance variable, while the right panel corresponds to the border dummy. The distance measure is the driving minutes between the region pairs. With possible product heterogeneity, and the resulting systematic price differences, we do not report the relative price dispersion measure $MAD(p)$.²⁴

Table 6: t -values of estimated β and γ parameters

Packsize 1	Median	distance parameter; $t_\beta \in$				Median	border parameter; $t_\gamma \in$			
		> 2	[0; 2]	[-2; 0]	< -2		> 2	[0; 2]	[-2; 0]	< -2
$SD(p)$	0.0050	34	3	2	8	0.037	47	0	0	0
$SD(\Delta p)$	0.0053	35	7	3	2	0.025	47	0	0	0
$MAD(\Delta p)$	0.0036	37	5	2	3	0.018	47	0	0	0

The table reports the median – among the 47 different regressions run separately on the most important pack size (“packsize 1”) of different brands – of the estimated β (on distance, in the left panel) and γ (on border, in the right panel) coefficients of regression (8), and the number of regressions in which they are positive significant, positive insignificant, negative insignificant and negative significant, respectively. The rows $SD(p)$, $SD(\Delta p)$ and $MAD(\Delta p)$ refer to different measures of relative price variation, and the numbers in each row of each panel add up to 47. The table shows that distance matters: for the majority of regressions (i.e. brands) it is positive and significant. The estimated border coefficient is always positive and significant; this, however, might be due to both true border effects or uncontrolled country-specific heterogeneity.

Consistently with the previous results, distance seems to be important: for the majority of brands, the estimated coefficient on distance is positive and significant (although for some brands now we have significantly negative esti-

²⁴Table 16 in the Appendix shows the same results as Table 6 for all available pack sizes in the data set. The results are similar.

mates). The estimated coefficients on border dummies are always positive and significant; this is the case despite the fact that the reported measures of relative price dispersion are not sensitive to systematic price level differences. Similarly to the distance parameters, the magnitude of the estimated border dummy is in line with previous estimates (0.010 and 0.065 in Engel and Rogers (1996) and Parsley and Wei (2001), respectively).

Note, however, that with the addition of international region pairs, we not only introduce possible product heterogeneity. As was shown by many papers following Engel and Rogers (1996),²⁵ the border dummy actually will pick up any unobserved difference between the countries (different price setting habits, different institutions, different country-specific shocks, even differences in country-specific sales habits and seasonality). One can control for some of these effects: for example, we can correct the data for any country-specific changes in value-added tax rates by considering only the net (as opposed of gross) prices. Therefore we should be cautious with the interpretation of significant border parameters: they might reflect true border effects as well as uncontrolled country-specific heterogeneity.

As a next experiment, we re-run regression (8) for those region pairs only, which are closer than 300 (or 180) driving minutes to each other. With this experiment we want to make sure that our results are not driven by very remote region pairs (like Tallinn-Porto or Dublin-Iraklion), where one would not expect any kind of co-movement between price levels, or small relative price variation. Essentially, we are “zooming” on the region pairs in which the two regions are just at the two sides of some national border in Europe. The choice of the distance threshold is arbitrary, and we face a trade-off between the number of observations in the regression and the closeness of the region pairs considered. Table 7 shows that the estimated border coefficient for the median product actually gets larger as we decrease the distance threshold in the regression, which means that our results are valid even for international region pairs that are next to each other (just the two sides of the same border).

4.4 Accounting for price level differences across regions

So far we have considered measures that are not sensitive to systematic cross-country differences in price *levels*, and therefore we have not investigated cross-country price level differences. In this sub-section we turn to this. If we want

²⁵For example, Gorodnichenko and Tesar (2009).

Table 7: Estimated β and γ parameters for different distance thresholds

Packsize 1	median distance parameter, β			median border parameter, γ		
	No threshold	$D < 300$	$D < 180$	No threshold	$D < 300$	$D < 180$
$SD(p)$	0.0050	0.0050	0.0051	0.037	0.041	0.050
$SD(\Delta p)$	0.0053	0.0033	0.0038	0.025	0.029	0.032
$MAD(\Delta p)$	0.0036	0.0025	0.0029	0.018	0.023	0.027

The table reports the median – among the 47 different regressions run separately on the most important pack size (“packsize 1”) of different brands – of the estimated β (on distance, in the left panel) and γ (on border, in the right panel) coefficients of regression (8), run for region pairs which are closer to each other than specific distance thresholds (no threshold, 300 minutes, 180 minutes). The rows $SD(p)$, $SD(\Delta p)$ and $MAD(\Delta p)$ refer to different measures of relative price variation. The table indicates that our previous results are robust to various distance thresholds, that is, they continue to hold even if we concentrate on international region pairs that are relatively close to each other and one might reasonably expect some kind of co-movement between prices.

to ensure that our results are meaningful, i.e. that price level differences do not simply reflect differences between products, we have to do this investigation on a sub-sample of products that are homogeneous across the countries.

To this end we have selected a sub-sample of 13 product varieties (from 7 different product categories) that are observed by Nielsen price collectors in at least six countries. The fact that we have only found 13 sufficiently international varieties, is in itself a good indication of how different are the most popular product varieties – which Nielsen observes – internationally. Table 8 summarizes these common varieties.²⁶

Figures 6-7 in the Appendix plot the times series of Coca Cola and Nivea shaving gel prices within European regions, for the product varieties listed in Table 8. At least two things are apparent from the figures. First, the cross-European price range is relatively wide: sometimes we can observe more than 80-100% difference between the prices of exactly same products. Second, countries matter: for the Coca Cola, one can clearly identify clusters, consisting of regions that belong to the same country, indicating that prices within the countries move together more strongly than between the countries. This clustering is perhaps less obvious on the graph on Nivea gel prices.²⁷

In terms of numbers, the average and median (calculated across the 13 prod-

²⁶For Coca Cola and Fanta, two types of varieties (e.g. 1.5l or 2l PET) are given. One of the varieties is available in roughly half of the countries, and other is in the other half. In subsequent analysis, we are considering only those country pairs that have the same varieties, and then pool together these observations.

²⁷These two figures are representative of the 13 homogeneous products of Table 8. The observations we have made about these two products are valid for other products as well. The figures on the other products are available from the authors upon request.

Table 8: Homogeneous products

Product	Brand	Variant	Countries	Regions
Chocolate	Milka	Alpenmilch 100g	9	47
Chocolate	Snickers	57g bar	6	36
Carbonated soft drinks	Coca Cola	1.5 or 2l PET	12	63
Carbonated soft drinks	Fanta	1.5 or 2l PET	11	59
Carbonated soft drinks	Fanta	0.33l can or 0.5l PET	9	44
Cereals	Kellogg's	Kellogg's Cornflakes 375g box	7	41
Cereals	Kellogg's	Kellogg's Cornflakes 500g box	10	54
Margarine	Becel	Becel Pro Active 250g	7	40
Deodorant	Nivea	For Men Roll-On 50 ml	8	49
Shave preps	Gillette	Sensitive Skin Gel 200ml	8	49
Shave preps	Nivea	Sensitive Gel 200ml	9	52
Shampoo	Elseve	Elseve Vive Color 250ml	8	36
Shampoo	Fructis	Fructis color 250ml	7	34

The table presents identical product varieties which are available in at least 6 European countries. For each variety, it gives the number of countries and Nielsen-regions for which there are data available.

ucts) of cross-country mean absolute price differences are 0.196 and 0.178 log points, respectively, with a range between 0.077 (Becel) and 0.326 (Nivea deodorant) – see the first column of Table 9. Note that these are averages across the international region pairs: for roughly half of the international region pairs, the mean absolute difference between price levels of identical products is more than around 20%. Taking into account that these products are identical, these are large numbers.

