

# On the Margins of Geographic Market Segmentation in the EU\*

Joris Hoste<sup>†</sup>

KU Leuven

Frank Verboven<sup>‡</sup>

KU Leuven

CEPR

March 1, 2023

[PRELIMINARY VERSION, PLEASE DO NOT CIRCULATE]

## Abstract

We analyze the degree of geographic market integration in European final goods markets. We propose a unifying framework to analyze geographic market segmentation both in terms of Law of One Price (LOP) deviations and choice set differences. To this end, we decompose regional cost-of-living differences into (1) LOP deviations, (2) pure taste differences and (3) choice set differences. In turn, we detect geographic market segmentation by considering terms whether LOP deviations and choice set differences are larger across international region pairs compared to intranational pairs. Using regionally disaggregate consumption data on 68 fast-moving consumer goods, we estimate that overall cost-of-living differences across Belgium, France, Germany and the Netherlands are more than seven times larger compared to intranational region pairs. Roughly 60% of the cost-of-living differences are explained by pure taste differences and the other 40% by the margins of geographic market segmentation. While choice set differences account for 38% of the variation, LOP deviations explain a mere 2%. Our results highlight the importance of choice set differences relative to LOP deviations as margins of geographic market segmentation and the presence of large cross-country taste differences across European countries.

**JEL codes:** D12, F15 and R32

**Keywords:** Market segmentation, Cost-of-living, Law of One Price deviations, Taste differences and Choice set variation

---

\*The data was obtained from Aimark.

<sup>†</sup>Electronic Adress: [joris.hoste@kuleuven.be](mailto:joris.hoste@kuleuven.be)

<sup>‡</sup>Electronic Adress: [frank.verboven@kuleuven.be](mailto:frank.verboven@kuleuven.be)

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Data</b>	<b>7</b>
<b>3</b>	<b>Reduced-form evidence</b>	<b>10</b>
3.1	LOP Deviations . . . . .	10
3.2	Choice set differences . . . . .	13
<b>4</b>	<b>Theoretical Framework</b>	<b>19</b>
4.1	Consumer preferences . . . . .	19
4.2	Decomposing Cost-of-living Differences . . . . .	20
<b>5</b>	<b>Structural Estimation</b>	<b>25</b>
5.1	Elasticities of Substitution . . . . .	25
5.1.1	Product variety-level elasticities - $\sigma_p$ . . . . .	25
5.1.2	Firm-level elasticities - $\eta_p$ . . . . .	28
5.2	Regional Cost-of-living Differences . . . . .	32
<b>6</b>	<b>Geographic Market Segmentation in Europe</b>	<b>33</b>
6.1	Empirical design . . . . .	35
6.2	Cost-of-living Differences for Comparable Regions . . . . .	38
6.3	Detecting Geographic Market Segmentation . . . . .	40
<b>7</b>	<b>Concluding Remarks</b>	<b>42</b>

# 1 Introduction

Understanding the presence of geographic market segmentation and the process towards geographically integrated international markets has been a question of central importance to both researchers and policymakers. For instance, the European Single Market (ESM) intends to forge a geographically integrated market from the national goods markets of its member states. A geographically integrated market is a market in which, beyond the costs associated with physically moving the product to the destination market, geography or the nationality of the buyer does not affect the terms at which producers sell their products (see [Flam \(1992\)](#) and [Goldberg & Knetter \(1997\)](#)). Conversely, if geography or the nationality of the buyer has residual predictive power for the terms at which transactions take place, markets are geographically segmented. Prior work has relied on two strategies to assess whether markets are geographically segmented. One strategy is to investigate whether prices of identical products differ significantly more across countries than across regions of the same country (e.g. [Engel & Rogers \(1996\)](#) and [Goldberg & Knetter \(1997\)](#)). Comparing Law of One Price (LOP) deviations across countries to LOP deviations across regions of the same country, [Beck et al. \(2020\)](#) and [Fontaine et al. \(2020\)](#) find considerably higher deviations across countries in Europe. Another strategy is to investigate to what extent trade shares discontinuously fall at national borders ([McCallum \(1995\)](#)). For instance, [Santamaría et al. \(2023\)](#) show that trade shares are substantially lower when goods cross a national European border compared to when they only cut across a regional border.

Conceptually, neither approach is fully satisfactory to measure the presence of goods market segmentation. First, the presence of LOP deviations is not a necessary condition for market segmentation. If fixed market entry costs were the only international trade costs, only the most competitive firms would sell outside their domestic market but it would not necessarily imply that these firms geographically price discriminate.<sup>1</sup> Thus, even though LOP deviations could be zero, market segmentation still arises through selection into geographic markets.<sup>2</sup> Second, the presence of a large border effect for trade shares is not a sufficient condition for market segmentation. When consumer tastes are home-biased, trade between international region pairs is lower compared to trade among intranational region pairs regardless of the size of border-related trade costs. This is because if consumer tastes are home-biased, trade between international region pairs is lower compared to trade among intranational region pairs regardless of the size of border-related trade costs.

This paper analyzes geographic market segmentation across final good markets of various European countries (Belgium, France, Germany and the Netherlands) and addresses the aforementioned measurement concerns. To this end, we construct a dataset from samples of households that report their expenditure on all product varieties within 68 fast-moving consumer goods (FMCG) categories. We observe product variety-level prices, quantities, and firm identifiers, all at a fine regionally disaggregate level. We measure geographic market segmentation with a two-step approach. First, borrowing insights from existing theory, we construct cost-of-living differences across all region pairs in our sample and decompose them into (1) LOP deviations,

---

<sup>1</sup>In a [Melitz \(2003\)](#) model with no variable trade cost, markets are segmented because choice sets differ, but firms do not price-to-market if consumers are equally price-sensitive.

<sup>2</sup>In fact, [Cavallo et al. \(2014\)](#) [Beck et al. \(2020\)](#) can only compute LOP deviations for a limited set of observations in their respective datasets.

(2) differences in consumer taste and (3) choice set differences. Second, we detect the presence of geographic market segmentation by estimating whether LOP deviations and choice set differences are larger for international region pairs, region pairs separated by a national border, relative to comparable intranational region pairs, which are region pairs within the same country. In this way, our approach addresses the aforementioned concerns by considering both LOP deviations and choice set differences as margins of market segmentation while filtering out differences in consumer taste as a residual source of cost-of-living differences.

Before outlining our two-step approach, we provide reduced-form evidence that both LOP deviations and choice set differences are larger across international region pairs compared to intranational pairs. In particular, the standard deviation of log price differences for identical product varieties within product category-region pair-year cells is, on average, 16% larger for international pairs relative to intranational pairs. While price differences are considerable, so are choice set differences. At the same time, the share in terms of the number of barcodes that are sold in both regions is on average 75% for intranational pairs but is below 25% in most cases for international pairs. Expressing choice set overlap in terms of expenditure or at the firm-level paints a similar picture.

To jointly analyze the role of LOP deviations and choice set differences as margins of geographic goods market segmentation, we specify a structural model of demand that guides the estimation of regional cost-of-living differences. We model preferences as a nested CES utility system, with one nest at the product variety level and the other at the firm level. We choose to work within the family of nested CES demand system as it is the workhorse framework to analyze the effects of changes or differences in choice sets<sup>3</sup> while providing a good fit for the type of scanner data we work with.<sup>4</sup> We rely on insights from Redding & Weinstein (2020) and obtain, under a normalization of the region-specific geometric average taste level, an expression for regional cost-of-living differences at the product category level that depends on (1) unweighted average LOP deviations, (2) average differences in market share on common varieties and common firms and (3) choice set differences. Importantly, even though average taste levels across regions are normalized, average differences in common market shares still capture mean-preserving differences in consumer tastes. To this expression, we add and subtract a restricted cost-of-living index for common firms and barcodes to which the general expression collapses when taste differences are zero. This allows us to decompose regional cost-of-living differences at the product category level into (1) LOP deviations, now corrected for substitution effects, (2) pure taste differences and (3) choice set differences.

The importance of taste differences and choice set differences crucially depends on the elasticities of substitution at the product variety and firm level, which we allow to vary across product categories. To estimate the product variety level elasticities, we capitalize on the fact that we also observe at which retail store purchases were made. While stores provide frequent sales and promotions over time, they tend to price uniformly across regions within the same country. After flexibly accounting for seasonal variation in prices

<sup>3</sup>See Feenstra (1994), Broda & Weinstein (2006), Handbury & Weinstein (2015), Romer (1990) and Jaravel (2019) in international trade, urban economics and macroeconomics respectively.

<sup>4</sup>More specifically, Dellavigna & Gentzkow (2019) shows that the implied log quantity-log price relationship fits the price and quantity data well. In addition, Hottman et al. (2016) shows how the nested CES-preference system is flexible enough to account for many empirical patterns that characterize an economy populated by heterogeneous multi-product firms.

and quantities, we instrument product variety level prices with prices of the same product in the same chain but in a different region. We recover a distribution of elasticities of substitution with a 10%-90% range of  $[-4.77, -1.15]$  across product categories. To estimate the elasticities of substitution at the firm level, we rely on the structure of the nested preference structure. In particular, the firm-level price index can be decomposed into a part that depends on the unweighted geometric average across product-variety level prices within the firm-level nest and a term that captures dispersion in market shares across product varieties within the same firm-level nest. Conditional on a specific fixed effect, the second term is uncorrelated with the firm-level demand shock and we use it to instrument firm-level prices. Doing so, we obtain an estimated distribution of firm-level elasticities of substitution with a range 10%-90% range of  $[-4.84, -1.71]$ .

With these elasticities in hand, we construct regional cost-of-living differences by first computing the three margins within product category-region pair-year cells and then by calculating the variance of the log cost-of-living differences across product categories within each region pair-year cell. The variance of regional cost-of-living differences is on average 0.12 for intranational pairs. In line with the motivating evidence, the variance of cost-of-living differences for international pairs is roughly 8 times larger compared to the average variance for intranational pairs. To understand the importance of each of the three components in determining the overall cost-of-living differences, we exactly decompose the variance of cost-of-living differences across the three components by distributing the covariance terms equally. Intranational cost-of-living differences are predominantly driven by differences in consumer taste. LOP deviations and choice set differences are quite muted for intranational pairs and explain less than 10% of the cost-of-living variation. For international regional pairs, LOP deviations, choice set differences and differences in consumer taste are all larger compared to intranational region pairs. Choice set differences and differences in consumer taste explain the bulk of elevated cost-of-living differences for international pairs.

To be able to detect geographic market segmentation from regional cost-of-living differences, we take two more steps. First, the decomposition of cost-of-living differences highlights that, besides LOP deviations and choice set differences, cost-of-living differences also capture pure taste differences. As heterogeneity in preferences is traditionally considered to be outside of the scope of international integration policies, we solely focus on LOP deviations and choice set differences as margins of geographic market segmentation.<sup>5</sup> Second, according to the definition of geographic market segmentation, we should only focus on cost-of-living differences that are not related to the costs associated with physically moving goods to different markets. As international region pairs are characterized by greater geographic differences, for instance in terms of physical distance, compared to intranational region pairs, one should compare cost-of-living differences between international and intranational pairs with similar geographic differences.<sup>6</sup> To this end, we follow [Santamaria et al. \(2021\)](#) and construct samples of comparable international and intranational region pairs by estimating the probability that a region pair is an international region pair, and thus separated by a national border, based on geographic determinants.<sup>7</sup>

<sup>5</sup>In our view, sustained LOP deviations and choice set differences can occur because of variable trade costs, such as transport costs, border-related costs or differences in distribution margins, and because of fixed market entry costs.

<sup>6</sup>The larger distance between international and intranational region pairs can lead to increased cost-of-living differences through transport-induced cost differences (e.g. [Donaldson \(2018\)](#)) or distance-induced selection (e.g. [Hummels & Skiba \(2004\)](#))

<sup>7</sup>With geographic differences we refer to differences in terms of distance, remoteness, elevation differences, shared river basins,

Controlling for geographic differences across region pairs reduces the differences in cost-of-living differences between intranational and international region pairs, but only to a limited extent. Whereas the cost-of-living differences are roughly 8 times larger unconditional on geographic differences, they are a little over 7 times as large conditional on geographic differences.<sup>8</sup> Above all, we find that a little under 60% of the increase in cost-of-living differences for international region pairs relative to intranational pairs is explained by pure taste differences. This result implies that trade data alone is unlikely to be very informative about geographic goods market segmentation. Also, a little over 40% of the increased cost-of-living differences can be attributed to margins associated with market segmentation. Out of the remaining 40% of the higher cost-of-living differences, 95% are due to choice set differences and only 5% of these differences are explained by LOP deviations.

We contribute to three strands of literature. Methodologically, our paper is related to the extensive literature on measuring inflation and cost-of-living differences using ideal price indices, using the family of CES preference systems in particular. Building on the work of [Sato \(1976\)](#) and [Vartia \(1976\)](#), estimating the gains from trade and welfare-relevant inflation figures have made considerable strides by accounting for changes in product variety (e.g. [Feenstra \(1994\)](#) and [Broda & Weinstein \(2006\)](#)) and by incorporating changes in consumer tastes ([Redding & Weinstein \(2020\)](#)), comparatively little is known about how important geographic market segmentation is in explaining spatial cost-of-living differences. [Handbury & Weinstein \(2015\)](#) and [Feenstra et al. \(2020\)](#) measure intranational variation in cost-of-living and [Argente et al. \(2021\)](#) and [Cavallo et al. \(2022\)](#) quantify international variation in cost-of-living levels. However, there is no prior work that combines intra- and international variation which is crucial to understand the role of geographic market segmentation ([Anderson & Wincoop \(2004\)](#)) and this paper fills this gap.

Second, we complement a substantial literature that infers international trade costs from international price differences, or LOP deviations, starting with [Engel & Rogers \(1996\)](#) and [Goldberg & Knetter \(1997\)](#). Moving from aggregate to increasingly detailed price data allowed the literature to overcome measurement issues due to aggregation and confirm the continued presence of large differences on the US-Canada border (see [Broda & Weinstein \(2008\)](#), [Gorodnichenko & Tesar \(2009\)](#) and [Gopinath et al. \(2011\)](#)) and across European countries (e.g. [Cavallo et al. \(2014\)](#), [Fontaine et al. \(2020\)](#) and [Beck et al. \(2020\)](#)). At the same time, focusing on increasingly detailed data also uncovered large underlying choice set differences. To date, there is no study accounting for both LOP deviations and choice set differences as margins of geographic goods market segmentation. We develop a unified framework encompassing both margins and illustrate that focusing on LOP deviations provides a very limited view of the current level of market segmentation across European consumer goods markets.

Finally, we contribute to a large literature that aims to measure market segmentation by estimating border effects for trade shares. In a seminal contribution, [McCallum \(1995\)](#) established that the US-Canada border led to substantially larger reductions in goods trade than regional borders. Increasingly sophisticated methods were developed to measure national border effects appropriately ([Anderson & Wincoop \(2003\)](#)) and

and their higher-order interactions.

<sup>8</sup>As cost-of-living differences intranational region pairs are quite similar conditional and unconditional on geographic differences, the main difference comes from excluding international region pairs which have propensity scores of being separated by a national border close to one.

[Santamaria et al. \(2021\)](#)), but it is unclear what these border effects imply for geographic market segmentation when heterogeneity in consumer taste is substantial. We measure border effects in terms of cost-of-living differences and separate differences in consumer taste from the two margins that reflect geographic goods market segmentation. We find that variation in consumer taste is substantial which implies that border effects for trade shares only partly reflect geographic goods market segmentation.

Section 2 introduces the dataset and section 3 provides motivating evidence for moving beyond LOP deviations when studying market segmentation. Section 4 introduces the structural framework we use to compute and decompose cost-of-living differences into the three components. Section 5 provides the estimates of the elasticities of substitution. Section 6 estimates which part of the elevated cost-of-living differences can be attributed to geographic market segmentation in goods markets and section 7 provides some concluding remarks.

## 2 Data

To execute the analysis we rely on household-level scanner data comprising 68 FMCG categories from Belgium, France, Germany and the Netherlands. In each country, a market research firm provides a panel of households with a scanning device to register for each purchased barcode the number of units bought, the total volume purchased and the total tax-inclusive monetary value of the transaction.<sup>9</sup> In addition, buyers report the retail chain in which the product was purchased. We focus on a relatively stable period from 2010 until 2019, omitting the trough of the financial crisis and the start of the COVID-19 pandemic. We now provide more detail on the construction of the sample, and the dimensions and suitability of the data.

We restrict attention to Belgium, France, Germany and the Netherlands for two reasons. First, this group of countries is probably among the most integrated countries in the Common Market. All countries were founding partners of the European Economic Area, they have always partaken in the successive integration efforts to form the Common Market and they use the Euro as their currency. For this reason, our results likely provide a conservative estimate of the degree of market segmentation across countries in the Common Market as other countries have either joined at a later stage or use a different local currency. Second, focusing on countries that share the same currency makes relying on cost-of-living differences, or real exchange rate variation, appealing to assess the degree of geographic goods market segmentation. When countries have fluctuating nominal exchange rates, imperfections in financial markets or monetary policy shocks can induce variation in nominal exchange rates which in turn spill into real exchange rates when nominal prices are sticky (see [Heathcote & Perri \(2014\)](#) and [Itskhoki \(2021\)](#)). Like [Berka et al. \(2018\)](#), we choose to focus on a set of countries that share the same currency to ensure that real exchange rate variation is mostly determined by real factors and not by financial or monetary ones.

We focus on a set of 68 FMCG product categories, ranging from food, alcoholic and non-alcoholic beverages to personal care items that jointly represent around 15% in total final consumer spending. In all countries, the raw data contain more than the 68 product categories we include, but we limit the set of product categories for

---

<sup>9</sup>The market research firm in Belgium, Germany and the Netherlands is GfK. In France, the data is gathered by Kantar. We were granted access to the data by AIMARK (Advanced International Marketing Knowledge).



two reasons (see Table ??). First, we only keep product categories that are consumed by more than 5% of the households in all countries. Second, product categories such as medicines and first aid products are omitted because the extent to which consumers can access them through retail stores differs across countries.<sup>10</sup> Still, our sample covers most of the recorded expenditure as the excluded product categories never account for more than 11% of total expenditures in any of the countries when taken together (Table ??). Figure ?? shows the distribution of expenditure shares across the included product categories.

