Federal Statistical Office of Germany 03.05.2017

Christian Blaudow, Florian Burg

15th Meeting of the Ottawa Group on Price Indices

Dynamic pricing as a challenge for Consumer Price Statistics

Abstract

The online market is increasingly gaining in importance. Consumers buy more and more goods on the online market due to the great variety of product offers and time saving. For the German Consumer Price Statistics, which comprises the National Consumer Price Index (CPI) and the Harmonised Index of Consumer Prices (HICP), the Federal Statistical Office (FSO) collects approximately 10,000 prices for products on websites of online retailers. The share of these products on the overall basket of goods and services amounts to approximately five per cent and will probably be rising in the forthcoming years. Furthermore, prices for services are collected online as well.

Thanks to fairly easy to adjust prices on the internet, shops are able to react to market conditions or consumer's behaviour by adjusting prices automatically in short intervals, applying algorithms that take into account different parameters. This phenomenon is known as dynamic pricing. First studies investigating dynamic pricing in Germany have shown that different variants of dynamic pricing exist and are very heterogeneous and not transparent. Dynamic pricing of online shops may lead to a bias in the index calculation since the traditional way of price collection via internet is done generally at one time during the month due to limited resources and therefore cannot capture rapidly changing prices.

For this reason, the FSO conducted a study concerning the incidence of dynamic pricing and used the technique of web scraping to collect online prices in short intervals. With the help of modern IT tools, automated price collections may be initiated at any time of the day, week or month with infinite repetitions. Therefore, a survey of roughly 2,700 products, distributed among 14 online shops, was compiled and their prices were collected hourly and seven days a week for nearly three months. Findings of the study relate to the frequency of and mechanism behind price changes of every single online shop, as well as main targets for dynamic pricing in terms of product groups. One main implication may be an increased use of web scraping techniques for the monthly price collection on websites for certain shops. A further solution would be the use of transaction data (scanner data) to capture all dynamic and individualized prices on the market.

Keywords: dynamic pricing, individualized pricing, internet purchases, web scraping

Introduction

What is dynamic pricing?

In general, dynamic pricing describes the application of automatic algorithms in order to change prices in short intervals due to market conditions or due to parameters indicating consumer's willingness to pay.

The phenomenon that prices change rapidly depending on market conditions is of course not new. Prior to the digital era this was limited to goods and services, for which the way of publishing the prices allowed to change the prices often and without significant costs (menu costs). For fuels, for instance, prices use to change several times a day due to electronic price signs at the fuel stations. Other services, for which prices use to change in relatively short intervals, are some transport and travel services like flights, package holidays and rental cars. For those services the price strongly depends on the time of booking as well as on the time the service is provided. Prior to the digital era, booking was mainly done through travel agencies, nowadays to a large extent via internet. This allows for immediate price adjustment due to factors like capacity and calendar effects. To take into account these price changes, the price collection for the German Consumer Price Statistics takes place not only once a month like for normal elementary product groups, but also at several times during the month according to pricing patterns that are usually known to price collectors and index compilers respectively. But of course this approach is time consuming and restricted by staff resources and difficult due to changes of pricing patterns. Therefore, for some selected goods prices are already collected with web scraping techniques.

Besides those special categories of goods and services, dynamic pricing seems to become of growing importance.[1] The phenomenon of dynamic pricing is even present in the media and is target of current discussions. Offering products on websites allows for immediate price changes to react to changes of market conditions. In fact, there has recently evolved a market for tools to optimize the prices in online shops. Major providers for example employ own specialists for optimizing their online pricing policy. On the other hand, there are tools which observe product prices in online shops and inform consumers when the price of a certain product falls below a certain pre-defined threshold. These developments indicate that the phenomenon of dynamic pricing plays an increasing role for online shops and for consumers.

Looking at this phenomenon from the perspective of price statisticians, dynamic pricing makes it more difficult to measure the price development of products in online shops and – as the online market has become more and more important in the last years – leads to new challenges when compiling consumer price indices. But actually we do not know how big the problem really is. One of the fundamental questions dealt in this paper concerns the extent of dynamic pricing. How large and how common is this problem nowadays? Therefore, the goal of the present study was to investigate the frequency and level of price changes of products relevant for the German CPI/HICP.

---

[1] Several articles in German magazines have recently picked up the topic of dynamic pricing. See for example: A. Jung, "Die Preis-Frage", in: *Der Spiegel*, No. 9 2017, pp.76-77 and M. Fischer et al., „Der Preis ist heiss", in: *Wirtschaftswoche*, No.10 2017, pp.18-24.

Dynamic versus individualized pricing

Dynamic pricing is often used as general term for price differences resulting from the application of algorithms, mainly observed in online shops. More precisely, there is a difference between dynamic pricing as referred to price changes in time and individualized pricing as referred to different prices offered to different consumers. In particular, individualised pricing describes the phenomenon of trying to charge every customer the individual price he/she is willing to pay for a certain good based on the (individual) value he/she places on the product.