Price level differences within the countries are much smaller. For the 13 varieties listed above, the average within-country price difference (measured by the mean absolute price difference between region pairs in the same countries) is between 0.014 log points (in Slovenia) and 0.064 log points (in Italy); for 9 countries, this figure is between 0.028 and 0.038 log points. This is about 7 times smaller than the same figure for cross-country price level differences, which is about the same order of magnitude that we have also found in the previous section for the within-and cross-country coefficient of variation calculations. All of these indicate that national borders remain significant even for the same product varieties.

To study numerically the significance of borders, we re-estimate regression (8) on the 13 homogeneous product varieties. The estimated border coefficients (γ , see column 1 of Table 10) are always significantly positive: the average and median are 0.170 and 0.139, respectively, with a range from 0.062 (Snickers) to 0.479 (Nivea deodorant).

As a next step, we try to explain (at least partially) the observed international price differences. We do so by correcting log nominal prices by differences between tax rates, income levels, local competition measures, consumption intensities and unemployment rates. For this, we estimate the following regression:

$$\log P_{irt} = \alpha + \beta_1 QSH_{irt} + \beta_2 \log GDP_{rt} + \beta_3 \log VAT_{irt} + \beta_4 \log DENS_{rt} + \beta_5 UE_{rt} + \varepsilon_{irt}, \quad (9)$$

where QSH and VAT are the value share and value-added tax of product variety i in region r at time t , and GDP , $DENS$ and UE are the per capita GDP, population density and unemployment rate in region r at time t ,²⁸ and ε_{irt} is an *i.i.d.* error term. We expect that larger volume share and population density within a region leads to lower prices ($\beta_1 < 0$, $\beta_4 < 0$) through larger competition; larger per capita GDP implies higher prices ($\beta_2 > 0$); larger value-added tax rate leads to higher prices $\beta_3 > 0$; and that larger unemployment rate decreases prices $\beta_5 < 0$. Ideally, we would like to add regional wages to control for local distribution costs; wages, however, are only available for 5 countries at the regional level, so we use per capita GDP as a proxy for this.

We estimated regression (9) for all 13 homogeneous product varieties. We also added time dummies, in order to make sure that the parameters of the above regression are identified from the cross-section (i.e. they reflect cross-sectional differences).²⁹ The results for the coefficients estimated with OLS are shown in Table 17 in the Appendix. We estimated a significantly negative β_1 parameter (on volume share) for 12 products out of 13, a significantly positive β_2 (on log GDP) for 9 products out of 13, a significantly positive β_3 parameter (on the log tax rate) for another 9 products out of 13, and a significantly negative β_4 parameter (on population density) for another 9 products out of 13. We had mixed results on β_5 , the parameter of the unemployment rate: it was significantly positive for 8 products, and significantly negative for 4 products (and not significant for 1 product). Overall, the signs of the estimated coefficients are mostly consistent with our initial expectations.

²⁸These regional variables are calculated from Eurostat regional statistics, but they are aggregated up to the Nielsen-regions (which are not identical to NUTS-2 or NUTS-3 regions of Eurostat).

²⁹Our results are essentially unchanged if we do not add these time dummies and just estimate a pooled regression. We also estimated regression (9) with IV, instrumenting QSH with its own third lag (to tackle the possible endogeneity of quantity with prices), but results are qualitatively unaltered.

The interpretation of the estimated coefficients is straightforward. The (cross-product) mean of the estimated β_1 -s is -0.141, which means that in regions where the region-specific volume share is, *ceteris paribus*, 10%-points higher, the price lever is smaller by approximately 1.41%. The mean of the estimated GDP-parameter (β_2) is 0.112, meaning that a in a region with 10% larger GDP per capita, nominal prices are expected to be larger by approximately 1.12%. For the VAT-rate, the mean estimated parameter is 0.382, meaning that if a region has, *ceteris paribus*, a larger VAT-rate by 10%-points, then that region can expect higher nominal prices, by 3.82% on average.³⁰ Population density and unemployment rates have numerically much smaller (but mostly statistically significant) effects on regional price levels.

After correcting nominal prices with the estimated regression coefficients, we re-calculated the cross-country mean absolute price differences from the residuals. The second column of Table 9 shows the results of this correction. The first observation from the table is that mean absolute cross-country price differences decrease for all the 13 products after the price correction. The (cross-product) mean of the absolute price differences decreases from 0.196 to 0.146, i.e. by somewhat more than 25%. This means that explanatory variables of regression (9) explain a bit more than a quarter of cross-country price level differences; but the remaining price differences are still large, around 15% for an average international region pair.

We reach a similar conclusion when looking at the estimated border dummies, based on corrected price levels (see the second column of Table 10). All but one estimated border dummies decrease (the exception is Kellogg's Cornflakes big), and the (cross-product) mean estimate goes down from 0.170 to 0.097, a decrease of approximately 43%. This is a substantial decline: 43% of the raw border estimates were due to different GDP per capita, value-added tax rates, and consumption intensities. Note, however, that all but one border dummies remain significant (the exception is Fanta small): we still have a substantial amount of unexplained international price level difference and thus border effect.

Finally, we turn to evaluating the impact of each explanatory variables in regression (9) in reducing cross-country mean absolute price differences. To do this, we performed alternative nominal price corrections by re-estimating regres-

³⁰This figure should *not* be compared with usual estimates of VAT pass-through, i.e. the effect of VAT-changes on prices. Pass-through figures are usually estimated from the time series, while this figure is estimated from the cross section.

Table 9: Average cross-country mean absolute price differences

Brand	Raw	Corrected	Correction with respect to				
			Volshare	GDP	VAT	Pop dens	Unemp
Milka	0.269	0.167	0.268	0.240	0.168	0.266	0.235
Snickers	0.151	0.114	0.148	0.150	0.146	0.151	0.129
Coca Cola	0.191	0.125	0.173	0.156	0.185	0.178	0.182
Fanta small	0.272	0.162	0.236	0.233	0.204	0.279	0.249
Fanta big	0.178	0.140	0.176	0.166	0.179	0.163	0.152
Kellogg's CF small	0.205	0.123	0.164	0.201	0.201	0.202	0.199
Kellogg's CF big	0.156	0.139	0.153	0.155	0.148	0.156	0.153
Becel	0.077	0.075	0.076	0.077	0.077	0.076	0.076
Nivea deodorant	0.326	0.256	0.323	0.326	0.272	0.324	0.317
Gillette gel	0.158	0.137	0.145	0.156	0.158	0.151	0.158
Nivea gel	0.171	0.151	0.170	0.166	0.165	0.170	0.166
Elseve shampoo	0.155	0.129	0.150	0.154	0.141	0.155	0.154
Fructis shampoo	0.237	0.181	0.210	0.233	0.235	0.237	0.234
Median	0.178	0.139	0.170	0.166	0.168	0.170	0.166
Mean	0.196	0.146	0.184	0.186	0.175	0.193	0.185

The table shows mean absolute cross-country price differences for the 13 homogeneous products. The first column shows calculations on raw (uncorrected) price level data. The second column is the result after price level correction, according to regression (9). Results in columns 3-7 are based on price level corrections with a single explanatory variable; these help evaluating the impact of each explanatory variable of regression (9).

sion (9) with only one explanatory variable each time, and calculated the figures of Tables 9-10 from these alternative price corrections. The results are shown in column 3-7 of Tables 9-10. Take, for example, the impact of value-added taxes on international price differences. Columns 5 show the results that are based on price level corrections that only filter out VAT-differences. According to column 5 in Table 9, the mean cross-country absolute price difference decreases to 0.175 (from 0.196) if we correct prices for differences in tax rates, i.e. different tax rates are responsible for a mean price level difference of around 2% points (or around 10% of the overall mean absolute difference). Similarly, column 5 of Table 10 informs us that the mean border dummy decreases to 0.137 (from 0.170), or by approximately 20%, if we correct for differences in VAT-rates.

