

---

# Central European Economic Journal

---

Kristóf Gyódi,<sup>1</sup> Maciej Sobolewski,<sup>2</sup> Michał Ziemiński<sup>3</sup>

## What Drives Price Dispersion in the European E-commerce Industry?

---

**1** Faculty of Economic Sciences, University of Warsaw and DELab, e-mail: kgyodi@wne.uw.edu.pl

**2** Faculty of Economic Sciences, University of Warsaw and DELab, e-mail: maciej.sobolewski@uw.edu.pl

**3** Faculty of Economic Sciences, University of Warsaw and DELab, e-mail: m.ziembinski@delab.uw.edu.pl

**Abstract:** An important aspect of economic integration of the European Union is price convergence on digital single market. In this study, we propose a novel way to measure price dispersion in the e-commerce industry, using a custom made web-scraping tool. We target all the major price comparisons sites in the 26 EU member states, which enables us to collect price signals from thousands of retail shops operating on-line. We analyse pricing data of 182 branded products sold on-line across the EU, representing the most popular categories: fashion, consumer electronics, gaming and software, and cosmetics. We find considerable dispersion of both pre and post-vat on-line prices ranging from 20% to 40%, depending on the product category. The observed on-line price dispersion is driven by both cost factors and the level of per capita income, which is consistent with the view that producers or large distributors might engage in strategic price discrimination induced by income heterogeneity.

At first look, our results point to the unexplored potential for cross-border trade, which could be released by policy interventions with regards to delivery, payment or law harmonization. However, under strategic price discrimination, reduced costs of arbitrage for consumers might induce discriminating firms to lower the magnitude of price dispersion between high and low income countries, bringing adverse welfare changes of a priori unknown net effect.

**Keywords:** price dispersion, price discrimination, e-commerce, web-scraping, price comparison service.

**JEL Codes:** L11, L12

## Acknowledgements

The article is based on a study commissioned by the Ministry of Foreign Affairs of the Republic of Poland. We would like to thank the anonymous reviewer for valuable comments and suggestions.

## 1 Introduction

On-line shopping is becoming an important channel of purchases besides the traditional off-line retail. According to the Eurostat data, the share of e-commerce in total turnover has already reached 15% in EU15, and in countries like Great Britain, it comprises more than 20%. The European Commission (EC), aware of the fundamental meaning of digital economy, made the completion of the Digital Single Market (DSM) one of its priorities. The aim of the DSM is to ensure the cross-border access of on-line activities under conditions of fair competition in the entire EU (European Commission 2015).

According to standard theories such as the Law of One Price, there should be no significant difference in prices on integrated markets. A potential impact of the stronger integration of digital markets between the EU member states is the decrease in the price dispersion across countries in on-line prices. However, this is not what is observed in the EU in recent years. According to the Eurostat (2016), prices converged in the EU28 between 2004 and 2008, but afterwards this process stopped, or even reversed. On the other hand, the level of price dispersion is fairly stable in the EU15. Recently, the European Commission (EC) has undertaken several legal initiatives, such as geo-blocking directive or vertical restraints inquiry to foster competition on digital single market. EC hopes that with increased volume of cross-border trade, further reduction of price dispersion will be achieved, even though there is still lack of understanding of all the factors and implications behind it. Importantly, lowering the costs of arbitrage between markets subject to third degree price discrimination could reduce price dispersion in many different ways, bringing adverse distributional effects in high and low income countries. As economic literature suggests, a-priori there is no conviction that the net welfare effect of such intervention will always be positive, unless prices in lower income country do not increase (Malueg and Schwartz 1994).

The objective of our study is to assess the scale and determinants of current on-line price dispersion within the EU. With respect to measurement, we use the web-scraped price data. Web-scraping is an innovative technique used to analyse the price dynamics, for example, by Cavallo (2015). The novelty of our approach lies in using data from price comparison services instead of working with great number of individual sites which often use incompatible web technologies. In this way, we are able to collect tons of individual price signals

from virtually all shops selling a particular item in each country. With such a large and diversified data, we can measure price dispersion in a robust way and make frequent updates at relatively low cost.

With respect to the pattern of observed on-line price dispersion among member states, we analyse three broad groups of factors that influence its magnitude: (a) supply side factors; (b) demand side factors and (c) institutional factors. Understanding sources of the price dispersion is crucial for correct targeting of the policy interventions.

Our paper is organized as follows: section 2 provides literature review, section 3 describes data collection technique and summarizes the data, section 4 sheds light on the price dispersion pattern, while section 5 summarizes and concludes.

## 2 Literature review

An important aspect of economic convergence is the decrease of price dispersion (the difference between prices for identical products). Faber and Stokman (2009) examine long run price convergence within the EU between 1960 and 2003. In order to gain such long time series, the authors scale harmonised indices of consumer prices (HICP) with occasional observations of absolute price levels. They conclude that during the observed time period, the price differences halved, and that the price convergence had been driven by indirect tax rate harmonisation, convergence in input costs due to exchange rate stability and openness of economy. Rogers (2007) also confirms price convergence between 1990 and 2004. Using data from the Economist Intelligence Unit (EIU), he claims that the price dispersion in tradable goods reached the level observed in the US. Bergin and Glick (2007) analyse the global price dispersion between using the EIU data and report a U-shaped pattern of price convergence between 1990 and 2005. A possible driver of the rising dispersion from 1995 is the price of oil and transport costs.

The introduction of the euro should have decreased price dispersion due to removing transaction costs, exchange rate risk and higher price transparency (Wolszczak-Derlacz 2010). However, the literature has been divided in the evaluation of the impact of the euro on price dispersion. Allington, Kattuman, and Waldmann (2004) finds that price convergence accelerated due to the euro, while others, including Parsley and Wei

(2008), Wolszczak-Derlacz (2010) or Fischer (2012) did not find such causal effect. Ogrokhina (2015) examines the impact of the common market and the EMU on price convergence of 120 traded goods. While the Single European Act decreased price differences by 5%, the single currency increased it by 2%. However, the euro had different impact considering product categories- price convergence had been observed for cars, while divergence for food and tobacco. Baye *et al.* (2006) study the impact of euro on prices, using data collected from a price comparison site (Kelkoo). The data covers prices for 28 products across 7 EU countries. The authors find an increase of prices relative to the non-EU countries.

The observed cross-country differences in prices may be caused by international price discrimination of the producers (differences in mark-ups), or by differences in costs (Verboven 1996). Mark-ups may be influenced by differences in consumer income and taste, while costs by transportation, service wages and so on. Additionally, structural causes, like market structure or differences in taxes (VAT) may cause prices to differ (DG Internal Market 2002).

Price dispersion has been widely analysed in the car market, due to the huge price differences between EU15 countries. Verboven (1996) shows that international price discrimination, measured by cross-country differences in relative wholesale mark-ups, is among the drivers of price-differences. Interestingly, car producers seem to charge a higher mark-up for their domestic market, than at the foreign markets. An analysis of price convergence in the car market is also provided by Goldberg and Verboven (2005), who find evidence on convergence to the Law of One Price between 1970–2000.

Alessandria and Kaboski (2007) show that income per capita explains half of the aggregate price dispersion across countries. Using a wide sample of the US export data, the authors suggest that mark-ups account for 40% of this relationship and that costumers in low income countries are more price sensitive. Crucini and Yilmazkuday (2014) analyse the determinants of price dispersion using a panel of retail prices from 79 countries in the period 1990–2005. The authors find that service-sector wage differences almost entirely explain price dispersion at the PPP level. When analysing price dispersion at the level of individual goods and services, wage differences explain around one third of dispersion, with an equal impact of trade related costs. However, the authors are aware of the fact that the prices of tradable goods are affected by non-tradable elements; therefore, it is difficult to distinguish between costs and mark-ups.

