

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Blaudow, Christian; Burg, Florian

Article Dynamic pricing as a challenge for consumer price statistics

Eurostat Review on National Accounts and Macroeconomic Indicators (EURONA)

Provided in Cooperation with: Eurostat, Luxembourg

Suggested Citation: Blaudow, Christian; Burg, Florian (2018) : Dynamic pricing as a challenge for consumer price statistics, Eurostat Review on National Accounts and Macroeconomic Indicators (EURONA), ISSN 1977-978X, Publications Office of the European Union, Luxembourg, Iss. 1, pp. 79-94

This Version is available at: https://hdl.handle.net/10419/309825

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

Dynamic pricing as a challenge for consumer price statistics (¹)

CHRISTIAN BLAUDOW, FLORIAN BURG (2)

Abstract: the use of automated algorithms allows online shops to change product prices at short notice, depending on various parameters; this type of price-setting is known as dynamic pricing. At present, roughly 10 000 online prices are collected each month by the German Federal Statistical Office (FSO) for the German consumer price index (CPI) and the harmonised index of consumer prices (HICP), but in many cases only one price is observed each month for each product, usually at a given point in time. Dynamic pricing poses a potential challenge to consumer price statistics: to capture the correct monthly average price and to process volatile prices. To understand this challenge, a study on dynamic pricing was conducted in 2017. It was limited to products and online shops that were included in the CPI/HICP sample and used web scraping to collect prices.

Keywords: consumer price index, harmonised index of consumer prices, dynamic pricing, online price collection, web scraping.

(') This paper has been presented at a meeting of the Ottawa Group on 10-12 May 2017 in

- Eltville am Rhein (Germany) and was updated in March 2018.
- (²) Destatis the Federal Statistical Office of Germany.

1. Introduction

A. What is dynamic pricing?

In general, dynamic pricing is the use of automatic algorithms to change prices at short notice due to changes in market conditions or due to parameters indicating a consumer's willingness to pay.

The phenomenon that prices change rapidly depending on market conditions is of course not new. Prior to the widespread use of e-commerce this was limited to goods and services for which price changes could be announced at little or no cost. For automotive fuels, for instance, prices could be changed several times a day, supported by the use of electronic price signs on the forecourts of petrol stations. Prices also used to change at relatively short intervals for some transport and travel services, such as flights, package holidays or rental cars. Their price strongly depends on the time of booking as well on the time when the service is provided. In the past, booking was mainly done through travel agencies, but nowadays this is done to a large extent using the internet. This development allows for immediate price adjustment due to factors like capacity, or calendar effects such as weekend or Christmas business. To take account of these price changes, price collectors for the German CPI/HICP collect prices several times each month according to a fixed schedule. However, this is time consuming and restricted by staff resources and hence, for some selected products, prices are collected using web scraping techniques.

Besides these special categories of goods and services, dynamic pricing would appear to be a practice of growing importance. Firstly, the phenomenon of dynamic pricing is present in (at least) the German media (Jung (2017); Fischer et al. (2017); Klemm (2017); Hoffmann (2018)). However, the incidence of dynamic pricing and the extent of its use in online markets have not yet been profoundly investigated in scientific literature. Secondly, a market has developed for tools to optimise prices displayed on online websites. Retailers who offer goods and services on websites have the potential to make immediate price changes in order to react to changing market conditions. Major online traders employ their own specialists for optimising their online pricing policy. There are also tools which can be used to observe product prices in online shops and inform small businesses and/or consumers when the price of a certain product falls below a pre-defined threshold. Indeed, the number of online price comparison portals and online pricing tool has increased steadily in recent years. These developments suggest that the phenomenon of dynamic pricing plays an increasing role for online traders and for consumers.

Looking at this phenomenon from the perspective of price statisticians, dynamic pricing makes it more difficult to measure the price developments of products in online shops and — as the online market seems to have gained importance in recent years — leads to new challenges when compiling consumer price indices. Therefore, one of the fundamental

questions dealt with in this paper concerns the extent of dynamic pricing. How large and how common is this issue nowadays? For this reason, the goal of the present study was to investigate the frequency and level of price changes of products that are relevant for the compilation of the German CPI/HICP.