According to this interpretation, we can say that out of the 25% decline in mean cross-country price differences (that we achieve with correcting prices), the relative contribution of volume share, per capita GDP and VAT-rate correction are 5 percentage points, 5 percentage points and 10 percentage points, respectively. Of the around 43% decline in the estimated border effects, the relative contribution of volume share, GDP and VAT-rates is 14 percentage points, 3 percentage points and 20 percentage points, respectively. The remain-

Table 10: Estimated border dummies

Brand	Raw	Corrected	Volshare	Correction with respect to			
				GDP	VAT	Pop dens	Unemp
Milka	0.134	0.057	0.149	0.092	0.047	0.133	0.131
Snickers	0.062	0.045	0.066	0.048	0.107	0.062	0.043
Coca Cola	0.149	0.070	0.150	0.131	0.121	0.119	0.117
Fanta small	0.215	0.004	0.003	0.222	0.014	0.210	0.010
Fanta big	0.139	0.092	0.157	0.151	0.148	0.110	0.088
Kellogg's CF small	0.216	0.104	0.119	0.230	0.165	0.206	0.189
Kellogg's CF big	0.114	0.118	0.165	0.110	0.117	0.114	0.102
Becel	0.064	0.059	0.063	0.064	0.065	0.063	0.064
Nivea deodorant	0.479	0.304	0.477	0.476	0.356	0.469	0.442
Gillette gel	0.144	0.086	0.096	0.143	0.144	0.127	0.142
Nivea gel	0.136	0.102	0.134	0.124	0.129	0.129	0.132
Elseve shampoo	0.103	0.090	0.112	0.103	0.087	0.102	0.105
Fructis shampoo	0.257	0.132	0.210	0.260	0.274	0.253	0.236
Median	0.139	0.090	0.134	0.131	0.121	0.127	0.117
Mean	0.170	0.097	0.146	0.166	0.137	0.161	0.139

The table shows estimated border coefficients (γ) from regression (8) for the 13 homogeneous products. The first column shows regression output on raw (uncorrected) price level data.

The second column is the result after price level correction, according to regression (9).

Results in columns 3-7 are based on price level corrections with a single explanatory variable; these help evaluating the impact of each explanatory variable of regression (9).

ing declines are due to the other two explanatory variables and interaction terms between the explanatory variables. We can thus conclude that among the explanatory variables of international price level differences we have considered, the most important one is value-added tax differences, and it explains around 10% of the absolute price differences and around 20% of the estimated border effects.

Several factors might explain that estimated borders are significantly positive even after the price correction. Most obviously, we may have omitted some important explanatory variables; an example is the share of discounters or petrol stations in the particular country or regions.³¹ Second, we might have mis-measured some explanatory variables in our current specification; as mentioned, ideally we should have local wages or disposable income measures instead of per capita GDP. Further, as Gorodnichenko and Tesar (2009) point out, estimated border coefficients might pick up country heterogeneity (with respect to price dispersion). Finally, and somewhat related, different seasonality and sales patterns of prices across the countries also leads to larger international relative price variation measures and will eventually be picked up by the larger

³¹Countries in our data set are very heterogeneous with respect to store types; see Leszczynska (2013).

estimated border coefficients. In our data of monthly average prices (as opposed to transaction prices), it is not straightforward how one should adjust for sales. Our estimated border coefficients, therefore, are a mixture of these factors and true border effects.

5 Comparison with the period 2000-2003

In addition to the dataset described in section 2, another dataset compiled by AC Nielsen was recently made available to us. It is structured in a similar way as the dataset of the ECB we used so far and was originally collected by AC Nielsen for the European Commission (EC). It contains a total of 50 product categories for 14 countries³² and 78 regions and spans a maximum time period from Jan 2000 to Nov 2003. For each product category we have 4 different brands (whenever possible pan-European brands) and for each brand we have a popular and a consistent packsize. There is a large overlap between the dataset compiled for the ECB and for the EC in terms of product coverage, although no complete overlap.³³ In effect, the selection of products in the ECB dataset was based on the products that were already present in the EC dataset to ensure a maximum level of comparability between the two datasets.

The similar structure and product coverage in both datasets allow us to compare our results so far with those obtained from a time period which includes the beginning of EMU and also the Euro cash changeover. To do so, we recalculate all descriptive measures of price dispersion and relative price variability for the EC dataset which will enable us to assess whether price dispersion in Europe has actually decreased since the inception of the Monetary Union.

To ensure a maximum degree of comparability, we base our calculations – like for the ECB data – on pan-European brands which are available in as many countries as possible. However, we consider only those countries which are available in both datasets. Table 11 lists all 61 pan-European brands of the EC dataset (for popular packsizes) which are available in at least 6 (common) countries. The table also contains the time-averages of the coefficients of varia-

³²The country composition of the EC dataset is somewhat different as it also contains countries which are not members of the Monetary Union like Denmark, Great Britain and Sweden but, on the other hand, it does not contain the countries which joined the Monetary Union later such as Estonia, Slovakia and Slovenia, and it additionally includes Finland.

³³There are 5 more product categories in the EC dataset: face care, flour, Gin, pasta sauce and tinned pineapples; and some products are named differently in the two datasets, e.g. nappies instead of diapers and meat extract instead of bouillon.

tion (CV) across and within countries for these pan-European brands over the (common sample) period 2000M11-2003M5 calculated the same way as for the ECB data.

Table 11: Average CV across and within countries (2000M11-2003M5)

Product	Brand	across	within	Product	Brand	across	within
Baby food	Nestle	0.23	0.02	Face care	Plenitude	0.24	0.03
Beer	Carlsberg	0.42	0.02	Face care	Nivea	0.46	0.03
Beer	Heineken	0.37	0.02	Face care	Oil of Olaz	0.36	0.03
Butter	President	0.24	0.02	Ice cream	Haagen Dazs	0.09	0.02
Carb. soft drink	Coca Cola	0.13	0.02	Ice cream	Carte D'Or	0.19	0.03
Carb. soft drink	Pepsi	0.17	0.02	Laundry det.	Ariel	0.31	0.02
Carb. soft drink	Fanta	0.16	0.02	Meat extract	Knorr	0.54	0.02
Carb. soft drink	Seven Up	0.37	0.02	Meat extract	Maggi	0.63	0.02
Carb. soft drink	Sprite	0.15	0.03	Olive oil	Bertolli	0.18	0.03
Cat food	Whiskas	0.26	0.02	Panty liners	Alldays	0.13	0.02
Cat food	Kitekat	0.11	0.03	Panty liners	Carefree	0.29	0.02
Cereals	Kellogg's	0.17	0.02	Pasta, sauce	Barilla	0.21	0.03
Cereals	Nestle	0.11	0.02	Peas, tinned	Bonduelle	0.33	0.03
Cereals	Weetabix	0.16	0.03	Pineapple, tin.	Del Monte	0.21	0.05
Chocolate (multi)	Mars	0.12	0.02	Shampoo	Elvital	0.30	0.03
Chocolate (multi)	Twix	0.19	0.02	Shampoo	Pantene	0.33	0.02
Chocolate (multi)	Kitkat	0.17	0.02	Shampoo	Fructis	0.31	0.03
Chocolate (multi)	Snickers	0.16	0.02	Shampoo	Head & Sh.	0.31	0.02
Chocolate (single)	Mars	0.13	0.04	Shave preps	Gilette	0.37	0.02
Chocolate (single)	Snickers	0.16	0.04	Shave preps	Nivea	0.13	0.03
Chocolate (single)	Twix	0.13	0.03	Soups, dry	Knorr	0.44	0.02
Cleaners	Ajax	0.30	0.03	Soups, dry	Maggi	0.56	0.03
Cleaners	Cif	0.39	0.03	Soups, wet	Knorr	0.38	0.03
Coffee, ground	Lavazza	0.22	0.01	Toilet paper	Kleenex	0.45	0.03
Coffee, instant	Nescafe	0.38	0.02	Toothpaste	Colgate	0.30	0.03
Condoms	Durex	0.22	0.02	Toothpaste	Sensodyne	0.22	0.02
Diapers	Pampers	0.17	0.02	Water, spark.	Perrier	0.43	0.02
Dishwasher	Calgonit pow.	0.36	0.02	Water, spark.	San Pellegr.	0.44	0.02
Dishwasher	Calgonit tab.	0.28	0.02	Water, still	Evian	0.28	0.03
Dog food	Pedigree	0.18	0.03	Water, still	Vittel	0.30	0.02
Dog food	Chappi	0.18	0.03				