Simonovska (2015) is tackling this issue analysing the relationship between the price of on-line tradable goods and per capita income. The author is using a dataset of prices on apparel products produced by a Spanish manufacturer, which are sold in 29 various countries (23 EU member states). The examined products are only available on-line, furthermore they are manufactured and shipped from a single place; therefore, destination-specific price contributions (such as wages) can be suppressed. The author finds that per-capita income differences account for one third, while shipping costs explain up to a third of the cross-country differences in prices. More specifically, doubling the per capita income of the buyer's country leads to an 18% rise in the price level of apparel and footwear. Additionally, the author finds that Eurozone membership decreases price dispersion relative to Spain.

On-line shopping is becoming an important channel of purchases besides the traditional off-line retail. According to Eurostat data, the share of e-commerce in total turnover (without the financial sector) has already reached 15% in EU15, and in countries like Great Britain comprises more than 20%. The European Commission, aware of the fundamental meaning of digital economy, made the completion of the Digital Single Market (DSM) one of its priorities. The aim of the DSM is to ensure the cross-border access of on-line activities under conditions of fair competition in the entire EU (European Commission, 2015). A potential impact of the stronger integration of digital markets between EU member states is the decrease in the price dispersion across countries in on-line prices.

To sum up, the literature suggests that price differences are indeed influenced by both mark-ups and costs. International price discrimination can be classified as third degree price discrimination, in which the seller discriminates among different groups of consumers based on observable heterogeneities (for example, the geographical location of the consumer or income). Theoretical literature shows that third degree price discrimination might increase or decrease overall social welfare compared to uniform pricing. Price discrimination is not harmful to consumers from low income countries and welfare effects of reducing price discrimination will have opposite signs in case of low and high priced markets. Even if the net effect is positive for the whole EU, some compensation mechanism should be in place.

For the analysis of monopoly market, see Cowan (2008), Aguirre, Cowan, and Vickers (2010), Galera, Kinateder, and Mendi (2014), while for duopoly markets

refer to Corts (1998), Stole (2003) and Adachi and Matsushima (2014). Malueg and Schwartz (1994) show the necessary conditions for price discrimination to increase the welfare of an economic union, which refer to the convexity of respective demand functions and the overall size of weak and strong markets.

This article addresses an important research gap in the literature. While price dispersion has been examined in the off-line retail across EU countries, on-line retail has been only partially analysed (e.g., Baye *et al.* 2006). This work provides the first analysis on the drivers of price dispersion across on-line retail in all the EU member states, using data on homogeneous products.

### 3 Scope of price dispersion

#### 3.1 Web-scraping technique and advantages and disadvantages of Scrapped-data

Cavallo (2015) introduces Scraped Data as a new source of micro-price information to measure price stickiness. It is only one of the examples of usage of the web-scraping techniques, which are getting attention as the new approach to analyse behaviour of prices. Lünemann and Wintr (2011), Gorodnichenko and Talavera (2014) and Gorodnichenko, Sheremirov, and Talavera (2014) find that on-line prices tend to be more flexible and have smaller price changes than the off-line prices. However current research was mainly focused on the price dynamics and their patterns. However, to the authors' best knowledge, there is still a gap in the empirical literature using the web-scraping and investigating the price dispersion between countries. To fill this gap we investigated the price differences in on-line retail in the European Union.

There are a few advantages of using on-line rather than off-line prices. Firstly, collecting data remotely is a much cheaper option than the survey based methods, especially in case of cross-country setting. Secondly, it is possible to include large number of retailers. This huge advantage is important, as the result won't be affected by the price strategies of a particular company or retail chain. However, in case of increasing the number of retailers, it would be necessary to design the web-scraping tool for every retailer separately, if they

are using different web-page designs. To address this issue, we run our web-scraping tool on the data provided by the price comparisons services.<sup>1</sup> It allows us to limit web-scraping tool complexity significantly with keeping the diversity of on-line retailers. Thirdly, we are able to collect many observations for one product, which should improve the quality of price information in our database. It is connected with one disadvantage – as the automated tool is collecting all information provided by the price comparison service, we observe 'heterogeneity' bias<sup>2</sup> extending tails of price distribution. To address this, we use the medians of prices for every country-product level in our analysis.

Our approach is similar to the one presented in Cavallo (2015). As we are not intending to analyse the dynamic of prices we are collecting our data in one point of time. Precisely, our procedure works in three steps. Firstly, we collect information about most popular products searched in price comparisons services and we choose only those, which have the most 'unified' names across the countries (i.e., the name of the product is the same in at least 9 countries). Secondly, we design a dedicated tool to visit the price comparison service web-pages and to collect information about the product, price and price information provider (i.e., online shop that is included in the price comparison service database). Our program is designed to read the webpage code, recognize proper tags defined separately for every price information provider<sup>3</sup> and store the data. Finally, we run our program in two different modes – using its own IP and using the external server placed in particular country. The last step is designed to ensure that we are getting the same price information as the local consumer on a particular market.

#### 3.2 Data set and results

We collect the data about five of the most popular categories in on-line retail, in particular: fashion (clothing and footwear), cosmetics and healthcare (cosmetics and

<sup>1</sup> The list of the price comparisons services is available upon request.

<sup>2</sup> As 'heterogeneity' bias, we refer to the situation when the price comparison service asked about the price of the particular product provides us with price information of the most similar products in case of not having proper data. Even though we use restrictive cleaning techniques to fully address this concern, we use the medians of the prices in further analysis.

<sup>3</sup> For more details about markup languages and their tag structure, please see Coombs, Renear, and DeRose (1987).

perfumes), computer games and software, electronics and computer hardware, and household appliances. The full list of products contains 657 different product names divided into categories, which results in 143,490 observations. The full split of the sample by products and by observations is presented in Tab. 1.

To ensure that we are collecting valid price information, we apply a few rigorist filters on our data, for example, every word from the product name (further called ‘product query’) is included in the description of our price record. These procedures limit the number of products to 382 and the number of observations to 32,116. Finally, we keep only the products, which are recorded in at least 9 countries in our sample. The final impact on the number of products after each step and in every category is presented in details in Tab. 1. As it can be seen, the cleaning procedure impacted the data significantly – finally we keep only around 19.1% of the sample. In the case of working with scrapped data, we face a trade-off between the number of data providers (price comparison services, on-line retailers’ webpages) and code complexity. In particular, if we want to extend the number of data providers, we need to define proper tag structure of every new one. Using data from price comparison services, we can address this problem; however, we need to pay more attention to its quality.

Even though at the beginning of data collection, we were trying to ensure that the data are equally distributed between countries; after all the cleaning procedures, we observe that for some of them, the total number of observations remains low. Total number of products and observations for each country is presented in Tab. 2. It has to be noted that only for two countries, the numbers remains low (Belgium and Portugal), while for others we get satisfying sample sizes.