However, this paper deals solely with dynamic pricing. Only the time dimension of price changes is analysed, individualised pricing - different prices for different consumers (price differentiation) - is not subject of this study.

Price collection via internet for the German CPI/HICP

The basket of goods and services for the German Consumer Price Statistics contains at the upper level approximately 600 published types of goods and services (elementary product group). For the bulk of these elementary product groups, more than 300,000 single prices are observed monthly within the scope of the traditional price collection at physical shops and service providers. The traditional price collection is usually conducted at one point during the month, i.e. for every single product in the respective elementary product group usually one price is used for the calculation of the average prices which are further used for the calculation of sub-indices up to the CPI/HICP overall. In addition to this traditional price collection, there are price collections for certain goods and especially services that require a more sophisticated survey design (for example flights or package holidays).

For online shops, the price collection is conducted centrally by the FSO for efficiency reasons. For the majority of goods, the price collection is conducted manually at one point in time during the month. The centrally collected prices are used to calculate elementary indices for online shops and subsequently provided to the German federal states for the further calculation of their CPIs.

For the German Consumer Price Statistics, an explicit weighting scheme for shop types on the level of goods and services and federal states was introduced in 2008. The weights for shop types used in the German Consumer Price Statistics allows for the calculation of the share of online shops as one of the shop type categories. Since the market share of online shops has increased in the last years, the number of elementary product groups relevant in online shops accounts to approximately 40% of all elementary product groups. Overall, the weight of online shops in the German CPI/HICP currently accounts to approximately 5%. This weight is expected to increase in the forthcoming years. Online prices are taken into account for almost all categories of goods and services. The number of prices of goods collected at online shops accounts to approximately 10,000. This number does not include prices for services. For some services, the price collection also takes place via internet. These numbers show that online prices have an increasing importance for the German CPI/HICP. But it has to be mentioned also that the much larger amount of prices still refers to goods and services not offered - or at least not offered exclusively - on the internet and therefore are not subject to dynamic pricing.

Set-up of the study

Sample design

The sample of products for the present study consists exclusively of products offered by online shops which are observed for the regular monthly price collection for CPI/HICP. Therefore, every product of the sample is assigned to a COICOP-10-digit[2] and the sample on the whole consists of products (only goods, no services) out of 242 COICOPs. At the beginning the sample started with 3,050 products, distributed among 15 online shops. Especially online shops with a certain amount of price observations and known to be important with respect to their market shares were selected for the study. Due to cancellation and non-availability which is explained below in the chapter "Results", the analysis only takes 2,680 products and 14 shops into account. The prices were collected via web scraping. In order to access the product's offer page directly, it was essential to find out the URLs of the respective products.

When taking a closer look at the products and their COICOP number, one may divide the products into products groups (see table 1). The bulk of observed products belong to the product groups of clothing and shoes (031 and 032), household accessories (051 to 056), recreation and culture (091 and 093) and other goods and services (121 and 123).

COICOP product group	Name	Number of Products
011	food products	11
012	nonalcoholic beverages	4
031	clothing	705
032	shoes	119
043	maintenance and repair of the dwelling	25
051	furniture, lights, carpets and other floor coverings	213
052	home textiles	132
053	household accessories	309
054	glassware, tableware and household utensils	175
055	tools and equipment for house and garden	180
056	goods and services for routine household maintenance	4
061	medical products, appliances and equipment	30
071	vehicles	8
082	telephones and mobile communication equipment	25
091	IT and audio/video devices	327
092	other major durables for recreation and culture	11
093	other items for garden, pets and leisure time	123
095	print products and stationery	33
121	personal care	140
123	personal effects	106
Total		**2,680**

Table 1: Number of products distributed among 3-digit COICOP product groups

[2] COICOP: Classification of Individual Consumption of Purpose. COICOP-10-digit is the lowest level of the classification used for the German Consumer Price Statistics and represents the elementary product groups for the price collection for CPI/HICP.

Technical instruments and procedure of the program

A relational database is used for the input and the retrieved output data. The main advantages of a relational database lie in the efficient management of the data with no redundancies, the relative small size of the input tables, high practicability and good performance even with some millions of records. The sample's product input data including product name, shop, COICOP number, COICOP name, unique identifier, product URL and product's article number was uploaded into an input table of the used relational database. Another input table was created which includes the XPaths of the offer pages of the respective online shops. Therein, XPaths for the positions of the product name, normal price, special price, article number and availability were stored only once per shop. Earlier studies have shown that the use of XPaths instead of using HTML-elements for the extraction of information is the most stable solution. Usually, XPaths include a unique path or an ID for the respective information. Therefore, the exact (visual) position of the information is irrelevant since an ID is unique on every page and the XPath searches for the information related to the ID in its path.