The table shows the average coefficient of variation (CV) calculated across and within countries for the 61 pan-European brands in the EC dataset for the period 2000M11-2003M5. It confirms that also in the EC dataset price dispersion across countries is of an order of magnitude larger than within countries.

The numbers in Table 11 confirm also for the EC dataset that price dispersion as measured by the coefficient of variation is of an order of magnitude larger across than within countries. It ranges from 3 times larger for Mars chocolate bars to 29 times larger for Maggi meat extract.³⁴

³⁴For Alldays panty liners we calculated the CVs disregarding the Netherlands because there is obviously a data problem for the price of Alldays in the Netherlands.

Table 12: Average CV across and within countries for 24 common brands

Product	Brand	EC		ECB	
		across	within	across	within
Beer	Heineken	0.37	0.02	0.25	0.04
Carbonated soft drink	Coca Cola	0.13	0.02	0.19	0.03
Carbonated soft drink	Fanta	0.16	0.02	0.20	0.03
Cat food	Whiskas	0.26	0.02	0.19	0.04
Cereals	Kellogg's	0.17	0.02	0.17	0.03
Cereals	Nestle	0.23	0.02	0.13	0.02
Cereals	Weetabix	0.16	0.03	0.14	0.03
Cleaners (all purpose)	Ajax	0.30	0.03	0.53	0.03
Condoms	Durex	0.22	0.02	0.15	0.03
Coffee, ground	Lavazza	0.22	0.01	0.23	0.03
Coffe, instant	Nescafe	0.38	0.02	0.22	0.05
Diapers	Pampers	0.17	0.02	0.19	0.04
Laundry detergent	Ariel	0.31	0.02	0.36	0.03
Panty liners	Alldays	0.13	0.02	0.31	0.03
Panty liners	Carefree	0.29	0.02	0.22	0.02
Peas, tinned	Bonduelle	0.33	0.03	0.43	0.06
Shampoo	Elvital	0.30	0.03	0.27	0.03
Shampoo	Pantene	0.33	0.02	0.25	0.03
Shampoo	Fructis	0.31	0.03	0.26	0.03
Shampoo	Head & Shoulders	0.31	0.02	0.29	0.03
Shave preps	Gillette	0.37	0.02	0.68	0.03
Shave preps	Nivea	0.46	0.03	0.33	0.04
Toothpaste	Colgate	0.30	0.03	0.40	0.03
Toothpaste	Sensodyne	0.22	0.02	0.40	0.02
unweighted average		0.32	0.02	0.28	0.03

The table shows the average coefficient of variation (CV) calculated across and within countries for 24 common pan-European brands in the EC dataset (2000M11-2003M5) and the ECB dataset (2008M12-2011M9). For 13 out of the 24 brands the CV is larger in the earlier than in the later dataset.

In order to be able to directly compare the measures of price dispersion between the periods 2000-2003 and 2009-2011, we identified 24 pan-European brands which are (identically) available in both datasets. These brands along with their coefficients of variation across and within countries for the EC and the ECB dataset are listed in Table 12. When comparing numbers between the two datasets, we realize that for 13 out of the 24 brands the coefficient of variation across countries is larger in the earlier than in the later period. The unweighted average of the cross-country CV over the 24 brands amounts to 0.32 for the period 2000-2003 and is thus also larger than the unweighted average for the period 2009-2011 amounting to 0.28. In contrast, for only 3 out of the 24 brands we find a higher within-country CV in the earlier compared to the later period. Thus, we may conclude that – unlike the within-country price dispersion – the price dispersion across countries in the euro area has not only

declined within the period 2009-2011 but also between 2003 and 2009.

As a next step, we repeated the calculations of Section 4.3 (about relative price variation measures for within- and cross-country region pairs) for both data sets. To ensure a maximum level of comparability, in these calculations we only used the 10 countries and the 24 pan-European products (most popular pack sizes) that are available in both periods. Table 13 presents the averages (across the 24 products) of the descriptive statistics and border regressions for the three alternative relative price measures ($SD(p)$, $SD(\Delta p)$ and $MAD(\Delta p)$).³⁵

Table 13: Comparison of relative price measures in 2000-2003 (EC) and 2008-2011 (ECB)

	Within		Cross		β_d		β_b	
	EC	ECB	EC	ECB	EC	ECB	EC	ECB
$SD(p)$	0.026	0.032	0.055	0.082	0.0011	0.0044	0.028	0.040
$SD(\Delta p)$	0.023	0.031	0.042	0.069	0.0005	0.0055	0.018	0.027
$SD(\Delta p)$	0.017	0.022	0.032	0.051	0.0005	0.0039	0.013	0.021

The table repeats the relative price calculations of Section 4.3 for both data sets (EC: 2000-2003, ECB: 2008-2011). It shows that both within- and cross-country relative price variation has *increased* over time (columns 2-5). Estimated border coefficients (columns 8-9) have also gone up, while estimated distance parameters (columns 6-7) were not significant in 2000-2003 – in contrast to 2008-2011.

Columns 2-3 show the within-country relative price variation in 2000-2003 (EC) and 2008-2011 (ECB). All three measures indicate that relative price variation – when calculated on the level of region pairs – actually increased within the countries (by 25-35%). Columns 4-5 show that relative price variation between international region pairs increased even more substantially (by 47-64%). Therefore, as can be seen from columns 8-9, the estimated border parameters (of regression (8)) also increased, from the range of 0.013-0.028 to 0.021-0.040. Interestingly, columns 6-7 show that our previous result about the importance of physical distance between regions, does not seem to hold for the period 2000-2003: estimated distance coefficients are close to zero, and for many products they are significantly negative.³⁶

This result about increasing border effect in relative price variation seemingly contradicts our earlier result based on the coefficients of variation (CVs).

³⁵Since we are not using identical products, we only present the calculations of those three measures that are not sensitive to this heterogeneity.

³⁶This also explains that estimated border coefficients in the EC data set are approximately equal to the difference between the cross- and within-country relative price variations, while in the ECB data set these border coefficients are much smaller than this difference: part of the difference between the raw measures is explained by larger distances between international region pairs.

Note, however, that by calculating CVs we are directly comparing price levels across the countries, while in the second exercise we investigated the relative prices, or the co-movement of price levels. This co-movement seems to have become weaker in 2008-2011 (relative to 2000-2003). A more directly comparable exercise with the CV calculations is the comparison of price levels, to which we turn now.