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 price collection from online shops, but also for price collection in physical (so-called 'bricks and mortar') shops. One solution to capture dynamic pricing that is applied in physical shops is to use transaction data (scanner data). Transaction data record the transactions of actual events when consumers buy goods or services. This is a promising, but also challenging approach which is not discussed in this paper. Furthermore, this study is not concerned with the collection of individualised pricing, namely, the practice of offering different prices to different consumers at different times (Schleusener (2016); Remmel (2016)). In particular, individualised pricing describes the phenomenon of trying to charge every customer the individual price that he/she is willing to pay for a certain good based on the (individual) value he/she places on the product at the moment of purchase. This paper deals solely with dynamic pricing and an analysis of dynamic price changes over time.

B. Price collection using the internet for the German CPI/HICP

The basket of goods and services used for the German consumer price statistics contains at its upper level approximately 600 different types of goods and services (Egner (2013)). For the vast majority of these goods and services, a traditional price collection exercise is conducted at physical shops and service providers: a total of more than 300 000 individual price observations are made each month. This traditional price collection is usually conducted at one point during the month, in other words, usually only one price observation is made each month for each product and this is then used for the calculation of average prices which in turn are further used for the calculation of sub-indices up to the overall CPI/HICP. In addition to this traditional price collection, prices are collected for certain goods and especially services that require a more sophisticated survey design (for example, the price of flights or package holidays).

For online shops, the price collection is conducted centrally by the FSO for efficiency reasons. For the majority of goods sold this way, prices are collected manually at one point in time during the month. For the German CPI, weights for different shop types at the level of products were introduced in 2008 (Sandhop (2012)). The overall weight of online shops in all shop types was approximately 5 % for the base year 2010, which is expected to increase in the coming years.

Although the online market is expected to expand further, it is important to note that a much larger amount of prices still refer to goods and services not offered online — or at least they are not offered exclusively online — and that they are therefore not subject to dynamic pricing.

2. Set-up of the study

A. 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. These products and online shops are specified by the CPI/HICP framework, with every product in the sample assigned to a COICOP-10-digit heading (³). The sample was composed solely of goods (no services) covering 242 COICOP headings. At the beginning of the study, the sample started with 3 050 products, distributed across 15 online shops: online shops with a relatively large number of price observations and known to have a large market share were selected for the study. As a result of data collection issues (non-availability of products, explained below), the final analysis was based on 2 680 products and 14 online shops. The prices were collected using web scraping techniques. Since a couple of years ago, these techniques have been used by the German FSO for automatically collecting prices from the internet, and an increasing number of national statistical offices make use of these techniques for their online price collection (Grifficen et al. (2014); Nygaard (2015)).

The target of the present study was to investigate the frequency of price changes of products relevant for the German CPI/HICP for a group of online retailers. The automatic collection of price observations was carried out with an hourly rhythm, allowing many price observations to be made. Consequently, a detailed picture emerged of price developments in online shops and the study was able to investigate the incidence of dynamic pricing and the extent of daily and hourly price changes.

B. Technical instruments and procedure for the data collection program (⁴)

The sample's product input data including product name, shop, COICOP number, COICOP name, unique identifier, product URL (⁵) and product's article number was uploaded to an input table of a relational database. In order to access each product's offer page directly, it was essential to find the URL for each specific product. Another input table was created which included the XPaths (⁶) of the offer pages for the respective online shops. As such, XPaths for the position of the product name, normal price, special price, article number and product availability were stored only once for each 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.

(3) COICOP: classification of individual consumption of purpose. COICOP-10-digit is the lowest (most detailed) level of the classification used for German consumer price statistics and represents the elementary product groups that form the basis for the price collection exercise for CPIs/HICPs.

⁽⁴⁾ For further explanations, see Blaudow (2018).

⁽⁵⁾ Uniform resource locator (reference to a web resource).

^(°) XPath is a syntax for defining parts of an XML document; it can be used to navigate through elements and attributes in an XML document.