As shown in the graph below, the procedure of the automated price collection is realized in Java. Java is a programming language which is basically used in the FSO and which is compatible with numerous scraping software and other IT tools. The programs are able to search for the relevant information in input databases and sends tasks to the scraping tool. In the present case, a Java program retrieves the necessary information from both input tables and creates combined lists. It transforms the information and equips the actual scraping programs with it. The actual scraping is conducted by the tool Selenium that controls a web browser. This tool Selenium may be installed as a plug-in for a common internet browser and can then be used for recording and replaying procedures with internet web browsers. Additionally, Selenium offers programming interfaces in various languages including Java. Then, the execution of the scraping processes as well as finding the relevant positions of information is done by the browser. Selenium only functions as a provider of information and an interface between Java and the browser. The usage of an internet browser for retrieving a web page is very efficient and flawless and has additional advantages. For instance, when using a common internet browser, it is possible to disable scripts on a page and to start and shut down a browser window within a short amount of time. Additionally, browser windows operate completely independent from each other. Therefore, it is easily possible to run several browser windows for a price collection simultaneously. These advantages lead to a huge increase in velocity of the price collection.

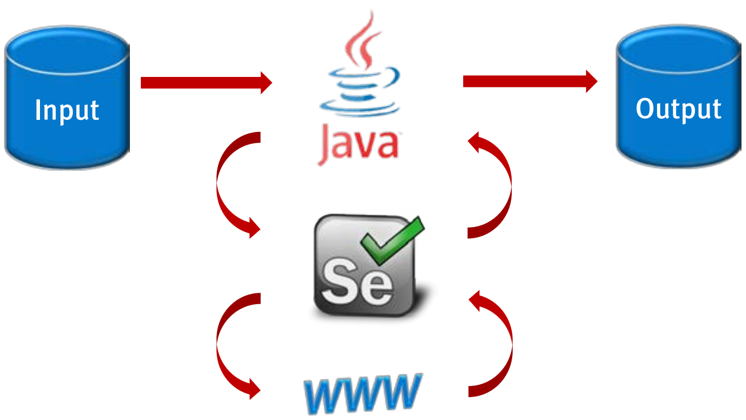

Chart 1: Technical Set-up

In particular, the Java program gives the URL of a product via Selenium to the browser and Selenium finds with the help of the XPaths the position of the product name, article number and prices. This retrieved information is send back to the Java program and the program, if required, cuts the scraped information (so called “strings”) and stores it in an output table of the database for subsequent use by the section of Consumer Price Statistics.

For the actual scraping of a page, the Java program first divides about 3,050 products into 12 groups. One window for price collection is opened by the program for each group. This means that 12 price collections run simultaneously during the whole price collection in one hour. The number of groups is easily adjustable; it only depends on the performance of the computer and its memory. Overall, with 12 jobs running simultaneously, the price collection of 3,050 products takes not more than 40 minutes on a business notebook. A certain time in between the collections is advisable since the developers perhaps need some time to adjust the program or adjust data in the database, e.g. XPaths or URLs. Additionally, in very few cases, a product page cannot be found by the program (due to a network issue or more likely a change on the web shop site). In such cases, the program tries to find the product up to 3 times before it gives up and continues with the next product in the list. In between the retrievals, the program includes an adjustable short break of about 5 seconds before retrying. Therefore, as temporarily more and more pages are not accessible in the whole period under consideration, the whole program took nearly 50 minutes in the end.

The scripts on every web page of a shop had to be adjusted separately. There are scripts which complicate the extraction of information but in some cases also give additional or even necessary information. Generally, the developers of the program tried to disable (forbid) as many scripts as possible to ensure that the process of price extraction runs without any delay due to page loading. The extraction of the information about the availability of a product is very crucial after extracting the price of a product. There are cases, in which several online shops show the price of a product although the product is actually not available at the moment of the extraction. Naturally, the price found on the product page without product availability may not be considered for investigating dynamic pricing of the shop.

Targets and limits of investigation

The target of the present study was to investigate the frequency of price changes of products relevant for the German CPI/HICP. An hourly rhythm of automatic price collection was used to obtain as many price observations as possible and consequently a detailed picture of price developments in online shops. In this way, this study was able to investigate the consequences of daily and hourly price changes. Due to the dependency on the internet connection and occasional delays in the price collection of individual browser windows, the extraction of a product’s price may not exactly be 60 minutes apart from each other. But this investigation assumes that all prices are equally collected and consequently only change up to once per hour.

The project used steadily changing IP addresses. Therefore, the influence of individual consumption behavior on the price of the product could not be observed. The prices shown on the page are definitely the prices of the shops for every customer and are not individualized prices. Additionally, this study is not able to investigate the influence of different electronic devices for the price collection since the prices were collected with the same electronic device.

Results

Preliminary remarks concerning analysis (treatment of data gaps)

The automated price collection for the present study started on December 9th 2016 and hourly scraped prices were taken into account until March 6th 2017 in order to have data for almost three months. The program of automated and hourly online price collection will continue to scrape prices and will allow for longer time series to be analyzed for future studies.