The transaction data records purchases at the barcode level and associates each barcode with an 8- or 13-digit EAN code.<sup>11</sup> We combine package information obtained from the barcode descriptions with information about units sold, volume sold and expenditure to compute quantity consumed in terms of metric units (number of liters, kilograms or units consumed) and unit prices in terms of prices per liters, kilograms and units.<sup>12</sup> When studying the degree of geographic market segmentation, scanner data offer three distinct advantages. First, observing barcode-level or product variety-level prices ensures that LOP deviations do not stem from differences in unobserved product characteristics. Since barcodes are inexpensive to acquire and retailers base their inventory systems on barcodes there are operational incentives to associate distinct product varieties with unique barcodes. Also, the EAN system is globally managed through GS1 and therefore two different firms will not be able to sell two different product varieties under the same EAN code.<sup>13</sup> Second, in addition to product variety-level prices, scanner data also record product variety-level data on physical quantities. Observing both variables at the same time is essential to estimate a structural model of demand and separate spatial differences in tastes from spatial differences in LOP deviations and choice sets. Finally, within the selected product categories, scanner data provide a complete picture of the overall consumption basket. This makes identifying both shared or traded varieties and purely local alternatives possible.<sup>14</sup> In addition, because all countries in the sample use the same barcode system to define product varieties, consumption baskets are comparable across both intranational and international region pairs.

While barcodes identify individual product varieties, it is possible that very similar product varieties, potentially even the same product variety, carry different barcodes across countries. Firms might deliberately offer different barcodes across countries to limit parallel imports by distributors or distributors might attach

<sup>10</sup>For instance, the distribution of these products is much more regulated in Belgium compared to the Netherlands.

<sup>11</sup>One example of a product variety is a 6-pack 330ML Can Coca-Cola Regular. Generally, barcodes carry a 13-digit identifier. However, there is a small set of product varieties that are sold in small packages, e.g. spices or small shampoo bottles or that are individually sold, e.g. small soda bottles. These product varieties will have a smaller 8-digit identifier.

<sup>12</sup>As barcode descriptions are provided by the local affiliate of the market research firms, there are a limited number of cases in which the exact barcode description for identical barcodes differs across countries. In these cases, we assume that this is due to measurement error and associate each barcode with one package size across countries such that price differences do not stem from differences in package size.

<sup>13</sup>GS1 is a global non-profit company that is in charge of, among other things, allocating unique EAN barcodes. They make sure that EAN codes are unique such that products can easily be identified by their EAN barcode

<sup>14</sup>Note that this is in stark contrast to the set of papers that rely on international trade data to study the relationship between market integration and differences in or changes in choice sets (e.g. Broda & Weinstein (2006), Kehoe & Ruhl (2013) and Cavallo et al. (2022)). International trade data usually provides a detailed picture of the choice set comprising shared or traded varieties, but measuring domestic choice set is much harder, even though it is often more important than the choice set comprising shared or traded varieties (see Eaton et al. (2011)). One reason why this is particularly hard is that trade data is usually reported according to the Harmonized System classification, and domestic production and sales data are usually reported according to industry classifications such as the NACE or ISIC systems. See Amity et al. (2019) for an application of matching trade and domestic price data.



different barcodes to product varieties when they repackage products before selling to final consumers.<sup>15</sup> Hence, relying solely on the set of common barcodes across countries to study choice set differences across international region pairs could paint an inconsistent picture. To overcome this issue, we associate each barcode with a firm identifier by matching barcodes to firms using data from GS1 who provided us with a mapping from barcodes to firm identifiers.<sup>16</sup> Table ?? indicates that we can allocate a firm identifier for around 75% to 85% of all expenditures depending on the country.<sup>17</sup> In this way, we can study choice set differences across geographic markets both at the product variety-level and at the firm-level. To check the quality of the firm identifier, we replicate the descriptive statistics on the firm size distribution documented by Hottman et al. (2016) in Tables ?? - ?.<sup>18</sup> We find that these empirical patterns are very similar across countries and closely replicate the patterns reported by Hottman et al. (2016) for US scanner data. By specifying a preference structure with nests at the product variety and firm level, our preference structure is flexible enough to capture the empirical regularities that characterize the heterogeneous multi-product firms that make up our dataset.

We link household characteristics to each purchase through a unique household identifier reported in transaction data. Crucially, we observe the region of residence and the ZIP code of the household. We combine the region of residence and ZIP-code data to establish their region of residence at the NUTS2 level and determine product variety-level prices, quantities and choice sets at a fine spatially disaggregate level. As we quantify the degree of geographic market segmentation across regions in terms of the yearly cost-of-living differences, we need to ensure that we have a representative picture of the consumption baskets of consumers in the different regions. To minimize the impact of measurement error through occasional consumption or consumers that rotate in and out of the sample in the middle of the year, we only include consumers in a given year if they register transactions in each quarter of the year. Depending on the country, the main sample includes between 3,200 and 23,348 households each year accounting for 60%-91% of total recorded expenditure within the select product categories (see Table ??). Figures ?? - ?? illustrate that the distributions of weekly shopping trips, the number of weekly purchases and the number of purchased barcodes are very similar across countries for the set of consumers that we focus on. This provides some assurance that we are comparing consumption baskets across countries for consumers with very similar overall purchase behavior. To ensure that we measure choice sets in a region as completely as possible, we use the full sample of households when determining the set of available product varieties and firms.

As mentioned before, the transaction data also record the retail chain at which the purchase was made. There are roughly six store types included in the data. Grocery stores (e.g. Carrefour Hyper, Hyper U) and hypermarkets (e.g. GB, Super U) are food and non-food supermarkets which are not discounters.

<sup>15</sup>A recent example is antitrust procedure AT.40134 by the European Commission against AB Inbev. The European Commission fined Ab Inbev after it changed the labeling on beer sold to the Netherlands to prevent it from being offered in Belgium. Besides, it is well-known that the Belgian retail chain Colruyt repackages many products before selling them in their stores.

<sup>16</sup>See Hottman et al. (2016) for a similar approach and appendix ?? for more detail on the exact procedure we use to overcome the aforementioned issues.

<sup>17</sup>Usually, when we cannot allocate a firm identifier this is because the barcode did not follow the 13-digit EAN standard or because product variety does not have an associated brand. Non-standard 13 EAN-digit are prevalent in Belgium, Germany and the Netherlands in product categories that contain a large share of fresh produce, e.g. fresh vegetables, fresh meat, etc. .

<sup>18</sup>More specifically, we document statistics about the overall firm size distribution, the upper tail of the distribution and within firm dispersion in expenditure across different barcodes.

Hypermarkets and grocery stores are large and mid-sized supermarkets respectively. Discounters are stores with an everyday low-price strategy (e.g. Aldi, Lidl). Convenience stores are smaller stores that have a limited assortment and focus on quick purchases. This group consists of the small store formats of larger retail chains (e.g. Carrefour Express, U Express), kiosks (e.g. 7-Eleven, Relay) and petrol stations (e.g. Total, Esso). Drugstores are stores that focus on mostly non-food items (e.g. Hema, MAC, KIKO). Finally, specialist stores are stores that tend to stock a limited number of product categories and are made up of bakers, liquor stores, butchers, etc. . Whereas specialist stores are often small independent stores that sell within one or two product categories, drugstores are often chains that sell across a wider set of product categories. We exclude cross-border transactions and online purchases that were made at an online retailer with no brick-and-mortar counterpart. Table ?? provides an overview of the importance of the different stores across countries and shows that most purchases are made at brick-and-mortar stores. Hence, omitting cross-border transactions and online purchases still leaves us with a very comprehensive picture of overall consumption baskets within the selected product categories.

To account for geographic differences across NUTS2 regions, we complement the consumption data with geographical data from three separate data sources. First, we use concordance tables from Eurostat to match ZIP codes and regions to their corresponding NUTS2-level.<sup>19</sup> In addition, we use data from Eurostat's GISCO services to obtain longitudes and latitudes for each of the ZIP codes. We determine the population-weighted centroids of each NUTS2 region and compute great circle distances between them. Second, we obtain elevation data from the US Geological Survey at the 30arc seconds level to determine the elevation differences between NUTS2 regions. Finally, the European Environmental Agency provides the region of river basins from which we construct whether the NUTS2 regions share a river basin.

### 3 Reduced-form evidence

In this section, we confirm the empirical relevance of LOP deviations and choice set differences as margins of geographic market segmentation. We confirm that price differences for identical product varieties are larger across international region pairs compared to intranational pairs. Importantly, there are also substantial differences in the set of barcodes and firms that are commonly available to consumers between intranational and international pairs.

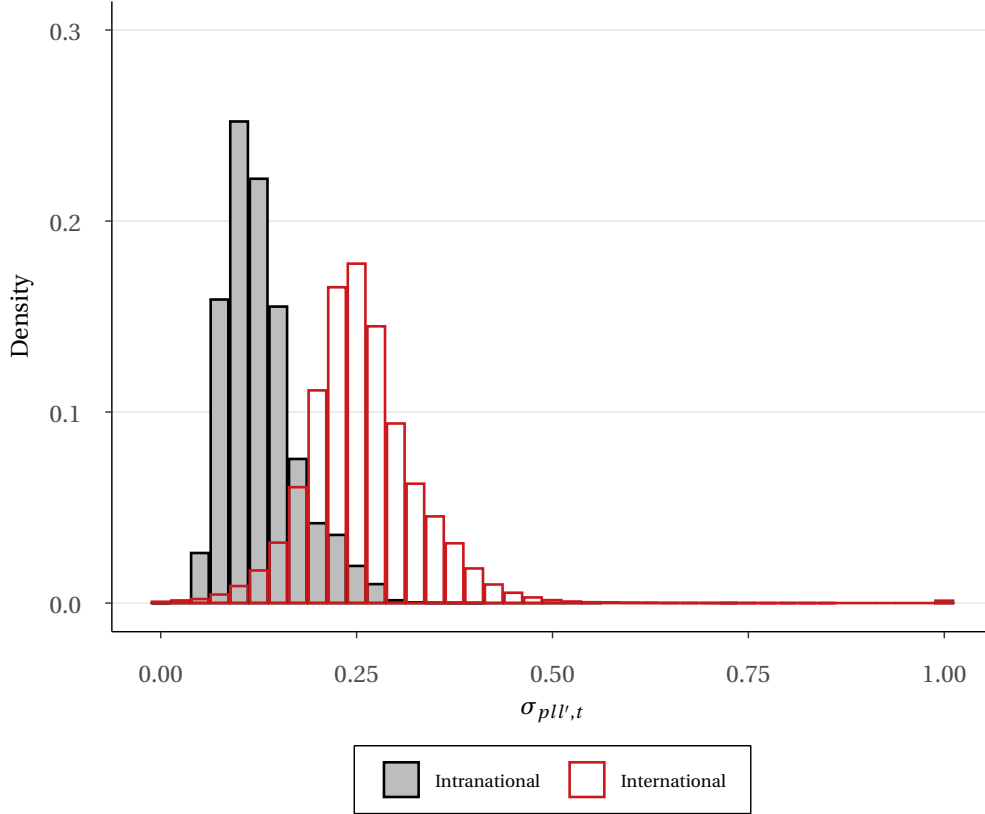
#### 3.1 LOP Deviations

To compute LOP deviations, we first aggregate the data to the product variety-NUTS2 region-year level and by computing average prices within each cell. Then, we compute for each product variety purchased in a given year the log price differences between all NUTS2 region pairs for which there exists a price observation in both regions. Figure ?? shows that the distribution of LOP deviations is centered around zero with most mass falling in between  $-0.25$  and  $0.25$  log points. To understand whether LOP deviations are larger across

<sup>19</sup>After administrative reforms in 2016, France changed their NUTS-classifications. Because the regional variable in the French dataset corresponds to the NUTS2-level of the pre-2016 NUTS2 version, we use the 2013 version of the NUTS regions throughout the paper.

international region pairs versus intranational pairs, we compute the standard deviation LOP deviations within product category-region pair-year cells across product varieties.<sup>20</sup> Figure 1 shows the conditional distribution of the standard deviations for international pairs in red and intranational region pairs in gray. Price dispersion is larger for international region pairs as the conditional distribution for international pairs as both its mode and tail are shifted to the right compared to the distribution for intranational pairs. While a little under 70% of LOP deviations fall between  $[-0.124, 0.124]$  for intranational pairs, this spread rises to  $[-0.281, 0.281]$  for international pairs.

**Figure 1:** Standard deviation of LOP deviations



*Notes:* This figure plots the conditional distributions of the standard deviation of log LOP deviations across NUTS2-region pairs. The unit of observation is a product category-NUTS2 region pair steps. We bin the standard deviation of log price differences within product category-region pairs-year cells into 40 separate bins and compute for each bin the number of observations that fall into each bin. Finally, we winsorize the standard deviations at a standard deviation of 1.

To see whether the increased price dispersion reflects geographic market segmentation, we need to account for the fact that international region pairs are also more distant from one another compared to intranational region pairs. To this end, we follow Engel & Rogers (1996) and estimate:

$$\sigma_{pll',t} = \beta B_{ll'} + \gamma d_{ll'} + \theta_l + \theta_{l'} + \lambda_{p,t} + \varepsilon_{pll',t} \quad (1)$$

<sup>20</sup>In contrast to Gopinath et al. (2011) and Beck et al. (2020) who focus on absolute log price deviations, we choose to compute standard deviations as the measure of dispersion to be consistent with section 6.

where

$$\sigma_{pll',t} \equiv \sqrt{\frac{1}{N_p^{ll'}} \sum_{i \in \mathcal{R}_p^{ll'}} (\log(p_{il',t}/p_{il,t}) - \widehat{\mathbb{E}_{pll',t}}[\log(p_{il',t}/p_{il,t})])^2}$$

is the standard deviation of log LOP deviations for region pair  $ll'$  within product category  $p$  in year  $t$ . We include a border dummy  $B_{ll'}$  which is one when region pair  $ll'$  is an international pair and zero otherwise. To control for the distance between region pairs, we include  $d_{ll'}$  which is the log of the population-weighted great circle distance between the regions. We focus on variation across region pairs within product category-year observations by including product category-year fixed effects. Finally, we add region fixed effects  $\theta_l$  and  $\theta_{l'}$  to account for the fact that certain regions might be characterized by always higher price dispersion.<sup>21</sup>

We present the baseline OLS estimates in columns (1)-(3) of Table 1. Column (1) shows the results when we estimate equation 1 and only include log distance  $d_{ll'}$ . In this case, a 100% increase in the distance between two regions leads to a statistically significant increase in the variance of log LOP differences of 8.4ppt. In line with Figure 1, column (2) shows that the standard deviation of log price differences is 15ppt larger for international pairs compared to international pairs when we only include the border dummy. Finally, in column (3) we include both distance and the border dummy. Whereas the distance effect drops considerably, the border effect remains large and precisely estimated at 15.4ppt.<sup>22</sup> Figure ?? and Figure ?? explore variation in the average border effect over time and across product categories respectively. In line with the results from Beck et al. (2020), there is little systematic variation over the years. There is, however, considerably more cross-category heterogeneity. Figure ?? shows that border effects estimates are significantly different from zero in all but one case and range from a low of 3.2ppt for packaged meat to a high of 42.5 for mineral water. Importantly, Figure ?? also shows that while there is significant cross-sectional variation in the border effects, there is little systematic cross-category variation in LOP deviations for intranational pairs.

Columns (4) - (6) present two robustness checks. First, column (4) considers cross-country heterogeneity in the border effect. If there is substantial cross-country heterogeneity in price dispersion for intranational region pairs, Gorodnichenko & Tesar (2009) shows that the border effect estimate from equation 1 is contaminated with average cross-country differences in price dispersion for intranational region pairs.<sup>23</sup> To isolate the former effect, we follow Gorodnichenko & Tesar (2009) and augment equation 1 with country dummies that are one if region  $l$  and  $l'$  belong to that country and zero otherwise for all countries apart from Germany, which functions as the baseline country. In this way, the border variable  $B_{ll'}$  measures the average difference in price dispersion for international region pairs relative to the average price dispersion for German intranational region pairs. The other country-specific border effects are obtained from  $\beta - \gamma_{c(l)}$ . Column (4) shows that the average border effect, now relative to average price dispersion for German intranational region pairs, is 12.5ppt. While the country dummies indicate that cross-country heterogeneity in price dispersion for

<sup>21</sup>Regions with disproportionately large urban centers might always have higher prices and therefore always larger price differences with any other region.

<sup>22</sup>Reassuringly, these border effects are quantitatively in line with the RDD estimates from Beck et al. (2020) which reports an average absolute price difference of around 18%.

<sup>23</sup>This point is related to the broader issue of a simple differences-in-means regression that suffers from heterogeneous treatment effect bias under non-random treatment assignment.

intranational pairs exists, column (4) also shows that all border effects are positive and precisely estimated relative to average price dispersion for intranational pairs in each country. Second, in columns (5)-(6) we consider an alternative level of aggregation. To obtain the baseline results, we collapsed the retail chain dimension. However, [Dellavigna & Gentzkow \(2019\)](#) provides evidence that LOP deviations are much larger across chains compared to within chains. Hence, by averaging across stores, we might artificially increase price dispersion, potentially leading to an inconsistent border effect. In columns (5) and (6), we re-estimate equation 1 and its augmented version after computing LOP deviations within retail chains.<sup>24</sup> We find that the average border effect falls from 15.4ppt to 13.1ppt (column (3) relative to (5)) and the spread changes from [0.125, 0.161] to [0.066, 0.262] (column (4) relative to (6)) but remains precisely estimated.