In this final set of calculations we have calculated the cross- and within-country mean absolute price deviations (a repetition of Section 4.4) both for the 2000-2003 and 2008-2011 sample periods. For 2000-2003, we used the consistent pack sizes of the EC data set (41 brands); and for 2008-2011, we used the homogeneous products identified earlier in the ECB data set (13 brands). To facilitate comparison, we calculated everything for the ten countries that are present in both data sets. Unfortunately, we were unable to select identical products that are present in both data periods.

The within-country mean absolute price differences are 0.035 for both sample periods (this is the simple unweighted average of the within-country mean absolute price deviations in the ten countries). The cross-country mean absolute price difference has slightly decreased: from 0.231 in 2000-2003, it decreased to 0.223 in 2008-2011. As a consequence, the estimated border coefficient also decreased: from 0.176 in 2000-2003, it decreased to 0.172 in 2008-2011. This result is similar to our earlier comparison of the coefficient of variation measures. Note, however, that these changes are very small and barely significant.

6 Summary

In this paper we have used a comprehensive and highly disaggregated data set on retail prices of 47 pan-European brands of 27 different products across the Euro Area between 2008-2011 (provided by AC Nielsen) to compare within- and cross-country price dispersion and relative price variation.

We find that cross-location price dispersion (measured by the coefficient of variation) changes little over time. For international region pairs, the trend is declining over time for more than half of the products, although the decline is numerically small. For intranational region pairs, however, there is a small but statistically significant increase over time for almost half of the products in our sample. Based on a similar dataset compiled by AC Nielsen for the European Commission from 2000 to 2003, we calculate comparable numbers for within-

and cross-country price dispersion and indeed find on average a higher degree of price dispersion for a comparable set of products in the earlier than in the later sample period.

Comparing within- and cross-country price dispersion and relative price variation, we find that all measures we considered indicate that price dispersion is by an order of magnitude larger across than within countries. We have separate results for relative price changes and price level differences. For relative price changes, distance is important: prices in closer locations move together more strongly, than prices in distant locations. National borders further decrease the co-movement of relative prices, even after controlling for distance. Interestingly, these statements do not hold in our earlier data period of the EC data set: there, distance is not an important determinant of the relative price variation, but national borders continue to decrease the extent of this co-movement.

As for the price level differences, for a small subset of homogeneous products we find that price differences across the countries are by an order of magnitude larger than within (20% vs 3.5%), and estimated border effects remain highly significantly positive. These raw numbers are a bit smaller than the numbers we obtain from the EC data set for 2000-2003, indicating a small decline in international price-level differences over time. For the 2008-2011 data period, we can account for around 25% of the difference between cross- and within-country price dispersion, and for around 43% of the estimated border effect by taking into account cross-location differences between income levels, tax rates, consumption intensities, population densities and unemployment rates. The most important factor of these is differences in value-added tax rates, which account for around 10% of international price differences and 20% of the estimated border effect. We emphasize, however, that the remaining, statistically and numerically significant border effect is a mixture of the true border effect and omitted variables, measurement error and unobserved country heterogeneity.

References

- [1] Asplund, Marcus and Richard Friberg (2001): “The Law of One Price in Scandinavian Duty-Free Stores,” *American Economic Review*, vol. 91(4), 1072-1083.
- [2] Braconier, Henrik and Mauro Pisu (2013): “Road Connectivity and the Border Effect,” *OECD Economics Department Working Papers* No. 1073, July 2013.
- [3] Broda, Christian and David E. Weinstein (2008): “Understanding International Price Differences Using Barcode Data,” *NBER Working Paper* No. 14017, May 2008.
- [4] Crucini, Mario J., Christopher I. Telmer and Marios Zachariadis (2005): “Understanding European Real Exchange Rates,” *American Economic Review*, vol. 95(3), 724-738.
- [5] Crucini, Mario J. and Christopher I. Telmer (2012): “Microeconomic Sources of Real Exchange Rate Variability,” *NBER Working Papers 17978*, April 2012.
- [6] Engel, Charles and John H. Rogers (1996): “How Wide Is the Border?” *American Economic Review*, vol. 86(5), 1112-1125.
- [7] Ghosh, Atish and Holger Wolf (1994): “Pricing in International Markets: Lessons from the Economist,” *NBER Working Papers*, No. 4806.
- [8] Gorodnichenko, Yuriy and Linda L. Tesar (2009): “Border Effect or Country Effect? Seattle May Not Be So Far from Vancouver After All,” *American Economic Journal: Macroeconomics*, vol. 1(1), 219-241.
- [9] Haskel, Jonathan and Holger Wolf (2001): “The Law of One Price—A Case Study,” *Scandinavian Journal of Economics*, vol. 103(4), 545-458.
- [10] Isard, Peter (1977): “How Far Can We Push the Law of One Price?” *American Economic Review*, vol. 67(5), 942-948.
- [11] Lach, Saul (2002): “Existence and Persistence of Price Dispersion: An Empirical Analysis,” *The Review of Economics and Statistics*, vol. 84(3), 433-444.

- [12] Leszczynska, Agnieszka (2013): “The Substitution Bias in the Outlet Structure – A Report Based on Nielsen Data,” *manuscript*, November 2013.
- [13] Meyler, Aidan (2013): “An Overview of the Nielsen Disaggregated Price Dataset Provided to the Eurosystem,” *manuscript*, November 2013.
- [14] Parsley, David C. and Shang-Jin Wei (2001): “Explaining the Border Effect: The Role of Exchange Rate Variability, Shipping Costs, and Geography,” *Journal of International Economics* 55, 87-105.

A Appendix: Additional Tables and Figures

Table 14: List of product categories

Product	Product
Baby food	Laundry detergent
Beer	Margarine
Bouillon	Milk, refrigerated
Butter	Milk, ultra-high-temperature
Cat food	Olive oil
Cereals	Panty liners
Chewing gum	Paper towels
Chocolate	Pasta, dry
Cigarettes	Peas, frozen
Cleaners (all purpose)	Peas, tinned
Coffee, ground	Rice
Coffee, instant	Shampoo
Condoms	Shave preps
Carbonated soft drinks	Soups, wet
Deodorant	Sugar
Diapers	Toilet paper
Dishwasher tablet	Toothpaste
Dog food	Tuna, tinned
Fabric softener	Vodka
Fish, frozen	Water, sparkling
Ice cream	Water, still
Jam, strawberry	Whiskey
Juice, 100% fruit	

Table 15: t -values of estimated β parameters, in within-country regression

All packsizes	Median β	driving minutes; $t_\beta \in$				Median β	driving kilometers; $t_\beta \in$			
		> 2	[0; 2]	[-2; 0]	< -2		> 2	[0; 2]	[-2; 0]	< -2
$SD(p)$	0.0035	30	15	2	0	0.0031	31	14	2	0
$MAD(p)$	0.0071	27	19	1	0	0.0062	27	18	2	0
$SD(\Delta p)$	0.0037	34	12	1	0	0.0032	34	12	1	0
$MAD(\Delta p)$	0.0026	37	8	2	0	0.0023	38	7	2	0
Packsize 1	Median	> 2	[0; 2]	[-2; 0]	< -2	Median	> 2	[0; 2]	[-2; 0]	< -2
$SD(p)$	0.0036	32	12	3	0	0.0031	33	11	3	0
$MAD(p)$	0.0040	19	25	3	0	0.0035	19	25	3	0
$SD(\Delta p)$	0.0030	32	12	3	0	0.0027	34	10	3	0
$MAD(\Delta p)$	0.0022	33	12	2	0	0.0021	34	11	2	0
Packsize 2	Median	> 2	[0; 2]	[-2; 0]	< -2	Median	> 2	[0; 2]	[-2; 0]	< -2
$SD(p)$	0.0032	27	15	5	0	0.0031	27	15	5	0
$MAD(p)$	0.0041	21	22	4	0	0.0037	22	21	4	0
$SD(\Delta p)$	0.0027	22	23	2	0	0.0023	24	21	2	0
$MAD(\Delta p)$	0.0020	28	17	2	0	0.0018	28	17	2	0
Packsize 3	Median	> 2	[0; 2]	[-2; 0]	< -2	Median	> 2	[0; 2]	[-2; 0]	< -2
$SD(p)$	0.0033	20	23	4	0	0.0029	19	24	4	0
$MAD(p)$	0.0056	24	18	5	0	0.0047	23	18	6	0
$SD(\Delta p)$	0.0039	25	18	4	0	0.0033	25	17	5	0
$MAD(\Delta p)$	0.0025	28	16	3	0	0.0022	28	16	3	0