We observe significant price dispersion in the European Union, with respect to both gross and pre-tax retail prices (see Tab. 3 and Tab. 4). We calculate average price dispersion in three steps. Firstly, as mentioned before, we calculate the median price on country-product level (for every product and for every country separately). It cleans our data from outliers and ‘heterogeneity’ bias issue. Secondly, we compute the mean of the median prices for every product. It gives us the average price of the product in European Union. Finally, we calculate the difference between median price and the product average price and scale it by the product average price. The results are presented in first column of Tab. 3. The absolute price deviation in the EU on an average is (+/-) 11.2%. In similar way, we calculate different dis-

**Tab. 1.** The number of products and observations in each product category

Category	# of products		
	scrapped	after cleaning	in final sample
Fashion	362	154	33
Cosmetics & Healthcare	87	33	24
Households Appliances	58	56	35
Electronics and Computer Hardware	84	75	43
Computer Games and Software	66	64	47
<b>Total</b>	<b>657</b>	<b>382</b>	<b>182</b>

Category	# of observations		
	scrapped	after cleaning	in final sample
Fashion	58,410	7,713	4,272
Cosmetics & Healthcare	25,472	5,419	4,736
Households Appliances	11,643	3,955	3,181
Electronics and Computer Hardware	23,697	7,503	6,666
Computer Games and Software	24,268	7,526	6,547
<b>Total</b>	<b>143,490</b>	<b>32,116</b>	<b>25,402</b>

Source: Authors’ own elaboration.

**Tab. 2.** The number of products and observations for EU countries

Country	# of products	# of observations	Country	# of products	# of observations
Austria	143	1,183	Italy	131	910
Belgium	16	37	Latvia	115	2,447
Bulgaria	101	678	Lithuania	92	494
Croatia	91	340	Netherlands	39	380
Czech Republic	144	2,427	Poland	137	1,805
Denmark	127	844	Portugal	36	158
Estonia	58	763	Romania	141	1,152
Finland	73	346	Slovakia	152	1,963
France	92	766	Slovenia	91	469
Germany	70	989	Spain	72	386
Greece	130	1,359	Sweden	144	2,177
Hungary	149	1,482	United Kingdom	131	1,342
Ireland	76	505	<b>Average</b>	<b>102.0</b>	<b>1,016.1</b>

Source: Authors’ own elaboration.

Tab. 3. The gross (retail) price dispersion in European e-commerce industry

Category	Subcategory	Average absolute dispersion	Average # of observation per product	Quantile coefficient of dispersion (0.9-0.1)	Quantile price dispersion (0.9-0.1) [EUR]	Quantile coefficient of dispersion (0.75-0.25)	Quantile price dispersion (0.75-0.25) [EUR]	Coefficient of variation (mean)	Coefficient of variation (median)
Fashion	Clothing	13.9%	9.91	39.3%	27.65	18.3%	12.03	0.35	0.34
Fashion	Footwear	12.3%	12.36	31.1%	27.91	18.0%	16.08	0.25	0.26
Fashion	Accessories	9.9%	7.92	26.4%	21.16	12.1%	10.52	0.54	0.49
Cosmetics & Healthcare	Cosmetics	12.9%	12.72	36.2%	11.93	19.4%	4.28	1.16	1.53
Cosmetics & Healthcare	Perfumes	12.5%	12.53	35.3%	22.91	20.5%	13.2	0.38	0.39
Households Appliances	House	9.6%	6.65	20.9%	12.05	8.7%	5.04	0.35	0.34
Households Appliances	House (>100EUR)	10.4%	7.02	21.1%	58.85	11.3%	30.42	0.78	1.31
Households Appliances	Kitchen	11.4%	7.06	35.5%	64.32	18.4%	35.05	0.78	0.91
Households Appliances	Beauty	13.1%	5.51	31.8%	17.24	17.8%	10.1	0.51	0.5
Electronics and Computer Hardware	Accessories	10.7%	10.05	31.6%	14.98	19.5%	9.78	0.51	0.5
Electronics and Computer Hardware	Laptops and PCs	11.4%	8.67	28.8%	162.74	16.3%	94.61	0.55	0.84
Electronics and Computer Hardware	Monitors	7.8%	9.18	20.6%	39.54	11.5%	19.81	0.95	1.46
Electronics and Computer Hardware	Tablets and E-readers	9.6%	7.36	24.1%	73.57	14.0%	43.34	0.55	0.59
Electronics and Computer Hardware	Consoles	9.6%	10.09	24.0%	67.89	11.7%	32.18	0.4	0.4
Computer Games and Software	Software	11.7%	9.11	32.4%	87.89	18.6%	55.35	0.7	0.81
Computer Games and Software	PC games	12.3%	7.46	35.5%	13.29	21.4%	8.09	0.39	0.36
Computer Games and Software	Console games	10.7%	7.44	31.7%	14.74	18.2%	8.5	0.26	0.25
<b>Total</b>		11.2%	8.93	29.8%		16.2%		0.55	0.66

Source: Authors' own elaboration.

Tab. 4. The pre-tax price dispersion in European e-commerce industry (VAT excluded)

Category	Subcategory	Average absolute price dispersion	Average # of observation per product	Quantile coefficient of dispersion (0.9–0.1)	Quantile price dispersion (0.9–0.1) [EUR]	Quantile coefficient of dispersion (0.75–0.25)	Quantile price dispersion (0.75–0.25) [EUR]	Coefficient of variation (mean)	Coefficient of variation (median)
Fashion	Clothing	14.50%	9.83	39.53%	21.66	20.38%	10.55	0.36	0.34
Fashion	Footwear	12.28%	12.34	31.35%	22.06	17.76%	12.36	0.26	0.27
Fashion	Accessories	9.85%	7.92	24.60%	15.62	13.45%	8.90	0.52	0.47
Cosmetics & Healthcare	Cosmetics	12.51%	12.86	34.11%	9.26	18.51%	3.70	1.17	1.49
Cosmetics & Healthcare	Perfumes	12.46%	12.57	35.35%	17.41	18.93%	8.99	0.37	0.38
Households Appliances	House	10.96%	6.38	25.05%	14.66	12.19%	7.90	0.42	0.39
Households Appliances	House (>100EUR)	8.83%	7.42	22.49%	52.58	13.12%	29.11	0.74	1.22
Households Appliances	Kitchen	11.58%	6.96	35.26%	52.13	19.79%	30.31	0.75	0.83
Households Appliances	Beauty	13.18%	5.61	33.12%	14.14	18.07%	7.91	0.50	0.49
Electronics and Computer Hardware	Accessories	10.38%	9.98	29.74%	11.47	18.55%	7.33	0.50	0.49
Electronics and Computer Hardware	Laptops and PCs	11.50%	8.67	28.34%	129.80	16.13%	77.04	0.55	0.84
Electronics and Computer Hardware	Monitors	7.69%	9.18	20.01%	29.64	10.79%	16.84	0.96	1.47
Electronics and Computer Hardware	Tablets and E-readers	9.93%	7.33	28.01%	67.71	13.59%	32.39	0.55	0.60
Electronics and Computer Hardware	Consoles	10.08%	10.07	25.05%	55.15	12.73%	28.22	0.41	0.40
Computer Games and Software	Software	11.74%	9.14	32.08%	67.40	17.00%	40.06	0.68	0.78
Computer Games and Software	PC games	12.63%	7.52	36.46%	10.63	21.04%	6.05	0.38	0.35
Computer Games and Software	Console games	11.04%	7.46	33.45%	12.04	19.06%	6.76	0.27	0.26
<b>Total</b>		11.2%	8.90	30.24%		16.53%		0.55	0.65

Source: Authors' own elaboration.

persion statistics, every time starting from median price on country-product level as proxy of the price in respective country for particular product. Price dispersion in the whole product sample, measured by interquartile range of median (post-tax) relative price distribution, amounts to 16% for 75–25 percentile interval. While this is a rather conservative measure, the figure almost doubles (30%) for 90–10 percentile range. Analysing the variation of the price differences, we see that both approaches – one based on quantiles and one based on the moments, show us that the possibilities for price arbitrage exist. In most extreme situations, they can reach even more than 160 EUR as in case Laptops and PCs category or more than 87 EUR for software, measured by quantile price dispersion (the difference between 90% quantile and 10% quantile). The scope of relative pre-tax dispersion is very similar, while in absolute terms it is obviously lower (see Tab. 4).