The procedure for automated price collection was programmed in Java (see Figure 4.1). The web scraping was carried out using a tool called Selenium that controls a web browser; this tool may be installed as a plug-in for a common internet browser and can then be used for recording and replaying procedures within a browser. Additionally, Selenium offers programming interfaces in various languages including Java. Once implemented, the browser is then used to execute the scraping processes as well as to find the relevant positions of information. Selenium only functions as a provider of information and an interface between Java and the browser. The use of an internet browser for retrieving the information from a webpage is very efficient and flawless and had additional advantages. For instance, when using a common internet browser, it is possible to disable scripts (7) on a page and to start and shut down a browser window within a short amount of time. Additionally, browser windows operate completely independently from each other. Therefore, it is possible to run several browser windows for price collection simultaneously. These advantages lead to a huge increase in speed of price collection.

Figure 4.1: Technical set-up



In addition, the project used steadily changing IP addresses. Therefore, the influence of individual consumer behaviour on the price of the product could not be observed. The prices shown on the page are definitely the prices of the shops for any customer and are not individualised prices. It should be noted that this study was not able to investigate the influence of different electronic devices for the price collection since the prices were collected with the same electronic device.

^(?) Online shops use different types of scripts, JavaScript foremost, to adjust features on their webpages easily or simply to enable interactive webpages. There are scripts which complicate the extraction of information but in some cases also give additional or even necessary information. In general, the developers of the program tried to disable (forbid) as many scripts as possible to ensure that the process of price extraction ran without any delay due to page loading.

C. Treatment of data gaps

The automated price collection exercise started on 9 December 2016 and lasted until 6 March 2017 in order to have data for almost three months. The main part of the study included an analysis of the frequency of price changes and the volatility of price developments. Originally, the program was supposed to collect prices for a whole year. Unfortunately, over the course of the year, a large number of price failed to be collected due to non-availability of the products. In order to evaluate a comparable sample after a period of one year, these products which were no longer available would have had to be replaced in a short time, as done for manual price collection. Replacements were not sought, as within the concepts of consumer price statistics they should not be automated at all (or at least not without insightful algorithms, which was not possible within the limits of this study). As a result, if products were not found for a certain period of time, the product was removed from the analysis. For this reason, the evaluation period was shortened to three months for the whole sample of products, as comparability would have been poor if online shops with a large proportion of failed products were included in the analysis alongside online shops with relatively few products that could not be found.

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

- As long as data gaps were assessed to be temporary (in other words, a price could be scraped again after a certain period of 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 online shop's webpages due to maintenance. In such cases, the last price found for the product was used to fill the gaps. Several shops use 'maintenance procedures' on their websites in order to make general price changes for their products. For this reason, when looking at the time of price changes of every single shop, there is often a cluster of price changes during the night (see Figure 4.3).
- Only price series with at least 50 % of the possible number of 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 price increases of at least 500 % as well as for products which experienced excessive price decreases of more than 80 %. In most of these cases, replacements occurred and, in a few cases, the article number of the old product was used for a new product. This was also the case for some products which were removed by an online shop, with customers being redirected to a new webpage that detailed a similar (but not the same) product. These series were not taken into account in the analysis as the products were not identical, even if they were similar from the online shop's point of view.
- One online shop closed down approximately two months after starting the automated price collection exercise. Since this shop was no longer relevant for the price collection of the CPI/HICP, its historical prices were also excluded from the analysis.

After these data cleansing processes had been applied, 2 680 price series for products in 14 online shops remained in the data set and formed the basis for the analysis. The results presented here for this one-off exercise are anonymised. This anonymisation is not crucial for the analysis. For regular price collection, however, it is very important to know in which shops dynamic pricing is conducted in order to be able to introduce more frequent survey intervals and additional checks at the end of each month.

3. Results

This chapter deals with the most important results of the study. At first, the number of price changes for the whole sample is evaluated. Secondly, we compare the online shops with respect to the number of price changes for each product and then continue to assess the volatility of the price changes of these products. The chapter ends with an analysis of the time when price changes are made.

A. Number of price changes during the observed period

We will first look at the number of price changes for each product during the observation period, split into four 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 figure shows the results:





The first two categories — either no price change or one to three price changes during the three-month period of the study — were chosen as (assuming evenly spread price changes) they reflect a similar number of price changes that might be captured if using manual price collection conducted at one point in time during the month.