### 3.2 Choice set differences

While LOP deviations are larger across international region pairs compared to intranational pairs, the set of product varieties for which we can compute LOP deviations in the first place is quite small across international region pairs. To illustrate this, we compute two intuitive choice set dissimilarity measures: one based on counts and another based on expenditures. First, consider the product variety level and define  $\mathcal{B}_{p,l,t}$  as the set of the consumed product varieties in product category  $p$  in region  $l$  at time  $t$  and  $\mathcal{B}_p^{ll'}$  as the set of product varieties that are available in region  $l$  and region  $l'$ .<sup>25</sup> We define the dissimilarity measures based on the counts  $N_{p,t}^{B,ll'}$  and based on expenditure  $E^{B,ll'}_{p,t}$  as follows:

$$N_{p,t}^{B,ll'} \equiv 1 - \frac{\sum_{i \in \mathcal{B}_{p,l,t}} \mathbb{1}(i \in \mathcal{B}_{p,l,t} \cap i \in \mathcal{B}_p^{ll'})}{\sum_{i \in \mathcal{B}_{p,l,t}} \mathbb{1}(i \in \mathcal{B}_{p,l,t})}, \quad \lambda_{p,t}^{B,ll'} \equiv 1 - \frac{\sum_{i \in \mathcal{B}_{p,l,t}} P_{il,t} Q_{il,t} \mathbb{1}(i \in \mathcal{B}_{p,l,t} \cap i \in \mathcal{B}_p^{ll'})}{\sum_{i \in \mathcal{B}_{p,l,t}} P_{il,t} Q_{il,t} \mathbb{1}(i \in \mathcal{B}_{p,l,t})}$$

The dissimilarity measures have bounded support between zero and one: if any two regions only consume common varieties, the dissimilarity measures are equal to zero and they are at their maximum value of one if the regions do not have any product variety in common. The only difference between the two measures is that the measure based on expenditures weights the barcodes at their importance in the consumption basket. Define  $\mathcal{F}_{p,l,t}$  as the set of the firms in product category  $p$  selling in region  $l$  at time  $t$  and  $\mathcal{F}_p^{ll'}$  as the set of firms that sell both region  $l$  and region  $l'$ , then the dissimilarity measures at the firm level are similarly defined as:

$$N_{p,t}^{F,ll'} \equiv 1 - \frac{\sum_{f \in \mathcal{F}_{p,l,t}} \mathbb{1}(f \in \mathcal{F}_{p,l,t} \cap f \in \mathcal{F}_p^{ll'})}{\sum_{f \in \mathcal{F}_{p,l,t}} \mathbb{1}(f \in \mathcal{F}_{p,l,t})}, \quad \lambda_{p,t}^{F,ll'} \equiv 1 - \frac{\sum_{f \in \mathcal{F}_{p,l,t}} P_{fpl,t} Q_{fpl,t} \mathbb{1}(f \in \mathcal{F}_{p,l,t} \cap f \in \mathcal{F}_p^{ll'})}{\sum_{f \in \mathcal{F}_{p,l,t}} P_{fpl,t} Q_{fpl,t} \mathbb{1}(f \in \mathcal{F}_{p,l,t})}$$

<sup>24</sup>The number of observations is lower in this sample compared to the initial sample. This is because, by computing price differences within retail chains, it occurs that we cannot compute LOP deviations anymore for some product varieties.

<sup>25</sup>Clearly, the set of commonly available  $\mathcal{B}_p^{ll'}$  product varieties is computed based on a set of  $\mathcal{B}_{p,l,t}$ 's. There are two dimensions along which this set could be defined. First, one could define the set of common varieties as a time-varying set (see e.g. [Broda & Weinstein \(2010\)](#); [Redding & Weinstein \(2020\)](#)) or as a set that does not vary across time (e.g. [Argente et al. \(2021\)](#)). Second, one could define the set of common varieties at the country (e.g. [Broda & Weinstein \(2010\)](#); [Redding & Weinstein \(2020\)](#)) or on the regional level (e.g. [Handbury & Weinstein \(2015\)](#); [Feenstra et al. \(2020\)](#)). Because of the stability in coefficients on the border dummy over time for the absolute lof price differences (see supra) and for choice set differences (see infra), we decide to abstract from variation in the choice sets over time. We choose to define the set of common product varieties on the regional level. In this way, we allow for within-country variation in choice sets across regions. Nevertheless, the results are robust to an alternative definition.

**Table 1:** Border effect: Standard deviation of log price differences

$\sigma_{p(i)ll',t}^p$	Pooled				Store	
	(1)	(2)	(3)	(4)	(5)	(6)
$d_{ll'}$	.084*** (.002)		.004*** (.001)	.004*** (.001)	.008*** (.001)	.005*** (.001)
$B_{ll'}$		.157*** (.001)	.154*** (.001)	.158*** (.001)	.131*** (.002)	.136*** (.001)
$\mathbb{1}(c(l) = c(l') = \text{BEL})$				.016*** (.002)		-.003** (.003)
$\mathbb{1}(c(l) = c(l') = \text{FRA})$				-.003** (.003)		-.126*** (.006)
$\mathbb{1}(c(l) = c(l') = \text{NLD})$				.033*** (.002)		.07*** (.002)
$\theta_l$	✓	✓	✓	✓	✓	✓
$\theta_{l'}$	✓	✓	✓	✓	✓	✓
$\lambda_{c(i),t}$	✓	✓	✓	✓	✓	✓
$\sigma_{\text{Within}}^p$	.124	.124	.124	.122	.13	.133
$\beta - \gamma_{\text{BEL}}$	-	-	-	.142***	-	.14***
$\beta - \gamma_{\text{FRA}}$	-	-	-	.161***	-	.262***
$\beta - \gamma_{\text{NLD}}$	-	-	-	.125***	-	.066***
Nr. obs	2,221,202	2,221,202	2,221,202	2,221,202	1,515,302	1,515,302
Within R <sup>2</sup>	0.15	0.25	0.25	0.25	0.21	0.23

*Notes:* This table presents the OLS estimates from equation 1. Columns (1)-(4) show the results from estimating the model when we compute price differences based on prices that are averaged across stores.  $\hat{\sigma}_{\text{within}}^p$  represents the average value of the left-hand side variable for the intranational pairs in all columns, except in columns (4) and (6), (9) and (12). In these columns, this number represents the average for the baseline country which is Germany. We cluster standard errors at the region pairs and present them in brackets below the coefficient estimates. Reported significance levels are at the  $p < 0.1^*$ ,  $p < 0.05^{**}$  and  $p < 0.01^{***}$  levels.

Figure 2 shows the conditional distributions of the dissimilarity measures across product category-year for intranational pairs in grey and international pairs in red. Starting with panel (a), which plots the count-based measure at the product variety level, there is a clear separation between the two conditional distributions. Whereas intranational pairs share more than 75% of product varieties, it is very uncommon for international pairs to share more than 25% of product varieties. Panel (b) confirms that this disparity continues to hold when we measure the dissimilarity in terms of expenditures. Moreover, Figures ?? - ?? plot the conditional distributions for a subsample of private label and branded product varieties and a subsample of branded product varieties only and confirm the same pattern.<sup>26</sup> Panels (c) and (d) show the conditional distributions of the dissimilarity measures at the firm level. While the count-based distributions remain equally separated, the expenditure-based conditional distributions overlap much more. Thus, in contrast to the set of common product varieties, the small set of common firms attracts a much larger share of total expenditure in each region. To control for increased geographic differences across international region pairs, we re-estimate equation 1 for the four dissimilarity measures:

$$y_{pll',t} = \beta B_{ll'} + \gamma d_{ll'} + \theta_l + \theta_{l'} + \lambda_{p,t} + \varepsilon_{p(i)ll',t} \quad (2)$$

where  $y_{pll',t} = \{N_{p,t}^{B,ll'}, \lambda_{p,t}^{B,ll'}, N_{p,t}^{F,ll'}, \lambda_{p,t}^{F,ll'}\}$ . As before, we include a border dummy  $B_{ll'}$  which is one when NUTS2-region  $l$  and NUTS2-region  $l'$  are an international region pair and zero otherwise,  $d_{ll'}$ , the log of the population-weighted distance between NUTS2-region  $l$  and NUTS2-region  $l'$  and product category-year  $\lambda_{p,t}$  and region fixed effects  $\theta_l$  and  $\theta_{l'}$ .

Table 2 presents the baseline OLS estimates. First, columns (1), (4), (7) and (10) show that more distant regions are characterized by larger choice set differences independent of the dissimilarity measure used. Even if the distance coefficient considerably drops when conditioning on national borders in columns (2), (5), (7) and (11), it remains positive and significantly different from zero. For instance, column (2) shows that a 100% increase in distance or a doubling of the distance between two regions increases choice set differences in terms of product variety counts by 5.1%. Second, in line with Figures 2a and 2b we estimate large border effects for the product variety level dissimilarity measures. Within product category - years, column (2) illustrates that the overlap in choice sets decreases by more than 71.7ppt for international pairs. In other words, intranational pairs have more than five times as many product varieties in common compared to international pairs.<sup>27</sup> The border effect for expenditure-based choice set differences in column (5) is similarly estimated at 73.8ppt. Although there is cross-country heterogeneity in the average choice set dissimilarity for intranational pairs, columns (3) and (6) confirm that the border effects are positive, independent of the chosen baseline country. Third, the border effects for the dissimilarity measures at the firm level are lower relative to the ones at the product variety level. Column (8) shows that intranational pairs have more than five times as many firms in common compared to international pairs.<sup>28</sup> However, consistent with Figure 2d the firm-level border effects in terms of expenditure are substantially lower. Column (10) illustrates that consumers across international pairs spend

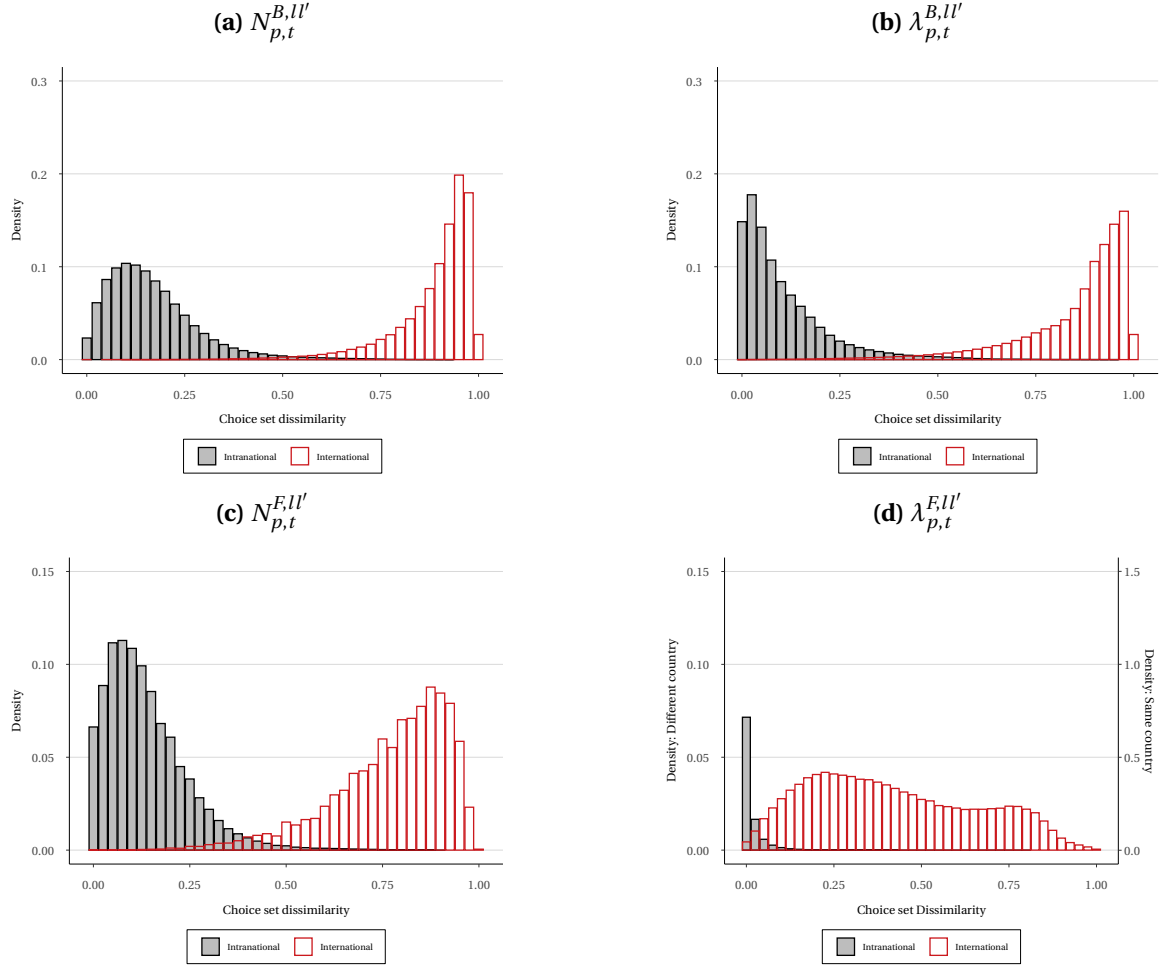
<sup>26</sup>The fact that the set of common barcodes across international pairs is small resonates well with prior work that studies LOP deviations across European countries (e.g. Cavallo et al. (2014); Beck et al. (2020))

<sup>27</sup>This number is obtained from the following calculation:  $\frac{0.178+0.717}{0.178} = 5.028$

<sup>28</sup>This number is obtained from the following calculation:  $\frac{0.149+0.649}{0.149} = 5.35$



**Figure 2: Choice set dissimilarity**



*Notes:* This figure plots the distribution for the count-based dissimilarity measures across NUTS2-region pairs. The unit of observation is at the product category-region  $l$ -region  $l'$ -year level. The grey bars plot the distribution for intranational pairs and the red bars do the same for international pairs. Panel (a) plots the count-based dissimilarity measure and panel (b) plots the expenditure-based dissimilarity measure for product varieties. Panel (c) and (d) do the same for firm-level dissimilarity measures. Note that panel (d) has two y-axes: the left y-axis scales with the different country region pairs and the right y-axis scales with the same country region pairs.

on average more than 50% on product varieties provided by firms common to both regions. Nevertheless, large 370  
differences remain as consumers in intranational pairs spend on average more than 97% on product varieties 371  
provided by common firms. Columns (9) and (12) confirm that these results are robust to accounting for cross- 372  
country heterogeneity in differences for intranational pairs. <sup>29</sup> 373

---

<sup>29</sup>Throughout the analysis, Belgium seems to consistently have lower estimated border effects. The country dummy indicates that variation across Belgian region pairs is consistently higher compared to other countries. Therefore, the border effects from the Belgian perspective are lower.

**Table 2: Border effect: Choice set dissimilarity**

$y_{p(i)l',t}$	$N^{B,l'}$			$\lambda^{B,l'}$			$N^{F,l'}$			$\lambda^{F,l'}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$d_{l'}$	.426*** (.009)	.051*** (.003)	.058*** (.002)	.449*** (.01)	.062*** (.003)	.068*** (.002)	.376*** (.008)	.037*** (.003)	.048*** (.002)	.255*** (.006)	.035*** (.003)	.043*** (.002)
$B_{l'}$		.717*** (.003)	.767*** (.004)		.738*** (.003)	.778*** (.004)		.649*** (.004)	.754*** (.004)		.421*** (.004)	.474*** (.004)
$\mathbb{1}(c(l) = c(l') = \text{BEL})$			.285*** (.009)			.251*** (.009)			.398*** (.008)			.362*** (.008)
$\mathbb{1}(c(l) = c(l') = \text{FRA})$			.037*** (.007)			.002 (.008)			.147*** (.008)			-.005* (.008)
$\mathbb{1}(c(l) = c(l') = \text{NLD})$			.138*** (.007)			.16*** (.009)			.256*** (.007)			.221*** (.007)
$\theta_l$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\theta_{l'}$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\lambda_{c(i),t}$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\hat{\mathbb{E}}(y_{c(i)l',t} \text{same})$	.178	.178	.191	.117	.117	.129	.149	.149	.142	.023	.023	.025
$\beta - \gamma_{\text{BEL}}$	-	-	.482***	-	-	.526***	-	-	.356***	-	-	.112***
$\beta - \gamma_{\text{FRA}}$	-	-	.73***	-	-	.776***	-	-	.607***	-	-	.479***
$\beta - \gamma_{\text{NLD}}$	-	-	.629***	-	-	.617***	-	-	.498***	-	-	.253***
Nr. obs	4,627,670	4,627,670	4,627,670	4,627,670	4,627,670	4,627,670	4,627,670	4,627,670	4,627,670	4,627,670	4,627,670	4,627,670
Within R <sup>2</sup>	0.51	0.94	0.94	0.50	0.91	0.92	0.47	0.88	0.90	0.35	0.64	0.66

*Notes:* This table presents the OLS estimates from equation . Columns (1) - (3) and columns (4) - (6) estimate the results for the count-based and expenditure dissimilarity measure on the product variety level respectively. Columns (7) - (9) and columns (10) - (12) estimate the results for the count-based and expenditure dissimilarity measure on the firm level respectively.  $\hat{\mathbb{E}}(y_{p(i)l',t}|\text{same})$  represents the average value of the left-hand side variable for the same country region pairs in all columns, expect in columns (3), (6), (9) and (12). In these columns, this number represents the average for the baseline country which is Germany. We cluster standard errors at the region pairs and present them in brackets below the coefficient estimates. Reported significance levels are at the  $p < 0.1$  \*,  $p < 0.05$  \*\* and  $p < 0.01$  \*\*\* levels.

In accordance with the results for the standard deviation of log price differences, there is much more heterogeneity across product categories compared to the time series. Figure ?? shows that the border effects are very stable for the count- and expenditure-based dissimilarity measures, independent of whether they are computed at the product variety or the firm level. In contrast, Figures ?? - ?? illustrate that, consistently across all the dissimilarity measures, there is substantial variation across product categories. For instance, Figure ??, which plots the expenditure-based dissimilarity measure at the firm level, shows that choice set overlap varies from a little under 20ppt for deodorants to 60ppt for toilet paper all the while both product categories have a very similar choice set overlap for intranational pairs.

## 4 Theoretical Framework

The previous section showed that both LOP deviations and choice set differences are larger across international region pairs compared to intranational pairs. To jointly analyze both margins of geographic market segmentation, we now specify a flexible model of consumer preferences that provides a theoretically-grounded expression for regional cost-of-living differences. We use the structure of the model to decompose this expression into (1) LOP deviations, corrected for substitution effects, (2) pure taste differences and (3) choice set differences.