As results presented above show significant price dispersion in on-line retail, we confront our data with the Eurostat price levels data (precisely we use *prc\_ppp\_ind* variable from Eurostat database for year 2014). We expect a significant difference between our and Eurostat levels due to the differences in the products included in the sample (see Fig. 1). However, in terms of magnitude, we see that the range of scale dispersion is similar in both datasets. This shows that the analysed phenomenon is not exclusive for on-line retail.

Our data confirms significant price dispersion in European e-commerce industry. Dispersion of prices was previously confirmed in the economic literature as is shown in section 2. Also, Eurostat provides comprehensive data confirming the phenomenon. As we can

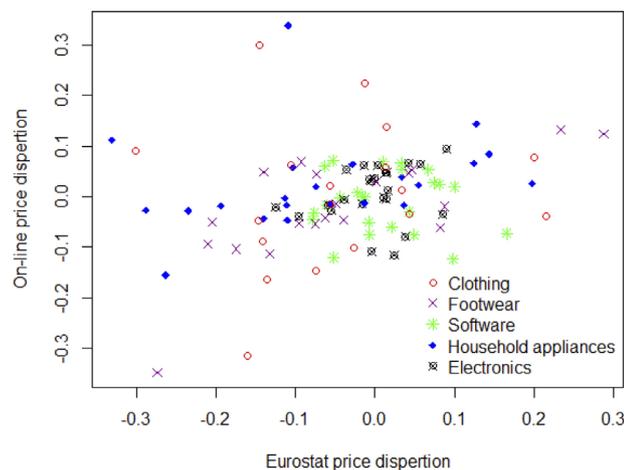


Fig. 1. Price dispersion in European e-commerce industry compared with Eurostat data

Source: Authors' own elaboration.

argue that in case of the off-line transaction making the cross-border transactions is connected with high transaction costs; for on-line retail, they are mostly limited. Therefore, to answer the question about the main factors behind price dispersion, usage of the on-line price data seems to be more appropriate.

## 4 Price dispersion pattern

Economic literature discussed in section two suggests at least three broad groups of factors that influence the magnitude of price dispersion: (a) supply side factors like the cost of manufacturing, sales and distribution of goods, (b) demand side factors such as preferences and more importantly purchasing power of consumers and (c) institutional factors such as differences in the level of competition in the market or taxation. It should be noted that although in most instances, items are produced in developing countries, and imported to the EU, factor prices in particular member states will still affect economic conditions of pursuing activity of on-line shops as sale, distribution and marketing generate bulk of costs, which are born domestically. Moreover, there is a large scope of differences among EU member states with respect to the contractual law regarding, for example, seller liability or scope of consumer protection that will be most likely reflected in prices.

In the following sections, we verify to what extent supply and demand side factors influence systematically on-line price dispersion among the EU states. We start this analysis by looking at various factor prices available in Eurostat, such as unit labour cost, unit cost of energy and fuels or capital costs. We expect that these supply side variables could affect prices in an obvious way. On top of it, two demand side variables are considered: annual net salary in the private service sector and GDP *per capita*. We also include standard VAT rates that generally apply to the product categories which we analyse.<sup>4</sup>

The scale of differences in all these variables among the Member States of the European Union is given in Tab. 5.

<sup>4</sup> Although VAT rules are not harmonized across the EU, reduced rates apply usually to books and baby clothing and equipment – all of which have not been considered in our study.

Tab. 5. Descriptive statistics of price dispersion determinants in the EU countries

Variable	mean	standard deviation	median	min	max	range	skew	kurtosis
Real GDP per capita [EUR, ths]	25.56	14.85	19.11	4.92	49.36	44.44	0.2	-1.64
Unit cost of labour [EUR/h]	18.67	12.01	15.6	3.8	40.3	36.5	0.45	-1.45
Average annual net salary [EUR/full time]	8.99	5.78	6.83	1.95	18.97	17.02	0.43	-1.51
Electric energy price [EUR/kwh]	0.18	0.06	0.17	0.09	0.3	0.21	0.64	-0.46
Euro95 petrol price [EUR/litre]	1.31	0.14	1.34	1.1	1.57	0.47	0.2	-1.21
Diesel fuel price [EUR/litre]	1.24	0.1	1.23	1.13	1.54	0.41	1.18	0.87
Cost of capital – interest rate [%] (a)	1.8	1.59	1.14	0.52	7.81	7.29	2.22	5.55
Standard VAT rate [%] (b)	22.04	2.07	21	19	27	8	0.6	-0.69
Population density	130.34	111.73	103.41	17.98	500.57	482.59	1.8	2.96

Source: Eurostat, average annual data (2014) except for (a) ECB (status as at 31/10/2015), (b) DG TAXUD (status as at 1/09/2015).

## 5 Results of the estimation

At present, in the EU, we can still observe considerable differentiation with regards to the factors presented in Tab. 4. To assess in a statistically meaningful way which of those factors have the strongest influence on dispersion of on-line prices in the EU countries, we have conducted linear regression analysis. As a dependent variable, we took relative price deviation, which is closely related to interquartile price dispersion but takes both negative and positive values around the average from median prices calculated for the countries where a given product was supplied. More specifically we have estimated the following model:

$$PD_{ij} = f(SS_{ij}, DS_{ij}, PC_k) \quad (1)$$

where  $PD_{ij}$  is percent deviation of median price of product  $i$  in country  $j$  from the average of median prices for all countries where the product was offered on-line.  $SS$  and  $SD$  are vectors of supply and demand side factors and  $PC$  are product subcategories. Detailed list of variables is provided in Tab. 6 below.

Prior to estimation, we inspected the relevance of our three categorical variables (*Category*, *Subcategory* and *Country*) by the analysis of the differences among group means of the dependent variable. We implement various techniques: visual assessment of boxplots (Fig. A1–A3 in Appendix, respectively), analysis of variance model (Tab. A1–A3 in the Appendix) and using Tukey’s range test (Fig. A4–A6 in the Appendix). All methods suggest that there are no significant differences between product categories and subcategories, while the values differ across countries, for example, in the case of Poland and Portugal, the average deviation of median prices is noticeably negative (Fig. A3 in the Appendix). We address this issue by including GDP *per capita* in the regression equation. The correlation between our dependent variable and GDP *per capita* in the whole sample is 18%. To guarantee that this variable captures differences across countries in the deviation in prices, we check correlation between differences in the dependent variable and in the GDP pc. It is significantly positive (54%), ensuring that including GDP pc into the regression addresses the cross-country differences in the dependent variable. We also check Levene’s test (Tab. A4–A6 in the appendix), which shows that we need to address the problem of heteroscedasticity.