Due to the large number of products and the long time period of data collection, there are some gaps in the series of prices. The vast majority of these gaps were caused by technical problems. Therefore, the following post-processing was done:

- As long as data gaps were assessed to be temporarily (i.e. a price was scraped again after a certain time), imputation was done to have complete series of prices. In most cases, an unsuccessful extraction of a product's price was caused by a temporary shutdown of the shop's page due to maintenance. In such cases, the last price found for the product was used to fill the gaps. Several shops use maintenance for a general price change of their products. Therefore, when looking at the time of price changes of every single shop, there is often a cluster of price changes during the night.

- Only price series with at least 50% of the possible price observations were taken into account. Products which had not been observed for at least half of the investigation period were deleted prior to the analysis.

- A manual check was done for products with either price increases of 500% or more than 500% as well as for products which experienced excessive price decreases. In most of these cases, replacements occurred and, in several cases, the article number of the old product was used for the new product. This also holds true for products which were removed by the shop and customers got redirected to the page of a similar product. These series were not taken into account in the analysis since the products were obviously not identical and only similar from the shop's perspective.

- One online shop closed approximately two month after starting the automated price collection. Since this shop is no longer relevant for the price collection of the CPI, the prices scraped from this shop were not considered in the analysis.

After this data cleansing 2,680 price series for products in 14 online shops remained for the analysis. The results presented in this chapter data were anonymized by not indicating the name of the online shops. This anonymization is not crucial for the analysis. For internal purposes of course it is very important to know in which shops dynamic pricing is conducted to introduce more frequent survey intervals and additional checks at the end of the month.

Number of price changes during the observed period

The present section deals with the number of price changes per product during the whole observation period, split into 4 categories. From the perspective of Consumer Price Statistics, it is valuable to know how many products change their prices in a way that the manual price collection is not able to capture. The following chart shows the results in the most comprehensive way:

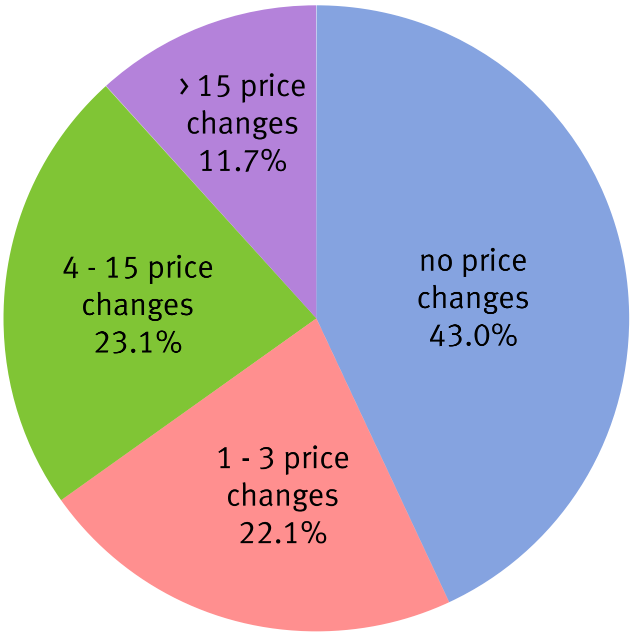

Chart 2: Share of products in the categories of price changes in the observed period

The first two categories - with up to 3 price changes in 3 month - were chosen due to the fact that up to 3 price changes in 3 month (on average 1 change per month) should usually be captured by the manual price collection conducted at one time during the month. If the price changes on average more than once per month, the manual price collection could perhaps lead to a bias in the index calculation since the price collector does not know whether this one price is representative. The category of products with 4 to 15 price changes sums up to nearly a quarter. These price changes are currently captured by the manual price collection only to a minor extent since price collections in general are only conducted once per month. With a price collection once per month there is a growing risk, that the observed price is not representative for the average price of that product in that month. A relatively small share of products in the sample (11.7%) changed its prices more than 15 times during the whole observation period. This means more than 5 price changes per month. In general, the higher the number of price changes per product per month, the more unlikely it is that the manual price collection is able to observe a representative price for the respective month. But moreover it is also important to consider the variation of the respective price changes.

Comparing the shops with respect to the number of price changes per products (table 2) shows that the share of products with 0 to 3 price changes is significantly higher for the majority of online shops (shops 2, 3, 5, 7, 8, 10, 11, 12, 13) than for the remaining shops. Therefore the results indicate that dynamic pricing is only applied broadly by some online shops whereas for other online shops price changes seem to be in a normal range and not caused by special techniques of dynamic pricing.