The table reports the median – among the 47 different regressions run on different brands separately – of the estimated β coefficient of regression (7), and the number of regressions in which it is positive significant, positive insignificant, negative insignificant and negative significant, respectively. In the left panel of the table, the distance variable between the region pairs is driving minutes; in the right panel, it is driving kilometers. We have different panels (in the vertical direction) for different pack sizes (pack size 1 being the baseline, also reported in the main text). Within the panels, $SD(p)$, $MAD(p)$, $SD(\Delta p)$ and $MAD(\Delta p)$ refer to different measures of relative price variation, as defined in the main text. The table shows that distance matters: for the majority of regressions it is positive and significant, while there is not a single combination of brand, pack size, relative price variation and distance measure for which it would be negative and significant.

Table 16: t -values of estimated β and γ parameters, in cross-country regression

All packsizes	Median β	distance parameter; $t_\beta \in$				Median γ	border parameter; $t_\gamma \in$			
		> 2	[0; 2]	[-2; 0]	< -2		> 2	[0; 2]	[-2; 0]	< -2
$SD(p)$	0.0034	31	6	3	7	0.034	47	0	0	0
$SD(\Delta p)$	0.0027	34	7	3	3	0.022	47	0	0	0
$MAD(\Delta p)$	0.0021	32	8	3	4	0.018	47	0	0	0
Packsizes 1	Median	> 2	[0; 2]	[-2; 0]	< -2	Median	> 2	[0; 2]	[-2; 0]	< -2
$SD(p)$	0.0050	34	3	2	8	0.037	47	0	0	0
$SD(\Delta p)$	0.0053	35	7	3	2	0.025	47	0	0	0
$MAD(\Delta p)$	0.0036	37	5	2	3	0.018	47	0	0	0
Packsizes 2	Median	> 2	[0; 2]	[-2; 0]	< -2	Median	> 2	[0; 2]	[-2; 0]	< -2
$SD(p)$	0.0040	29	10	1	7	0.038	47	0	0	0
$SD(\Delta p)$	0.0027	30	9	4	4	0.024	47	0	0	0
$MAD(\Delta p)$	0.0021	33	6	6	2	0.020	47	0	0	0
Packsizes 3	Median	> 2	[0; 2]	[-2; 0]	< -2	Median	> 2	[0; 2]	[-2; 0]	< -2
$SD(p)$	0.0041	30	9	3	5	0.035	47	0	0	0
$SD(\Delta p)$	0.0043	32	8	3	4	0.028	47	0	0	0
$MAD(\Delta p)$	0.0034	34	5	2	5	0.020	46	1	0	0

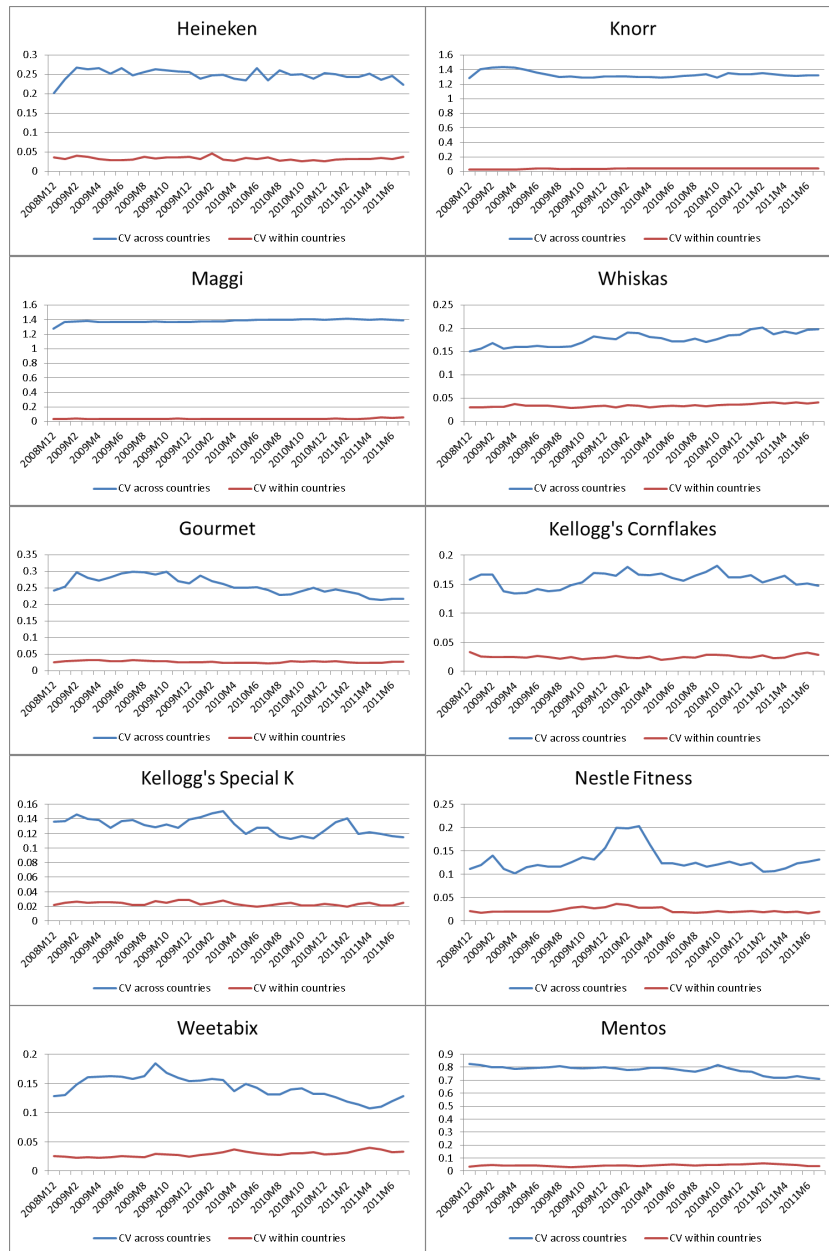
The table reports the median – among the 47 different regressions run separately on the most important pack size (“packsize 1”) of different brands – of the estimated β (on distance, in the left panel) and γ (on border, in the right panel) coefficients of regression (8), and the number of regressions in which they are positive significant, positive insignificant, negative insignificant and negative significant, respectively. The different vertical panels refer to different pack sizes (the baseline “packsize 1” is reported in the main text). The rows $SD(p)$, $SD(\Delta p)$ and $MAD(\Delta p)$ refer to different measures of relative price variation, and the numbers in each row of each panel add up to 47. The table shows that distance matters: for the majority of regressions (i.e. brands) it is positive and significant. The estimated border coefficient is always positive and significant; this, however, might mean both true border effects or uncontrolled country-specific heterogeneity.