### 4.1 Consumer preferences

As mentioned before, we model preferences as a nested CES utility system for three reasons. First, nested CES utility systems have shown the ability to fit empirical patterns in international economics while maintaining empirical tractability. In particular, [Atkeson & Burstein \(2008\)](#) and [Hottman et al. \(2016\)](#) show that nested CES system can rationalize the positive correlations between markups and firm size uncovered in [Loecker & Warzynski \(2012\)](#) and [Loecker et al. \(2016\)](#) and between markup elasticities and firm size (see [Berman et al. \(2012\)](#); [Amiti et al. \(2019\)](#)). At the same, with data on prices and quantities, the elasticities of substitution are the only structural parameters one needs to estimate. Second, the nested CES utility system is part of a family of preference systems that admits an intuitive way to appraise the importance of taste differences for cost-of-living differences. [Redding & Weinstein \(2020\)](#) show how, given a normalization of the average taste level across regions, the importance of taste differences can be uncovered by comparing cost-of-living differences obtained from the general nested CES price index to the Sato-Vartia CES price index, which restricts tastes to be equal across regions. Finally, CES systems are the workhorse model to understand differences in product variety in international trade (e.g. [Feenstra \(1994\)](#); [Broda & Weinstein \(2006\)](#)), urban economics (e.g. [Handbury & Weinstein \(2015\)](#)) and macroeconomics (e.g. [Romer \(1990\)](#); [Jaravel \(2019\)](#)). To gauge the role of choice set differences, it intuitively depends on expenditure shares on the common set of product varieties and firms and estimates of the elasticities of substitution at each level of aggregation.

**Preferences** Within each region consumers derive utility from a triple nested utility system. We assume that the final good aggregator is separable across the set of product categories  $\mathcal{P}$ , but we leave its particular

functional form unspecified:

$$U(C_{l,t}) = F_{l,t} \left( \{C_{pl,t}\}_{p=1}^{\mathcal{P}} \right)$$

where  $F_{l,t}(\cdot)$  is the final good aggregator which can be region-specific and time-varying.<sup>30</sup>  $C_{pl,t}$  is the real consumption level in region  $l$  of product category  $p$  at time  $t$ . Each product category-level consumption bundle  $C_{pl,t}$  comprises two CES-utility nests that sequentially aggregate consumption of individual product varieties. In the middle nest, consumers allocate  $C_{pl,t}$  across different firms that supply at least one product variety in that product category and region subject to the following aggregator:

$$C_{pl,t} = \left( \sum_{f \in \mathcal{F}_{pl,t}} (\phi_{fpl,t} C_{fpl,t})^{\frac{\eta_p - 1}{\eta_p}} \right)^{\frac{\eta_p}{\eta_p - 1}}$$

where  $\mathcal{F}_{pl,t}$  is the set of firms that supply at least one product variety in product category  $p$  in region  $l$  at time  $t$ .  $\phi_{fpl,t}$  represents consumer taste for product varieties supplied by firm  $f$  in product category  $p$  in region  $l$  at time  $t$  and  $\eta_p$  denotes the constant elasticity of substitution across firms which is allowed to vary across product categories. In the lower nest, consumers allocate  $C_{fpl,t}$  across product varieties, denoted by  $i$  subject to another CES-utility aggregator:

$$C_{il,t} = \left( \sum_{i \in \mathcal{B}_{fpl,t}} (\phi_{il,t} C_{il,t})^{\frac{\sigma_p - 1}{\sigma_p}} \right)^{\frac{\sigma_p}{\sigma_p - 1}}$$

where  $\mathcal{B}_{fpl,t}$  is the set of product varieties supplied by firm  $f$  in product category  $p$  in region  $l$  at time  $t$ .  $\phi_{il,t}$  captures consumer taste for product variety and  $\sigma_p$  is the elasticity of substitution across product varieties which is also allowed to vary across product categories.  $C_{il,t}$  is the corresponding consumption level of variety  $i$  in region  $l$  at time  $t$ .

**Cost-of-living** Given the preference structure and the assumption that consumers maximize utility subject to their budget constraint  $P_{l,t} C_{l,t} \leq I_{l,t}$ , the product category-level cost-of-living for consumers in region  $l$  at time  $t$  is given by:

$$P_{pl,t} = \left( \sum_{f \in \mathcal{F}_{pl,t}} \left( \frac{P_{fpl,t}}{\phi_{fpl,t}} \right)^{1-\eta_p} \right)^{\frac{1}{1-\eta_p}}, \quad \text{where} \quad P_{fpl,t} = \left( \sum_{i \in \mathcal{B}_{fpl,t}} \left( \frac{P_{il,t}}{\phi_{il,t}} \right)^{1-\sigma_p} \right)^{\frac{1}{1-\sigma_p}}$$

and where  $P_{il,t}$  is the price of product variety  $i$  in region  $l$  at time  $t$ .

## 4.2 Decomposing Cost-of-living Differences

We decompose cost-of-living differences across two regions  $l$  and  $l'$  into three structural components: (1) LOP deviations, corrected for substitution effects (2) pure taste differences and (3) choice set differences. To this end, we define the share spent in region  $l$  at time  $t$  on firms that sell both in region  $l$  and region  $l'$  in product category  $p$ ,  $\lambda_{pl,t}^{F,l'}$ , and the share spent in region  $l$  in at time  $t$  on common product varieties sold by firm  $f$

<sup>30</sup>A Cobb-Douglas aggregator would satisfy this condition.

between region  $l$  and region  $l'$  in product category  $p$ ,  $\lambda_{fpl,t}^{B,ll'}$ , as:

$$\lambda_{pl,t}^{F,ll'} \equiv \frac{\sum_{f \in \mathcal{F}_p^{ll'}} P_{fpl,t} C_{fpl,t}}{\sum_{f \in \mathcal{F}_{pl,t}} P_{fpl,t} C_{fpl,t}}, \quad \lambda_{fpl,t}^{B,ll'} \equiv \frac{\sum_{i \in \mathcal{B}_{fp}^{ll'}} P_{il,t} C_{il,t}}{\sum_{i \in \mathcal{B}_{fpl,t}} P_{il,t} C_{il,t}} \quad (437)$$

where  $\mathcal{F}_p^{ll'}$  is the set of firms that sell both to region  $l$  and region  $l'$  in product category  $p$ ,  $\mathcal{F}_{pl,t}$  is the set of all firms selling to region  $l$  in product category  $p$  at time  $t$ . Likewise,  $\mathcal{B}_{fp}^{ll'}$  is the set of product varieties sold by firm  $f$  that are available in region  $l$  and region  $l'$  in product category  $p$  and  $\mathcal{B}_{fpl,t}$  is the set of all product varieties that are available in region  $l$  sold by firm  $f$  in product category  $p$  at time  $t$ . In addition, define for all firms and for all product varieties that sell in region  $l$  and  $l'$  in product category  $p$  and the common market share in region  $l$  at time  $t$  as:

$$S_{fpl,t}^{ll'} \equiv \frac{P_{fpl,t} C_{fpl,t}}{\sum_{f \in \mathcal{F}_p^{ll'}} P_{fpl,t} C_{fpl,t}}, \quad S_{il,t}^{ll'} \equiv \frac{P_{il,t} C_{il,t}}{\sum_{i \in \mathcal{B}_{fp}^{ll'}} P_{il,t} C_{il,t}} \quad (444)$$

This allows us to write the market shares of firms and product varieties that are available in region  $l$  and region  $l'$  at time  $t$  as a combination of their common market share and the share spent on common firms and product varieties in region  $l$  at time  $t$ :

$$S_{fpl,t} = S_{fpl,t}^{ll'} \lambda_{pl,t}^{F,ll'} \quad \forall f \in \mathcal{F}_p^{ll'}, \quad S_{il,t} = S_{il,t}^{ll'} \lambda_{fpl,t}^{B,ll'} \quad \forall i \in \mathcal{B}_{fp}^{ll'} \quad (448)$$

Using the definitions of the common market shares and the definitions of the expenditure shares on common firms and product varieties, we can write the difference in the cost-of-living between region  $l'$  and  $l$  for product category  $p$  as the sum of two parts:

$$\ln \left( \frac{P_{pl',t}}{P_{pl,t}} \right) = \frac{1}{N_p^{F,ll'}} \sum_{f \in \mathcal{F}_p^{ll'}} \left[ \ln \left( \frac{P_{fpl',t}}{P_{fpl,t}} \right) - \ln \left( \frac{\varphi_{fpl',t}}{\varphi_{fpl,t}} \right) + \frac{1}{\eta_p - 1} \ln \left( \frac{S_{fpl',t}^{ll'}}{S_{fpl,t}^{ll'}} \right) \right] + \frac{1}{\eta_p - 1} \ln \left( \frac{\lambda_{pl',t}^{F,ll'}}{\lambda_{pl,t}^{F,ll'}} \right) \quad (452)$$

where  $N_p^{F,ll'} \equiv |\mathcal{F}_p^{ll'}|$  is the number of firms in product category  $p$  that sell to region  $l'$  and  $l$ . The first part captures cost-of-living differences between the two regions that stem from price and expenditure differences within the set of firms that sell to both regions. This term intuitively depends on the differences in the unweighted geometric average price level across region  $l'$  and region  $l$ . This term is given by  $\prod_{f \in \mathcal{F}_p^{ll'}} \left[ P_{fpl',t} / P_{fpl,t} \right]^{\frac{1}{N_p^{F,ll'}}}$  and captures the fact that if firm-level prices are on average higher in region  $l'$ , the cost-of-living in region  $l'$  should be higher.<sup>31</sup> However, to translate average price level differences into welfare-relevant numbers, two correction terms are required. The first is the difference in the unweighted geometric average taste levels across regions. Still, as in Redding & Weinstein (2020), we rule out cost-of-living differences that solely reflect differences in the average taste level across regions. The reason for this normalization is related to the fact we only observe choices and not the underlying utility levels. Hence, to avoid arbitrary shifts in cost-of-living levels, the cost-of-living levels should be the same across regions if

<sup>31</sup>This price index is also known as the Jevons price index.

prices and market shares are identical across regions. In practice, we make the following normalization:

$$\tilde{\varphi}_{pl',t} \equiv \prod_{f \in \mathcal{F}_{pl',t}} [\varphi_{fpl,t}]^{\frac{1}{N_p^{ll'}}} = \prod_{f \in \mathcal{F}_{pl,t}} [\varphi_{fpl,t}]^{\frac{1}{N_p^{ll'}}} \equiv \tilde{\varphi}_{pl,t}$$

Importantly, the normalization does not rule out taste differences across regions altogether. This is because the second correction term  $\prod_{f \in \mathcal{F}_p^{ll'}} \left[ \left( S_{fpl',t}^{ll'} / S_{fpl,t}^{ll'} \right)^{\frac{1}{\eta_p - 1}} \right]^{\frac{1}{N_p^{ll'}}}$ , which is the difference in the unweighted geometric average of the firm-level common market share across regions, represents a substitution effect and differences in consumer taste. The substitution effect captures the idea that in the presence of LOP deviations, consumers in different regions will have different expenditure shares as they substitute away from expensive bundles to bundles with lower relative prices. To reflect the relative importance of each product in total expenditure in each region, we need to correct the average price differences with a term that is a function of the market shares on common firms in each region.<sup>32</sup> To see why the second correction term also captures taste differences across regions, suppose that consumer tastes are more dispersed in region  $l'$  relative to region  $l$  and that there are no LOP deviations. Such a difference in dispersion in consumer taste leads consumers in  $l'$  to allocate a greater share of expenditure firms for which they have a high taste which results in a lower cost-of-living level for consumers in region  $l'$ . As the greater dispersion in consumer taste also increases the dispersion in firm-level common market shares, it lowers the geometric average of common market shares reflecting the lower cost-of-living level. In addition to the difference in geometric average common market shares, the second correction term also depends on the firm-level elasticity of substitution. To see how, consider the limit when goods are perfect substitutes,  $\eta_p \rightarrow \infty$ . In this case, goods are perfect substitutes and common market shares are completely detached from taste-adjusted relative prices. Therefore, changes in taste-adjusted prices, originating from LOP deviations or taste differences, do not matter anymore for cost-of-living differences. In contrast, when firm-level bundles are very differentiated,  $\eta_p \rightarrow 1$ , consumers value taste-adjusted price differences which translates into a much larger correction term given the same difference in the geometric average common market share.

The second term in the expression,  $\left( \lambda_{pl',t}^{F, ll'} / \lambda_{pl,t}^{F, ll'} \right)^{\frac{1}{\eta_p - 1}}$ , accounts for choice set differences across regions. This term is the cross-sectional firm-level equivalent of the product variety term derived in [Feenstra \(1994\)](#) and [Broda & Weinstein \(2006\)](#) and depends on two moments of the data. For a given elasticity of substitution, the expenditure share on common firms in region  $l'$  region relative to region  $l$  captures the relative importance of common firms in total expenditure. If the expenditure share on common firms is lower in region  $l'$  it must mean that consumers in region  $l'$  have a better option outside of the set of common firms compared to consumers in region  $l$  which translates into a lower cost of living in region  $l'$ . The extent to which relative differences in expenditure on common firms lead to a lower cost of living depends on the firm-level elasticity of substitution. If this elasticity of substitution is high, bundles offered by firms are considered close substitutes. In this case, an increase in the choice set adds little additional substitution possibilities and the effect on the cost of living

<sup>32</sup>Note that this is also the effect that gives rise to the traditionally-used Sato-Vartia weights. However, these weights are only the correct weights for CES preferences under the assumption that there are no differences in consumer taste:  $\varphi_{fpl',t} = \varphi_{fpl,t} \quad \forall f \in \mathcal{F}_{p,t}^{ll'}$ . The Sato-Vartia price index is, for instance, used by [Handbury & Weinstein \(2015\)](#) to study cost-of-living differences across US cities.



should be small. 497

The differences in the unweighted geometric average price level still depend on  $P_{fpl,t}$  which are firm-level 498  
aggregators across product varieties part of the sets  $\mathcal{B}_{fpl,t}$ . To eliminate these firm-level price indices, we 499  
further decompose them like before: 500

$$\begin{aligned}
 \ln\left(\frac{P_{pl',t}}{P_{pl,t}}\right) &= \underbrace{\frac{1}{N_p^{F,ll'}} \sum_{f \in \mathcal{F}_p^{ll'}} \left[ \frac{1}{N_p^{B,ll'}} \sum_{i \in \mathcal{B}_p^{ll'}} \ln\left(\frac{P_{il',t}}{P_{il,t}}\right) \right]}_{\text{LOP deviations}} \\
 &+ \underbrace{\frac{1}{N_p^{F,ll'}} \sum_{f \in \mathcal{F}_p^{ll'}} \left[ \frac{1}{\eta_p - 1} \ln\left(\frac{S_{fpl',t}^{ll'}}{S_{fpl,t}^{ll'}}\right) + \frac{1}{\sigma_p - 1} \frac{1}{N_p^{B,ll'}} \sum_{i \in \mathcal{B}_p^{ll'}} \ln\left(\frac{S_{il',t}^{ll'}}{S_{il,t}^{ll'}}\right) \right]}_{\text{Common expenditure share differences}} \\
 &+ \underbrace{\frac{1}{\eta_p - 1} \ln\left(\frac{\lambda_{pl',t}^{ll'}}{\lambda_{pl,t}^{ll'}}\right) + \frac{1}{\sigma_p - 1} \frac{1}{N_p^{F,ll'}} \sum_{f \in \mathcal{F}_p^{ll'}} \ln\left(\frac{\lambda_{fpl',t}^{ll'}}{\lambda_{fpl,t}^{ll'}}\right)}_{\text{Choice set differences}}
 \end{aligned} \tag{3} \quad 501$$

Equation 3 illustrates that we can express cost-of-living differences across regions into three terms: (1) LOP 502  
deviations (2) common expenditure share differences and (3) choice set differences. Like before, we have 503  
normalized the unweighted geometric average of product variety-level taste differences such that 504  
cost-of-living levels are the same if prices, expenditure shares and choice sets are the same across regions: 505

$$\tilde{\varphi}_{fpl',t} \equiv \prod_{i \in \mathcal{B}_{fpl',t}} [\varphi_{il,t}]^{\frac{1}{N_{fp}^{ll'}}} = \prod_{i \in \mathcal{B}_{fpl,t}} [\varphi_{fpl,t}]^{\frac{1}{N_{fp}^{ll'}}} \equiv \tilde{\varphi}_{fpl,t} \tag{3} \quad 506$$

The common firm-level and product variety-level market share terms capture both substitution effects 507  
and taste differences. Isolating taste differences from LOP deviations, choice set differences and substitution 508  
effects is needed for the following reason. In a world without geographic market segmentation, LOP 509  
deviations and choice set differences only occur because the cost of physically moving goods to specific 510  
regions differs. Conversely, once the cost of physically moving goods is appropriately controlled for, LOP 511  
deviations and choice set differences should vanish. Importantly, also substitution effects zero in the absence 512  
of LOP deviations, but differences in taste differences remain. 513