Tab. 6. List of variables used in regression analysis

Name of variable in the model	Variable description
PD: [percent_deviation]	Percentage deviation of the median of product prices in a given country from the average of medians for all countries based on the web <i>scraping</i> study of results from price comparison websites
DS.: [GDP_pc]	Real GDP <i>per capita</i> in EUR thousands, at the end of 2014
DS: [net_earnings]	Average annual net salary in 2014
SS: [pop_density]	Population density – proxy for transportation costs
IE: [vat_stdrate]	Current level of standard VAT rate
SS: [ele_cost]	Unit cost of electric energy, average annual value in 2014
SS: [E95_price]	Euro95 petrol price, average annual value in 2014
SS: [diesel_price]	Diesel oil price, average annual value in 2014
SS: [unit_lab_cost]	Unit cost of labour, average annual value in 2014
SS: [interest_rate]	Current interest rate on 10-year Treasury bonds in the secondary market– capital price index
PC: [Games for PC], [Games for consoles], [Consoles], [Cosmetics], [Laptops and computers], [Small household appliances for the home < EUR 100], [Small household appliances for the home > EUR 100], [Small household appliances for the kitchen], [Monitors], [Footwear], [Perfumes], [Software], [Tablets and e-book readers], [Clothing], [Body care appliances] – Binary variables for subcategories of products.	

Source: Authors’ own elaboration.

Starting with the full model specification (Model 1), we face the problem of collinearity of independent variables. Based on the variance inflation factor test (VIF test), we drop the unit labour cost from the model (Model 2, see Tab. A7 in the Appendix).<sup>5</sup>

Next, we carry out the model selection procedure using two approaches: frequentist model averaging and stepwise regression. Model selection uncertainty can lead to the underreporting of variability and too optimistic confidence sets. Frequentist model averaging accounts for this uncertainty and incorporates it into the estimation process (Wang, Zhang, and Zou 2009). The results are reported in Tab. A8. We include only variables with relative importance > 0.5 to the restricted form of the model (Model 3).

For robust model selection, the stepwise regression procedure is implemented. Starting with the full model and using the AIC procedure, the same variables are chosen as in the previous approach (reported as Model 4). The standard diagnostic tests<sup>6</sup> show no heteroscedasticity, however, we notice auto-correlated errors. Therefore, the final model is estimated with a robust

enhanced White variance-covariance estimator. The coefficients of the final model are presented in Tab. 7.

Regression results shed light into the pattern of observed price dispersion among the EU countries. Its magnitude is conditional on the level of GDP *per capita* (‘GDP\_pc’ variable), *population density* (‘pop\_density’) and *petrol price* (‘E95\_price’). The relationship between price dispersion and GDP per capita follow the expected direction: an increase in GDP *per capita* by EUR thousand is connected with an increase in deviation of the price in the given country from the median of prices in the EU by 0.14 percentage point (the estimated coefficient: 0.0013864). The influence of the GDP variable can be illustrated on the example of Poland and United Kingdom, whose GDP levels are, respectively, EUR 11 thousand and 41 thousand. Based on the model results, prices in Poland will be, due to the differences in income, on average 4.2 percentage point lower than in the United Kingdom. The other main driver of price dispersion is petrol price: a 1 EUR increase of average petrol price leads to 10.5 percentage points higher deviation from EU average prices (coefficient: 0.1049988399).

<sup>5</sup> To address the problem of collinearity we could conduct the factor analysis; however, we would lose the interpretability of the model.

<sup>6</sup> We use Breusch-Pagan test and Goldfeld-Quandt test against heteroscedasticity (Tab. A9–A10 in the Appendix), and the Durbin-Watson test for autocorrelation (Tab. A11).

Tab. 7. Results of linear regression 'percent\_deviation' variable for the

	Model 1			Model 2			Model 3 /4			Final Model		
	Estimates	Confidence Intervals	p-values	Estimates	Confidence Intervals	p-values	Estimates	Confidence Intervals	p-values	Estimates	Confidence Intervals	p-values
(Intercept)	-0.0928	-0.20 – 0.01	0.077	-0.1419	-0.24 – -0.05	0.003	-0.1397	-0.20 – -0.08	<.001	-0.1517	-0.21 – -0.09	<.001
pop_density	-0.0002	-0.00 – -0.00	<.001	-0.0002	-0.00 – -0.00	<.001	-0.0002	-0.00 – -0.00	<.001	-0.0002	-0.00 – -0.00	<.001
netearnings	-0.0064	-0.01 – -0.00	0.039	0.0003	-0.00 – 0.00	0.829						
ulab_cost	0.0035	0.00 – 0.01	0.02									
vat_stdtrate	-0.0013	-0.00 – 0.00	0.385	-0.0003	-0.00 – 0.00	0.831						
ele_cost	-0.2056	-0.39 – -0.02	0.033	-0.1069	-0.28 – 0.06	0.217						
E95_price	0.0733	-0.02 – 0.16	0.108	0.08	-0.01 – 0.17	0.079	0.087	0.03 – 0.14	0.003	0.105	0.05 – 0.16	<.001
diesel_price	0.0093	-0.10 – 0.11	0.862	0.0227	-0.08 – 0.13	0.668						
GdpR_pc_rep	0.0021	0.00 – 0.00	<.001	0.0018	0.00 – 0.00	<.001	0	0.00 – 0.00	<.001	0.0014	0.00 – 0.00	<.001
interest_rate	0.0057	0.00 – 0.01	0.019	0.0032	-0.00 – 0.01	0.139	0.0027	-0.00 – 0.01	0.156			
Observations	2551			2551			2551			2551		
R <sup>2</sup> / adj. R <sup>2</sup>	.047 / .043			.045 / .042						.043 / .042		

Source: Authors' own elaboration.

## 6 Summary and conclusions

This study applies a new way of collecting price data to measure on-line price dispersion across the EU for a broad variety of **homogeneous** commodities. We have obtained a unique snapshot picture of price dispersion based on the data collected from price comparison sites using the web-scraping technique. We have analysed the pattern of observed price dispersion with selected economic characteristics of member states to gain better insights into factors, which determine on-line prices. Our study brings two main results.

First, price dispersion in the whole product sample, measured by interquartile range of median (pre-tax and post-tax) relative price distribution, amounts to 16% for 75–25 percentile interval. While this is a rather conservative measure, the figure almost doubles (30%) for 90–10 percentile range. We note that both figures show average price dispersion for the whole product sample. The sample itself is composed of a large variety of branded items representing all main product categories. We believe that this large heterogeneity adds to the robustness of our results. The magnitudes of price dispersion differ quite a lot between product categories and subcategories, ranging from roughly 20% for electrical household appliances to 40% for clothing.

Secondly, price levels grow with GDP *per capita* and petrol price. According to the results of regression analysis, an increase in the real GDP *per capita* by one thousand EUR increases the deviation of median price by 0.14 percentage point relative to the average level for the EU. In turn, an increase of average petrol price by 1 EUR increases the price dispersion by 10.5 percentage points. We note that the observed link between *per capita* income and price dispersion might be consistently explained with at least two thorough mechanisms. The first is classical third degree price discrimination based on differences in demand elasticity induced by income levels. The second is supply side mechanism related to cost differences. Unfortunately, due to lacking data on quantities, we could not isolate both effects.<sup>7</sup> Nevertheless, we argue that it is reasonable to assume that both mechanisms are likely in force. Consequently, at least part of observed price dispersion is induced by strategic price discrimination between country markets within

the EU. This has important bearing for new policy interventions aiming to reduce price dispersion in the Digital Single Market.