In general, the unit value price over the entire month is the target price to be collected. If the price changes on average more than once each month, the traditional methods for price collection could perhaps lead to the inclusion of less representative prices (outliers) compared with the unit value; when performing the index calculation the price collector and the index compiler do not know whether the single observed price is representative.

Almost one quarter of the sample was made up of products where the price changed 4-15 times during the course of the study. This frequency of price changes is not entirely captured by the current manual price collection techniques. With a price collection once each month

there is a growing risk that the observed price is not representative for the monthly average price of a particular product. A relatively small share of products in the sample (11.7 %) changed price more than 15 times during the observation period; this means that on average they had more than five price changes each month. In general, the higher the number of price changes for each product each month, the more unlikely it is that the current manual price collection is able to observe a representative price for the respective month. Moreover, it is also important to consider the variation of the respective price changes (as we will see in the following section).

Table 4.1 compares online shops with respect to the frequency with which they changed their prices for various products. It shows that the share of products with no price changes or with 1-3 price changes was significantly higher for a majority of online shops (shops 2, 3, 5, 7, 8, 10, 11, 12 and 13) than for the remaining shops. These results indicate that dynamic pricing is only broadly applied by some online shops, whereas for other online shops the frequency of price changes seem to be in a range that is usually captured by the current manual price collection and is not caused by special techniques of dynamic pricing.

	Price changes									
Online shop	0		1-3		4-15		>15		Tatal	
	(number of price series)	(%)	number of price series							
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	
Total	1 153	43.02	593	22.13	620	23.13	314	11.72	2 680	

Table 4.1: Number and share of price series analysed by the frequency of price changes during the observation period, by online shop

B. Volatility of prices

The following analysis is restricted to only those cases in which the price of a product changed more than three times during the observation period; this subset is used to analyse the volatility of prices (see Table 4.2). To measure the volatility of prices, the coefficient of variation was calculated for each price series. The coefficient of variation is a measure of dispersion and was chosen because it is a relative measure that does not depend on the absolute level of prices.

The analysis has an important impact on the assessment of the current way of collecting prices manually. When the volatility of prices is low, dynamic pricing might not be a major problem for measuring inflation (in other words, for the calculation of the CPI/HICP). For example, more than four fifths of the observed products offered by Shop 1 saw their price change at least four times during the observation period (in other words more than once each month on average), while the price of almost half the products sold by online Shop 1 changed more than 15 times (more than 5 times each month on average). By contrast, the coefficient of variation for more than four fifths of the observed products offered by Shop 1 was below 0.1 and was even below 0.05 for 55 % of the observed products. This indicates that dynamic pricing was applied in Shop 1, but that the volatility of prices is in a range that could be captured by an extension of the number of points in time of the current manual price collection method. When the fluctuation in the price of a product has a low coefficient of variation, it is unlikely that an incorrect measurement of the monthly price development will occur when using current methods, as it is unlikely that one observed price is an extreme price for the product in question. For instance, one of the observed products offered by Shop 1 changed its price 1 304 times during the three-month period of the study with a coefficient of variation of just 0.02.

Across all online shops in the sample the coefficient of variation for prices was below 0.1 for two thirds (66 %) of the price series where the price changed more than three times during the three month period of the study, while for 42 % of these price series the coefficient of variation was below 0.05.

By contrast, there are also cases for which the price of a product changed almost every hour and did so with a high coefficient of variation. Two examples are shown in the appendix (see Figure 4.4 and Figure 4.5). Without the use of modern techniques such as web scraping, frequent price changes as shown in both of these figures are impossible to capture. Considering the two examples, seasonality is not apparent and the range of price changes is not explainable.

There were only a few online shops for which a relatively high number of price changes was observed with a relatively high coefficient of variation. For example, in Shop 14, prices changed frequently (see Table 4.1) and with a relatively high coefficient of variation (see Table 4.2).

1

	Coefficient of variation							
Online shop	<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				
Total	394	223	233	84				

 Table 4.2: Number of price series for which more than three price changes were

 recorded during the observation period, by online shop and coefficient of variation

C. Time of price changes

Looking at the time of day when price changes occurred (see Figure 4.3), there was a clear pattern with many price changes taking place during the first third of the day, especially between midnight and 1 a.m. This is probably caused by technical reasons: for example, some (more simple) tools used for dynamic pricing only allow price changes once a day; online shops may try to change their prices at a time when there are less consumers shopping online. By contrast, during those periods when most consumers are shopping — the afternoon and evening — it seems clear that online shops avoid changing prices as this is assumed to annoy and/or discourage consumers (Remmel (2016)).