Online Shop	Number of price series with ... price changes								
	0		1 - 3		4- 15		> 15		all
	no.	in %	no.	in %	no.	in %	no.	in %	
Shop 1	17	7.91%	20	9.30%	76	35.35%	102	47.44%	**215**
Shop 2	268	47.94%	109	19.50%	114	20.39%	68	12.16%	**559**
Shop 3	105	50.72%	54	26.09%	44	21.26%	4	1.93%	**207**
Shop 4	15	20.55%	13	17.81%	34	46.58%	11	15.07%	**73**
Shop 5	19	32.76%	37	63.79%	2	3.45%	0	0.00%	**58**
Shop 6	21	25.61%	21	25.61%	30	36.59%	10	12.20%	**82**
Shop 7	483	54.39%	174	19.59%	145	16.33%	86	9.68%	**888**
Shop 8	21	84.00%	4	16.00%	0	0.00%	0	0.00%	**25**
Shop 9	6	4.88%	36	29.27%	60	48.78%	21	17.07%	**123**
Shop 10	23	31.51%	19	26.03%	31	42.47%	0	0.00%	**73**
Shop 11	16	29.63%	19	35.19%	15	27.78%	4	7.41%	**54**
Shop 12	59	79.73%	11	14.86%	1	1.35%	3	4.05%	**74**
Shop 13	75	68.81%	33	30.28%	1	0.92%	0	0.00%	**109**
Shop 14	25	17.86%	43	30.71%	67	47.86%	5	3.57%	**140**
all	**1,153**	**43.02%**	**593**	**22.13%**	**620**	**23.13%**	**314**	**11.72%**	**2,680**

Table 2: Price series per shop divided by number of price changes in the observed period

Comparing the products divided by product categories using the COICOP, no clear patterns of price changes are obvious. There are no product categories for which price changes are observed more often than for others. However, this analysis is limited since it was done at an upper level (COICOP-3-digit) due to the number of products observed. For a more detailed analysis the number of products is too low and too unevenly distributed among the product categories.

Product category	Number of price series with … price changes								
	0		1 - 3		4- 15		> 15		all
	no.	in %	no.	in %	no.	in %	no.	in %	
011	9	81.82%	2	18.18%	0	0.00%	0	0.00%	11
012	0	0.00%	2	50.00%	2	50.00%	0	0.00%	4
031	313	44.40%	175	24.82%	150	21.28%	67	9.50%	705
032	47	39.50%	33	27.73%	30	25.21%	9	7.56%	119
043	18	72.00%	4	16.00%	3	12.00%	0	0.00%	25
051	136	63.85%	28	13.15%	31	14.55%	18	8.45%	213
052	42	31.82%	20	15.15%	24	18.18%	46	34.85%	132
053	95	30.74%	61	19.74%	117	37.86%	36	11.65%	309
054	83	47.43%	27	15.43%	37	21.14%	28	16.00%	175
055	100	55.56%	57	31.67%	12	6.67%	11	6.11%	180
056	0	0.00%	1	25.00%	1	25.00%	2	50.00%	4
061	11	36.67%	7	23.33%	10	33.33%	2	6.67%	30
071	6	75.00%	0	0.00%	2	25.00%	0	0.00%	8
082	9	36.00%	6	24.00%	5	20.00%	5	20.00%	25
091	87	26.61%	77	23.55%	107	32.72%	56	17.13%	327
092	8	72.73%	3	27.27%	0	0.00%	0	0.00%	11
093	76	61.79%	31	25.20%	15	12.20%	1	0.81%	123
095	6	18.18%	4	12.12%	13	39.39%	10	30.30%	33
121	50	35.71%	28	20.00%	47	33.57%	15	10.71%	140
123	57	53.77%	27	25.47%	14	13.21%	8	7.55%	106
Total	**1,153**	**43.02%**	**593**	**22.13%**	**620**	**23.13%**	**314**	**11.72%**	**2,680**

Table 3: Price series per product category divided by number of price changes in the observed period

Volatility of prices

Considering cases in which the price changed more than 3 times during the observed period (more than once per month on average), it is worth looking at the volatility of the prices (see table 4). In order to measure the volatility of prices, the variation coefficients of the single price series were calculated. The variation coefficient as dispersion measures was chosen, because it is a relative measure that does not depend on the absolute level of prices. This analysis has an important impact on the assessment of the traditional way of price collection. When the volatility of prices is low, dynamic pricing might not be a major problem for measuring inflation. For example, in shop 1 for over 80% of the observed products the price changed more than once per month on average (for almost 50% more than 5 times per month on average), but the variation coefficient for more than 80% was below 0.1 and for 55% even below 0.05. This indicates that dynamic pricing is applied in this shop, but the volatility of prices is in a range that could be captured by an extension of the dates of traditional price collection. In case of a low variation coefficient, it is unlikely that a bias in the index calculation occurs since it is unlikely that one observed price is an extreme price for this product. For instance, for one of the observed product the price was changed 1,304 times during the 3 months with a variation coefficient of only 0.02. Of course there are also cases, in which the prices of products changed nearly every hour with a high variation coefficient. Two examples are shown in the appendix (chart 5 and chart 6). Considering all shops with more than 3 price changes in the 3 month of the study, 66% of the variation coefficients are below 0.1 and 42% are even below 0.05.

There are few shops in which a relatively high number of price changes was observed with relatively high variation coefficients. In shop 14, for instance, prices were changed frequently (table 2) with a relatively high variation coefficient (table 4).