To isolate taste differences from the other terms, we follow Redding & Weinstein (2020) and add and 514

subtract:  $\sum_{f \in \mathcal{F}_p^{ll'}} \omega_{fpl,t} \left[ \sum_{i \in \mathcal{B}_p^{ll'}} \omega_{il,t} \ln \left( \frac{P_{il',t}}{P_{il,t}} \right) \right]$  from equation 3:

$$\begin{aligned}
 \ln \left( \frac{P_{pl',t}}{P_{pl,t}} \right) &= \underbrace{\sum_{f \in \mathcal{F}_p^{ll'}} \omega_{fpl,t} \left[ \sum_{i \in \mathcal{B}_p^{ll'}} \omega_{il,t} \ln \left( \frac{P_{il',t}}{P_{il,t}} \right) \right]}_{\text{LOP deviations + Substitution Effect}} \\
 &+ \underbrace{\frac{1}{N_p^{F,ll'}} \sum_{f \in \mathcal{F}_p^{ll'}} \left[ \frac{1}{N_p^{B,ll'}} \sum_{i \in \mathcal{B}_p^{ll'}} \left( \frac{P_{il',t}}{P_{il,t}} \right) \right] - \sum_{f \in \mathcal{F}_p^{ll'}} \omega_{fpl,t} \left[ \sum_{i \in \mathcal{B}_p^{ll'}} \omega_{il,t} \ln \left( \frac{P_{il',t}}{P_{il,t}} \right) \right]}_{\text{Taste differences}} \\
 &+ \underbrace{\frac{1}{N_p^{F,ll'}} \sum_{f \in \mathcal{F}_p^{ll'}} \left[ \frac{1}{\eta_p - 1} \ln \left( \frac{S_{fpl',t}^{ll'}}{S_{fpl,t}^{ll'}} \right) + \frac{1}{\sigma_p - 1} \frac{1}{N_p^{B,ll'}} \sum_{i \in \mathcal{B}_p^{ll'}} \ln \left( \frac{S_{il',t}^{ll'}}{S_{il,t}^{ll'}} \right) \right]}_{\text{Taste differences (ctd.)}} \\
 &+ \underbrace{\frac{1}{\eta_p - 1} \ln \left( \frac{\lambda_{pl',t}^{ll'}}{\lambda_{pl,t}^{ll'}} \right) + \frac{1}{\sigma_p - 1} \frac{1}{N_p^{F,ll'}} \sum_{f \in \mathcal{F}_p^{ll'}} \ln \left( \frac{\lambda_{fpl',t}^{ll'}}{\lambda_{fpl,t}^{ll'}} \right)}_{\text{Choice set differences}}
 \end{aligned} \tag{4}$$

517

where

$$\begin{aligned}
 \omega_{fpl,t} &\equiv \frac{\frac{S_{fpl',t}^{ll'} - S_{fpl,t}^{ll'}}{\ln S_{fpl',t}^{ll'} - \ln S_{fpl,t}^{ll'}}}{\sum_{f \in \mathcal{F}_p^{ll'}} \frac{S_{fpl',t}^{ll'} - S_{fpl,t}^{ll'}}{\ln S_{fpl',t}^{ll'} - \ln S_{fpl,t}^{ll'}}}, & \omega_{il,t} &\equiv \frac{\frac{S_{il',t}^{ll'} - S_{il,t}^{ll'}}{\ln S_{il',t}^{ll'} - \ln S_{il,t}^{ll'}}}{\sum_{i \in \mathcal{B}_p^{ll'}} \frac{S_{il',t}^{ll'} - S_{il,t}^{ll'}}{\ln S_{il',t}^{ll'} - \ln S_{il,t}^{ll'}}},
 \end{aligned}$$

518

This expression decomposes cost-of-living differences into (1) LOP deviations, corrected for substitution effects, (2) pure taste differences and (3) choice set differences. The first part is the Sato-Vartia price index and captures only LOP deviations and substitution effects. The reason why it captures LOP deviations and substitution effects and not taste differences is that it is the ideal price index for common firms and product varieties for our nested CES preference system in the absence of taste differences.<sup>33</sup> The second part captures pure taste differences. This is because this term is defined as the difference between the generalized price index, which allows for pure taste differences, and the Sato-Vartia price index, which abstracts from taste differences. This term is the cross-sectional analog to the taste-bias term derived in Redding & Weinstein (2020). The final part is the same as the one in equation 3 and therefore still captures differences in choice sets between region  $l'$  and  $l$ . This expression forms the basis to detect geographic market segmentation in section 6 and to separate taste differences from LOP deviations and choice set differences.

529

<sup>33</sup>This follows immediately from the derivation of the common market share terms when setting  $\varphi_{fpl',t} = \varphi_{fpl,t} \quad \forall f \in \mathcal{F}_p^{ll'}$  and  $\varphi_{il',t} = \varphi_{il,t} \quad \forall i \in \mathcal{B}_p^{ll'}$

## 5 Structural Estimation

The expressions to compute the terms capturing taste and choice set differences crucially depend on the product variety- and firm-level elasticities of substitution. This section starts with estimating the product variety level elasticities of substitution and uses these estimates to construct firm-level price and quantity indices used to estimate the firm-level elasticities of substitution. In turn, we use these elasticities of substitution to construct cost-of-living differences and provide a first look into the size of cost-of-living differences and the importance of the three terms of the decomposition.

### 5.1 Elasticities of Substitution

#### 5.1.1 Product variety-level elasticities - $\sigma_p$

**Estimation strategy** To estimate the variety-level elasticity of substitution, we return to the product variety-level demand system. Demand for product variety  $i$  in region  $l$  at time  $t$  is given by:

$$C_{il,t} = \varphi_{il,t}^{\sigma_p-1} \left( \frac{P_{il,t}}{P_{fpl,t}} \right)^{-\sigma_p} C_{fpl,t}$$

as before  $C_{il,t}$  is the consumed quantity in terms of metric units,  $\varphi_{il,t}$  is the taste parameter and  $P_{il,t}$  is the unit price for product variety  $i$  in region  $l$  at time  $t$ . In addition,  $P_{fpl,t}$  and  $C_{fpl,t}$  are the firm-level price index and consumption index in region  $l$  at time  $t$  and  $\sigma_p$  is the product variety-level elasticity of substitution. Taking logs, we have:

$$c_{il,t} = -\sigma_p p_{il,t} + \sigma_p p_{fpl,t} + c_{fpl,t} + (\sigma_p - 1) \ln(\varphi_{il,t})$$

where small letters indicate logarithmic transformations of the level variables, e.g.  $c_{il,t} \equiv \ln(C_{il,t})$ . To estimate elasticities of substitution, we rely on the fact that after collapsing the household dimension in the transaction data the unit of observation is at the product variety-region-retail chain-time level. In practice, we consider the following model:

$$c_{icl,t} = -\sigma_p p_{icl,t} + \theta_{icn(l),y(t)} + \theta_{icn(l),w(t)} + \lambda_{fpcl,t} + \varepsilon_{icl,t} \quad (5)$$

where  $\varepsilon_{icl,t}$  subsumes the structural residual  $\varphi_{il,t}$ . Estimating the elasticities of substitution is subject to two endogeneity concerns. First, the price and consumption index  $P_{fpl,t}$  and  $C_{fpl,t}$  depend on the structural residuals that determine the quantity level. Second, if price variation is due to temporary discounts and promotions, product variety-level prices are likely correlated with the structural residual as well. We deal with the first issue by including firm-category-chain-NUTS2-time fixed effects  $\lambda_{fpcl,t}$ . The inclusion of a full set of fixed effects at the level of the indices is a common strategy to soak up this variation (e.g. [Atkin et al. \(2018\)](#); [Fajgelbaum et al. \(2020\)](#)). Moreover, the addition of  $\lambda_{fpcl,t}$  also controls for any time-varying region-chain-specific demand shock that affects the product varieties of a specific firm in the same way. Collapsing the data at the variety-region-chain-time dimension is important to deal with the second identification concern. [Dellavigna & Gentzkow \(2019\)](#) illustrates that retail chains tend to follow uniform pricing strategies in which they frequently change prices over time through temporary discounts while

limiting spatial variation to a minimum.<sup>34</sup> The idea behind the identification strategy is that once we condition on the seasonal variation in prices and quantities, the lower-frequency variation in prices and quantities should reflect variation due to cost factors. To this end, we include highly-detailed product variety-chain-country-year fixed effects,  $\theta_{icn(l),y(t)}$ , and product variety-chain-country-week-of-the-year fixed effects,  $\theta_{icn(l),w(t)}$ . By interacting product variety-retail chain-country fixed effects with week-of-the-year and year effects, we capture seasonal variation in prices and quantities across variety-store cells at the weekly frequency and allow this pattern to shift across the years. However, even conditional on these fixed effects, there might still be price variation that is potentially correlated with occasional local demand shocks outside of this recurring seasonal pattern at the country level. Therefore, we instrument product variety-chain level prices with the average within-chain product variety level price across the other regions. The suitability of this Hausman (1996)-type instrument relies on the exclusion restriction that chain-level pricing strategies are assumed not to be correlated anymore with local demand shocks after conditioning on the seasonal pattern of promotions and discounts.<sup>35</sup>

**Baseline results** We estimate the elasticities using data at the product variety-retail chain-NUTS2-week level. We restrict the sample by placing restrictions on the frequency by which product varieties appear with positive sales. This is because there is widespread evidence of the existence of many zeros in scanner data which might potentially downwardly bias the elasticity estimates (e.g. Dubé et al. (2021); Gandhi et al. (2022)). Given our broad focus on many product categories, it is hard to obtain exogenous variation in choice set determination for each product category as in Dubé et al. (2021). Instead, we choose to only include product varieties that are frequently purchased and thus suffer less from zero market shares. Below, we discuss the sensitivity of the estimates to alternative sample restrictions. Figure ?? and Table ?? present the baseline OLS and IV estimates and report clustered standard errors at the product variety level. As expected, all OLS estimates have a negative sign and are precisely estimated but also represent quite inelastic residual demand curves. Their distribution is centered around  $-1.02$  and has a 10%-90% spread of  $[-1.96, -0.22]$ . As the OLS estimates may still suffer from the simultaneous determination of prices and quantities, we instrument prices with the Hausman-type instrument. Figure ?? shows that the IV-estimates are quite precisely estimated.<sup>36</sup> In addition, they are almost always more negative compared to the OLS estimates. The IV estimates have a

<sup>34</sup>Dellavigna & Gentzkow (2019) point to managerial inertia, agency costs and threats to brand image as possible explanations for why retail chains adopt (potentially suboptimal) uniform pricing strategy.

<sup>35</sup>In the estimation, we rely on the following moment condition  $\mathbb{E}_t[\varepsilon_{icl,t}|\bar{p}_{ic-l,t}, \boldsymbol{\theta}, \boldsymbol{\lambda}] = 0$  and minimize the following GMM-objective function to obtain:

$$\widehat{\sigma}_p = \underset{\sigma_p}{\operatorname{argmin}} \mathbf{M}(\sigma_p)' \mathbf{W} \mathbf{M}(\sigma_p) \quad \forall p \in \mathcal{P}$$

where

$$M_{icl}(\sigma_p) = \mathbb{E}_t[\bar{p}_{ic-l,t} \varepsilon_{icl,t}(\sigma_p)], \quad \bar{p}_{ic-l,t} \equiv \frac{1}{N_{lc}} \sum_{k \in \mathcal{L}_c \setminus l} p_{ick,t}$$

and  $\mathbf{W}$  is a weighting matrix that weights the product variety-region moment conditions using the number of transactions associated with that product variety in that region. For this reason, our estimator is very similar to the one developed in Dellavigna & Gentzkow (2019) but different from Faber & Fally (2021) which estimates brand-level elasticities in the US using only regional variation and no variation across retail chains and different from Atkin et al. (2018) which use it to estimate store-level elasticities in Mexico by collapsing the product variety dimension.

<sup>36</sup>The precision of the IV-estimates is due to the generally high first-stage F-statistics. The Kleibergen-Paap statistic has an unreported 10%-90% range of [12.35, 1098.44] across product categories.

10%-90% spread of  $[-4.77, -1.15]$  with a median estimate of  $-2.77$ . While this product variety level estimates are more inelastic than the ones reported in [Hottman et al. \(2016\)](#), the estimates are quantitatively in line with the estimates reported in different strands of literature. For instance, [Dellavigna & Gentzkow \(2019\)](#), [Faber & Fally \(2021\)](#) and [Döppler et al. \(2022\)](#) (for comparable US scanner data) and [Fajgelbaum et al. \(2020\)](#) (for US trade data at the product variety level) estimate very similar median estimates.<sup>37</sup> In addition, we reject the null hypothesis that the elasticities are equal to the theoretical constraint of  $-1$  for all but two product categories.<sup>38</sup> Nevertheless, when estimating cost-of-living differences we experiment with more elastic elasticities of substitution that are closer to the ones from [Hottman et al. \(2016\)](#) and show that qualitative conclusions are robust.

**Robustness** We consider three different robustness checks. First, the baseline results are obtained by placing restrictions on how frequently positive sales are observed across weeks in a given year. Table ?? shows that the IV estimates are less elastic when we do not place any restrictions on the sample. In this case, the distribution of IV estimates has a 10%-90% range of  $[-3.45, 0.52]$ . By placing ever stricter restrictions on the sample in terms of the frequency of positive sales and on the actual importance of different product varieties, the elasticity estimates become more elastic. When we restrict the frequency at 26 weeks and the product variety-level market share at 0.1%, the distribution of elasticities is almost identical. Hence, we view the restriction on the frequency of more than or equal to 50% as striking a good balance. Second, the baseline specification uses data at the weekly frequency. Figure ?? and Table ?? shows the results when we estimate the elasticities using monthly data. The monthly IV estimates are almost always precisely estimated but they are also generally more inelastic. In addition, Table ?? indicates that there are slightly more product categories for which the theoretical constraint is not satisfied. For this reason, we stick with the weekly estimates as input for the subsequent analyses. Finally, the theoretical framework does not have a retail chain dimension and therefore there is some leeway as to how we deal with regional time-varying demand shocks. Table ?? provides an overview of the results when we experiment with alternative fixed effects. When we replace the firm-chain-category-region-time fixed effects with firm-chain-category-country-time fixed effects, we recover a more elastic median demand elasticity of  $-3.89$ , but also with a much wider range of  $[-8.89; 4.20]$ .<sup>39</sup> In addition, the number of product categories where the theoretical restriction is not satisfied rises from 2 to 10. When we only include firm-category-region-time fixed effects instead of the firm-chain-category-region-time fixed effects, the 10%-90% range is  $[-5.01, -1.15]$  centered around  $-3.12$ . This specification also leads to more estimates that violate the theoretical constraint. The robustness exercise of cost-of-living differences to alternative elasticities includes the implied elasticities with country-level fixed effects and the results

<sup>37</sup>These papers report median elasticities of  $-2.6$ ,  $-2.2$ , slightly below  $-2$  and  $-2.53$  respectively

<sup>38</sup>Unfortunately, we are unable to estimate elasticities of substitution for the Skincare - Makeup and Infant food categories because they have too few observations, conditional on the fixed effects. However, failing to obtain IV-estimates is common (see e.g. [Hottman et al. \(2016\)](#); [Jaravel \(2019\)](#)) and we explain in section 6 how we deal with this to obtain the structural components.

<sup>39</sup>One potential explanation for the more elastic demand estimates when including time-varying fixed effects at the country level is that if promotional activities come with temporary price reductions and if there is regional variation in promotional activity, only including time-varying fixed effects at the country level will overestimate the elasticity of demand. This is because in this case the increase in consumption and the fall in the price are both driven by the same promotion. As  $\sigma_{c,prom} > 0$  and  $\sigma_{p,prom} < 0$  the elasticity of substitution would be downward biased.

presented below are also robust to using alternative fixed effects in the estimation of the elasticities.

621

### 5.1.2 Firm-level elasticities - $\eta_p$

622

**Estimation strategy** To estimate the firm-level elasticities of substitution, consider the residual demand curve for firm  $f$  in product category  $p$  in region  $l$  at time  $t$ :

623

624

$$C_{fp,l,t} = \varphi_{fp,l,t}^{\eta_p-1} \left( \frac{P_{fp,l,t}}{P_{p,l,t}} \right)^{-\eta_p} C_{p,l,t}$$

625

where  $C_{fp,l,t}$  and  $P_{fp,l,t}$  are the firm-level quantity and price level,  $P_{p,l,t}$  and  $C_{p,l,t}$  are the aggregate price and consumption index and  $\eta_p$  governs how consumers substitute across firms in product category  $p$ .  $\varphi_{fp,l,t}$  is the firm-level taste parameter that shifts demand across all product varieties in product category  $p$  offered by firm  $f$ . At any level of aggregation, the utility functions are homogeneous of degree 1 in the taste parameters. Therefore, without an additional normalization, it is impossible to distinguish between the firm-level tastes  $\varphi_{fp,l,t}$  and the aggregated product-variety level taste parameters within each product category-firm-region-time combination. For this reason, we follow [Hottman et al. \(2016\)](#) and make two useful normalizations regarding the taste parameters:

626

627

628

629

630

631

632

633

$$\tilde{\varphi}_{fp,l,t} \equiv \left( \prod_{i \in \mathcal{B}_{fp,l,t}} \varphi_{il,t} \right)^{\frac{1}{N_{fp,l,t}}} = \left( \prod_{f \in \mathcal{F}_{fp,l,t+1}} \varphi_{fp,l,t+1} \right)^{\frac{1}{N_{fp,l,t+1}}} \equiv \tilde{\varphi}_{p,l,t+1}$$

634

where  $N_{fp,l,t}$  is the number of product varieties that firm  $f$  sells in product category  $p$  in region  $l$  at time  $t$  and  $N_{p,l,t}$  is the associated number of firms. We normalize the product variety level taste parameters  $\varphi_{il,t}$  such that the geometric mean of the product variety level taste parameters within each product category-firm-region-time does not change over time.<sup>40</sup> In this way, we ensure that differences in firm-level consumption levels conditional on relative prices and aggregate consumption are captured by  $\varphi_{fp,l,t}$  and any shifts in product variety-level consumption levels conditional on relative prices and firm-level consumption levels is captured by the  $\varphi_{il,t}$ 's. Consider the log transformation of the firm-level residual demand curve:

635

636

637

638

639

640

641

$$c_{fp,l,t} = -\eta_p p_{fp,l,t} + \eta_p p_{p,l,t} + c_{p,l,t} + (\eta_p - 1) \ln(\varphi_{fp,l,t})$$

642

and its empirical counterpart are given by:

643

$$c_{fp,l,t} = -\eta_p p_{fp,l,t} + \theta_{pfl} + \lambda_{p,l,t} + \varepsilon_{pfl,t} \quad (6)$$

644

where  $\varepsilon_{pfl,t}$  subsumes  $\varphi_{pfl,t}$ . We take the following steps to estimate equation 6. First, we obtain the