Figure 4.3: Average number of price changes during the observation period, by hour (number)

4. Implications for consumer price statistics

The information presented in this study about online shops and their price setting behaviour with respect to the frequency of price changes and the volatility of price series is very important for price collectors and price index compilers. With this information, resources for price collection can be managed in a different way, assigning more resources to the collection of prices from online shops which display a high frequency of price changes and a high volatility of prices. Certainly, price setting behaviour has to be monitored on a regular basis which may be done with the help of web scraping.

Applying the results of this study to consumer price statistics provides the following insights:

- There is evidence that at the moment dynamic pricing is only applied to a large extent by relatively few online shops. Dealing with 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 current monthly price collection methods.
- The study has revealed that for two thirds of the products surveyed, the price did not change more than three times during the observation period covering almost three months. Assuming that these price changes are evenly spread over time (in other words, no more than one change in price each month) these price changes are captured by the current manual price collection methods for consumer price statistics (one price observation each month). For the remaining one third of products in the sample, consumer price statistics will have to find ways to collect prices more frequently.
- For the products identified as having 4-15 price changes during the three month period, more frequent manual price collection and additional checks could be an acceptable solution. For the products where prices changed more than 15 times during the three month period, web scraping provides a suitable tool for price collection.
- The frequency with which prices change is not of major concern for the calculation of consumer price indices as long as the coefficient of variation remains below a certain level. Considering products for which the price changed more than three times during the three month period (an average of more than once each month) and where simultaneously the coefficient of variation was more than 0.05, to be not manageable by the means of current manual price collection methods, leads to the outcome that 20 % of the observed prices in this study may lead to an incorrect measurement of the monthly price development.
- Finally, the timing (during the day) of the price changes has to be considered as well. Manual price collection should preferably take place during periods when consumers typically make their purchases, assuming that during those periods the prices are more representative as they tend to be relatively stable. However, this assumption has to be proved through a more in-depth analysis than has done so far.

Moreover, it has to be mentioned that the extension of automated price collection may also pose methodological challenges to statisticians. Replacement products and the corresponding quality adjustments will likely be difficult to automate, especially for technical products. Following the traditional bilateral approach of comparing single products over time, Δ

replacement products and quality adjustments will likely remain a task for price collectors/ product experts. An alternative approach would be the use of multilateral methods, which are not discussed in this paper. Nevertheless, in the near future, the majority of the workload of a price collector will cover the task of plausibility checking, the implementation of replacement products and quality adjustments rather than the actual collection of price data.

Moreover, calculating average prices based on prices scraped at regular intervals (for example, 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 the calculation of price indices (Mayhew (2017); Breton et al. (2016); Cavallo (2018)). Indeed, consumers may face prices at a level at which a purchase is unlikely. Against this background, calculating average prices over time based on scraped prices requires the elimination of such outliers and the imputation of missing prices. One possible approach to solve the problem of detecting unusual price quotes is to use transaction data. By doing so, there is no need to calculate average prices or to collect more prices (in case of products with volatile prices). More precisely, only prices of products which actually formed the basis for a purchase by a consumer would enter into the average price calculation; in this case, a complex method of outlier recognition is not necessary.

5. Conclusions and outlook

The main result of the study presented in this paper is that dynamic pricing is in fact currently applied in some online shops. One may draw the conclusion that only larger online retailers can afford the application of complex and high-capacity algorithms. This result can be interpreted positively: it makes it possible to concentrate on the application of new tools (like web scraping) and the introduction of more frequent manual price collection and checks for certain online shops which have 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 current manual price collection method seems to be sufficient.

The use of web scraping presents an efficient way of conducting price collection using the internet in the future, since fewer resources are needed in the medium-term to set up these techniques in comparison with collecting prices manually. However, the extension of automated price collection also poses methodological challenges to statisticians. Replacement products and the corresponding quality adjustments will likely be difficult to automate, especially for technical products.