While price dispersion between EU countries is unavoidable because of objective differences in taxes, wages and capital costs, as well as lack of harmonization of contract laws, the main policy question remains whether its scope is not too large and how can it be effectively reduced? Some policy interventions are now being debated within the EC, in response to public consultations on geographically based-restrictions initiated in September 2015. The new regulation will reinforce equal treatment of cross-border consumers with so called ‘shop like a local’ principle – and perhaps will go even further and impose an obligation to serve customers across the EU giving the trader the right to charge extra delivery costs. What would be the effects of such regulation?

Lowering barriers for cross border trade will lower costs of arbitrage for consumers, and result in their ability to exercise price discrimination by distributors on the European level. It is, however, unlikely that equal treatment principle would have any effect on non-strategic component of price dispersion, because on-line retailers act in competitive environments and hence, all existing differences in operational costs between countries will be passed on retail prices.

EC interventions to lowering barriers for cross-border trade such as reducing geo-blocking restrictions or liberalizing parcel delivery market would most likely decrease price dispersion due to reduction of incentives for strategic price discrimination on the wholesale level. This would eventually decrease prices in high income countries bringing positive welfare effects for customers. On the other hand, prices in lower income countries are expected to increase, because in this way, traders can reduce potential benefits from arbitrage to customers. Hence, welfare effect for customers from lower income countries will most likely be negative. Producers’ surpluses are expected to change in the opposite directions.

The major conclusion from this discussion is that if price dispersion is a result of strategic (income induced) price discrimination on producer and/or distributor level, then any policy intervention reducing trade barriers would essentially bring adverse distributional effects on customer and producer welfare in particular member states. As economic literature suggests, a-priori there is no conviction that the net effect is guaranteed to be positive (Malueg and Schwartz 1994). Therefore, we argue that any policy interventions targeting price

<sup>7</sup> Lacking sales data makes it impossible to account for market competition, which is also an important determinant for price dispersion. There is some evidence that market power of merchants from off-line channel is often transmitted to e-commerce.

dispersion in the Digital Single Market should be carried out very carefully.

## References

- [1] Adachi, Takanori, and Noriaki Matsushima. 2014. "The Welfare Effects of Third-Degree Price Discrimination in a Differentiated Oligopoly." *Economic Inquiry* 52(3): 1231–1244. <https://doi.org/10.1111/ecin.12075>.
- [2] Aguirre, Iñaki, Simon Cowan, and John Vickers. 2010. "Monopoly Price Discrimination and Demand Curvature." *The American Economic Review* 100(4): 1601–1615.
- [3] Alessandria, George, and Joseph P. Kaboski. 2007. "Pricing-to-Market and the Failure of Absolute PPP." *FRB of Philadelphia Working Paper* No. 07-29.
- [4] Allington, Nigel F., Paul Kattuman, and Florian A. Waldmann. 2004. "One Market, One Money, One Price? Price Dispersion in the European Union." *University of Cambridge Judge Institute of Management Working Paper*. <http://dx.doi.org/10.2139/ssrn.633581>.
- [5] Barry, Frank. 2003. "Economic Integration And Convergence Processes in the EU Cohesion Countries." *Journal of Common Market Studies* 41(5): 897–921. <https://doi.org/10.1111/j.1468-5965.2003.00468.x>.
- [6] Baye, Michael, J. Rupert J. Gatti, Paul Kattuman, and John Morgan. 2006. "Did the Euro foster online price competition? Evidence from an International Price Comparison Site." *Economic Inquiry* 44 (2): 265–279. <https://doi.org/10.1093/ei/cbj015>.
- [7] Bergin, Paul R., and Reuven Glick. 2007. "Global Price Dispersion: Are Prices Converging or Diverging?" *Journal of International Money and Finance* 26(5): 703–729. <https://doi.org/10.1016/j.jimonfin.2007.04.007>.
- [8] Cavallo, Alberto. 2015. "Scraped Data and Sticky Prices." *NBER Working Paper* No. 21490.
- [9] Coombs, James H., Allen H. Renear, and Steven J. DeRose. 1987. "Markup Systems and the Future of Scholarly Text Processing." *Communications of the ACM* 30 (11): 933–947.
- [10] Corts, Kenneth S. 1998. "Third-Degree Price Discrimination in Oligopoly: All-out Competition and Strategic Commitment." *The RAND Journal of Economics* 29 (2): 306–323.
- [11] Cowan, Simon. 2008. "When Does Third-Degree Price Discrimination Reduce Social Welfare, and When Does It Raise It?" *Department of Economics, University of Oxford Working Papers*.
- [12] Crucini, Mario J., and Hakan Yilmazkuday. 2014. "Understanding Long-Run Price Dispersion." *Journal of Monetary Economics* 66: 226–240. <https://doi.org/10.1016/j.jmoneco.2014.04.012>.
- [13] DG Internal Market. 2002. *Price Differences For Supermarket Goods in Europe Internal Working Document of the Internal Market Directorate General*. [http://ec.europa.eu/internal\\_market/economic-reports/docs/2002-05-price\\_en.pdf](http://ec.europa.eu/internal_market/economic-reports/docs/2002-05-price_en.pdf) (accessed 28.01.2018).
- [14] European Commission. 2015. *A Digital Single Market Strategy for Europe – Analysis and Evidence*. <https://ec.europa.eu/digital-single-market/en/news/digital-single-market-strategy-europe-analysis-and-evidence-swd2015-100-final> (accessed 28.01.2018).
- [15] Eurostat. 2016. *Comparative Price Levels of Consumer Goods and Services*. [http://ec.europa.eu/eurostat/statistics-explained/index.php/Comparative\\_price\\_levels\\_of\\_consumer\\_goods\\_and\\_services](http://ec.europa.eu/eurostat/statistics-explained/index.php/Comparative_price_levels_of_consumer_goods_and_services) (accessed 28.01.2018).
- [16] Faber, Reimer P., and AD C.J. Stokman. 2009. "A Short History of Price Level Convergence in Europe." *Journal of Money, Credit and Banking* 41 (2–3): 461–477.
- [17] Fischer, Christoph. 2012. "Price Convergence in the EMU? Evidence from Micro Data." *European Economic Review* 56(4): 757–776. <https://doi.org/10.1016/j.euroecorev.2012.01.008>
- [18] Galera, Francisco, Markus Kinaterder, and Pedro Mendi. 2014. "The Cost Effect in Third-Degree Price Discrimination." *Journal of Economics* 113(2): 175–186. <https://doi.org/10.1007/s00712-013-0368-y>.
- [19] Goldberg, Pinelopi, and Frank Verboven. 2005. "Market Integration and Convergence to the Law of One Price: Evidence from the European Car Market." *Journal of International Economics* 65(1): 49–73.
- [20] Gorodnichenko, Yuriy, Viacheslav Sheremirov, and Oleksandr Talavera. 2014. "Price Setting in Online Markets: Does It Click?" *NBER Working Paper* No. 20819.
- [21] Gorodnichenko, Yuriy, and Oleksandr Talavera. 2014. "Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration." *NBER Working Paper* No. 20406.
- [22] Halmai, Péter, and Viktória Vásáry, V. 2010. "Real Convergence in the New Member States of the European Union (Shorter and Longer Term Prospects)." *The European Journal of Comparative Economics* 7(1): 229–253.
- [23] Kim, Sukkoo. 1998. "Economic Integration and Convergence: US Regions, 1840–1987." *The Journal of Economic History*, 58(03): 659–683.
- [24] Lünemann, Patrick, and Ladislav Wintr. 2011. "Price Stickiness in the US and Europe Revisited: Evidence from Internet Prices\*." *Oxford Bulletin of Economics and Statistics* 73(5): 593–621. <https://doi.org/10.1111/j.1468-0084.2011.00657.x>.
- [25] Malueg, David A., and Marius Schwartz. 1994. "Parallel Imports, Demand Dispersion, and International Price Discrimination." *Journal of International Economics* 37(3): 167–195. [https://doi.org/10.1016/0022-1996\(94\)90044-2](https://doi.org/10.1016/0022-1996(94)90044-2)
- [26] Massmann, Michael, and James Mitchell. 2003. "Reconsidering the Evidence: Are Eurozone Business Cycles Converging." *ZEI Working Papers* B 05-2003.
- [27] Ogrokhina, Olena. 2015. "Market Integration and Price Convergence in the European Union." *Journal of International Money and Finance* 56: 55–74. <https://doi.org/10.1016/j.jimonfin.2015.04.004>.
- [28] Parsley, David, and Shang-Jin Wei. 2008. "In Search of a Euro Effect: Big Lessons from a Big Mac Meal?" *Journal of International Money and Finance* 27(2): 260–276. <https://doi.org/10.1016/j.jimonfin.2007.12.008>.
- [29] Rogers, John H. 2007. "Monetary Union, Price Level Convergence, and Inflation: How Close Is Europe to the