Online Shop	Variation coefficient			
	‹ 0.05	0.05 - ‹ 0.1	0.1 – 0.25	› 0.25
Shop 1	99	46	27	6
Shop 2	67	44	52	19
Shop 3	26	19	2	1
Shop 4	38	7	0	0
Shop 5	0	0	0	2
Shop 6	24	13	3	0
Shop 7	75	40	67	49
Shop 8	0	0	0	0
Shop 9	47	24	9	1
Shop 10	2	18	11	0
Shop 11	11	5	3	0
Shop 12	0	1	1	2
Shop 13	0	0	1	0
Shop 14	5	6	57	4
all	394	223	233	84

Table 4: Price series with more than 3 changes in the observed period per shop divided by level of variation coefficients

<u>Time of price changes</u>

Looking at the time when price changes occurred, there is a clear pattern of conducting changes at the first third of the day, especially between midnight and 1 am. This is probably caused by technical reasons: Firstly, it is known that some (more simple) tools used for dynamic pricing allow price changes only once a day. Another reason might be that online shops try to change prices at a time when there are less consumers shopping in online shops. At times when presumably most of the consumers are shopping (afternoon and evening) it seems clear that online shops try to avoid price changes as this is expected to annoy and discourage consumers.

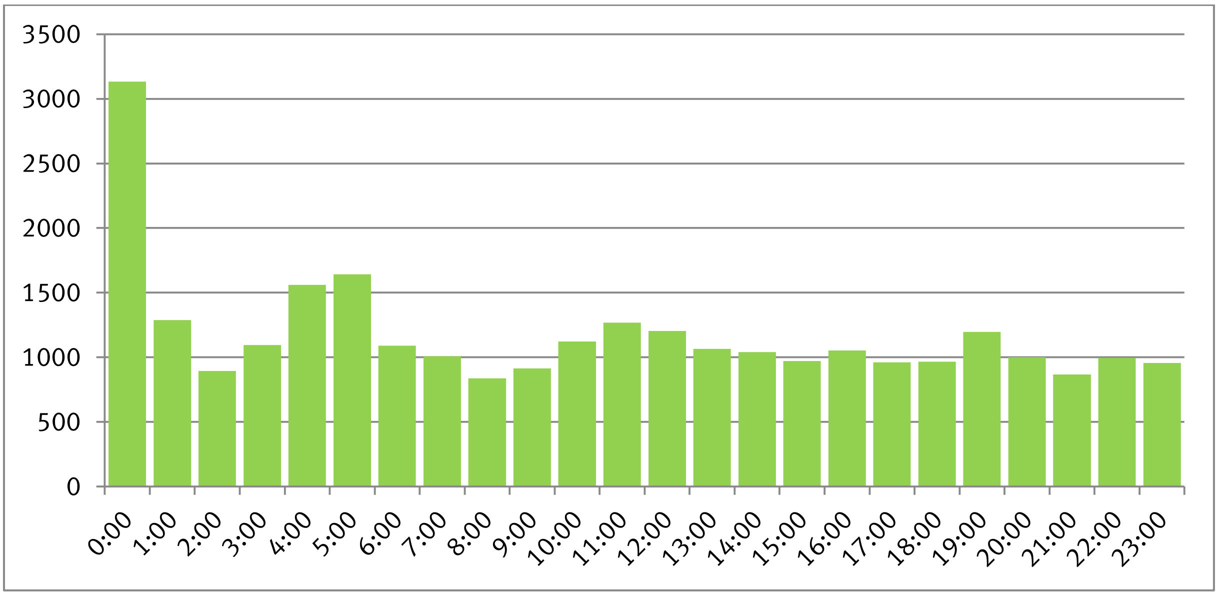

Chart 3: Price changes per hour

Calendar effects

In order to identify the existence of calendar effects, the number of price changes was related to the day they were observed (see chart 4). There is a pattern with respect to the day of price changes. Thursday is the day with the largest amount of price changes whereas Monday and Sunday are the days with the smallest amount of price changes. For a more profound analysis the period of the study is too short.