There are numerous aspects worth analysing in more detail in future studies. These include, for instance, the incidence of individualised pricing. In order to analyse this particular phenomenon, it is necessary to create different designs for automated price collection. Exemplary designs may be the use of different devices that may be used when navigating to online shops or the inclusion of different consumption patterns indicated by the pages visited before making a purchase (⁸). Furthermore, a long-term analysis of online prices collected through manual price collection methods compared with online prices collected using web scraping seems an obvious and instructive area to explore. For this reason, the German FSO is currently working on the implementation of a similar survey, including 10 000 products at roughly 400 online shops. Prices will be collected hourly for an observation period of 12 months and then analysed. The actual price collection will again be performed using web scraping techniques. But this time, where possible, manual and automated replacements will be put in practice for cases where a product disappears. In the end, this long-term analysis will make possible a comparison between online prices collected by manual price collection methods and online prices collected using web scraping.

To sum up, on the basis of the present study, dynamic pricing is an issue to be tackled in the coming years. However, from the perspective of consumer price statistics, two out of three price developments are still captured using traditional price collection methods, while the remaining price developments will have to be captured either through additional manual price collection (more frequent than once a month) or by an increased use of modern tools such as web scraping. There is no need to overstate the influence of dynamic pricing of individual online shops or products on the calculation of the monthly CPIs/HICPs.

^(*) For instance, it may make a difference whether a consumer visits a website for a certain product several times before making a purchase. The shop then assumes increased interest for a certain product. The way of reaching the website may also have an influence on the pricing behaviour, as customers can either reach an online shop by directly typing in the URL or by using a search engine. The latter may be interpreted as more price conscious consumer behaviour which results in a lower price being offered. During the manual price collection exercises that are conducted for the German CPI/HICP, cases such as those detailed above have been observed.

References

Blaudow, C. (2018), 'Fortschritte und Herausforderungen beim Web Scraping — Automatisierung von Preiserhebungen im Internet', *Methoden - Verfahren - Entwicklungen*, Issue 1, pp. 3.

Breton, R., T. Flower, M. Mayhew, E. Metcalfe, N. Milliken, C. Payne, T. Smith, J. Winton and A. Woods (2016), 'Research indices using web scraped data: May 2016 update', Office for National Statistics (ONS).

Cavallo, A. (2018), 'Scraped data and sticky prices', *Review of Economics and Statistics*, Issue 1, pp. 105ff.

Egner, U. (2013), 'Verbraucherpreisstatistik auf neuer Basis 2010', *Wirtschaft und Statistik*, Issue 5, pp. 329ff.

Fischer, M., J. Goebel and H. Hielscher (2017), 'Der Preis ist heiß', *Wirtschaftswoche*, No. 10, pp. 18ff.

Griffioen, R., J. De Haan and L. Willenborg (2014), 'Collecting clothing data from the Internet', Statistics Netherlands, The Hague.

Hoffmann, C. (2018), 'Schwer zu fassen', Süddeutsche Zeitung from 4 January 2018, pp. 18.

Jung, A. (2017), 'Die Preis-Frage', Der Spiegel, No. 9, pp. 76ff.

Klemm, T. (2017), 'Misstraut der Inflation!', Frankfurter Allgemeine Sonntagszeitung, No. 34, pp. 29.

Mayhew, M. (2017), 'A comparison of index number methodology used on UK web scraped price data', ONS methodology working paper series, No. 12.

Nygaard, R. (2015), 'The use of online prices in the Norwegian consumer price index', paper presented at the 14th meeting of the Ottawa group, Tokyo.

Remmel, J. (2016), 'Die verbraucherpolitische Perspektive: aktuelle Entwicklungen im Online-Handel', *Wirtschaftsdienst*, Issue 12, pp. 875ff.

Sandhop, K. (2012), 'Geschäftstypengewichtung im Verbraucherpreisindex', *Wirtschaft und Statistik*, Issue 3, pp. 266ff.

Schleusener, M. (2016), 'Dynamisch und personalisiert: Wie entwickelt sich die Preissetzung im Online-Handel?', *Wirtschaftsdienst*, Issue 12, pp. 868ff.

Appendix