645

<sup>40</sup>Note that this normalization is without loss of generality as any geometric aggregator would be suitable (see [Redding & Weinstein \(2020\)](#)). In addition, this normalization is different from the normalization we performed in the derivation of the cost-of-living differences across regions in two ways. First, the previous normalization was about ruling out that pure changes in tastes at the firm or product variety level across regions can induce changes in cost-of-living differences. In this case, the normalization is about how the product variety- en firm-level shifters are related within each region and is based on the fact that the utility aggregators are homogeneous of degree 1 in the taste parameters. Second, this normalization is about the full set of product varieties offered in region  $l$ . The previous normalization only covers the set of product varieties that is common across the two regions.

firm-level consumption and price indices by aggregating product variety-level quantity and price using the baseline product-variety level elasticities of substitution and the backed-out product variety-level taste parameters.<sup>41</sup> We include two sets of fixed effects to deal with omitted variables and use an instrument to construct moment conditions. First, including the product category-region-time fixed effects  $\lambda_{pl,t}$  captures variation stemming from the unobserved product category-level price and consumption indices, given by  $p_{pl,t}$  and  $c_{pl,t}$  respectively. In addition, these fixed effects pick up time-varying regional demand shocks that affect all firms similarly in product category  $p$  in region  $l$ . Second,  $\theta_{pfl}$  are category-firm-region fixed effects and pick up persistent differences in firm-level tastes across regions, such as the unweighted average product variety-level taste parameters  $\varphi_{pfl,t}$ . As the local demand shocks might still drive both consumption and prices, we rely on an instrument that follows from the structure of the demand system and the normalization imposed on the average product variety-level demand shifters. In particular, the firm-level price index can be written as a product of two intuitive terms (see Appendix ?? and Hottman et al. (2016)):

$$P_{fpl,t} = \tilde{P}_{fpl,t} \left( \sum_{i \in \mathcal{B}_{fpl,t}} \frac{S_{il,t}}{\tilde{S}_{fpl,t}} \tilde{\varphi}_{fpl,t}^{\sigma_p-1} \right)^{\frac{1}{1-\sigma_p}} \quad (658)$$

The first part is the unweighted geometric average across product variety-level prices offered by firm  $f$  in product category  $p$ , region  $l$  at time  $t$ . When firms simultaneously set the average firm-level price and taste parameter, this first part is correlated with the firm-level taste shifter  $\varphi_{pfl,t}$ . The second part of this expression depends on the dispersion in product variety level market shares within each product category-firm-region-time cell. Intuitively, more dispersion in market shares, reflecting greater dispersion in taste-adjusted prices, leads to a fall in the geometric average and thus an increase in the sum of market shares relative to their geometric average. However, as this sum is raised to a negative power  $(1 - \sigma_p)$  greater dispersion within product category-firm-region-time cells leads to a fall in the firm-level price index. As the relative within product category-firm-region-time market shares do not depend on the firm-level taste shifter and the unweighted geometric average across product varieties is partialled out with the inclusion of  $\theta_{pfl}$ , the second part is uncorrelated with the firm-level taste parameter and therefore a suitable instrument.<sup>42</sup>

<sup>41</sup>We back out the product variety level demand shifters using:

$$\varphi_{il,t} = \frac{P_{il,t}}{\tilde{P}_{fpl,t}} \left( \frac{S_{il,t}}{\tilde{S}_{fpl,t}} \right)^{\frac{1}{\sigma_p-1}} \tilde{\varphi}_{fpl,t}$$

where  $\tilde{P}_{fpl,t} \equiv \left( \prod_{i \in \mathcal{B}_{fpl,t}} P_{il,t} \right)^{\frac{1}{N_{fpl,t}}}$  and  $\tilde{S}_{fpl,t} \equiv \left( \prod_{i \in \mathcal{B}_{fpl,t}} S_{il,t} \right)^{\frac{1}{N_{fpl,t}}}$ .

<sup>42</sup>This instrument relies on the presence of multi-product firms and imperfect substitutability across product varieties which. When only one product is supplied, the dispersion in market share is zero. If product varieties are perfect substitutes ( $\sigma_p \rightarrow \infty$ ), market shares are disconnected from taste-adjusted prices leading to no dispersion in market shares. In the estimation, we rely on the following moment condition  $\mathbb{E}_t [\varepsilon_{fpl,t} | p_{fpl,t}^D, \theta, \lambda] = 0$  and minimize the following GMM-objective function to obtain:

$$\hat{\eta}_p = \underset{\eta_p}{\operatorname{argmin}} \mathbf{M}(\eta_p)' \mathbf{W} \mathbf{M}(\eta_p) \quad \forall p \in \mathcal{P}$$

where

$$M_{pfl}(\eta_p) = \mathbb{E}_t \left[ p_{fpl,t}^D \varepsilon_{fpl,t}(\eta) \right], \quad p_{fpl,t}^D \equiv \frac{1}{1 - \hat{\sigma}_p} \ln \left( \sum_{i \in \mathcal{B}_{fpl,t}} \frac{S_{il,t}}{\tilde{S}_{fpl,t}} \tilde{\varphi}_{fpl,t}^{\hat{\sigma}_p-1} \right)$$



**Baseline results** We estimate the elasticities using data at the firm-NUTS2-week level and restrict the baseline sample to include all purchases on product varieties that register positive sales in more than 50% of the time within a year. We present the baseline in Figure ?? and report an overview of all specifications in Table ?? with clustered standard errors at the firm level. The OLS estimates are all negative, precisely estimated and almost always fall within the  $[-2, -1]$  range. Given that firm-level prices conditional on the fixed effects might still be correlated with the structural residual, we turn to the IV estimates. First, the instruments are strong as the F-statistics of the first-stage regressions are almost always larger than the conventional rejection levels for weak instruments.<sup>43</sup> Second, the IV estimates imply more elastic residual demand curves as they are centered around  $-3.10$  and have a 10%-90% range of  $[-4.84, -1.71]$ .<sup>44</sup> This range is still lower, but comparatively closer to the estimates in Hottman et al. (2016) which estimate firm-level elasticities between  $[-7.3, -2.6]$  centered around  $-3.9$ . The estimation routine successfully ends for all product categories and we reject the null hypothesis that the elasticities are different from the theoretical constraint of  $-1$  for all the product categories.

**Robustness** We consider four robustness checks. First, Table ?? shows that the elasticities are very similar across different sample restrictions. The reason for this is that the data becomes much less sparse when collapsing both retail chain and product variety dimensions such that imposing the same sample restrictions does not result in markedly different samples. At this level of aggregation, it does not matter whether we consider the full sample or the frequent sample. Second, similar to the product-variety estimates, the distribution of monthly firm-level elasticities is shifted upwards when we collapse the data at the monthly level. While the elasticities are still precisely estimated, Table ?? shows that the distribution of monthly estimates is centered around  $-1.66$  with a range  $[-3.20, -1.32]$ . Third, the baseline estimation includes product category-firm-region fixed effects and thus only controls for persistent differences across firms within regions. However, if retail chains and firms coordinate on seasonal price changes and promotion, a better identification strategy could be to use time variation conditional on seasonal shocks. For this reason, we re-estimate equation 6 by replacing the  $\theta_{pfl}$  fixed effects with product category-firm-region-year fixed effects,  $\theta_{pfl,y(t)}$ , and product category-firm-region fixed effects-week-of-the-year  $\theta_{pfl,w(t)}$ . This set of fixed effects also flexibly controls for seasonal demand shocks that could drive both firm-level demand and prices. Nevertheless, Table ?? shows that the distribution of elasticities is largely unaffected.

**Cross-country Heterogeneity** The assumption that elasticities of substitution are common across countries is necessary to exactly decompose cost-of-living differences across regions. To check the validity of this assumption for the firm-level elasticities of substitution, we estimate country-level elasticities of substitution by interacting the price variable in equation 6 with country dummies. Tables ?? - ?? present the elasticity of substitution estimates at the country level and Table ?? provides a comparison of the distribution of the IV-estimates across the different countries. Tables ?? - ?? confirm that the elasticities are still precisely

and  $W$  is a weighting matrix that weights the product variety-region moment conditions using the number of transactions associated with that firm in that region.

<sup>43</sup>More precisely, the first-stage Kleibergen-Paap F-statistics have a 10%-90% range of  $[15.60; 5,830.67]$  while the smallest F-statistics is 7.24.

<sup>44</sup>The range of firm-level elasticities is very close to the range of product-variety level elasticities. Given that we do not impose any restriction on their relative magnitude, this does not pose a problem for our analysis per se.

estimated and satisfy the theoretical constraint in almost all cases. Even though there is some variation across countries for a couple of product categories, Table ?? shows that these differences do not imply large systematic differences in the country-level distributions. Therefore, the assumption of equal elasticities does not seem to be very restrictive.

**Implied markups** An additional check on the estimated elasticities of substitution is to see whether they imply firm-level markups when combined with an assumption on how firms set prices. If we assume that firms set prices in a simultaneous move Bertrand game, firm-level markups depend on the firm-level elasticity of substitution and on the firm-level market share in location  $l$  at time  $t$ .<sup>45</sup> Figure 3 shows the full distribution of recovered firm-level markups across product category-firm-country-year observations. We recover a median firm-level markup of 1.5 which implies that the median firm charges a 50% price premium over its marginal costs.

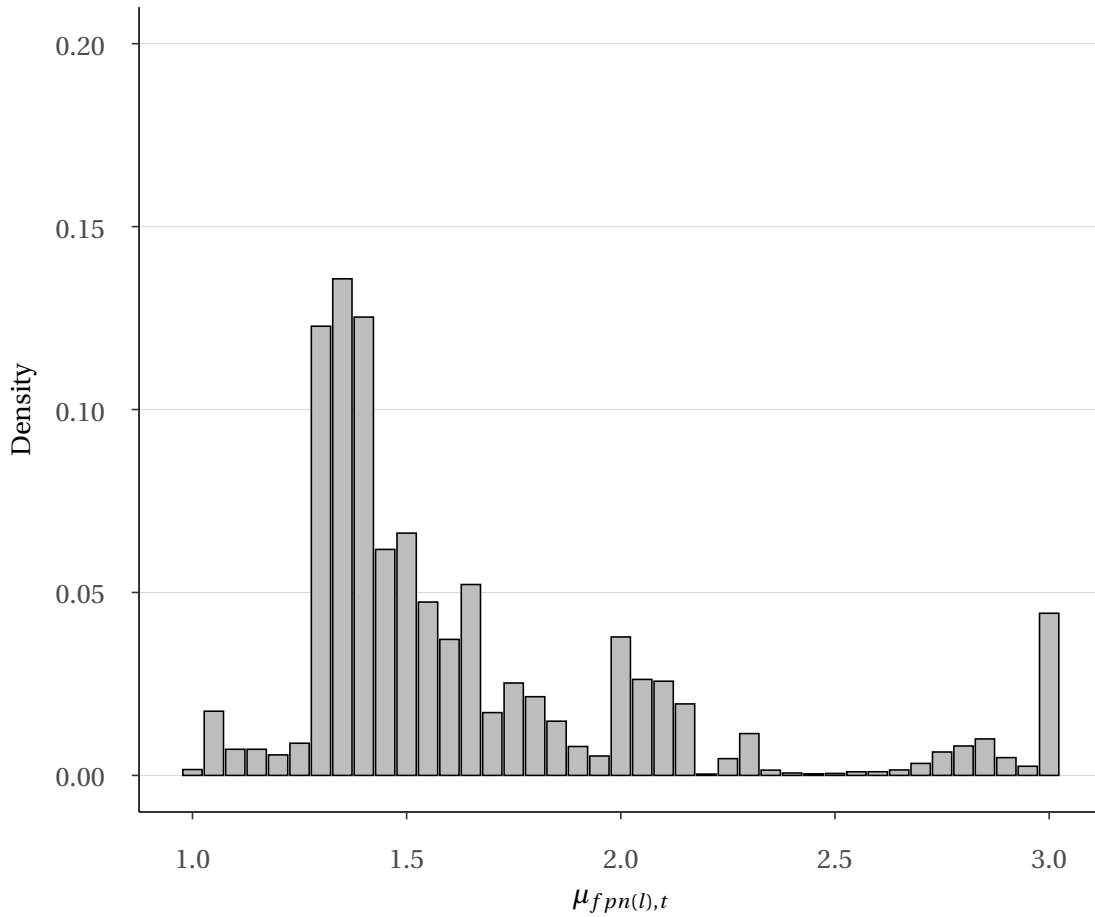
How sensible are these markup estimates? Because there is little empirical evidence on the level of firm-level markups for a comparable set of product categories in Europe, we benchmark our estimates to the broader literature on markup estimation. There are two broad strands in this literature. First, the demand approach estimates markups by specifying a model of demand and market structure. Markups are computed based on the estimated demand parameters and the optimal markup function as implied by the assumed market structure. Our approach falls in this strand. Other papers that take the demand approach to estimate markups for a broad set of product categories are Hottman et al. (2016) and Döpper et al. (2022). While Hottman et al. (2016) finds a median markup of 1.31, Döpper et al. (2022) reports a median markup of 2.08.<sup>46</sup> Second, Loecker & Warzynski (2012) pioneered the production approach in which the combination of cost minimization and an estimated output elasticity on a completely variable input in production yields markup estimates. Loecker et al. (2016) and Loecker et al. (2020) report a median elasticity of 1.6 for Indian manufacturing firms and an average markup of 1.6 for public US companies. Given that our estimates are close to estimates from both of these strands in the literature, we consider our markup estimates as sensible.

<sup>45</sup>In particular, under Bertrand competition in prices, markups are given by:

$$P_{il,t} = \frac{\varepsilon_{fnp,t}}{\varepsilon_{fnp,t} - 1} MC_{il,t}, \quad \text{where} \quad \varepsilon_{fnp,t} \equiv \eta_p - (\eta_p - 1) S_{fnp,t} \quad (7)$$

<sup>46</sup>These papers report different measures of the markup. Hottman et al. (2016) report a median  $\frac{P-MC}{MC}$  of 0.31, which results in a median  $\frac{P}{MC}$  of 1.31. Döpper et al. (2022) report a median Lerner index  $\frac{P-MC}{P}$  of 0.48 which results in a median  $\frac{P}{MC}$  of 2.08.

**Figure 3: Firm-level markup distribution**



*Notes:* This figure plots the distribution of firm-level markups. To account for the sampling variation in the elasticities of substitution, we bootstrap the markup distribution. In practice, we draw from the limiting distribution of the firm-level elasticities of substitution and for each bootstrap sample, we compute firm-level markups at the product category-firm-country-year level. Hereafter, we bin the absolute markup estimates into 40 separate bins and compute for each bin the number of observations that fall into each bin. Finally, we winsorize the markup distribution at a markup of 3.

## 5.2 Regional Cost-of-living Differences

We construct regional cost-of-living differences by computing each product category-region pair-year cell the log cost-of-living differences according to Equation 4. To obtain a measure of variation in regional cost-of-living differences, we take the variance of log cost-of-living differences across product categories for each region pair-year. Table 3 shows the mean and 95% bootstrapped interval of the estimated variance of cost-of-living differences for intranational and international region pairs separately.<sup>47</sup> Table 3 also provides the mean and

<sup>47</sup> Bootstrapped confidence intervals are obtained by drawing blocks of households within regions with replacement and computing the cost-of-living differences for each of 100 bootstrap samples. To account for estimation in the elasticities of substitution, sampling and design-based uncertainty, we construct block-bootstrapped confidence intervals. In practice, we first draw region pairs with replacement and within each region pair, we draw a sample of consumers with replacement as well. Then, we draw firm- and product variety-level elasticities of substitution from their respective empirical distributions. Finally, for each bootstrap sample, we compute

95% bootstrapped intervals of the three structural components presented in equation 4. To construct each of these components, we rely on Equation 3 and note that log cost-of-living differences can be written as the sum of three structural components. Therefore, the variance can be written as the sum of the variances of three terms and the covariances between them. As there is ex-ante no reason to favor any of the three terms, we then decompose the overall variance into these three terms by allocating the covariance terms equally.<sup>48</sup>.

First, consistent with the reduced-form evidence, cost-of-living differences for international pairs are larger compared to intranational pairs. On average, cost-of-living differences are eight times larger for intranational region pairs. While cost-of-living differences for international region pairs have a more dispersed distribution compared to international region pairs, it is mostly a shift in the conditional mean that drives the elevated cost-of-living differences. Nevertheless, even within countries, cost-of-living differences exist which is consistent with the results reported in Handbury & Weinstein (2015) and in Feenstra et al. (2020) for the USA and China respectively. Second, decomposing cost-of-living differences into the three structural components provides more insight into the relevant margins for cost-of-living differences. For instance, pure taste differences have large explanatory power for regional cost-of-living differences. On average, they account for roughly 95% and 64% of the variation in cost-of-living differences for intranational and international region pairs respectively. This underscores the importance of filtering out taste differences when assessing the presence of geographic market segmentation. LOP deviations and choice set differences are relatively small for intranational region pairs but choice set differences matter much more for international pairs. A little over 40% of the variation in regional cost-of-living differences across intranational region pairs is driven by choice set differences.

## 6 Geographic Market Segmentation in Europe

There are two reasons why a simple comparison of conditional means in cost-of-living differences between international and intranational region pairs will lead to an overestimation of the level of geographic market segmentation. First, Table 3 illustrated that the conditional distribution of pure taste differences for international pairs is shifted upwards relative to the distribution for intranational pairs. As pure taste differences are traditionally considered to be outside of market integration policies, we will filter out pure taste differences when assessing the presence of geographic market segmentation. Second, geographic differences between international pairs are larger compared to intranational pairs. As geographic differences are likely positively correlated with the cost of physically moving goods to their final destination markets, simply comparing intranational and international region pairs will lead to an overestimation of cost-of-living differences due to geographic market segmentation. For this reason, we first elaborate on an empirical design that controls for these geographic differences across region pairs.

the overall cost-of-living differences, the different margins and their contribution.

<sup>48</sup>Hottman et al. (2016) decomposes differences in sales or sales growth across firms into structural components by successive OLS estimations of the component on total sales. By doing so, the variance decomposition distributes the covariance terms equally across the different components.