- USA?" *Journal of Monetary Economics* 54(3): 785–796. <https://doi.org/10.1016/j.jmoneco.2006.01.004>.
- [30] Simonovska, Ina. 2015. "Income Differences and Prices of Tradables: Insights from an Online Retailer." *The Review of Economic Studies* 82(4): 1612–1656. <https://doi.org/10.1093/restud/rdv015>.
- [31] Stole, Lars A. 2003. "Price Discrimination and Competition." In: *Handbook of Industrial Organization* 3, eds. Mark Armstrong and Robert Porter, 2221–2299. Amsterdam: Elsevier.
- [32] Verboven, F. 1996. "International Price Discrimination in the European Car Market." *The RAND Journal of Economics* 27 (2): 240–268.
- [33] Wang, Haiying, Xinyu Zhang, and Guohua Zou, G. 2009. "Frequentist Model Averaging Estimation: A Review." *Journal of Systems Science and Complexity* 22(4), 732–748.
- [34] Wolszczak-Derlacz, Joanna. 2010. "Does One Currency Mean One Price? An Analysis of the Euro Effect on Price Dispersion and Convergence." *Eastern European Economics* 48(2): 87–114.

## Appendix

Note: all figures and tables prepared by the Authors.

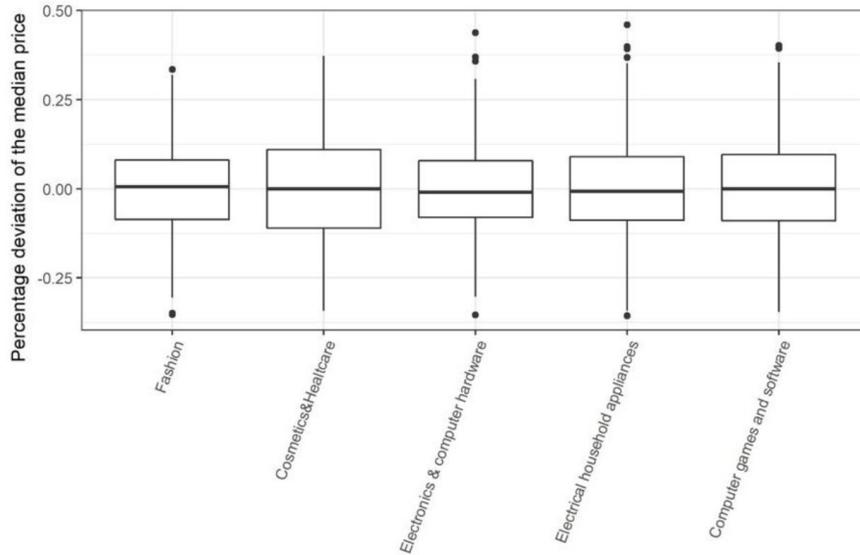


Fig. A1. Boxplots - Percentage deviations of median prices – Product Category Variables

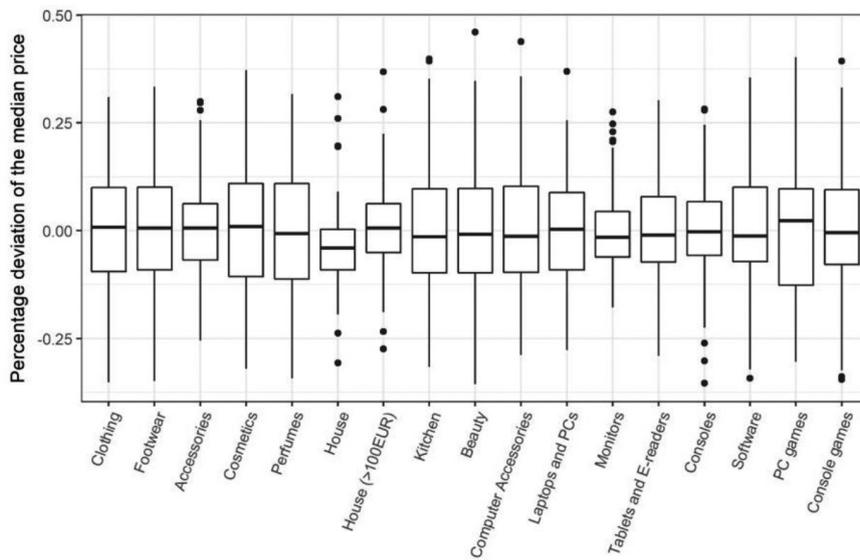


Fig. A2. Boxplots - Percentage deviations of median prices – Product Subcategory Variables

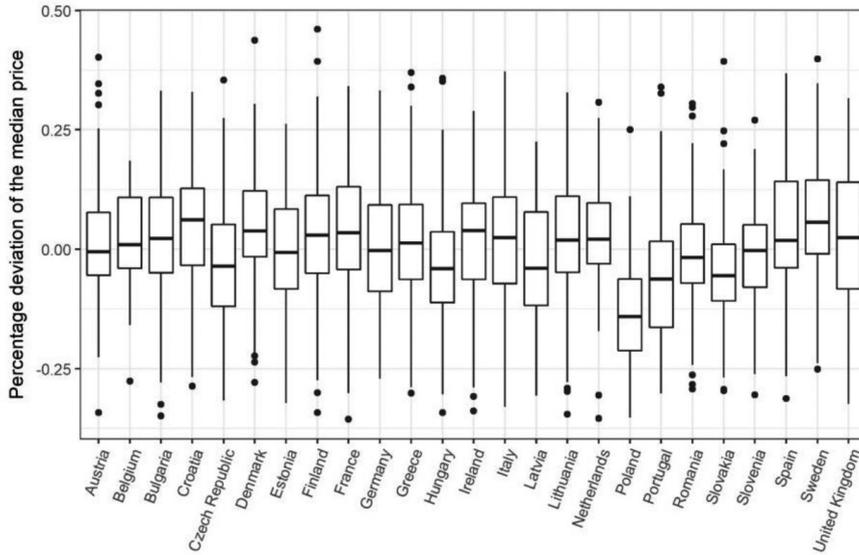


Fig. A3. Boxplots – Percentage deviations of median prices – Country Variables

Tab. A1. Result of the analysis of variance model – Product Category Variables

	<b>Df</b>	<b>Sum Sq</b>	<b>Mean Sq</b>	<b>F value</b>	<b>Pr(&gt; F)</b>
<b>Category</b>	4	0.00	0.00	0.00	1.0000
<b>Residuals</b>	2546	46.95	0.02		