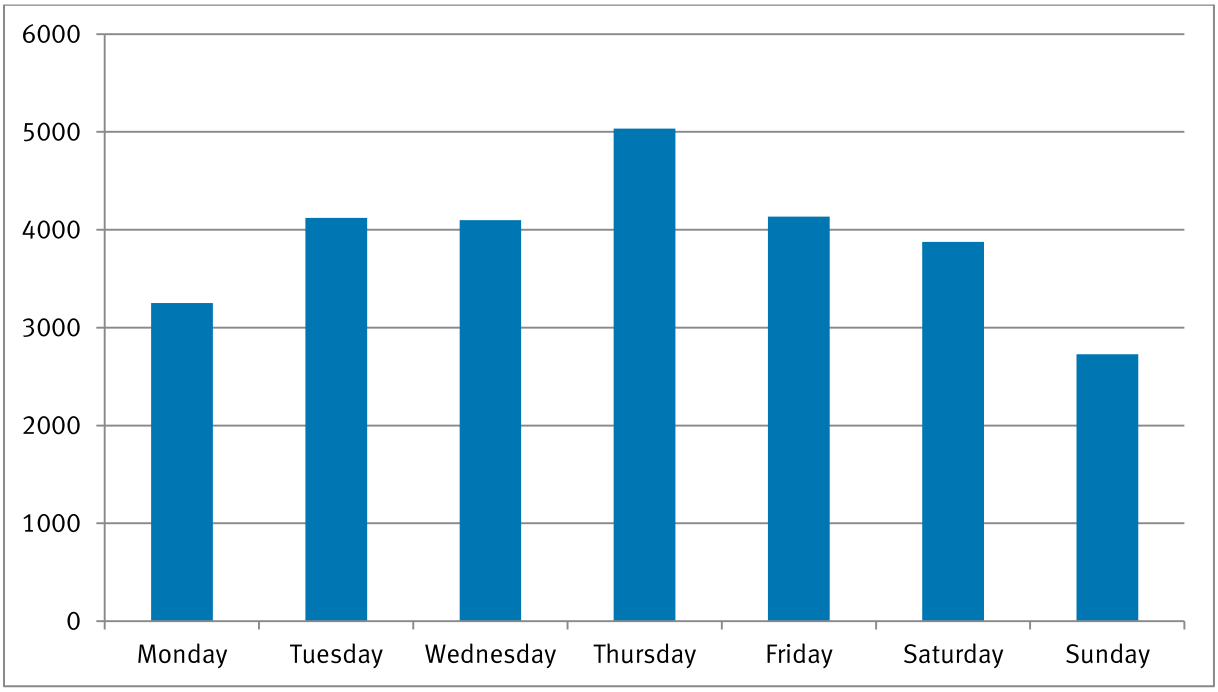

Chart 4: Price changes per weekday

Implications for Consumer Price Statistics

Applying the results of this study to Consumer Price Statistics provides the following insights. There is evidence that at the moment dynamic pricing is applied by few online shops in a remarkable extent. This is a major task to tackle in the near future with improved methods and tools. Web scraping which was used in the present study is a suitable tool to improve the monthly price collection.

The study has revealed that for two thirds of the products the price was only changed up to 3 times in 3 months (once per month in average) which is captured by the traditional price collection of Consumer Price Statistics (one price observation per month). For the remaining one third of products of the sample, Consumer Price Statistics will have to find ways to collect prices more frequently.

For the share of products with 4-15 price changes in 3 months more frequent dates of manual price collection and additional checks could be an acceptable solution. For the share of products with more than 15 price changes in 3 months, web scraping provides a suitable tool for price collection.

In general, frequently changing prices are not of major concern for the calculation of the CPI as long as the variation coefficient is below a certain level. Considering products for which the price was changed more than 3 times in 3 months (more than once per month in average) with a variation coefficient of more than 0.05 to be not manageable by the means of traditional price collection, leads to the outcome that 20% of the observed prices of this study can lead to a bias in index calculation.

The information about online shops and their price setting behavior in respect of the frequency of price changes and the volatility of price series is very important for price collectors and index compilers. With this information the resources for price collection can be managed in a way that more resources are used for price collection in shops with a high frequency of price changes and a high volatility of prices. Price collection in these shops could take place more often and checks could be extended respectively. In this respect, one can use the results for the timing of price changes described above. The manual price collection should preferably take place at times during the typical time of purchasing when the prices tend to be relatively stable assuming that those are the more representative prices. This assumption of course has to be proved by more profound analysis that has not been done so far.

Conclusion and Outlook

The main result of the study presented in this paper is that dynamic pricing currently is applied in some online shops that seem to be big enough to afford the application of complex and high-capacity algorithms. Taking into account that Statistical Offices cannot react immediately on more frequently changing prices by using for example new tools like web scraping for all online shops, this result can be interpreted positively: It allows to concentrate on the application of new tools (like web scraping) and the introduction of additional manual price collection and checks for certain online shops with a high frequency of price changes and a high volatility of price series. For online shops with low numbers of price changes and low volatility of price series, the traditional price collection seems to be sufficient.

Although the use of web scraping seems to be an efficient way of conducting price collection via internet in future, since less resources are needed in the medium term to set up techniques in comparison to collect prices manually, it has to be mentioned that the use of this tool also poses methodological challenges to statisticians. Replacements and the corresponding quality adjustments will likely be difficult to automate – especially for technical products. Thus, replacements and quality adjustments will likely remain a task for price collectors/product experts. Nevertheless, in the near future, the majority of workload of a price collector will predominantly cover the task of plausibility checking, the implementation of replacements and quality adjustments rather than the actual price collection. Moreover, calculating average prices based on prices scraped in regular intervals (e.g. every hour as in the present study) is straightforward only at first sight. It is questionable whether all scraped prices are representative and therefore suitable for index calculations. Consumers may face prices at a certain level for which the purchase is unlikely. Against this background, calculating average prices based on scraped prices requires the elimination of outliers. A possible solution to approach this particular problem is to use transaction data. Then there is no need of calculating average prices or collecting more prices in case of volatile prices.