**Table 3:** Variance decomposition of Cost-of-living differences

$y_{ll',t}$	CLE $_{ll',t}$	LOP $_{ll',t}$	Taste $_{ll',t}$	Choice $_{ll',t}$
	(1)	(2)	(3)	(4)
PANEL A: Intranational pairs ( $B_{ll'} = 0$ )				
$\hat{E}[\hat{V}_{ll',t}(y_{pl'l,t})]$	.1458	.0003	.1345	.011
5%-95% Bootstrapped CI	[.1096, .186]	[.0002, .0003]	[.1032, .1699]	[.0061, .017]
$\hat{E}[\hat{V}_{ll',t}(y_{pl'l,t})] / \hat{E}[\hat{V}_{ll',t}(\ln(\frac{P_{pl',t}}{P_{pl,t}}))]$	1	.0019	.9244	.0737
5%-95% Bootstrapped CI	-	[.0014, .0027]	[.9016, .9416]	[.056, .0968]
PANEL B: International pairs ( $B_{ll'} = 1$ )				
$\hat{E}[\hat{V}_{ll',t}(y_{pl'l,t})]$	1.1718	.0166	.6769	.4782
5%-95% Bootstrapped CI	[.826, 1.6442]	[.0149, .0193]	[.5246, .838]	[.3015, .7979]
$\hat{E}[\hat{V}_{ll',t}(y_{pl'l,t})] / \hat{E}[\hat{V}_{ll',t}(\ln(\frac{P_{pl',t}}{P_{pl,t}}))]$	1	.0147	.5841	.4012
5%-95% Bootstrapped CI	-	[.0104, .0198]	[.4953, .6406]	[.3437, .4951]

*Notes:* This table presents the overall variance and decomposition of log cost-of-living differences for intranational region pairs and international region pairs separately. Column (1) shows the average log cost-of-living difference for intranational region pairs in Panel A and for international region pairs in panel B. Below the estimate of the average log cost-of-living difference we display the 5%-95% confidence interval. We compute these statistics in four steps. First, within each bootstrap sample, we compute log cost-of-living differences following the expression displayed in Equation 3. Second, within each bootstrap sample, we compute the variance of log cost-of-living differences across product categories within each region pair-year cell. Third, within each bootstrap sample, we compute the average cost-of-living difference by averaging across region pair-year cells. Finally, the  $\hat{E}[\hat{V}_{ll',t}(y_{pl'l,t})]$  represents the average across bootstrap samples and the 5%-95% confidence intervals are 5<sup>th</sup> and 95<sup>th</sup> percentiles of the distribution across bootstrap samples. Columns (3) - columns (8) present the decomposition of the variance of log cost-of-living differences into the different margins. To compute the variance of each of the margins, we run the variance operator through the log-linear expression for log cost-of-living differences and allocate the covariance terms equally across each of the margins. In turn, we compute the average and the 5%-95% confidence interval for each of the components in the same way. For each margin, we also provide an estimate and a confidence interval of the share that each margin explains in the overall variance of log cost-of-living differences. For each bootstrap sample and for each margin, we compute this share as the ratio of the variance of the margin, averaged across region pair-year cells, to the variance of overall cost-of-living differences, averaged across region pair-year cells as well. In this way,  $\hat{E}[\hat{V}_{ll',t}(y_{pl'l,t})] / \hat{E}[\hat{V}_{ll',t}(\ln(\frac{P_{pl',t}}{P_{pl,t}}))]$  presents the average of this ratio across bootstrap samples and the 5%-95% confidence intervals are 5<sup>th</sup> and 95<sup>th</sup> percentiles of the distribution of this share across bootstrap samples.

## 6.1 Empirical design

To condition on geographic differences across regions, we introduce a potential outcomes framework that draws heavily on the ideas developed in [Santamaria et al. \(2021\)](#).

**Potential outcomes framework** We consider the following potential outcomes framework

$$\mathbb{V}_{ll'} = \begin{cases} \mathbb{V}_{ll'}(0) & \text{if } B_{ll'} = 0 \\ \mathbb{V}_{ll'}(1) & \text{if } B_{ll'} = 1 \end{cases}$$

where  $\mathbb{V}_{ll'}$  is the variance of cost-of-living differences between region  $l$  and region  $l'$ . Define  $\mathbb{V}_{ll'}(0)$  as the value of the variance that would have materialized if region  $l$  and region  $l'$  were intranational region pairs and  $\mathbb{V}_{ll'}(1)$  as the variance if they were an international region pair. We define the effect of being separated by a national border as:

$$\tau \equiv \mathbb{E}[\mathbb{V}(1) - \mathbb{V}(0) | B_{ll'} = 1] \quad (8)$$

By defining the effect in this way, the border effects measure the increase in the variance of cost-of-living differences for international pairs that are currently being by a national border compared to when they would only be separated by a regional border.<sup>4950</sup> Consistent estimation of the border effect requires individualistic, probabilistic, unconfounded assignment and compliance with the assignment. We discuss the plausibility of these assumptions now.

First, consistent estimation requires the SUTVA assumption that requires that (1) separating a region pair by a national border does not affect the potential outcomes of other region pairs and (2) that separating region pairs by a national border cannot be executed in multiple ways. The first part of this assumption requires that the national border assignment is individualistic and thus rules out that separating one region pair affects the potential outcomes of other region pairs. In general, spatial spill-overs might occur if changes in national border assignment changed the size of the relevant market. However, the size of these spill-overs might be small given the disaggregated spatial level at which we define the treatment variable. There are 3,403 region pairs we consider. If we were to allocate a Belgian region to the Netherlands, there would be 9 additional borders with Belgium and 12 fewer borders with the Netherlands. This amounts to a 0.6% change in the number of treated units. While this number is not zero, it is small and it seems reasonable to assume that the change in the aggregate economic determinants of the market segmentation would be small. The second part requires that separating regional pairs by a national border does not vary across region pairs. Given the clear separation of national and supranational regulatory, legal and tax responsibilities between member states and

<sup>49</sup>In terms of the causal inference vocabulary, our estimand corresponds to an average treatment effect on the treated. We believe that considering how cost-of-living differences would have looked if there were no national border seems like a more sensible approach than an average treatment effect or an average treatment effect on the untreated. This is because in this case, we would have to impose national border assignments on certain intranational region pairs and have to take a stance on what it would do to other adjacent intranational pairs. Instead, taking the current national border assignment as given and estimating the impact of those existing national borders seems like a more reasonable thought experiment.

<sup>50</sup>In addition, the bias due to treatment effect heterogeneity as mentioned by [Gorodnichenko & Tesar \(2009\)](#) will not affect our estimates. This is because by defining the estimand as an average treatment effect on the treated (ATT) and not as an average treatment effect (ATE) differences in the treatment effect on the treated and on the untreated (ATU) do not enter the estimand anymore.

the European institutions and the lack of additional bilateral agreements across member states, it seems plausible to assume that what matters up to a first-order is the separation by a national border and not the specific bilateral border.<sup>51</sup>

Second, border assignment needs to be probabilistic. In other words, every region pair needs to have a probability of being separated by a border strictly different from zero and one. This assumption is likely to hold in our setting given that both contiguous regions and very spatially separated region pairs are separated by region and national borders in the data.

Third, to rule out geographic differences as a source of cost-of-living differences, we require that national border assignment was not chosen with the potential outcomes or the potential degree of market segmentation in mind. The results from section 3 show that conditional on a national border distance affects price dispersion and choice set differences. Hence, geographic differences are correlated with the degree of market segmentation. If national border assignment is also determined by geographic differences, a simple comparison of international and intranational regions, along the lines of Engel & Rogers (1996), will not yield a reliable estimate (see Gorodnichenko & Tesar (2009)). To estimate a border effect that is not confounded by geographic characteristics, we follow the approach taken in Santamaria et al. (2021) and estimate the border effect by conditioning on a set of geographic covariates that are correlated with national border assignment:

$$\{\mathbb{V}_{ll'}(0), \mathbb{V}_{ll'}(1)\} \perp\!\!\!\perp B_{ll'} | \mathbf{X}_{ll'}$$

where  $\mathbf{X}_{ll'}$  includes great-circle distance, remoteness, shared river basins and altitude differences between region  $l$  and  $l^{w'}$ . This approach has merit for two reasons. First, geography can be considered as a pre-treatment variable that determines border assignment. It is well-known that mountainous areas and rivers have shielded nations from invasions (e.g. Nunn & Puga (2012)) and more distant and larger populations are more difficult to govern Alesina & Spolaore (1997). Second, even if there are many potential explanations for border effects and we are unable to isolate one particularly important source, this approach eliminates geographic frictions, such as physical distance, as a potential explanation for the estimated border effects.

Finally, in the presence of LOP deviations or choice set differences, consumers may engage in cross-border shopping. If so, this would result in non-compliance with national border assignments. To understand the importance of such non-compliance, we use the Belgian data for which the variable containing the store name indicates whether the store is located in Belgium or in one of the neighboring countries.<sup>52</sup> As previous studies have shown that Belgium appears to have higher consumer prices for the products we study (e.g. Beck et al. (2020)) and Belgium is well-connected to its neighboring countries, cross-border shopping would manifest itself, especially in Belgium. While there is some cross-border shopping, Table ?? shows that over 97% of expenditure by Belgian households is made in stores located in Belgium. Since this table pools across all Belgian households, Figure ?? focuses on the importance of cross-border shopping close to the

<sup>51</sup>While the border assignment is assumed to be equal across region pairs, this assumption does not rule out the possibility that border effects are heterogeneous across region pairs.

<sup>52</sup>To be precise, the variable indicates whether the store is a Belgian, French, German, Dutch store or whether the store is a foreign one.



Belgian-French and the Belgian-Dutch border.<sup>53</sup> We plot the propensity that households have engaged at least once in cross-border shopping and the overall expenditure share of cross-border shopping as a function of the distance to the border. While the propensity that households engage at least once in cross-border shopping over the sample period tends to increase when we approach a border, the overall expenditure share on cross-border transactions in very close proximity to the border remains low at a little over 5% and 10% for the French and Dutch borders respectively. We conclude that non-compliance with national borders is not a major issue in our data.

**Propensity score and subclassification** Table ?? shows the conditional distributions of the geographic variables across intranational and international pairs. The t-tests and overlap tests highlight that there exist important imbalances between these conditional distributions. To ensure that geographic differences are not responsible for the estimated border effects, we implement a two-step approach that ensures that we only use variation across region pairs equally likely to be separated by a national border. In the first step, we estimate the probability that two regions are separated by a national border based on geographic determinants. In the second step, we trim the sample and construct groups of intranational and international pairs with very similar estimated probabilities of being separated by a national border. We estimate the propensity score that two regions are separated by a national border by assuming that the probability of being separated by a national border takes the following logistic form:

$$P(B_{ll'} = 1 | \mathbf{X}_{ll'}) = \frac{e^{\beta_0 + \mathbf{X}_{ll'} \boldsymbol{\beta}_1 + \mathbf{X}_{ll'} \cdot \mathbf{X}_{ll'} \boldsymbol{\beta}_2 + \varepsilon_{ll}}}{1 + e^{\beta_0 + \mathbf{X}_{ll'} \boldsymbol{\beta}_1 + \mathbf{X}_{ll'} \cdot \mathbf{X}_{ll'} \boldsymbol{\beta}_2 + \varepsilon_{ll}}} \quad (9)$$

where  $\mathbf{X}_{ll'} = \{\text{Distance, Remoteness, Altitude difference, Shared River basin}\}$  is the set of geographic determinants. We consider a second-order polynomial in the geographic variables as the set of possible explanatory variables. We follow Imbens & Rubin (2015) by iteratively selecting from this set the variables that improve the explanatory power of the model the most and stop when we cannot improve the predictive power of the model meaningfully.<sup>54</sup> Table ?? presents the outcome of this iterative selection procedure and shows how we arrive at the final set of covariates. Columns (1) and (2) of Table ?? present the estimation results after estimating 9 with maximum likelihood. Adding the selected second order-terms to the restricted first-order model leads to an increase in the pseudo- $R^2$  from 0.2 to 0.28.

Figure ?? presents the conditional distribution of estimated propensity scores across international and intranational regions. While there is considerable overlap in the conditional distributions, there is much less overlap in the tails. This is problematic for two reasons. First, sampling variance increases when we were to compare a small number of intranational pairs with a large set of international pairs as it puts a very large weight on a few control units. To limit the contribution of region pairs in the tail ends of the distribution, we implement the algorithm developed in Crump et al. (2009) and exclude observations with an estimated

<sup>53</sup>Cross-border shopping does not seem to be relevant for the Belgian-German border at all.

<sup>54</sup>In practice, we start from a first-order polynomial in the geographic variables and check which second-order term adds the most explanatory power to the model by comparing the restricted model with an augmented model that includes one second-order term. Hereafter, we add this variable to the restricted and search for the next variable that improves the explanatory power the most. We stop when we cannot further increase the explanatory power by a given threshold. This threshold is a likelihood ratio test score of 2.71 as recommended by Imbens & Rubin (2015).

propensity score outside of a critical interval of  $[\alpha, 1 - \alpha]$ .<sup>55</sup> The algorithm reaches its critical point at  $\alpha = 0.093$  such that we remove a little over 20% of the region pairs (see Table ??). Figure ?? and Figure ?? show the improved overlap in the conditional distributions of the estimated propensity scores once after trimming the sample and once after trimming the sample and re-estimating the propensity scores after trimming. Second, comparing intranational pairs with low probabilities of being separated by a national border and international pairs with a high probability of being separated requires substantial extrapolation. Even after trimming the sample appropriately, we still need to ensure that we only compare intranational and international pairs with roughly equal probabilities of being separated by a national border. To this end, we subdivide the intranational and international pairs into groups in which the normalized difference in the estimated propensity between intranational and supranational is below a threshold value. We implement the algorithm developed in Imbens & Rubin (2015) and subdivide the region pairs into 7 groups. Table ?? provides an overview of the number of international and shows that within each block the standardized differences in the estimated propensity score are small between international and intranational pairs. Given these groups, we estimate the border effect through the following estimator:

$$\hat{\tau} = \sum_{j=1}^{J=7} \frac{N_{\text{int}}(j)}{N_{\text{int}}(j) + N_{\text{nat}}(j)} \hat{\tau}(j), \quad \hat{\tau}(j) = \arg \min_{\tau(j)} \sum_{ll'} \mathbb{1}(ll' \in \mathcal{G}(j)) \cdot (\mathbb{V}_{ll'} - \beta_0 - \tau B_{ll'} - \mathbf{X}_{ll'} \boldsymbol{\beta}_1)^2 \quad (10)$$

where  $\mathcal{G}(j)$  indicates the set of region pairs included in group  $j$  and  $N_{\text{int}}$  and  $N_{\text{nat}}$  are the number of national and international region pairs in group  $j$ .<sup>56</sup> We include the geographic variables to improve precision and to remove the potential bias from any remaining unbalance in conditional distributions of the geographic determinants.

## 6.2 Cost-of-living Differences for Comparable Regions

Before discussing the results in which we detect the presence of geographic market segmentation, we provide an estimate of the overall cost-of-living differences by comparing intranational and international region pairs with similar geographic differences.

**Baseline results** Table 4 presents the border effect estimates of cost-of-living differences at the baseline elasticities of substitution estimates for different specifications. We report the average variance of log cost-of-living differences for intranational pairs in the row indicated by  $\mathbb{V}(|l, l'' \in \mathcal{L}_{\text{intra}})$  for each of the specifications. Column (1) shows the results when we estimate the border effect as we did in section 3. We regress the variance in cost-of-living differences across region pairs and years on a national border dummy and include time fixed effects. Consistent with Table 3, the variance of cost-of-living differences is on average over 8 times larger for international pairs relative to intranational pairs. When we control for geographic differences in column (2), cost-of-living differences are more than 7 times larger. However, these estimates

<sup>55</sup>The idea behind the algorithm is to search for the cut-off value that balances the increase in variance by reducing the sample size and the decrease in variance because of excluding observations in the tail of the distribution. Figure ?? shows the output of the grid search algorithm.

<sup>56</sup>This weighting scheme ensures that we obtain the Average Treatment Effect on the Treated.

potentially compare very geographically different intranational and international pairs and could overestimate the border effect. To this end, we first trim the sample by the critical value established in the previous and keep only region pairs with a propensity score in the range of [0.093,0.907]. In this case, the cost-of-living differences fall in magnitude for both international and intranational region pairs, but the difference in cost-of-living differences is still close to over 7 times larger for international pairs. Second, in column (4) we apply the subclassification estimator from equation 10 on the trimmed sample. In this way, we estimate the border effect by only comparing international and intranational pairs with very similar geographic differences. Compared to column (3), the difference in cost-of-living differences is essentially unchanged and the variance in cost-of-living differences remains 7 times larger for international pairs. Comparing the estimate in column (4) to the one in column (1), geographic differences account for around 12% of the estimated cost-of-living differences across the countries in our sample.

**Robustness** We check the robustness of the results regarding the elasticities of substitution used. While the baseline product variety level elasticities of substitution are in line with some recent papers, they are lower in absolute value than the estimates reported in [Hottman et al. \(2016\)](#). As the product variety level expenditure share and choice set differences fall with the elasticity of substitution, we check the robustness of our results when we consider higher values for the product-variety level elasticities. We re-estimate the specification used in column (4) by shifting the distribution of estimated product-variety level elasticities by 1 in column (5) and by 2 in column (6).<sup>57</sup> As expected the absolute size of the border effect falls when the elasticities of substitution used to compute the two product-variety level components rise. However, because the estimated cost-of-living differences for the set of intranational pairs also fall, the ratio of the two variances remains relatively stable at 6.8 and 6.7 in columns (5) and (6) respectively.

---

<sup>57</sup>In this way, we respect the absolute level of heterogeneity in the estimated elasticities, but shift the location of the distribution.