Table 17: Estimated coefficients of regression (9)

Brand	QSH	log GDP	log VAT	log DENS	UE
Milka	-0.108*	0.002	0.375*	0.019*	-0.011*
Snickers	-0.350*	0.188*	0.125*	0.004	0.011*
Coca Cola	-0.037*	0.214*	0.103*	-0.012*	-0.001*
Fanta small	-0.364*	0.267*	0.297*	-0.073*	0.002
Fanta big	-0.103*	0.033*	0.029*	0.027*	-0.011*
Kellogg's CF small	-0.070*	0.267*	0.009	-0.036*	0.008*
Kellogg's CF big	-0.114*	0.142*	0.076*	-0.028*	0.013*
Becel	0.010*	0.031*	0.014	-0.010*	0.001*
Nivea deodorant	-0.198*	0.128*	1.962*	-0.031*	0.006*
Gillette gel	-0.038*	0.015	-0.119*	-0.027*	0.002*
Nivea gel	-0.012*	0.016	0.738*	-0.024*	0.010*
Elseve shampoo	-0.144*	-0.031*	1.838*	0.014*	-0.006*
Fructis shampoo	-0.298*	0.184*	-0.478*	-0.054*	0.021*
Median	-0.108	0.128	0.103	-0.024	0.002
Mean	-0.141	0.112	0.382	-0.018	0.004

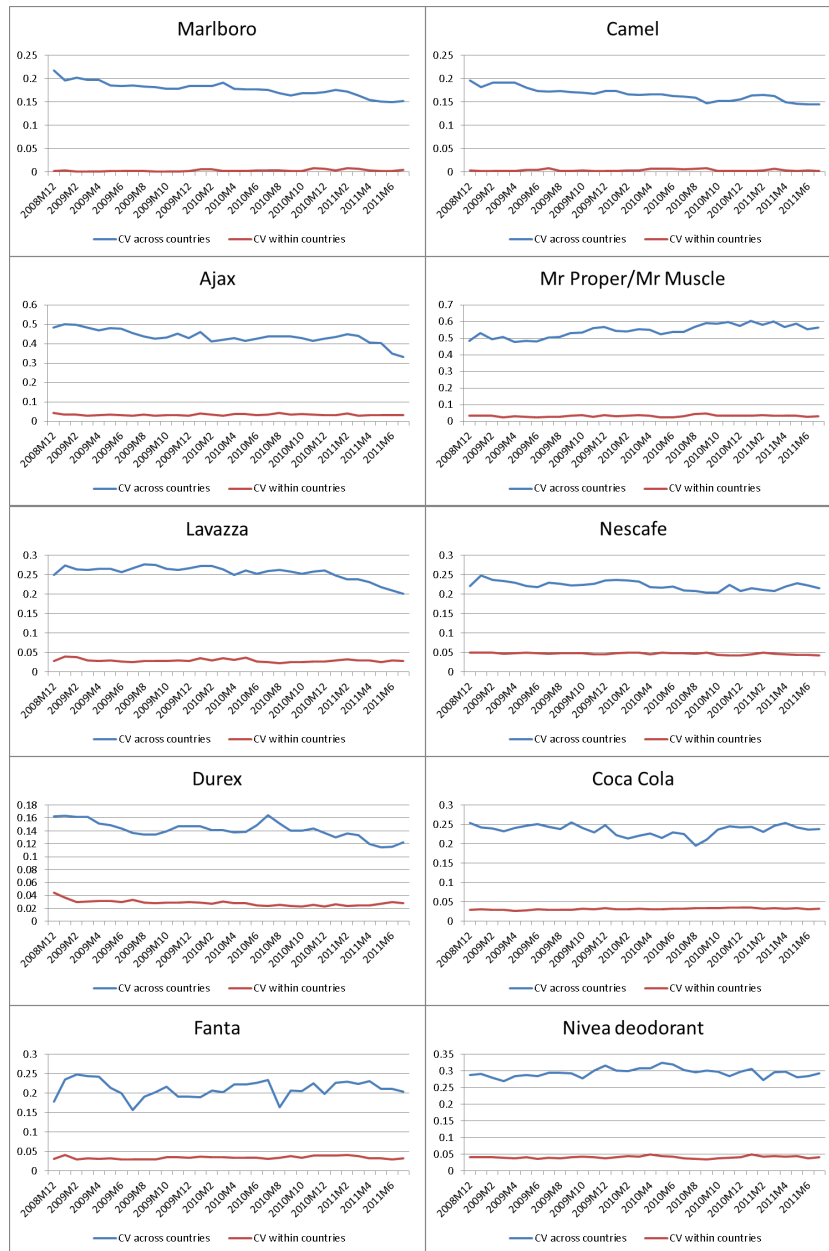
The table shows estimated coefficients of regression (9) for the 13 homogeneous products.
Stars denote significant parameter estimates at the 5% level.

Figure 1: Coefficient of variation within and across countries



The figure plots the coefficient of variation (CV) within and across countries, calculated over time, for ten pan-European brands. It shows that the CV is usually much larger across than within countries.

Figure 2: Coefficient of variation within and across countries



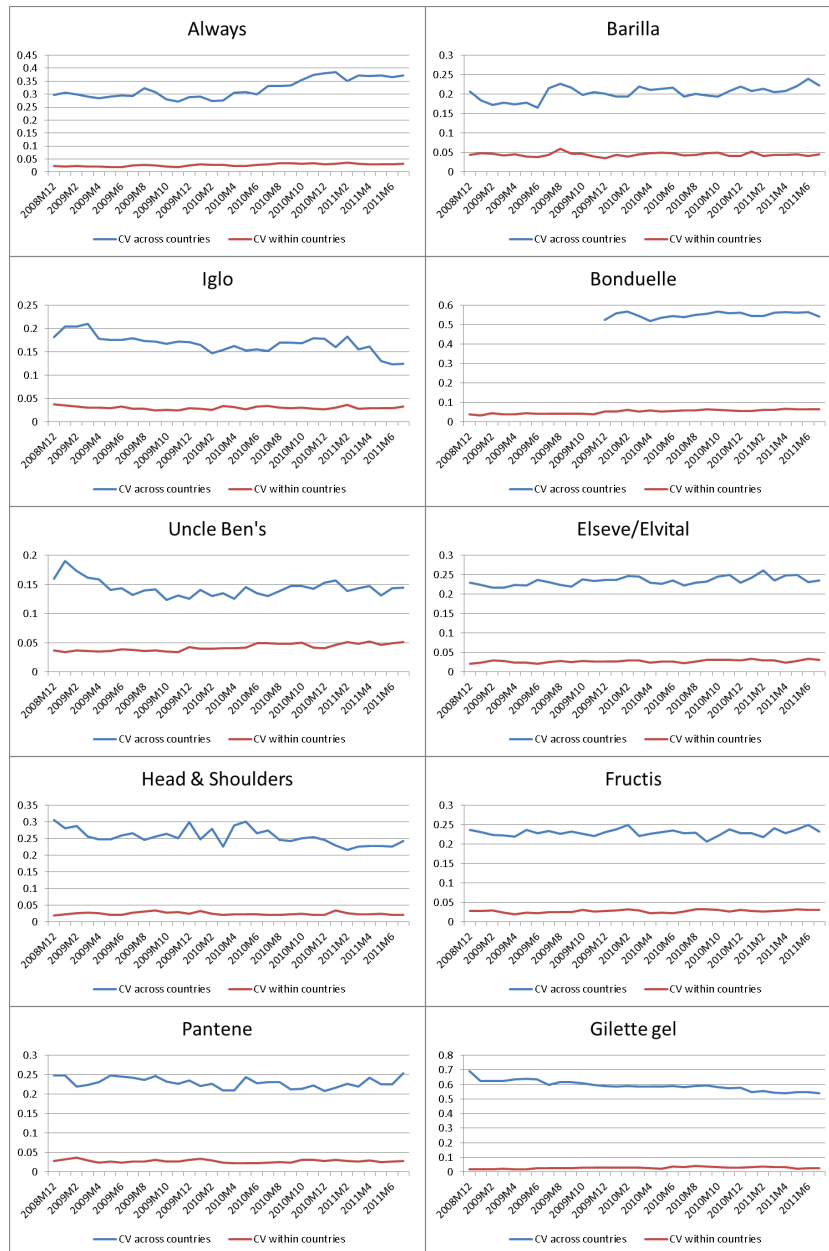
The figure plots the coefficient of variation (CV) within and across countries, calculated over time, for ten pan-European brands. It shows that the CV is usually much larger across than within countries.