Various shops in Germany have recently introduced digital price signs, at least on an experimental basis. If the share of digital price signs further increases, dynamic pricing will not only be an issue for the price collection in online shops anymore, but also for the price collection in physical shops. A solution to capture dynamic pricing applied in physical shops is the use of transaction data (scanner data). Transaction data are data which record the transactions of the actual event of a consumer buying a good. This is a promising, but also challenging approach which is not discussed in this paper.

There are numerous aspects worth being analyzed in more detail in future studies. This refers, for instance, to the phenomenon of individualized pricing. For an analysis of this particular phenomenon it is necessary to create different designs of automated price collection. Exemplary designs are the use of different devices for visiting online shops or the inclusion of different consumption patterns indicated by the websites visited before.[3] Furthermore, a long-term analysis of online prices collected by traditional price collection, compared to the online prices collected via web scraping, is obvious and revealing.

To sum up, on the basis of the present study, dynamic pricing is an issue to tackle in the forthcoming years but, from the perspective of Consumer Price Statistics, two out of three price developments are still captured by the traditional price collection and the remaining price developments will have to be captured either by additional manual price collections or by an increase use of modern tools as web scraping. There is no need to overestimate the influence of dynamic pricing of individual shops and products on the calculation of the monthly CPI/HICP.

[3] For instance, it may make a difference whether a consumer visits a website for a certain product several times before. The shop then assumes increased interest for a certain product. The way of reaching may have an influence on the pricing behavior as well. Customers either reach the shop directly by typing in the URL or by using a search engine. The latter may be interpreted as more price conscious consumption behavior which results in a lower price. During the manual price collection for the German CPI/HICP, such cases have been noticed.

Appendix

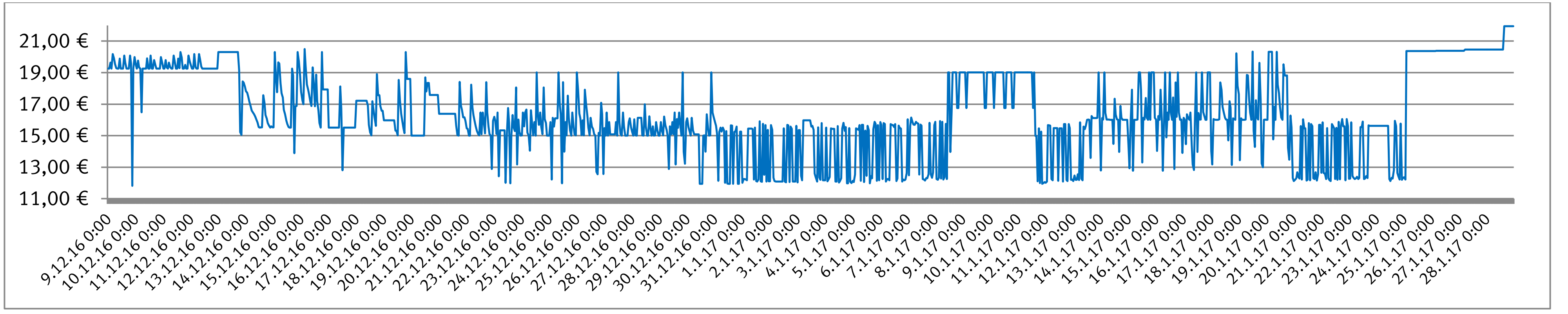

Chart 5: Example of extreme frequent price changes: aftershave

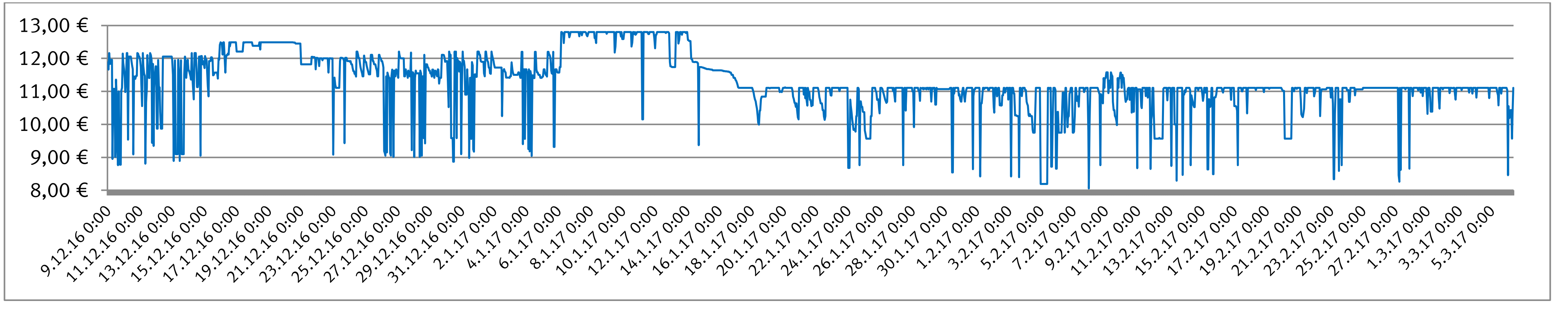

Chart 6: Example of extreme frequent price changes: condoms