Tab. A2. Result of analysis of variance model – Product Subcategory Variables

	<b>Df</b>	<b>Sum Sq</b>	<b>Mean Sq</b>	<b>F value</b>	<b>Pr(&gt; F)</b>
<b>Subcategory</b>	16	0.04	0.00	0.15	1.0000
<b>Residuals</b>	2534	46.91	0.02		

Tab. A3. Result of analysis of variance model – Country Variables

	<b>Df</b>	<b>Sum Sq</b>	<b>Mean Sq</b>	<b>F value</b>	<b>Pr(&gt; F)</b>
<b>Country</b>	24	4.84	0.20	12.09	0.0000
<b>Residuals</b>	2526	42.11	0.02		

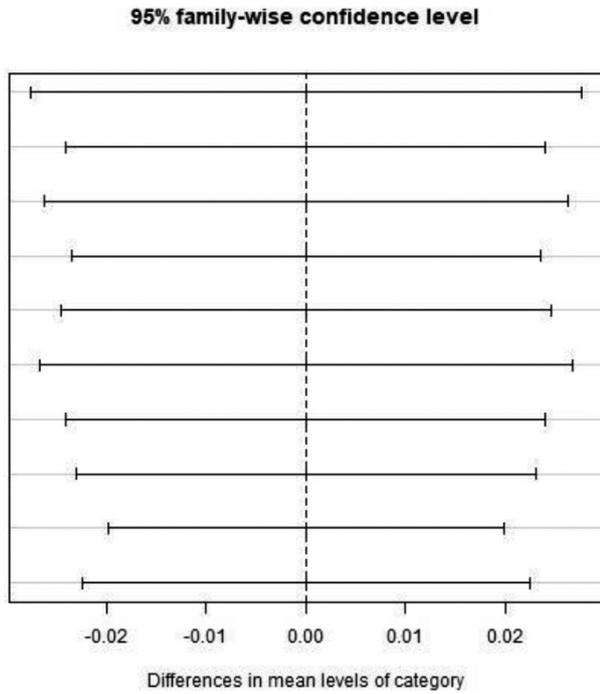


Fig. A4. Tukey's range test – Product Category Variables

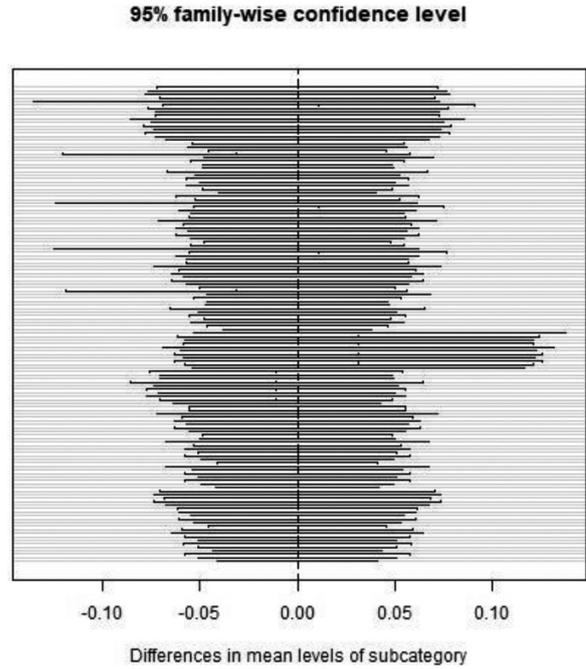


Fig. A5. Tukey's range test – Product Subcategory Variables

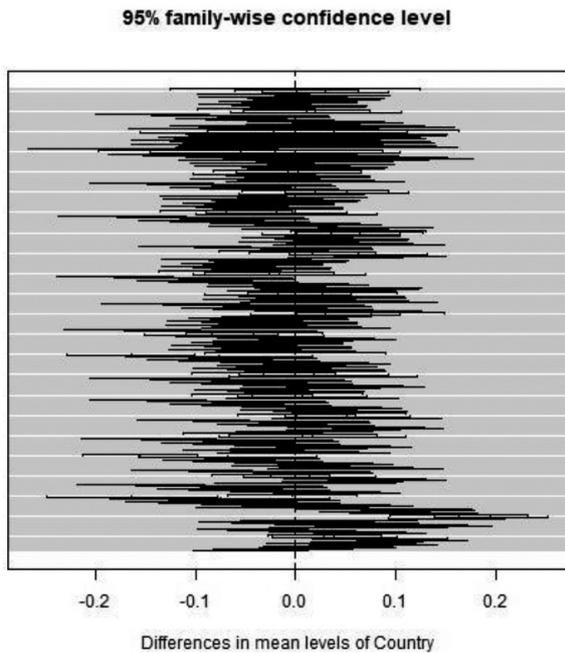


Fig. A6. Tukey's range test – Country Variables

Tab. A4. Result of Levene's Test for Homogeneity of Variance – Product Category Variables

	<b>Df</b>	<b>F value</b>	<b>Pr(&gt; F)</b>
<b>group</b>	4	10.53	0.0000
	2546		

Tab. A5. Result of Levene's Test for Homogeneity of Variance – Product Subcategory Variables

	<b>Df</b>	<b>F value</b>	<b>Pr(&gt; F)</b>
<b>group</b>	16	5.50	0.0000
	2534		

Tab. A6. Result of Levene's Test for Homogeneity of Variance – Country Variables

	<b>Df</b>	<b>F value</b>	<b>Pr(&gt; F)</b>
<b>group</b>	24	2.23	0.0005
	2526		

Tab. A7. Results of the VIF Test

variable	Model 1	Model 2	Model 3	Model 4
diesel_price	4.84	4.78		
E95_price	5.58	5.56	2.25	2.25
ele_cost	4.02	3.24		
GdpR_pc_rep	7.70	7.29	2.23	2.23
interest_rate	2.43	1.97	1.50	1.50
netearnings	38.90	5.78		
pop_density	1.55	1.55	1.35	1.35
ulab_cost	40.29			
vat_stdrate	1.56	1.44		

Tab. A8. Frequentist model averaging – relative variable importance

	Importance:	N containing models
GdpR_pc_rep	1	128
pop_density	1	128
E95_price	0.92	128
interest_rate	0.55	128
ele_cost	0.45	128
diesel_price	0.34	128
vat_stdrate	0.29	128
netearnings	0.28	128

Tab. A9. The results of the Breusch–Pagan test

---

**Studentized Breusch-Pagan test**  
**data: fit4**  
**BP = 1.1911, df = 3, p-value = 0.7551**

---

Tab. A10. The results of the Goldfeld–Quandt test

---

**Goldfeld-Quandt test**  
**data: fit4**  
**GQ = 0.92688, df1 = 1272, df2 = 1271, p-value = 0.912**

---

Tab. A11. The results of the Durbin–Watson test

---

**Durbin-Watson test**  
**data: fit4**  
**DW = 1.6737, p-value < 0.00000000000000022**

---