Figure 3: Coefficient of variation within and across countries



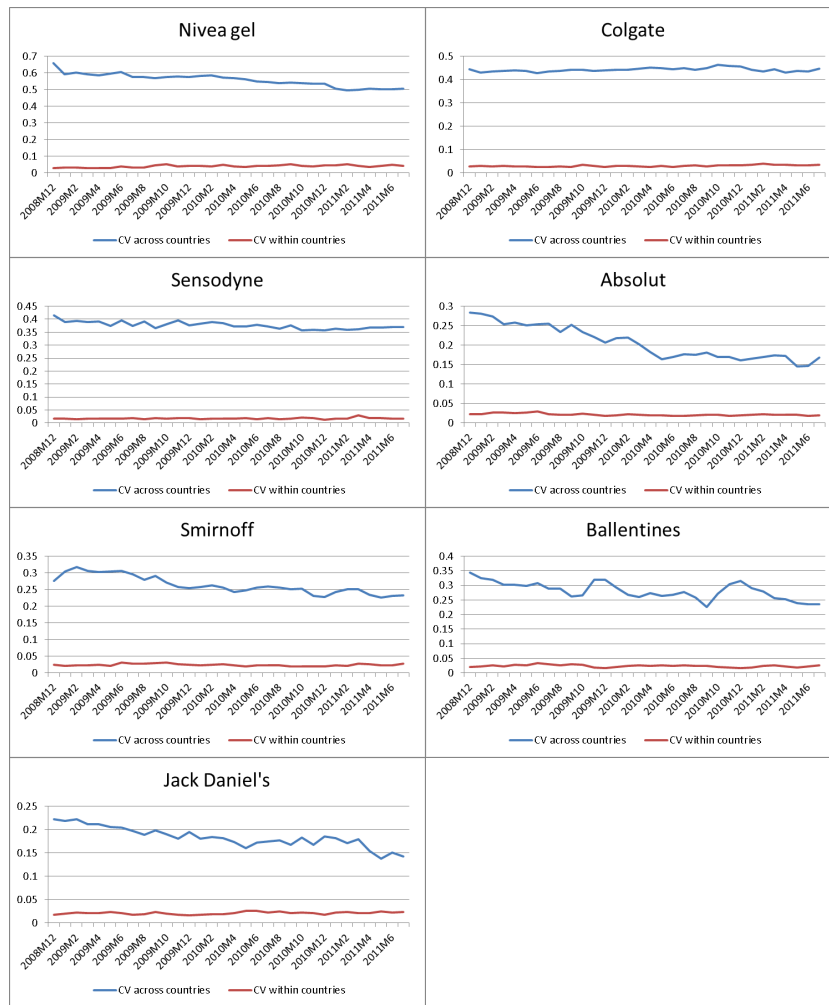
The figure plots the coefficient of variation (CV) within and across countries, calculated over time, for ten pan-European brands. It shows that the CV is usually much larger across than within countries.

Figure 4: Coefficient of variation within and across countries



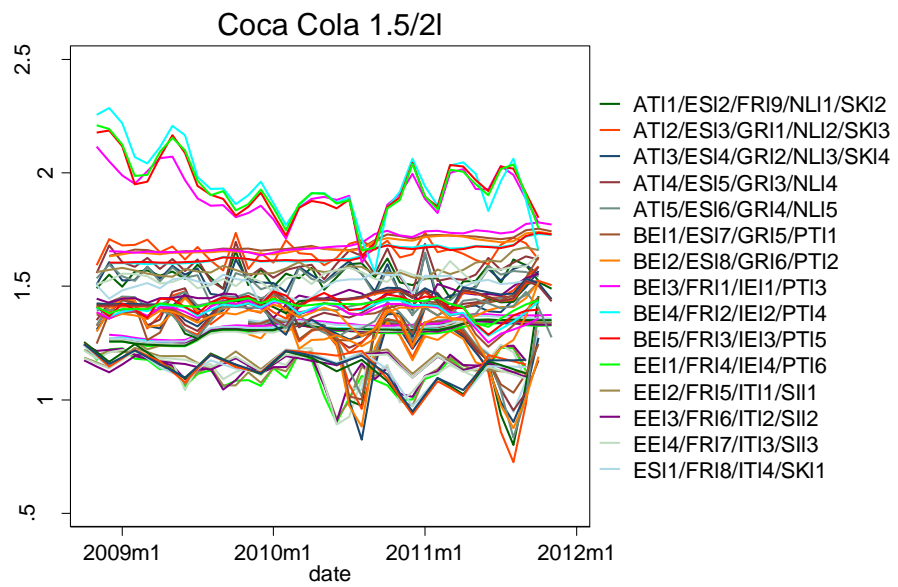
The figure plots the coefficient of variation (CV) within and across countries, calculated over time, for ten pan-European brands. It shows that the CV is usually much larger across than within countries.

Figure 5: Coefficient of variation within and across countries



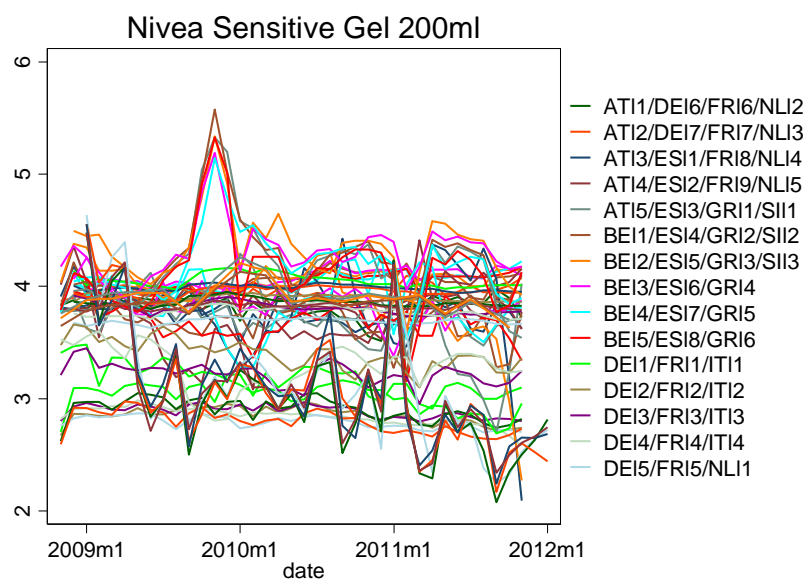
The figure plots the coefficient of variation (CV) within and across countries, calculated over time. It shows that the CV is usually much larger across than within countries.

Figure 6: Coca Cola prices within the euro area



The figure plots time series of Coca Cola prices (either 1.5l or 2l PET plastic bottle) in 63 European regions. Prices are sometimes substantially different, sometimes exceeding 100%. One can clearly identify country-specific clusters.

Figure 7: Nivea sensitive shaving gel prices within the euro area



The figure plots time series of Nivea sensitive shaving gel prices (200ml) in 52 European regions. The range of prices is quite wide, between 2.5 euros and 4.5 euros, approximately. Country-specific clusters are perhaps less apparent than for the Coca Cola.