**Table 4:** Border effect: Cost-of-living

$\hat{\mathbb{V}}[p_{l',t} - p_{l,t}]$	(1)	(2)	(3)	(4)	(5)	(6)
$B_{ll'}$	1.027*** [.704, 1.487]	.9354*** [.642, 1.316]	.9198*** [.63, 1.279]	.9273*** [.648, 1.279]	.5582*** [.504, .62]	.5263*** [.472, .585]
$\lambda_t$	✓	✓	✓	✓	✓	✓
Controls		✓	✓	✓	✓	✓
Trimmed			✓	✓	✓	✓
Subclassification				✓	✓	✓
Variety-level Elasticity	$\hat{\sigma}_p$	$\hat{\sigma}_p$	$\hat{\sigma}_p$	$\hat{\sigma}_p$	$\hat{\sigma}_p + 1$	$\hat{\sigma}_p + 2$
$\hat{\mathbb{V}}(\cdot   l, l'' \in \mathcal{L}_{\text{intra}})$	.1456	.1456	.1415	.1415	.0959	.0917
Nr. obs	34,030	34,030	24,450	24,450	24,450	24,450

*Notes:* This table presents the border effect estimates for cost-of-living differences across international and intranational pairs. Cost-of-living differences are computed based on Equation 3. Hereafter, we take a logarithmic transformation of these level differences and compute the variance across product categories within region pair-year cells. Column (1) estimates the border effect by regressing the variance of log cost-of-living differences on a border dummy and year fixed effects. Column (2) adds geographic controls. Column (3) takes the regression from column (2) and estimates the border effect on the trimmed sample (based on the admissible propensity score range). Column (4) applies the subclassification estimator on the trimmed sample. Columns (5) and (6) re-compute cost-of-living differences by shifting the product variety level elasticity of substitution distribution by 1 and 2 respectively. Hereafter, border effects are estimated using the subclassification estimator applied to the trimmed sample. Beneath each specification, we disclose the simple average of the variance of log cost-of-living differences for the intranational region pairs included in the estimation sample. We compute cluster standard errors at the region pairs and present them in brackets below the coefficient estimates. Reported significance levels are at the  $p < 0.1^*$ ,  $p < 0.05^{**}$  and  $p < 0.01^{***}$  levels.

### 6.3 Detecting Geographic Market Segmentation

910

The final step to detect the presence of geographic market segmentation is to filter out regional differences in consumer taste from overall cost-of-living differences between comparable international and intranational region pairs. Table 5 presents the results from applying the subclassification estimator to the trimmed sample of region pairs to each of the structural components from equation 4 evaluated at the baseline elasticities of substitution. Because the variance decomposition is exact, the sum of the effects for each of the individual structural components sums to the overall effect, which we replicate in column (1). To obtain the relative importance of LOP deviations and choice set differences, we compare each of their individual effects to the overall effect. We also report the mean of the variance of each of the structural components for intranational pairs in the row indicated by  $\mathbb{V}(\cdot | l, l'' \in \mathcal{L}_{\text{intra}})$ .

919

First, we detect considerable geographic market segmentation across international region pairs. This is because both LOP deviations and choice set differences are significantly higher for international pairs relative to comparable intranational pairs. Taken together, they account for a little over 40% of the overall cost-of-living differences across international region pairs. Given that overall cost-of-living differences across international regional pairs are seven times larger compared to cost-of-living differences across international region pairs, this result implies a substantial degree of geographic market segmentation across international region pairs. At the same time, intranational region pairs seem to be characterized by a limited amount of

920

921

922

923

924

925

926

geographic market segmentation. Across intranational region pairs, LOP deviations and choice set differences jointly lead to a variance in cost-of-living differences of 0.0105. Therefore, a back-of-the-envelope calculation implies that around 70% of cost-of-living differences for intranational region falls within a range  $[-10.2\%, 10.2\%]$  around its mean.<sup>58</sup>

Second, while both LOP deviations are statistically significantly higher for international region pairs, choice set differences quantitatively dominate LOP deviations as a margin of geographic market segmentation. Out of the 40% of regional cost-of-living differences attributable to geographic market segmentation, choice set differences explain more than 95% of the variation. The important role of differences in choice sets to understand cost-of-living differences is qualitatively in line with [Argente et al. \(2021\)](#) and [Cavallo et al. \(2022\)](#) which decompose differences between ICP price levels and cost-of-living differences across countries.<sup>59</sup> [Argente et al. \(2021\)](#) find that not accounting for choice set differences can lead to a difference of 25% between a welfare-based price measure and the ICP measure. [Cavallo et al. \(2022\)](#) measure cost-of-living differences across countries by combining the number of firms and barcodes across countries with a Melitz-Pareto model to estimate the role of differences in variety (broadly defined). They find that up to 36% of the difference in cost of living relative to a GDP-deflator can be attributed to differences in variety. Relative to these papers that focus on a set of trading partners that still have many formal trade barriers, our results show that differences in choice sets still matter greatly across trading partners that are part of both a currency and customs union. In addition, the fact that choice set differences seem a much more important margin of geographic market segmentation compared to LOP deviations is remarkable in light of the predominant focus on LOP deviations as a measure of geographic market segmentation.

Finally, column (3) confirms that more than 50% of the overall cost-of-living differences can be accounted for by taste differences. The large role of taste differences is consistent with the evidence provided in [Coşar et al. \(2018\)](#). They document that home-biased preferences are the main driver of home market advantage in terms of market shares for automobile manufacturers. So, not only do we find that taste differences are key to explaining within-country cost-of-living differences (see Table 4), but they are also even higher across countries. This has two implications. On the one hand, this implies that even if geographic market segmentation would fall to levels seen for intranational region pairs, large cost-of-living differences across European countries would remain simply because of cross-country differences in consumer tastes. On the other hand, the existence of large taste differences across countries casts doubt on whether only relying on trade shares is very informative about geographic market segmentation. Nevertheless, while both trade shares and cost-of-living differences will yield differences between international and intranational region pairs when consumer tastes vary across countries, our decomposition allows us to separate which part of cost-of-living differences is due to taste differences and which part can be attributed to geographic market segmentation.

<sup>58</sup>The back-of-the-envelope calculation we perform is taking the square root of 0.0115 and assuming that this distribution is approximately normal. If so, roughly 70% of the mass then falls within one standard deviation around a zero mean.

<sup>59</sup>Our exercise decomposes the variance of level differences in cost of living into different components and therefore our results do not translate one-to-one to these papers.

**Table 5: Geographic Market Segmentation**

$\hat{\mathbb{V}}_{ll',t} \left[ y_{pl',t} - y_{pl,t} \right]$	CLE $_{ll',t}$ (1)	LOP $_{ll',t}$ (2)	Taste $_{ll',t}$ (3)	Choice $_{ll',t}$ (4)
$B_{ll'}$	.9288*** [.632, 1.306]	.0194*** [.018, .022]	.5466*** [.394, .716]	.3628*** [.226, .6]
$\hat{\mathbb{V}} \left[ y_{l',t} - y_{l,t} \right] / \hat{\mathbb{V}} \left[ p_{l',t} - p_{l,t} \right]$	1 [1, 1]	.022 [.015, .031]	.593 [.512, .648]	.385 [.327, .472]
$\hat{\mathbb{V}} \left( \cdot   l, l'' \in \mathcal{L}_{\text{intra}} \right)$	.1391	.0003	.1286	.0102
Nr. obs	24,450	24,450	24,450	24,450

*Notes:* This table decomposes the border effect of cost-of-living differences into the structural components. The log of cost-of-living differences is the sum of the structural components. Therefore, it can be subdivided into the variance of each of the structural components and the covariances and we allocate these covariances equally across the components. In this way, the sum of the individual border effects sums to the associated border effect overall for the variance of log cost-of-living differences. We compute each of the components at the baseline elasticities and apply the subclassification estimator at the trimmed sample of region pairs (based on the admissible propensity score range). Column (1) displays the border effects for marginal cost differences and column (2) for markup differences. Columns (3) and (4) display the border effects for expenditure share differences at the product variety level and firm level respectively. Columns (5) and (6) present the border effects for choice set differences at the product variety level and at the firm level. Below the border effects, we display the average variance for the intranational pairs included in the estimation panel and the ratio of the component-specific border effect and the border effect for overall cost-of-living differences. We compute cluster standard errors at the region pairs and present them in brackets below the coefficient estimates. Reported significance levels are at the  $p < 0.1^*$ ,  $p < 0.05^{**}$  and  $p < 0.01^{***}$  levels.

## 7 Concluding Remarks

Understanding the margins of geographic market segmentation is central to discerning whether there is still scope for continued integration efforts across European markets and if so which margin should be targeted by policymakers. Recent studies have reiterated the continued existence of large price differences and differences in trade shares across European countries. However, only studying LOP deviations is potentially a too narrow view and solely looking at regional variation in trade shares risks convoluting taste differences with limits to geographic market integration.

This paper combines household-level scanner data across European regions with a structural model of preferences to detect geographic market segmentation in terms of regional cost-of-living differences. By relying on an intuitive decomposition of the cost-of-living differences, we decompose cost-of-living differences into (1) LOP deviations, (2) pure taste differences and (3) choice set differences. In this way, we detect geographic market segmentation by considering whether LOP deviations and choice set differences are higher for international region pairs relative to intranational region pairs with similar geographic differences.

We find that geographic market segmentation is still large among European countries as LOP deviations and choice set differences lead to cost-of-living differences that are more than three times larger for international region pairs relative to intranational pairs. While LOP deviations contribute to these larger

cost-of-living differences, they are of second-order importance. Differences in choice sets are responsible for 95% of geographic market segmentation in Europe.

Importantly, we estimate that roughly 60% of the increased cost-of-living differences across international pairs are related to differences in consumer taste. This implies two things. First, large cost-of-living differences across European countries will likely remain even after the sources of geographic market segmentation have been removed. As the same implication holds for trade shares, border effects for trade shares should be interpreted with caution.

Our results further imply that to reduce geographic market segmentation, policymakers should likely focus on stimulating cross-country entry of firms and product varieties to reduce choice differences. Unfortunately, while being very rich, our data does not allow us to relate cost-of-living differences stemming from choice set differences to specific policies. Still, our approach can rule out certain reasons for their continued existence. First, all countries we study are part of the European customs and currency union. This ensures that the remaining level of geographic market segmentation is not due to conventional barriers to trade such as tariffs, quotas or currency conversion costs. Second, the estimated level of geographic market segmentation is unlikely to be driven by transport costs as we compare quite small and homogeneous geographic regions. This leads us to speculate that certain informal barriers are still segmenting European markets, but we leave it to future work to help uncover those.

## References

- Alesina, A., & Spolaore, E. (1997). On the number and size of nations. *The Quarterly Journal of Economics*, 112, 1027-1056.
- Amiti, M., Itskhoki, O., & Konings, J. (2019). International shocks, variable markups, and domestic prices. *The Review of Economic Studies*, 86, 2356-2402.
- Anderson, J. E., & Wincoop, E. V. (2003). Gravity with gravitas: A solution to the border puzzle. *The American Economic Review*, 93, 170-192.
- Anderson, J. E., & Wincoop, E. V. (2004). Trade costs. *Journal of Economic Literature*, 42, 691-751.
- Argente, D., Hsieh, C.-T., & Lee, M. (2021). Measuring the cost of living in mexico and the us. *American Economic Journal: Macroeconomics*.
- Atkeson, A., & Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices. *The American Economic Review*, 98, 1998-2031.
- Atkin, D., Faber, B., & Gonzalez-Navarro, M. (2018). Retail globalization and household welfare: Evidence from mexico. *Journal of Political Economy*, 126, 1-73.
- Beck, G. W., Kotz, H. H., & Zabelina, N. (2020). Price gaps at the border: Evidence from multi-country household scanner data. *Journal of International Economics*, 127, 1-25.
- Berka, M., Devereux, M. B., & Engel, C. (2018). Real exchange rates and sectoral productivity in the eurozone. *The American Economic Review*, 108, 1543-1581.
- Berman, N., Martin, P., & Mayer, T. (2012). How do different exporters react to exchange rate changes? *The Quarterly Journal of Economics*, 127, 437-492.
- Broda, C., & Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly Journal of Economics*, 121, 541-585.
- Broda, C., & Weinstein, D. E. (2008). Understanding international price differences using barcode data. *NBER Working Paper Series*.
- Broda, C., & Weinstein, D. E. (2010). Product creation and destruction: Evidence and price implications. *The American Economic Review*, 100, 691-723.
- Cavallo, A., Feenstra, R. C., & Inklaar, R. (2022). Product variety, cost-of-living and welfare across countries. *American Economic Journal: Macroeconomics*, 1-60.
- Cavallo, A., Neiman, B., & Rigobon, R. (2014). Currency unions, product introductions, and the real exchange rate. *The Quarterly Journal of Economics*, 129, 529-595.
- Coşar, A. K., Grieco, P. L., Li, S., & Tintelnot, F. (2018). What drives home market advantage? *Journal of International Economics*, 110, 135-150.
- Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009, 3). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96, 187-199.



- Dellavigna, S., & Gentzkow, M. (2019). Uniform pricing in u.s. retail chains. *The Quarterly Journal of Economics*, 134, 2011-2084. 1027  
1028
- Donaldson, D. (2018, 4). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108, 899-934. 1029  
1030
- Dubé, J. P., Hortaçsu, A., & Joo, J. (2021, 7). Random-coefficients logit demand estimation with zero-valued market shares. *Marketing Science*, 40, 637-660. 1031  
1032
- Döpfer, H., Mackay, A., Miller, N. H., & Stiebale, J. (2022). *Rising markups and the role of consumer preferences*. 1033
- Eaton, J., Kortum, S., & Kramarz, F. (2011). An anatomy of international trade: Evidence from french firms. *Econometrica*, 79, 1453-1498. 1034  
1035
- Engel, C., & Rogers, J. H. (1996). How wide is the border? *American Economic Review*, 86, 1112-1125. 1036
- Faber, B., & Fally, T. (2021). Firm heterogeneity in consumption baskets: Evidence from home and store scanner data. *The Review of Economic Studies*. 1037  
1038
- Fajgelbaum, P. D., Goldberg, P. K., Kennedy, P. J., & Khandelwal, A. K. (2020). The return to protectionism. *The Quarterly Journal of Economics*, 135, 1-55. 1039  
1040
- Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *American Economic Review*, 84, 157-177. 1041  
1042
- Feenstra, R. C., Xu, M., & Antoniadis, A. (2020). What is the price of tea in china? goods prices and availability in chinese cities. *The Economic Journal*, 130, 2438-2467. 1043  
1044
- Flam, H. (1992). Product markets and 1992: Full integration, large gains? *Journal of Economic Perspectives*, 6, 7-30. 1045  
1046
- Fontaine, F., Martin, J., & Mejean, I. (2020, 5). Price discrimination within and across emu markets: Evidence from french exporters. *Journal of International Economics*, 124. 1047  
1048
- Gandhi, A., Lu, Z., & Shi, X. (2022). *Estimating demand for differentiated products with zeroes in market share data*. 1049  
1050
- Goldberg, P. K., & Knetter, M. M. (1997). Goods prices and exchange rates: What have we learned? *Journal of Economic Literature*, 35, 1243-1272. 1051  
1052
- Gopinath, G., Gourinchas, P.-O., Hsieh, C.-T., & Li, N. (2011). International prices, costs, and markup differences. *American Economic Review*, 101, 2450-2486. 1053  
1054
- Gorodnichenko, Y., & Tesar, L. L. (2009). Border effect or country effect? seattle may not be so far from vancouver after all. *American Economic Journal: Macroeconomics*, 1, 219-241. 1055  
1056
- Handbury, J., & Weinstein, D. E. (2015). Goods prices and availability in cities. *The Review of Economic Studies*, 82, 258-296. 1057  
1058
- Hausman, J. (1996). *Valuation of new goods under perfect and imperfect competition* (Vol. I; T. F. Bresnahan & R. J. Gordon, Eds.). 1059  
1060
- Heathcote, J., & Perri, F. (2014). Assessing international efficiency. *Handbook of International Economics*, 4, 523-584. 1061  
1062

- Hottman, C. J., Redding, S. J., & Weinstein, D. E. (2016, 8). Quantifying the sources of firm heterogeneity. *Quarterly Journal of Economics*, 131, 1291-1364. 1063  
1064
- Hummels, D., & Skiba, A. (2004). Shipping the good apples out? an empirical confirmation of the alchian-allen conjecture. *Journal of Political Economy*, 112, 1384-1402. 1065  
1066
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social and biomedical sciences: An introduction* (1st Edition ed.). Cambridge University Press. 1067  
1068
- Itskhoki, O. (2021). The story of the real exchange rate. *Annual Review of Economics*, 13, 1-33. 1069
- Jaravel, X. (2019). The unequal gains from product innovations: Evidence from the us retail sector. *The Quarterly Journal of Economics*, 134, 715-783. doi: 10.1093/qje/qjy031.Advance 1070  
1071
- Kehoe, T. J., & Ruhl, K. J. (2013). How important is the new goods margin in international trade? *Journal of Political Economy*, 121, 358-392. 1072  
1073
- Loecker, J. D., Eeckhout, J., & Unger, G. (2020, 5). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135, 561-644. 1074  
1075
- Loecker, J. D., Goldberg, P. K., Khandelwal, A. K., & Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, 84, 445-510. 1076  
1077
- Loecker, J. D., & Warzynski, F. (2012, 10). Markups and firm-level export status. *American Economic Review*, 102, 2437-2471. 1078  
1079
- McCallum, J. (1995). National borders matter : Canada-u.s. regional trade patterns. *The American Economic Review*, 85, 615-623. 1080  
1081
- Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71, 1695-1725. 1082  
1083
- Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94, 20-36. 1084  
1085
- Redding, S. J., & Weinstein, D. E. (2020). Measuring aggregate price indices with taste shocks: Theory and evidence for ces preferences. *The Quarterly Journal of Economics*, 135, 503-560. 1086  
1087
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98, 71-102. 1088
- Santamaria, M. A., Ventura, J., & Yesilbayraktar, U. (2021). Borders within europe. *NBER Working Paper Series*, 1-87, 1689-1699. 1089  
1090
- Santamaría, M., Ventura, J., & Yeşilbayraktar, U. (2023). Exploring european regional trade. *Journal of International Economics*. 1091  
1092
- Sato, K. (1976). The ideal log-change index number. *The Review of Economics and Statistics*, 58, 223-228. 1093
- Vartia, Y. O. (1976). Ideal log-change index numbers. *Scandinavian Journal of Statistics*, 3, 121-126. 1094